Final Project Report

Trustworthy Computing

(Professor : Mike Shaw)

Trustworthy System Development : Grid Computing

Sunghee Cho (MS-Tech)

May 4, 2006

- Table of Contents -

- **Overview**

Although computing power of individual computer is not significant, a cooperative and distributive work combined with thousands of computers in a grid computing environment results in substantial performance that can be comparable to that of a super computer enabling high-level research or mass data analysis of cutting-edge sciences. Although grid computing has developed due to such a high performance, many security issues have been noted due to the fact that locally distributed computer resources are tied together for sharing and usage. Some of the factors that contribute to such security issues include diversity in computers which include hardware and software, different security policies, various user demands, and security pitfalls that may occur while transferring data.

I will discuss grid computing in terms of security issues, the demand relevant to the security issues, and the fundamentals of the grid security in detail. Also, I will address the future direction with suggestion.

- **Review**

We had a lecture related to Grid Computing about "TCIP (Trustworthy cyber infrastructure for the Power Grid)". Roy Campbell, the speaker pointed out many vulnerabilities and risks of information sharing like Grid system and mentioned that TCIP should provide the fundamental science and technology to create an intelligent, adaptive power grid.

I will address several aspects of Grid computing concept such as what is Grid and what kinds of managerial issues we have as well as any benefits and business risks.

Grid means a type of parallel and distributed system that enables the sharing, selection, and aggregation of geographically distributed "autonomous" resources dynamically at runtime depending on their availability, capability, performance, cost, and users' quality-of-service requirements. It should be noted that Grids aim at exploiting synergies that result from cooperation--ability to share and aggregate distributed computational capabilities and deliver them as service. If you are still fuzzy about Grid concept, think about Napster which is a basic grid computing concept that tremendous PC users share their MP3 files and upload and download their files at the same time without rein.

Ian Foster (picture on the left)*, a University of Chicago professor and pioneer in the grid computing field, provides an example of the technology's application. A personal computer in 2006 is as fast as a supercomputer of 1990. But 15 years ago, biologists were happy to compute a single molecular structure. Now, they want to calculate the structures of complex assemblies of macromolecules and screen thousands of drug candidates. Personal computers now carry up to 100 gigabytes (GB) of storage – as much as the storage capacity of the entire 1990 supercomputer center. To satisfy this kind of computing demand, a new technological concept is necessary other than just a traditional distributed model. That is the grid computing.

Currently, almost 90% capacity of most of the office PCs is not used. Grid computing which is connected to network offers a model for solving massive computational problems pertaining to such work as science and technology research and analysis that requires a tremendous amount of calculation. Such a powerful grid computing could be also used in high tech fields such as

Ian Foster *is a senior scientist and associate division director in Argonne National Laboratory's mathematics and computer science division, a professor in the University of Chicago's department of computer science, and a senior fellow in the Argonne-Chicago Computation Institute.*

biotechnology, environmental engineering and aerospace industry, which are also in need of high computing power. Whenever we need electricity and water, we can use them and we do not need to know which power plants and dams generate electricity and water and what kind of processes are involved. However, for many times, we need to know CPU (Central Processor Unit) speed and memory size of a computer and install appropriate software for a certain job. Using grid computing also requires a certain application and standard security solution to secure implementation, plug-in, and compatibility.

Moreover, most of grid computing technologies are used in academia rather than normal business world such as financial service industry. Like most of IT technologies, grid computing contains lots of difficult technical problems and uncertainties. However, many experts forecast the optimistic future of the grid computing.

- **Evolution**

It seems that grid computing is one of state-of-the art technologies but takes a while to make it prevalent in reality. In the early 1970s when computers were first linked to networks, the idea of harnessing unused CPU cycles was born. A few early experiments with distributed computing - including a pair of programs called Creeper and Reaper - ran on the Internet's predecessor, the ARPAnet.(refer annotation in detail)

Distributed computing scaled to a global level with the maturation of the Internet in the 1990s. Two projects in particular have proven that the concept works extremely well — even better than many experts had anticipated.

The most successful and popular of distributed computing projects in history, is the SETI@home project(Exhibit 1). Over two million people — the largest number of volunteers for any Internet distributed computing project to date — have installed the SETI@home software agent since the project started in May 1999. This project conclusively proved that distributed computing could accelerate computing project results while managing project costs.

The term "grid computing" suggests a computing paradigm similar to an electric power grid-a variety of resources contribute power into a shared "pool" for many consumers to access on an as-needed basis. This ideal is still a few years off. However, there are key efforts underway to define standards that would allow the easy pooling and sharing of all computing resources, including cycles, storage, and data in a way that can promote mass adoption of grid computing.

Today we approach the grid computing from a value-to-customer perspective. Desktop PCs are usually the most plentiful computing resource in a company or organization — and they are consistently the most underutilized. By addressing this area of the computing arena first and completely, we provide both immediate value and provide a foundation for an IT organization to build on as they move forward with their grid infrastructure plans.

- **Emerging Trends**

Why grid computing is important? Because of *Time and Money*. Organizations that depend on access to computational power to advance their business objectives often sacrifice or scale back new projects, design ideas, or innovations due to sheer lack of computational bandwidth. Project demands simply outstrip computational power, even if an organization has significant investments in dedicated computing resources.

Even given the potential financial rewards from additional computational access, many enterprises struggle to balance the need for additional computing resources with the need to control costs. Upgrading and purchasing new hardware is a costly proposition, and with the rate of technology obsolescence, it is eventually a losing one. By better utilizing and distributing existing computing resources, Grid computing will help alleviate this problem.

Many companies want to take advantage of the cost and efficiency benefits that come from a grid infrastructure today, without being locked in to a system that will not grow with their needs.

As a matter of fact, many schools and companies are trying to develop and research grid programming environment, middleware, system, scheduler, monitoring tool and application.

- **Examples and cases of grid computing usage**

Not only SETI@home I mentioned above but also Internet is also one of the distributed severs. I can connect to hundreds million servers from my computer. Actually the servers are distributed each other, Internet leading us to a misunderstanding and makes us feel like using one server and be able to access everything. If I search information on Google, Google collects the appropriate information and show them from distributed servers but connected to Google.

In fact, many projects which utilize grid computing strategy are on the way in world wide. Late last year, World Community Grid (WCG) and IBM launched a new research effort to help battle AIDS using the massive computational power of the World. We can indirectly participate in discovering a new drug which remedies AIDS when we go to the WCG homepage and download a specialized program and install it. Whenever a PC is idle for ten minutes, it is automatically used for the grid. When the owner comes back to work, the PC will retreat and carry out its own work. Thus, there are no troubles.

Moreover, in 2003, a Swiss-based pharmaceutical giant, Novartis, needed a new supercomputer for designing drug. The company found it already had one. It was hidden in the unused computing power the company had available in the thousands of PC's that were already being used in its offices. Novartis used software by United Devices of Austin, Tex., to link 2,700 desktop personal computers to help create drugs. The cost to make this grid system was just $400,000.

National Aeronautics and Space Administration (NASA, USA) and National Science Foundation (NSF, USA) make progress in grid project. Ministry of Education, Culture, Sports, Science and Technology (MEXT, Japan) is also trying to promote four grid projects. In Korea, Korea Institute of Science and Technology Information (KISTI) has on the way of business, ‘Government Grid Constituency Establishment’ since year 2002.

Acute entrepreneurs and investors who understand huge potential power of grid computing already take advantages of grid as means to their business. As we see the real examples, United Devices Inc., headquartered in Austin, Texas, is a technology company that develops and manages the infrastructure required to aggregate idle computation, storage, and bandwidth resources on the Internet or corporate intranets. United Devices delivers this unprecedented power in the form of application services for a number of industries, including Bioinformatics, Web Testing and Monitoring, and Content Search and Management. Intel-United Devices Cancer Project were getting a lot of attention and raised awareness of the robustness and value of PC-based grids like as SETI@home and others mentioned previously. The Meta

Processor (MP) generated by United Devices grid computing solution is a virtual supercomputer that can be harnessed to power computational research and analysis projects on a massive scale and is resources used for cancer medical research.

Entropia is a privately held company headquartered in San Diego, is creating the world's largest Internet and enterprise distributed computing service by converting the idle time of PCs into large-scale computer power via the Internet. Equally by downloading Entropia software from the company's Web site, people and organizations can rescue their computers' otherwise wasted processor time for important projects that significantly accelerate progress and production in the medical, scientific, research and entertainment industries. Interested parties can play a personal role in accelerating AIDS research by running FightAIDS@Home (www.fightaidsathome.org , Exhibit 2). The FightAIDS@Home Project has been of enormous benefit on drug resistance in HIV therapeutics.

DataSynapse Inc., headquartered in New York, offers a business solution, WebProc, a P2P distributed computing application based on the Java API targeted towards financial data processing. WebProc can be configured to use idle processing ability within an intranet, across extranets or over the Internet. The WebProc system works by installing a downloadable low level resource with a footprint of less than one megabyte on each computer in a company's intranet. DataSynapse secures enough finance and energy business cutomer and is one of companies which run active grid business.(Exhibit 3)

Most companies just started their business and study and has not promoted right prospects of grid computing as it is. Due to extravagant technical potential, they might not know how to start advertising.

Many successful grid cases result from lots of projects in the academic area. Large IT companies such as SUN MICOSYSTEMS ・ HP ・ IBM and ORACLE are interested with grid computing. Especially IBM plans to have compatibility among their hardware, software and grid computing. For having wide grid applications, IBM provides an open tool-kit, so-called 'Globus', with customers and communities which use this tool-kit spread out. (Exhibit 4)

- **Issues**

Grid computing is a well-known and rosy paradigm for distributed computing. However, to universalize grid computing, many components should be complemented and secured. Grid experts expect that transferring grid computing environment needs 2~3 years and more at least.

Software license as well as data management matters are obstacles for companies to adopt grid. To extend grid computing technology, the technology needs to be proved to help database and transaction environments.

There are no standard pricing for grid computing service and a lack of application packages for grid. Moreover, grid computing security and trustworthy insurance is surely needed for being willing to share servers and storages which companies and departments own. Actually, the biggest problem is related to people when companies adopt grid system. Many of large enterprise companies have not tended to share their hardware resources and data. Wolfgang Gentzsch, an executive of non-profit organization, MCNC, which establishes grid system for government, university and other organizations, said that "People do not want to share. He had been a senior director of Grid computing for Sun Microsystems since July 2000 and join MCNC last year.

- **Conclusion**

IBM's Globus tool kit for using grid system and other grid technical recommendations aim to address computing tasks suited to the business realm rather than academia, where computing grids have been used for years. Perhaps more significantly, these efforts seek to establish industry wide grid standards that industry experts say are still lacking. Many grid technologies are still underway in university laboratories.

The grid computing industry today is roughly at the same stage the Internet was about 10 years ago, experts say. Before commercial customers can share their computing resources more effectively across widespread networks, they need a wide variety of standardized products. Many parts of grid computing are in the stage of development where business customers build grid, they do not buy it, because these are all separate tools. It is commercially viable, but it still has to put a lot of things together. Today, most examples of grid computing are done using vendor-specific tools within a single company.

Even though there still exist many obstacles to get over and are tasks to commercialize, many companies acknowledge that grid computing has a titanic potential and has shown many feasibilities. Recently, various business areas attempt to take grid computing in their business and in fact, they bring about successful results.

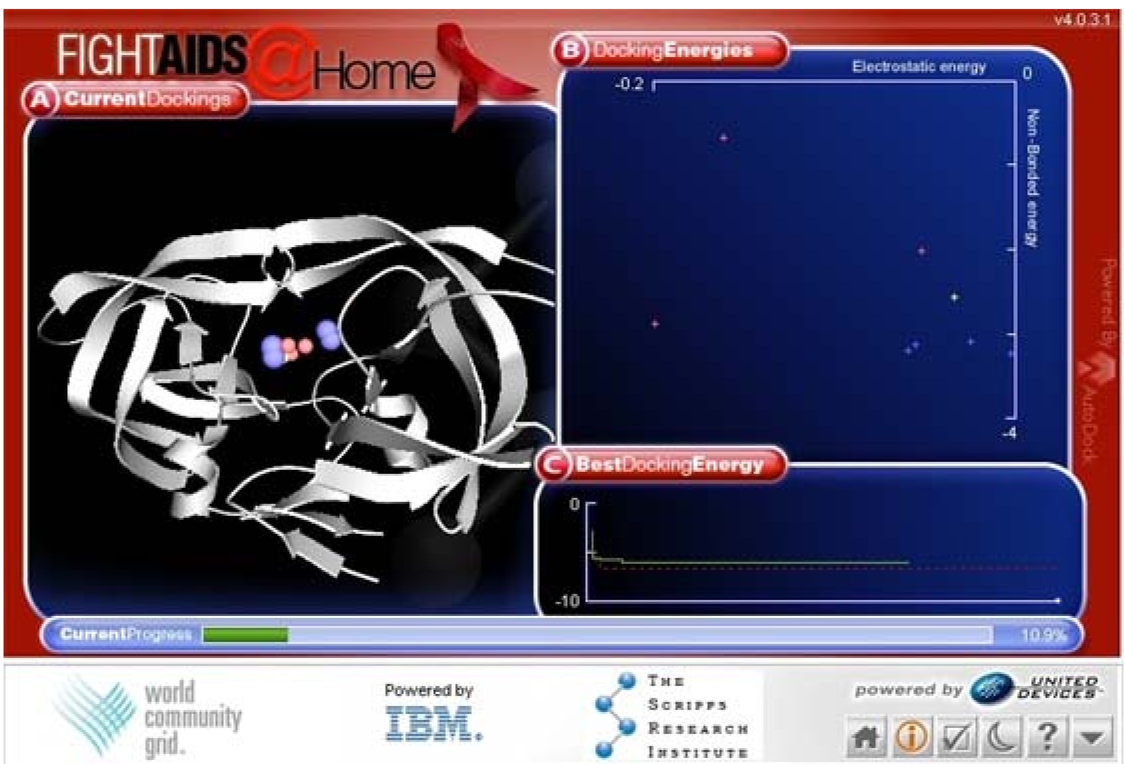
In conclusion, grid computing is not enough mature technology. For this reason, the challenge does not have to stay in laboratories and it is getting the ideas out of the lab and into commercial use. Thus, the balance between using and developing the technology can have influence on technology improvement.

<Exhibit 1> Arecibo Radio Telescope in Puerto Rico (the world's largest radio telescope)

SETI@home launched in May 1999 to search through signals collected by Arecibo Radio Telescope

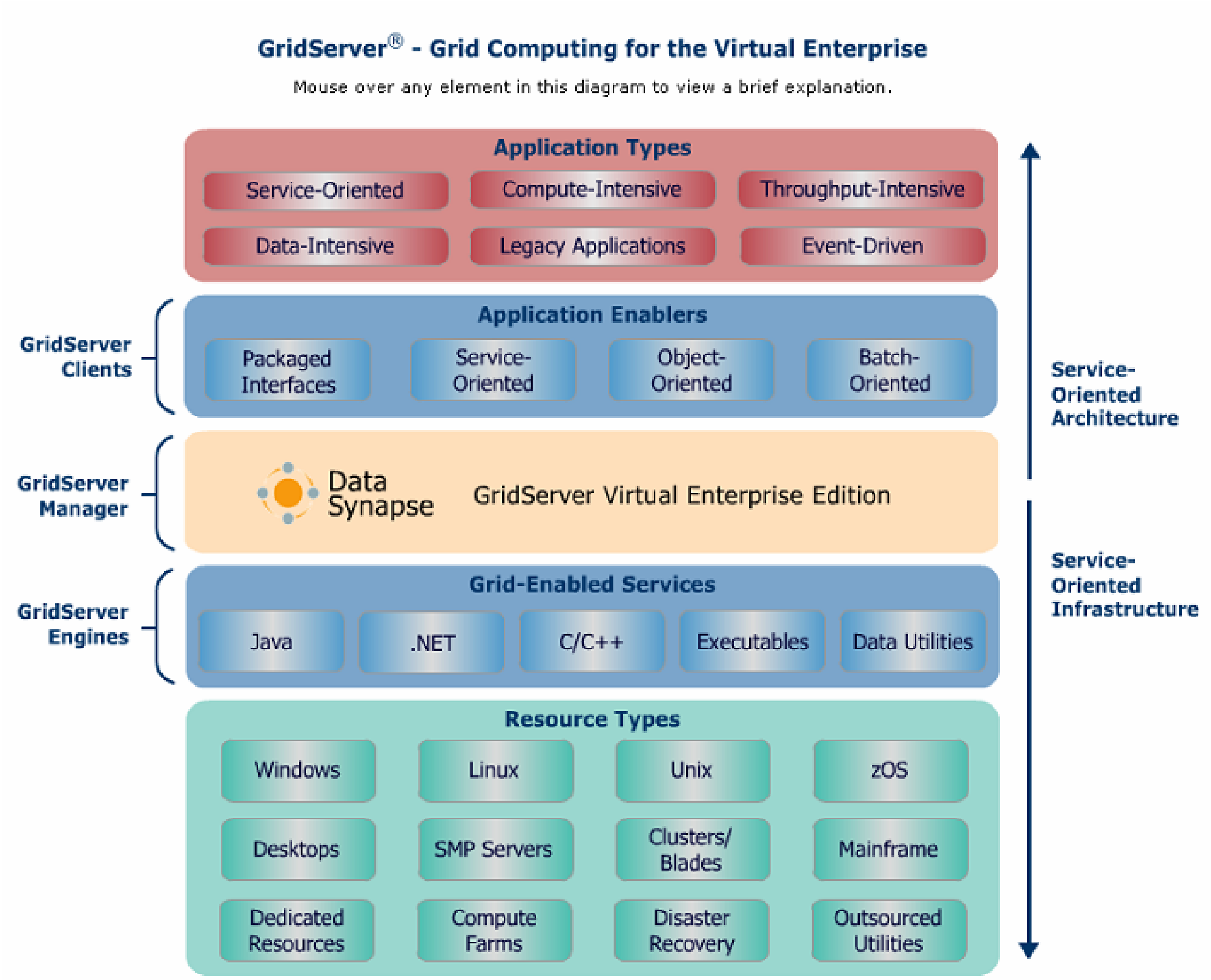
Source : SETI@home homepage - http://setiathome.ssl.berkeley.edu/

<Exhibit 2> Agent application



Source : FightAIDS@home homepage - http://www.worldcommunitygrid.org/

<Exhibit 3> Data Sysnapse GridServer Architecture Model



Source : http://www.datasynapse.com/

<Exhibit 4> Grid Computing model with Globus toolkit

Source : http://www.dlr.de

- **Annotated References**

http://www.nic.funet.fi/index/FUNET/history/internet/en/arpanet.html (ALPARNET VIDEO CLIP & HISTORY)

http://www.globus.org/toolkit/

http://www.grid.org

http://www.gridtoday.com

http://www.gridcomputing.com

http://www.gridforum.org/

http://www-1.ibm.com/grid/

http://www.nytimes.com/

http://www.ud.com/

http://fightaidsathome.scripps.edu/

http://www.businesswire.com/

http://www.gridworldhome.com/

http://setiathome.ssl.berkeley.edu/

http://www.worldcommunitygrid.org/