

SNF-center for Grid Computing

1 Introduction

Grid computing is based on the idea of joining individual computers and clusters of computers and organizing them into a single logical entity with a common interface. This interface will act as a meta-computer offering, for example, uniform access control and resource locator services to the user applications. By using these services, applications can be developed and tested on local machines and subsequently submitted to the meta-computer without modifications when a significant increase in computer resources are needed for the project.

Grid-based systems fall into two basic categories, depending on the requirements of the applications. Those that exploit the availability of large quantities of computing power distributed over a network are usually denoted *Computational Grids*, while those that focus on accessing and displaying large quantities of information, typically to a scientific or business community, are denoted *Access Grids*. These two types of Grid systems complement one another in giving their users access to globally available information. Both types of Grids will be covered by the proposed center.

Many Danish research groups have created local computer clusters, ranging in size from a few to several hundred machines. Each cluster is equipped with disk systems and other expensive storage elements. Because isolated computer clusters are designed to match peak performance rather than the average load, these systems often have an unused spare capacity. By introducing a cost-sharing model like the Grid, each research group could have access to all the IT-capacity needed, for an investment only slightly higher than the cost of its average consumption. At the same time new investments can be made at a slower rate and "just in time", as long as the total resource on the Grid matches the peak load.

In this project we envision building an inter-departmental Grid research center that will continue and coordinate the Danish research within the field of Grid computing. It will facilitate sharing by interconnecting clusters of computers in a Grid structure thus participating in international efforts to build worldwide Grids. The center will establish a group of *Cluster Collaborators* who will add their clusters to the Grid.

This project aims to make a significant research contribution to the development of the next generation of computing infrastructure that will provide increased availability and utilization of computing resources in a secure manner.

To enable the experimental verification of our research in a realistic Grid environment, the center will establish a group of *Core Collaborators*, which will consist of research groups who are willing to participate in the development and use of experimental Grid software.

2 Research Areas

Large-scale Grid deployment are currently being initiated in most countries in Europe and in North America, amongst the most ambitious is the British e-Science Programme[1] with a funding in excess of 150M euro. In addition to the national initiatives, international projects are currently being structured, most significantly the EGEE project in Europe, which seeks to include more than 100 groups from all European countries. At the Nordic level NOS-N has established the Nordic DataGRID Facility, which the proposed center will collaborate closely with.

Grid research is a very active field, however, much of what is being developed is dominated by an ad-hoc approach to solving very concrete problems. Amongst the serious Grid-software projects are Globus[3], Legion[4] and NorduGrid[5]. The following research focal points represent areas in which we expect to make significant contributions to fundamental and still unsolved Grid related issues. The research findings will be incorporated into the NorduGrid software-package.

Previous research from the group of applicants has proved to provide significant contributions to the field. A detailed listing of recent activities can be found in the appendix.

2.1 Security Models and Authorization

The importance of security in large meta-computing systems cannot be emphasized too strongly, as it is essential that organizations offering computing resources can be given an assurance that these resources will not be misused. In the proposed project, attention will be focused on issues of authentication, confidentiality, and protection from risks posed by ill-formed programs.

In order to guarantee that the application of Grid systems can be protected from attacks and misuse, a proven security model has to be defined, and also efficient mechanisms for runtime checking of potential threats. The applicants have extensive experience with these subjects for distributed systems in general, and additional work is needed in order to extend it to Grid systems. In particular, it becomes critical to handle key management in a scalable manner and also to define a flexible way for allowing execution of trusted applications without introducing additional performance penalty caused by security checks.

The development of a suitable trust model for use in global scale computer systems also merits attention. Current trust models are almost all based on pre-knowledge of the subjects who will be attempting to use the system. For systems on a global scale, this is not always a satisfactory approach, and models based on the incremental development of trust, incorporating reputation, recommendation and past experience will be investigated.

2.2 Data Management on Grid

The fact that latency in communication networks cannot be improved (because of the limitations due to the speed of light) is troubling, because processors in the time frame 2004-2008 will increase performance by close to an order of magnitude, and the number of CPU-cycles that are wasted while waiting for data will grow proportionally. The principal method for reducing the idle-time, and thus boosting performance, in computer systems is to store copies of vital information close to the processor, via the techniques of caching and replication.

Caching and replication have been studied extensively by the on-line algorithms community, using the terminology paging, distributed paging, distributed data management and page replication. The standard measure for on-line algorithms, the competitive ratio, has proven to be inadequate at differentiating between paging algorithms, so there have been several attempts at finding other measures. Recently, Boyar and Favrholdt have defined a promising new measure, the relative worst order ratio, which they and Larsen have used to show the effectiveness of look-ahead with respect to paging. Further study of this ratio with respect to distributed paging and page replication is expected to produce new algorithms and new performance guarantees concerning known algorithms.

Replication is related to caching, but differs enough in its nature to present an autonomous problem area. Replication has been used for close to a decade in the World Wide Web model but only in a rough approach. Unsolved problems within replication management include division of the available replica space between the potential replica targets and location of the nearest replica of a given dataset in an unordered network such as the Grid model. It will be important to consider some aspects of this in the current project.

Systems science aspects of caching and replication will be emphasized heavily in this project. This includes establishing a set of constants that the formal models will need to be usable in a production Grid and investigating the problems of security barriers that currently limit the growth of the Grid.

Finally, we will make several application investigations in order to establish a set of typical behaviors that the final caching and replication models should handle well.

2.3 Scheduling

Based on the characteristics of its current collection of application jobs and its available resources, a Grid must maintain a schedule for job execution, which on the one hand satisfies the resource requirements of the jobs and on the other hand optimizes system performance. Even in a traditional static resource environment, this objective has been shown to be computationally intractable, and it is even hard to find a reasonable sub-optimal, but satisfactory, solution. In a heterogeneous, multilevel Grid environment involving both logical and physical distribution and diverse networks, the problem clearly becomes even harder because of the dynamics and the scale of the system. Hence, there is a need for an extension of classical models and adaptive algorithms of the scheduling problem. The project will focus on this issue by experimenting with different multilevel scheduling prototypes based on (at least) the following levels:

- The local cluster level, where each individual cluster applies its own preferred strategy for local optimizations
- The global level, where a Grid adapts its current job schedule to the resource dynamics (load, availability, cost, price, etc.).
- The planning level, where a Grid decides whether or not a submitted job can be executed according to its needs. This involves negotiation with the other levels.

The planning level has not been investigated much in Grid-like contexts. A major issue will be the coordination between the different levels.

Recently, the complexity of large scale peer to peer systems has been attacked through fault-tolerant, self-organizing, self-repairing overlay networks. Where the peer to peer community has focused on routing [6, 8] and storage systems [7], we wish to apply overlay networks to the area of resource discovery and scheduling in Grid systems.

2.4 Monitoring and Resource Management

Monitoring and management of available resources on a Grid are prerequisites for controlling a number of basic services such as accounting, surveillance and scheduling. However, there are a number of aspects that are unique to Grid environments:

Due to its highly distributed nature and sheer size, the collection of resources forming a Grid is constantly changing. Therefore in order to maintain an up-to-date snapshot, the monitoring system must support mechanisms for failure detection and recovery. Moreover, due to the running time of most applications, program analysis must be supported at run-time, in contrast to existing systems, where post-mortem analysis is the typical approach. Hence there is a need for interactive exchange of monitoring information on a much larger scale than traditionally seen. Finally, it is necessary to coordinate information from a (heterogeneous) collection of monitoring systems and to make this information available on the Grid in a uniform way.

Compared to traditional distributed systems, the resource allocation part of a Grid is confronted with the complication that, even if resources are granted to a job, these resources may be withdrawn or disappear before the job actually uses them. In traditional systems, this would be an exception or even a failure, whereas we consider it as a normal and common event in a Grid.

Thus the monitoring and resource management subsystems:

- Maintain knowledge about the identity and quantity of available resources based on information obtained from the applications and from monitoring of the system.
- Maintain knowledge about the application jobs requested and jobs currently being executed.

- Maintain knowledge about the current state of the Grid.
- Maintain error statistics.
- Monitor the state of system services and save this information in system logs.
- Gather and analyze accounting (value and cost) information.

Execution monitoring related to the above aspects will form part of the project, and experiments with selected services and programming paradigms will be made. The project will investigate existing proposals both theoretically and through experiments with emphasis on scalability and versatility, and based on this investigation develop techniques adequate for the above mentioned challenges.

3 Education and Dissemination Aspects

An important aspect of the center is support of a number of Ph.D. research project as well as Masters projects.

The budget includes partial funding (50%) of four Ph.D. students—the participating universities are expected to provide the other half of each Ph.D. stipend. We already have firm commitments from most of the involved universities which assuming that we present well qualified candidates pledges support supplementing each of these stipends so that there will be a total of four full stipends. These pledges are included in the letters of support in the appendix.

We will strengthen our Ph.D. effort by applying for a Ph.D. school in Grid computing. Typically, the students hired as research programmers will be Computer Science Masters students.

We have started a Danish Grid Forum initiative (see <http://www.lebox.dk/grid>) which will create an industrial Grid forum for the exchange of ideas and for disseminating research results and information on using Grid. To get the Grid Forum started we will arrange the first seminar in September 2003.

4 Operation – The Danish Production Grid

An even better cost-benefit ratio can be obtained by sharing resources internationally, as already demonstrated with the NorduGRID production Grid [5]. The bandwidth between nodes sets a limit for the system performance. NORDUNet and GEANT have plans for the development of the Nordic and European networks which fulfill most of the requirements from known academic users, even those planed by astronomy and particle physics and later by bio-sciences and earth observation. With the new developments of efficient net-caching and replication all computers will be seen as one large resource.

We propose to organize the resource sharing at the national level as a Danish production Grid, providing access to extensive computer resources for Danish researchers over the Danish Research Network. The Danish production Grid Center will be responsible for the organization and day-by-day operation. The center will be located at the Niels Bohr Institute – University of Copenhagen, together with the KU-Grid center. Among the center's highest priorities is to

1. construct and maintain a PC-cluster, sufficiently versatile to become the natural entrance point for new users of Grid technology. The cluster has a capacity big enough to be the buffer that smoothens the work load on the production Grid
2. run a 5 days by 8 hours help desk. It is assumed that large research groups relatively quickly will acquire sufficient experience to solve the most common problems on their own. Each major university will provide help for the single users and only if the local experience becomes

insufficient the problem is forwarded to the help desk. At Copenhagen University the help desk and the local expert will be placed in the same section of offices, together with visitors to the center

3. provide access to test facilities where Danish IT-researchers and students can test new ideas for Grid and associated technologies. The test benches shall be made available for developers working for Danish industry
4. become the natural host for developers who want to contribute to the Grid. Office space and access to test clusters will be available for short and long term visitors.
5. be the authentication authority for the Danish Academic Grid Virtual Organization.
6. be the international contact organization for Grid and negotiate authentication with virtual organizations outside Denmark

5 Collaborators

Besides the international collaboration mentioned above, the center will establish the following collaborator groups:

Core Collaborators who are research groups that are willing to participate in development and use of experimental Grid software.

Cluster Collaborators who are research groups that have clusters that they are willing to add to the Grid.

Danish Grid Forum which is an industrial outreach initiative that the center will foster. The Forum is open to industrial contacts that wish to participate in the exchange of ideas concerning the use of Grid technology. The center will help establish the Forum by arranging an initial seminar open to all.

6 Organization

The center has a management with a member from each of the participating research groups, Eric Jul is chairman.

The four computer science research groups at AAU, DTU, SDU, and KU will participate on an equal footing. A small research group at the University of Aarhus will participate via the research group at SDU.

The management appoints members of an advisory board from the core collaborators' group. The advisory board collaborates with the management on defining research directions and experiments.

7 Initial Project Milestones and Events

The following milestones and recurring events have been defined:

- **Sept 03:** Introduction seminar for *Danish Grid Forum*. The purpose of the Danish Grid Forum is to create an organization where stakeholders in Grid technology can meet and discuss the development.
- **Oct 03:** DK-Grid experimental cluster operational.
- **Dec 03:** 5 Grid sites operational on DK-Grid.
- **Jan 04:** Identify and analyze critical areas of two core collaborators' applications.

- **Feb 04:** Helpdesk operational.
- **March 04:** Submission for *Supercomputing 04* of paper on the results of the analysis of applications.
- **June 04:** Demonstration of Grid integrated applications.
- **Aug 04:** 5 additional sites on DK-Grid.

A public dissemination seminar will be held annually where the group will present recent research results and software development. Further, an annual workshop will be held where ongoing work will be discussed.

Further milestones will be defined throughout the project by the management in collaboration with the advisory board.

References

- [1] *The e-Science Core Programme*, <http://www.escience-grid.org.uk/>
- [2] *The Grid—Blueprint for a New Computing Infrastructure*, Ed. I. Foster, C. Kesselman, Morgan Kaufmann, 1998
- [3] *The Globus Project*, <http://www.globus.org>
- [4] *Legion*, <http://legion.virginia.edu>
- [5] *NorduGRID*, <http://www.nordugrid.org>
- [6] Antony Rowstron and Peter Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, pages 329–350, November 2001.
- [7] Antony Rowstron and Peter Druschel. Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility. In *18th ACM Symposium on Operating Systems Principles*, pages 188–201, October 2001.
- [8] B. Y. Zhao, J. D. Kubiatowicz, and A. D. Joseph. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report UCB/CSD-01-1141, UC Berkeley, April 2001.

8 Budget by Year

Expenses are distributed uniformly over the project period.

<i>Item</i>	<i>Note</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>Total</i>
Salaries						
1 post doc	1	200.000	480.000	480.000	280.000	1.440.000
4x1/2 Ph.D.		347.222	833.333	833.333	486.111	2.500.000
Programmers		122.917	295.000	295.000	172.083	885.000
Equipment						
Hardware/operation	2	125.000	300.000	300.000	175.000	900.000
Operational expenses						
Ph.d.-workshop		20.833	50.000	50.000	29.167	150.000
Travel expenses		52.083	125.000	125.000	72.917	375.000
Total before overhead		868.056	2.083.333	2.083.333	1.215.278	6.250.000
Overhead (20%)						1.250.000
Total						7.500.000

Notes:

1. A post. doc. with qualifications corresponding to a Ph.D. in computer science (i.e., including “fagligt tillæg”)
2. Establishing operations center, maintaining existing clusters and integrating existing applications. (Expenses will cover establishing a center cluster the first year, and its maintenance the following.)
3. The four half Ph.D. stipends will be used to secure matching funding from the participating institutions and/or from industry.