Global CyberBridges 2008 Interim Assessment

September 5, 2008
Institute for Public Opinion Research * Florida International University

Overview

This interim report is a formative assessment of ongoing activities of the four project groups. It is primarily based on project members' own reports on their activities and accomplishments. Moodle communications were also assessed by the IPOR team. For the three project groups with international participation, a map of Moodle message exchanges was constructed to measure the amount of participation by each group member.

In this interim report we are primarily assessing ongoing activity. Based on this, we conclude that all four groups are applying knowledge of parallel computing and/or distance collaboration tools (e.g EVO collaboration tool, http://evo.caltech.edu/evoGate/) to collaboratively solve problems and make programming/scientific progress. Three of the groups involve successful international participation. The groups are to be particularly commended for their perceptive reports and awareness of their own progress. This awareness increases the effectiveness of Global Cyberbridges as a productive environment for developing and testing cyberinfrastructure procedures for collaboration and learning.

In addition to this interim report the assessment team is working on an analysis of GCB objectives this year, individual participant objectives, accomplishments, and concerns as revealed through interviews with them, and review by the external scientific evaluators. This will be the basis for the final summative evaluation to be completed after the four project groups finish their work together.

1. A Novel Environment for Collaboration and Learning Based on SAGE

Description: This project involves enhancing the Scalable Adaptive Graphics Environment (SAGE) as a collaborative platform.

GCB Fellow and Participants:

- Javier Delgado (Florida International University, USA) - Project leader
- Juan Zhao (Computer Network Information Center, Chinese Academy of Sciences) - SAGE Web browser development
- Shuren Bi (Computer Network Information Center, Chinese Academy of Sciences) - SAGE Web browser development
- Javier Figueroa (Florida International University and University of Miami, USA) - SAGE Web browser development

- Mark Eirik Scortegagna Joselli (University of Rio de Janeiro, Brazil) - Enhancement of SAGE interface by adding support for Wiimote input device
- Silvio Luis Stanzani (Catholic University of Santos, Brazil) - Enhancement of remote desktop sharing functionality of SAGE

Summary from Project Leader:
(As of August 20, 2008)

This project has had and continues to have two major impacts on its members. For one, it is allowing us to work on a cutting-edge platform that could be used for remote collaboration and e-learning. As part of the process, it is helping us to develop our skills in collaborating with members who are in different parts of the globe. In a way, GCB is giving us a motive for improving this platform so that we can improve our ability to improve it. For the last five or six months, we have dealt with several problems. For starters, we had to wait for students in different parts of the world to start their school year, since their vacation schedules are different. We had to learn to work around very different time zones in order to agree on meeting times (which often took place at very unconventional hours). We dealt with software and hardware (e.g. network) problems in the classroom which have made it difficult for students who were attending the classroom "virtually" to follow (and thus get the most out of the class). As students, we have learned to deal with these problems. But we have also learned the needs of a collaboration environment, and could apply those to our project of creating a novel platform for such an environment.

The project itself addresses the problem of remote collaboration and/or learning by using the Scalable Adaptive Graphics Environment (SAGE). Advances in communication technology have made it possible for individuals in widely distributed geographical locations to collaborate by means of audio (e.g. telephone) or text (e.g. e-mail). These means of communication have had a great impact on globalization. Major scientific breakthroughs have been achieved as a result of the possibilities that are opened by this. By allowing diverse people in different disciplines, from different cultures, and with different points of view to seamlessly collaborate, scientific breakthroughs are made more possible.

One link missing from the above collaboration modalities is visual collaboration. There have been developments to this end as well. This includes video conferencing, remote desktop sharing, etc. However, this is still an area of ongoing research. Several software solutions exist, including advanced remote desktop sharing tools such as VNC and NX, videoconferencing solutions such as Access Grid and "Enabling Virtual Organizations" (EVO), etc. These tools address many important issues that are critical in advancing research by helping with globalization and remote collaboration.

We chose to use SAGE as a platform for this kind of collaboration due to its promising features that allow for a fully-integrated visualization environment. At the most basic level of visual collaboration is videoconferencing. With SAGE, this is made possible without any changes to the software. Multiple streams can be going on at the same time on a SAGE display. Another common need is the sharing of ones work. SAGE allows this as well. It includes built-in Virtual Network Computing (VNC) software. One typical use of such software is for showing presentation slides. A "virtual" learning environment such as our GCB class is one example of this. However, upon trying this on a SAGE display in the case that students are in far away places, the performance was inadequate. This is

detrimental to the learning experience. As a result, we have begun working on porting a high-performance remote desktop protocol to SAGE. The protocol of choice is NX (Network X). Our team consists of several members with experience with computer networking who are able to develop a solid platform for this.

It is not possible to predict any particular application requirement for a remote visualization environment. SAGE, by its nature, requires applications to be written to work with it. This presented a conflict for us as we wanted to come up with a solution that would be flexible for the needs of different types of "customers." A high-performing remote desktop approach helps this issue somewhat, but not entirely. As a result, we decided to port the most popular application platform today to SAGE: a web browser. A web browser allows a vast array of applications to be developed without needing to know anything about SAGE. Whether the purpose of the visualization environment is e-learning, medical visualization, geoscience, etc. a web browser can be used for portraying the application.

The final missing link that we found was the human interface to the SAGE display. A tiled display is a relatively new concept that requires new developments in human-computer interaction. Since users of a tiled display are usually not sitting down at a desk, a keyboard and mouse do not always provide the best or most efficient user experience. We therefore chose the Nintendo Wiimote as a possible interface device. We have a student who has experience working with such a controller and porting it for use by a standard computer workstation. The fact that this controller works in a way that is natural to the user makes it an ideal candidate for a SAGE display, in which many times application windows are near life-size and benefit from such a device.

Moodle Communication Review:
(As of August 12, 2008)

Discussions began on March 11, 2008 and are continuing through the present time. There were 32 total discussions or announcements initiated by the group leader (13), group members (12) or the instructors (7).

The SAGE group began with a welcome message from the group leader and followed with discussions regarding a weekly meeting time and type of communication venue. It was decided by the group to hold their meetings Wednesdays at 9 PM (USA), 10 PM (Brazil) and 9 AM (China). The group had booked their meeting time with EVO, but switched to SKYPE because there was no video or desktop sharing at this time.

The group leader directed the team to the SAGE WIKI page for general information about the project and a list of group roles. The group leader has instructed that the team should continue posting on Moodle, but they also find it necessary to use WIKI and personal email for additional correspondence.

The group posted many comments regarding the status of SAGE 3 at the various locations and was directed to WIKI for installation and configuration instructions. As of April 1st, all sites had SAGE 3 running.

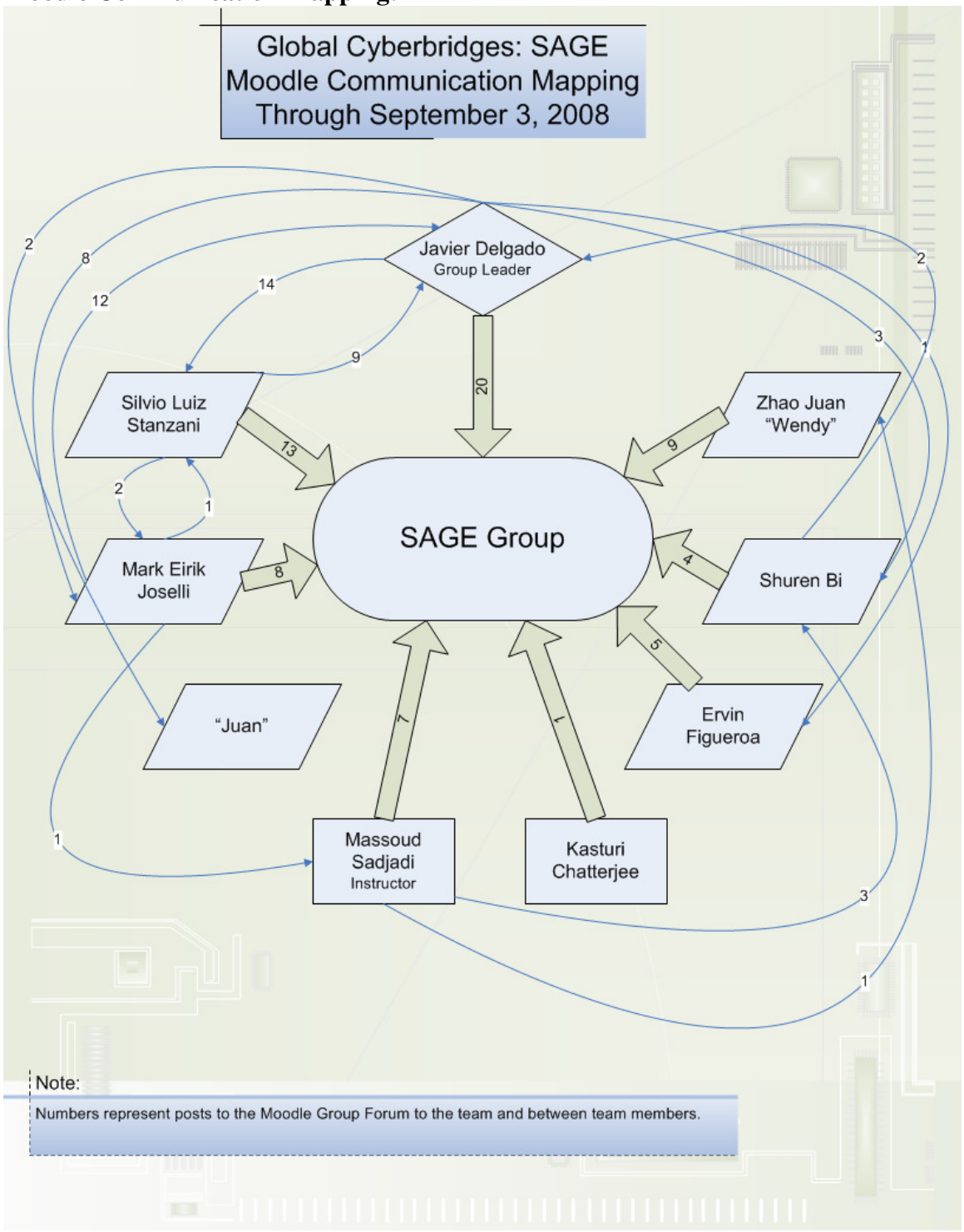
The instructor requested additional input and postings from the CNIC students and they replied that they would post their progress on the WIKI page.

It was noted that there were hardware problems with the FIU tile displays, but there were no follow up postings to this issue. Some of the points to be addressed by the group were the progress and/or findings regarding the VNC viewer, browser, freeNX and Wii. There were many discussions about problems and bugs and how they could be resolved, and the group leader was efficient in responding to team members. However the last post was on April 30th and there was no indication on Moodle what progress had been made. The instructor requested a project update on July 14th.

The University of Rio de Janeiro student replied that he had implemented the basic functionalities of a Wiimote pointer for SAGE in C and was working on other pointer functions. He wants to test the WiiSage pointer over the tile wall at FIU. He indicated that weekly meetings had not happened for at least one month and that the group leader was in Europe.

Group discussions resumed in August, and the group leader rescheduled weekly meetings for Tuesdays at 9 AM (USA), 10 AM (Brazil) and 9 PM (China). The last postings deal with drafting a project update report.

Moodle Communication Mapping:



Project Tasks and Achievements:

The latest version of SAGE was deployed in two display walls: one at CNIC and one at FIU (CATE). It is also working on two Windows desktop machines at the Brazilian

universities. Regarding the SAGE-enabled Web Browser, the team is analyzing the Firefox source code to determine how it works, how it could be parallelized for high-performance rendering on a large tiled display and what modifications are necessary to its source code in order to project an otherwise unmodified Firefox application to a SAGE display.

Two of the team members have successfully compiled Firefox from source and the WIKI page was created in the LA-Grid. The team is analyzing the X11 source code and learning how users interact with the X11 interface. The team is also working towards compiling Firefox with SAGE libraries to visualize on the SAGE display.

The SAGE group is also analyzing the SAGE VNCViewer code to improve existing or create new remote desktop software. They are attempting to modify FreeNX to be workable with SAGE, and have discovered that compilation of the stock source code is possible, but are unable to modify it. The team is considering different approaches to using FreeNX alone. The team members are analyzing Luc Renambot's Display sharing code for Windows and Mac to learn how it works and possibly apply the concepts to FreeNX.

The group is developing an extension to SAGEUI for Wiimote functionality as a pointer for a SAGE display. After some limitations were discovered with a Java implementation of the remote interface, a C version was written and is performing well. Several functionalities have been implemented and will be tested on high-resolution displays.

Goals for the future:

The team members will have the next steps and implementation details prepared by the end of August. In September the team plans to assess the performance and display quality of Firefox on a SAGE display, the functionality of Wiimote on SAGE (local and remote), and the progress of the implementation of NX. Also, planned for September is to begin working on the project poster and to assess the progress of Firefox parallelization scheme, Wiimote functionality with different applications and progress of implementation.

By October 15th, the team will begin to work on the conference paper. In November they will submit their poster for the IWIC 2009 conference, finish their technical report and submit the project paper for the CSEDU 2008 conference.

2. WRF Grid Portal: Hurricane Modeling

Description: This project deals with developing an innovative grid-enabled multiple scale hurricane modeling system.

GCB Fellow and Participants:

- Konstantinos Menelaou (Florida International University, USA) – Project Leader

- Sean P. Leslie (Florida International Univerity, USA) – Developer – PIRE student
- Allison M. Lana ger (Florida International Univerity, USA) – Developer – PIRE student
- Marlon Bright (Florida International Univerity, USA) – Developer
- Javier Delgado (Florida International Univerity, USA) – Developer
- Javier Munoz (Florida International Univerity, USA) – Developer
- Diego Lopez (Florida International Univerity, USA) – Developer
- Eric Meyer (Florida International Univerity, USA) – Developer
- Javier Figueroa (Florida International Univerity /University of Miami, USA) – Developer – REU student
- Seychelles Martinez (Florida International Univerity, USA) – Developer – REU student
- Elias Rodriguez (Florida International Univerity, USA) – Developer – REU students

Summary from Project Leader:
(As of August 20, 2008)

Although there has been a substantial improvement in concern regarding the problems faced in atmospheric science meteorologists still spend much of there time dealing with the construction of the basic tools and not totally focused on understanding and explaining the principles of the atmosphere. Having this in mind WRF Grid portal project is challenged with the task to develop a web base portal that will provide people who want to study the weather with easy access and usage of the Weather Research and Forecasting model (WRF).

To better approach this project and be able to provide accurate results good knowledge of both atmospheric and computer science is needed. Based on this, the project team is a combination of people with strong backgrounds in either meteorology or computer science. Throughout this collaboration we've been able to not only enhance our knowledge within our own fields but also to achieve basic knowledge on different fields.

With part of the team being out of the country we had to deal with the issue of communication. Though with the tools introduced in the GCB class these challenges were overcome. Working with people form different countries was the key fact that gives this project a spherical prospective and provided guidance on how to achieve a solution that would be available to a broader majority of people.

Building an accurate meteorological tool not only demands indispensable knowledge but also lots of money. Many people may have the ability to become good meteorologists but they lack money, facilities, or instruments. WRF Grid portal is intended to be a primary tool for many underrepresented groups and provide them with the opportunity to contribute and be active in the prediction of the weather.

Predicting the weather was and still is mandatory and with significant socioeconomic impact for every country. Easy and cheap access to accurate meteorological forecasting gives the opportunity for more people to learn and submit research. With Grid WRF Portal schools and Universities of varying sizes will have the ability to generate meteorological

programs and teach people about the amazing phenomenon that is the weather. As for meteorologists they can now focus to their job.

The purpose of WRF Portal is to provide a simplified interface with which Meteorologists can work with WRF in a simplified manner. This objective by itself is not the immediate goal of our research, but is rather more of a mission statement. During this project we hope to kick start the development of what we hope will be a web tool with full WRF simulation creation capabilities. This will be eventually attained by fully integrating multiple tools into one simplified tool accessible via a web browser.

Moodle Communication Review:
(As of August 12, 2008)

Discussions began on April 2, 2008 and have continued through the present time. There were 37 total discussions or announcements initiated by the group leader (4), group members (25) or the instructors (8).

It appears that the Hurricane modeling group began its discussions via a communication device other than Moodle. The group began using Moodle on April 2nd, and began discussions relating to WRF. The postings gave no indication of a meeting schedule or project goals and timeline. However on April 2nd, there was an inquiry from a group member whether there would be a meeting in person on Saturday. The assessment team assumes that this is the group meeting time; however some of the team members indicated schedule conflicts.

The main topic of discussion in the Moodle forum is ensuring that WRF is compiled correctly on the GCB and MIND clusters. The team created a WIKI page for WRF on the LA Grid and posted links to WRF website and publications on the WIKI pages.

Once WRF was successfully compiled to GCB cluster, the instructor assigned project members tasks relating to WRF and clarified assignments. He requested assistance from Diego Lopez and Javier Munoz with the problems the team was having running benchmarks on stock WRF installation.

The team seemed to still be working out a meeting schedule for Saturdays to resolve issues of technical problems while running a simulation, running “Bob” scripts on compiled WRF and the integration of the newly compiled WRF with Bob scripts. The decision was made to compile WRF again with live documentation to make sure there are no missing steps.

The instructor reminded the group to update their WIKI page with details. It was noted that the group needed to clarify the goals of the project and their plan. The instructor posted a message clarifying items for the group including scheduling, assignments, and meetings and wanted to verify that all were on the same page. Again, the instructor reminded the team to use Moodle instead of direct email.

The group was still working out issues with WRF and decided to work with WRF-2 and to not use Bob Scripts because they were too slow. The team was working on the WRF database, file storage and API definitions. The instructor requested group input.

The team began discussions regarding visualization ideas for the project, verifying WRF output files, portal design and a user ensemble creation template.

The team identified that it takes as much time and effort to install/compile WPS as it does to install/compile WRF and that this may change the project goals.

There were a number of posted discussions that dealt with the WRF portal diagram regarding users, history, ensemble components and namelist property tables. Also, the team noted that they were working on a web service to run jobs on GCB and the possibility of creating a dummy web service for the portal to test the API calls.

The remainder of the posted discussions related to the Web Portal Architecture including diagrams, flow charts, structure and ideas.

Moodle Mapping not completed for this group since all members were at FIU and had many alternative modes of communication.

Project Tasks and Achievements:
The team first met with past and present WRF project participants to exchange knowledge and be introduced to WRF. A very general concept of WRF as a system was introduced and some of the problems faced with WRF grid-enablement were discussed. The team then created a Software Requirement Specification and learned the basics of the tools required to complete the project.

A mock UI was created and project and database design was determined by the team. The team developed the User History page which will dynamically create a table based on the number of jobs a user has created and will display the jobs for a specific user.

Namelist variable editing was enabled via the creation of a page using web forms. This allowed a user to submit data from the form on a web browser and into the database. Sessions were added to the Portal enabling access to user specific histories. A logged in user will now see his/her history of jobs only instead of a default person's history. The java code for parsing a namelist file was also created but not launched on the web.

The team created the capability to upload files and added the ability of a logged in user to change their password as well as their email address, and create new accounts for new users. The addition of the option to create Field-Of-Study specific single runs was also added.

Goals for the future:

At present the goal of this team is to provide the basics needed to accomplish job creation and submission. The goal will be attained by providing a combination of features and web functionality including creating a basic website interface which allows the creation of WRF simulations, allowing files to be uploaded, maintaining a job history, enabling the exposure of namelist variables and the capability to edit them, enabling login and session features for users, and to visualize the WRF simulation output. The team is currently working on job submission to the portal, researching visualization and adding visualization to the portal.

It is through the creation of these features that will enable the basic functionality of the WRF Portal. Session enablement is an essential functionality that will preface the addition of features such as job history, personal identifications, and job tracking. The exposure of namelist variables and their storage will play an essential role in making WRF Portal more meaningful to its users. Finally the visualization of output received from a WRF simulation will lessen the post-processing burden on the user.

3. Finding Discriminating Probes in Bacteria

Description: To create complex and time consuming Bioinformatics experiments in a Parallel Computing environment.

GCB Fellow and Participants:
Michael Robinson (Florida International University, USA) - Project Leader
Guangyuan (Gary) Liu (Computer Network Information Center, Chinese Academy of Sciences) - Developer
Camilo Silva (Florida International University, USA) – Developer

Summary from Project Leader:
(As of August 20, 2008)

> One of the most challenging problems in Bioinformatics is the large amount of data that needs to be analyzed. Many problems in Bioinformatics can only be done using statistical methods, many others take months and years of processing, and many other can not be done at this time because of the amount of data and time that is needed to get any meaningful results.
>
> My Bioinformatics advisor Dr. Giri Narasimhan designed over 36 possible experiments that we could implement in a cluster environment. After developing several of those experiments in single user mode, we selected a project called “Finding Discriminating Probes” (Project 18) as the most suitable for a cluster. We also found that we could implement this project in many different ways, and we decided to select five bacterias of the same family to find all the discriminating probes of length 10, in all 20 possible combinations. Because of the large amount of work ahead of us, we also decided to process each of these 20 jobs, individually of each other as far as sharing results.

After creating many programs that produced initial data and processed it, we found all the discriminating probes, and validated the results. We then developed two mpi programs that have gone thru many modifications, and used them to run the jobs in the two clusters using different size data files, with great results.

Finally we merged the two mpi programs into one final program that ties all fully tested single user programs without modifications, to be processes in two independent clusters at FIU, GCB cluster and Starscream cluster, which is being done at this time.

All of the programs were done in native c, except for the mpi programs that use mpi functions.

Finding Discriminating Probes in multiple bacterias is the task of finding what probes and their corresponding reverse complements are in one bacterium and not in the other. This information is very important because when a strain of a bacterium mutates producing a new strain, this process will allow us to find the differences in multiple bacterias, of all probes of any length.

Because of the complexity of these processes, implementing them in parallel processing will allow us to embark in larger and more detailed research. These will allow us to create tools for all areas of Biology, Chemistry and all medical fields.

I, Michael Robinson have been doing Bioinformatics research for almost four years. Guangyuan Liu has been working in parallel processing for the last two years and has done some Bioinformatics implementations. Camilo Silva is a Computer Engineer student, with a tremendous amount of energy, curiosity and wiliness to work. I have worked in group projects for more years that I want to admit, but this group without a doubt has been the best.

The large amount of projects that Dr. Giri Narasimhan gave us, as well as his constant advice, and the permanent help and guidance from Dr. Sadjadi, and our group made it very easy to conceived and organized the project activities.

We were also fortunately in that each member of the group had its own laptop computer, and we had access to the GCB cluster at all times. Although we had access to many different modes of communications, we elected to use EVO as a method of doing our weekly meetings, as well as the Moodle system and email for overall communications. Access to resources has been excellent.

Learning to establish and implement international projects like ours is one of the broader impacts of these project activities. It has greatly help us in advancing discovery and understanding what is being done in other countries while promoting teaching, training, and learning. This experience has taught me how science transcends all gender, ethnicity, disability, geographic, and other areas.

Creating projects like this one at Cyberbridges extents and enhances the infrastructure for research and education, it also facilitates the infrastructure needed to create partnerships between schools and research groups allowing us to broadly disseminated and enhance scientific and technological understanding.

It makes it possible to bring technology and discovery to any corner of the world, which obviously benefits society greatly.

A very important by-product of this project, for me, is that I have been asked to be a visiting professor by a university in Latin America, to be part of implementing a Bioinformatics class for the school of Veterinary Medicine, nine PhD and two Master students have already signed up. I will be using the communications techniques, and

running all projects in the same manner as I have been doing since January, in Cyberbridges.

Moodle Communication Review:
(As of August 12, 2008)

Discussions began on March 12, 2008 and have continued through the present time. There were 99 total discussions or announcements initiated by the group leader (76), group members (19) or the instructors (4).

The Genomic Sequences group began their Moodle posts with a discussion of weekly meeting time, description of project goals and verification of class requirements. The group determined that EVO was the best communication device for their meetings because the recorded log makes it easier to create the meeting minutes document. The group is implementing the WIKI pages through the LA Grid and is using personal email for additional correspondence.

This group had the most postings to the Moodle group forum, and thus had the most detailed account of the project's progress as compared with the other three groups. If the assessment team were to only use Moodle as a method to evaluate the groups' projects, this team would stand out in their communications, organization, research methods, priorities and division of labor. The Moodle posts walk the assessment team through the project's development, including technical problems, reasons for team members' absences, schedule changes, outside consulting, resolved issues and any revisions to the team's goals.

The group leader posted files to start developing a program that will allow them to do grid computing. There were discussions about their project's contribution to the field and input was given from the group's mentor, Dr. Giri Narasimhan, as to the project's direction and 18 subprojects.

It was discovered that Moodle could only accept one attachment, so the group was using email to send documents and data files. The group eventually used a website to post large data files for team member downloading. At this stage the group had done the following: 1) tested data output from existing Perl programs over 2.5 million searches and java programs with no errors found.
2) perl program to c language, 3) created search.o to make applications writing easier and faster, 4) application sample to find probes in data, 5) provided a complete DNA bacteria pa01 with data, indexes and application in C, 6) downloaded complete record for PA01, to create a new file with 4 fields from each gene. The group was then ready to implement the first application in the GRID and plans to create a new application that will expand the information given to the users, using the above programs and data.

There was a note that the CNIC student was having trouble receiving Moodle posts, but this was eventually resolved by using the most recent version of Moodle. The two FIU students were meeting in person every other day, but are keeping the CNIN team member

up to date on their progress via Moodle and email. The group revisited their goals and timeline and decided they wanted to have a useful parallel application running in the grid by August 15th. In addition, they determined that their subproject #18 ("find discriminating probes present in some genomes, absent in others") would have the most impact and contribution to the field, so this was made the primary focus.

The group identified that they were using an outside evaluator to spot check their results and that one team member was temporarily relocating to Mexico. Despite this change, the group continued to have their meetings (increased to 2 per week), frequent communication and forward progress.

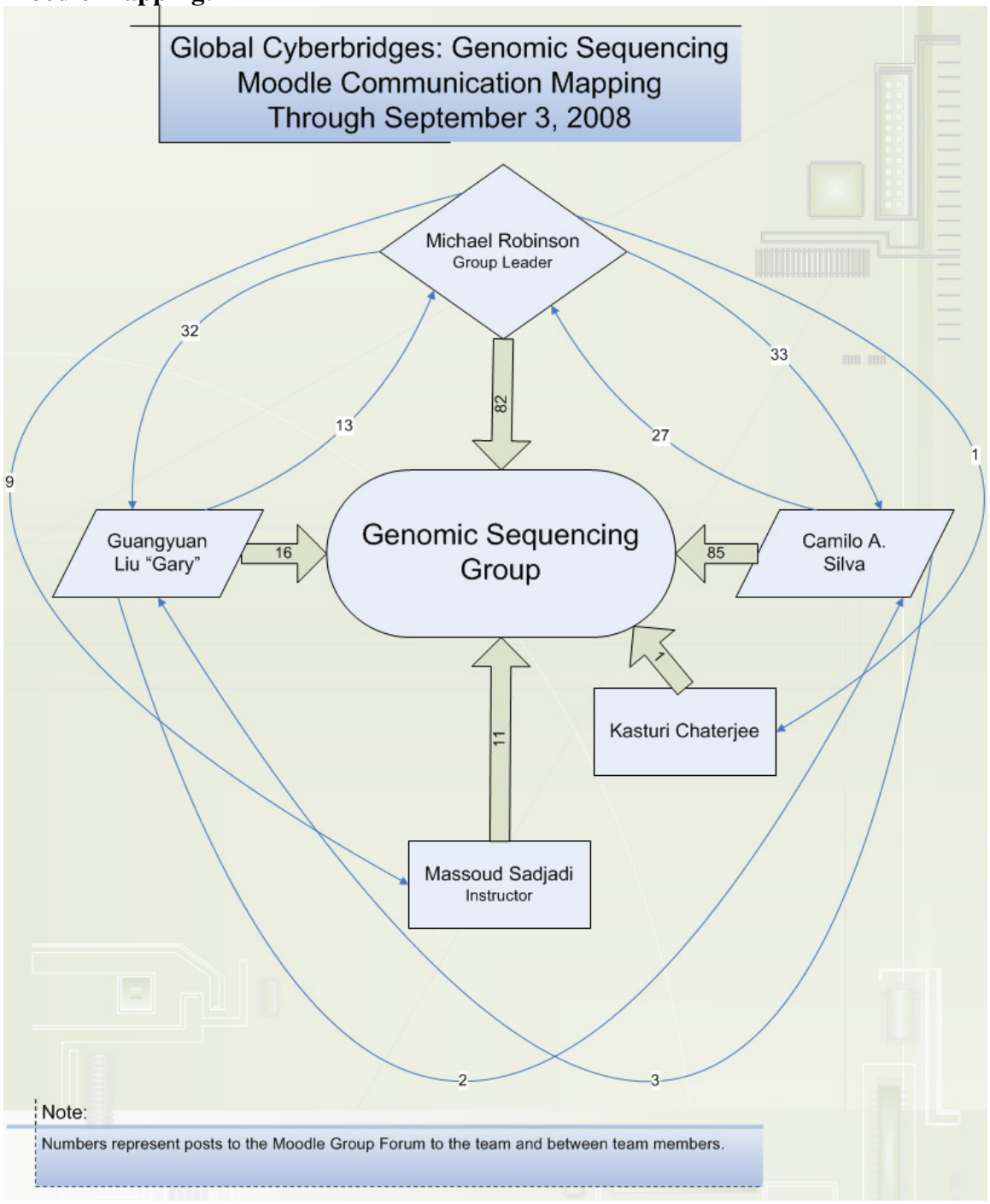
The team was checking test results for locating probes using algorithms and was modifying the code to run in a cluster. The MPI guide was posted and there was discussion regarding the steps to running a parallel program. One of the team members posted a guide on how to submit jobs to GCB and a flow chart describing how they are finding discriminating probes using a cluster with MPI. Data was posted on the GCB cluster and all nodes have project information files. The team did three tests of MPI-10 and the tools for parallelization were acquired, so they could move forward with the parallelization of the code. They found a bug in the project, but corrected it and reran the program.

The group had many discussions about testing data and checking results. The data being used was posted at the cs.fiu.edu website. The team created a document describing the methods used to test the data and a PowerPoint presentation was posted describing the parallelization of project #18.

The instructor has recommended that the group rename their project with more identifiers.

The group found an error in the operating environment and discovered and fixed the problem with the directory. They modified the system so no manual input was needed and then were ready to test all jobs to see if they matched the manual results. The remainder of the posts dealt with running all 20 jobs in full size files. This required 30 hour run times and was producing some errors. They tested the GCB cluster and SS cluster with no errors, but found "bugs" in the MPI program that was preventing the running of 20 jobs. At this point they were rerunning a test with 5K files and trying to figure out a memory leak.

Moodle Mapping:



Project Tasks and Achievements:

The team developed all bacteria data needed for the project as well as all programs to run on a single computer. They developed all validation programs and tested all data results. The group created an mpi interface that used all single user programs and corresponding data to run in the GCB cluster and created another implementation of an mpi interface using advanced error checking techniques as well as implementing queue subroutines to

control the nodes processes. One mpi version was run in the GCB cluster and the other version was run in the Starscream cluster: both clusters used the same data and the data files results where the same. Data validation was conducted on both data sets.

Goals for the duration of the project:

The most immediate goals for this project include finalizing the results of all jobs using the final mpi program implementation and to have these results validated. At the present time is collecting statistics from the results obtained and creating the graphical representations of those results for their paper submission in September. By October the team plans to have another paper ready to submit for publication and by November have a list of future additions to the project.

4. Distributed Multimedia Data Management over Grid

Description: To design and develop an efficient multimedia data management framework over Grid with a capability to support popular multimedia retrieval strategies like Content-Based Information Retrieval (CBIR).

GCB Fellow and Participants:
Kasturi Chatterjee (Florida International University, USA) – Project Leader
Dongri Luo (Computer Network Information Center, Chinese Academy of Sciences) – Developer
Fabio Lopes (University of Sao Paulo, Brazil) – Developer

Summary from Project Leader:
(As of August 20, 2008)

The project proposes an efficient multimedia data management over the grid environment and especially concentrates on image storage and retrieval across a distributed environment. Multimedia data is voluminous and has an atypical representation which calls for special attention. Multimedia data is usually represented as multidimensional feature vectors where the features are extracted by processing the contents of the multimedia objects like images or videos. Content processing for multimedia data like images/videos requires a lot of computation power and effort is made to improve it as much as possible as the subsequent quality of retrieval results directly depend on the multimedia object representations. Retrieving multimedia objects from a huge multimedia storage require one to search the multimedia database based on some similarity criteria and provide a ranked set of available results. Grid provides the perfect platform for storing and managing such voluminous data which can harness its computing ability. The grid data management framework mean for file based text data can be extended and improved for managing multimedia data. In this manner the storage space as well as the computing power of each node can be utilized to improve the overall performance and extend the capability of the multimedia data management system.

The project requires an in-depth knowledge of multimedia data base, content-based retrieval as well as expertise of Grid architecture and Grid Data Management framework. Thus, it provides the perfect diverse collaboration platform to exchange domain specific

knowledge among the team members having varied expertise. This team has three members. The lead, Kasturi, is in her final stage of PhD and her thesis concentrates on developing a universal multidimensional tree-based index structure to organize multimedia data like images and videos and devise similarity search methods based on their contents. The other two members, Dongri and Fabio, work mainly on Grid architecture and distributed computing. Thus, they share the knowledge from one another to solve different problems. The outline of the project i.e. proposing a multimedia index structure is a PhD thesis topic. Almost three years of research work has been conducted on it with several published papers and verified results. Thus, the attempt to implement it over the grid environment and tailoring the grid database management tools to achieve it is a novel approach and as far as our knowledge, was not attempted before. The group, located in three different geographic locations i.e. China, Brazil and USA, communicate via skype and online course management site, Moodle regularly to exchange ideas. Though, managing and organizing such meetings become a difficulty at times but the group makes an effort to keep in touch as closely as possible. Communications as well as implementation of the project require plenty of access to computing as well as internet resources. Also, a detailed knowledge of the specification of the Grid environment, need to be mastered.

Since the project includes a multidisciplinary collaboration, some training/teaching is required to transfer the knowledge. Also, we have a varied ethnic representation like a female student, a South American as well as a Chinese national. The project enabled an environment where a partnership was developed among three participating institutes viz. FIU, CNIC and USP. The GCB cluster and the Grid environment are located at FIU and the resources are used by all the students from different participating institutes. This opens the gateway for future resource as well as idea sharing among these institutes and each can leverage the best of the other. The project also has an undeniable effect on the society as firstly, people from different ethnic and cultural background collaborate with each other overcoming the communication problem, secondly, a successful implementation of the project will help people to share and utilize the storage and computing facility of each other to access multimedia data, one of the most popular form of expression, easily and efficiently.

Moodle Communication Review:
(As of August 12, 2008)

Discussions began on March 11, 2008 and have continued through the present time. There were 29 total discussions or announcements initiated by the group leader (13), group members (13) or the instructors (3).

The Multimedia Data Management over Grid group began with a test topic and sorting out the modes in Moodle. The group leader posted a suggested roadmap and timeline for the project and requested that the team members draft a document of their related experience in the project's topic. The group scheduled weekly meetings for Thursdays at 8:30 AM (USA), 9:30 AM (Brazil) and 8:30 PM (China) using SKYPE.

It is evident that the group is using WIKI and personal emailing for additional correspondences. Detailed meeting minutes were posted on Moodle, which allowed the assessment team to follow the progress of the group. There were eight meetings beginning April 4th and finally on July 22nd.

The instructor directed the group to the WIKI page for their literature review and citation postings and reminded them to use the Moodle forum.

According to the minutes of meeting 4, it was decided that both the CNIC and University of Sao Paolo students would share the lead role of the project. It is unclear why the former group leader changed position.

Some of the points being addressed by the group were the progress and/or findings regarding setting up a stream media server, the feature extraction code, installation of matlab in Linux, launching f and p server and how to install GT4. The group made note that image file names must be specific, for example, "9.jpg" works, "09.jpg" does not.

The MDM group set goals to deploy the index structure in the grid and develop the application, organize the grid (3) to use the application, test 100 pictures after deploying the prototype and set up a GCB cluster account.

There were some communication problems noted by the team members. The University of Sao Paolo student lost SKYPE and Moodle communications and the CNIC student was having trouble with Moodle. It seems these issues were somehow resolved.

The group's working discussions and meeting minute posts describe some of the projects the team were working on as well as issues that arose and were resolved. The group spent time testing c++ codes for image retrieval and discovered that they need to set the c++ executable at each node, set up at grid service at the front-end node at the GCB cluster. Then in the grid service, can call the c++ executables to extract the image.

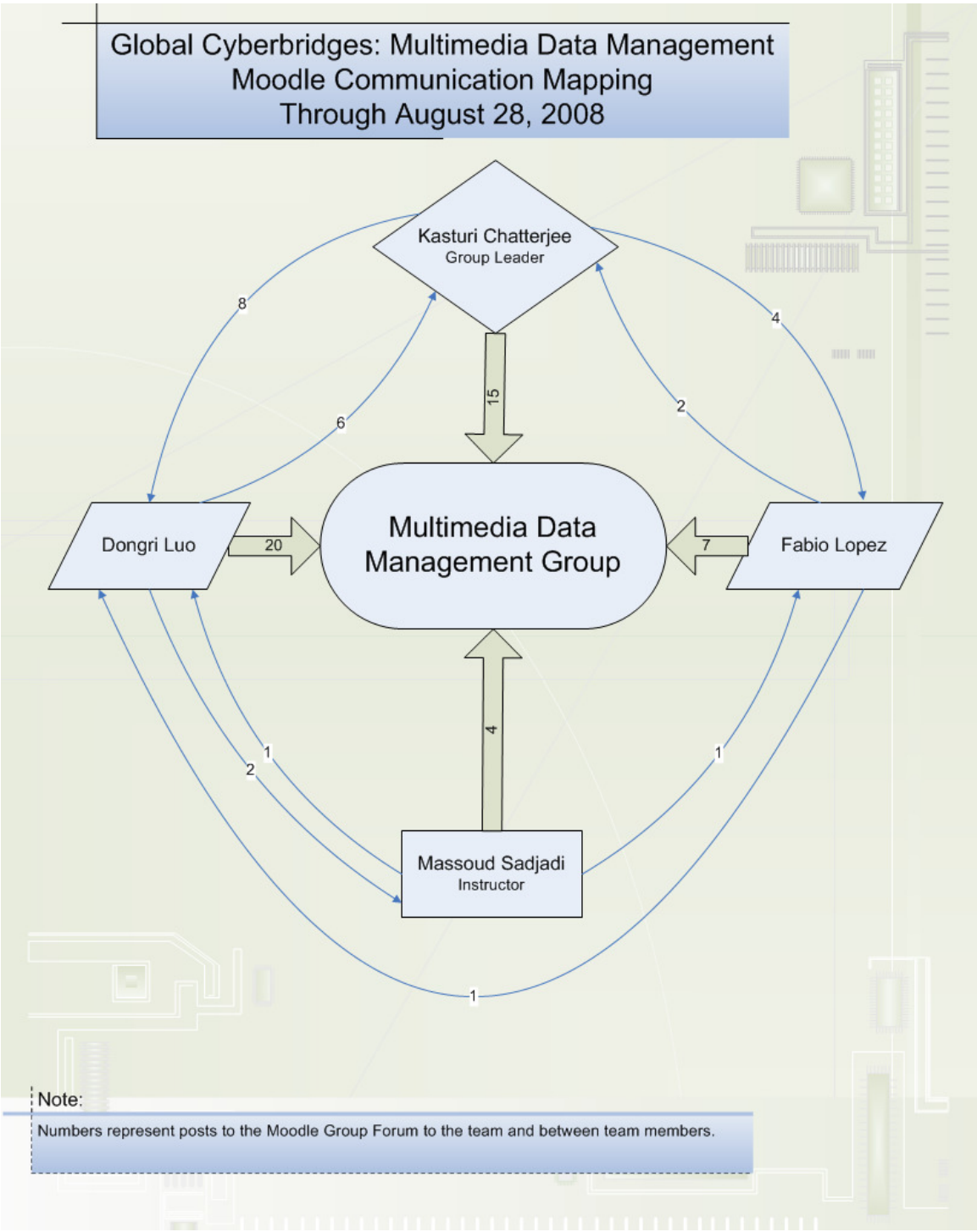
The team stated they were having trouble installing GT4 and that they were working with "David" to get matlab installed correctly at the GCB cluster. Eventually, the team installed GT4 successfully and matlab at the Linux machine. They ran tests and produced "**.mat", which allows for jobs to be submitted to the GT4 server using matlab codes as the application.

The group spent some time revisiting the goals of the project and decided they want to describe how the application will perform and see how difficult it is to use for image sharing among different nodes or over the internet, and to see if it can be extended to videos and multimedia objects. They plan to write a c++ code to invoke the retrieval algorithm at each node one after another like a batch execution/parallel.

For clarification, a team member drew up a flow chart to describe how the user retrieves images from the grid service which executes the application from other nodes. Another team member drew a diagram of UML showing deployment, components, use case and sequence for discussion by the group. These tools seemed helpful to the group and initiated discussion.

The next items the team worked on were make files, GIST code and MTREE. The team solved their problems with the installation of MTREE and make file modifications and posted the steps they took for modification of the codes on Moodle. The last posts identified the group's action items to extract the features of images to build an index structure using matlab code and to use MTREE to get a txt file with ids of query images.

Moodle Mapping:



Project Tasks and Achievements:

After executing a detailed literature survey on existing state-of-the-art multimedia data management and grid data management techniques, the group leader held training sessions for the team members regarding Multimedia Data Management and Multidimensional index structure knowledge. The Image Content Extraction code was made available to the group and a proxy cluster node was set up at a CNIC computer to test the code on sample images.

The team made a sample of the multidimensional index tree code and changes were made to the environment variables as well as to the code to make it run successfully on the computing environment at the test setup at CNIC.

The team designed the frontend and backend for the project and created a software engineering document with the help of UML Diagrams like Class Diagrams and Sequence Diagrams to lay down the flow and idea of the project. They also set up the necessary environment at the GCB cluster; installing Matlab, testing the c++ compiler and its capability of handling the code for the Index Structures.

Goals for the duration of the project:

The main goal of the project is to provide an application over the Grid which will provide a front end that will let a user upload an image of a choice or select an image from the database of existing images. Then the user can submit the query which will be received by bridge application that communicates between the nodes and the GUI. The user can refine the query results by submitting his/her feedback and the search method may be repeated for improving the relevance.

This group is finalizing the data replication policies on the Grid and by September will develop the front end application consisting of a GUI and deploying it on the master node of the GCB cluster. By October, the team will implement an application that will communicate with the code for the multidimensional index structure and the front end. The application will take care of the different grid data management policies. During this time the team will be testing and conducting performance evaluations. The team will begin the final report and paper and prepare for the workshop in November.

Moodle Assessment - News Forum (as of August 18, 2008):

The News Forum was used ten times by the instructors over the course of the semester. These postings related to announcements regarding class requirements and assignments as well as meeting time/ place changes.

***The assessment team would like to note that all the groups developed visual aids such as flowcharts for project clarification and posted them to their respected group forums. These diagrams sparked much discussion between the team members and pointed out discrepancies in the team's concept of the plan. Thus, many versions were drafted and modified by the team. If this is not already a class requirement, we recommend adding this exercise to the curriculum of future GCB classes.