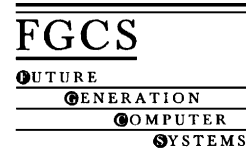




ELSEVIER

Future Generation Computer Systems 18 (2002) 1143–1153



www.elsevier.com/locate/future

GridLab—a grid application toolkit and testbed

Ed Seidel^a, Gabrielle Allen^a, André Merzky^b, Jarek Nabrzyski^{c,*}

^a *Albert Einstein Institut, Potsdam, Germany*

^b *Zuse-Institut, Berlin, Germany*

^c *Poznań Supercomputing and Networking Center, ul. Noskowskiego 12/14, 61-704 Poznan, Poland*

Abstract

In this paper we present the new project called GridLab which is funded by the European Commission under the Fifth Framework Programme. The GridLab project, made up of computer scientists, astrophysicists and other scientists from various application areas, will develop and implement the grid application toolkit (GAT) together with a set of services to enable easy and efficient use of Grid resources in a real and production grid environment. GAT will provide core, easy to use functionality through a carefully constructed set of generic higher level grid APIs through which an application will be able to call the grid services laying beneath in order to perform efficiently in the Grid environment using various, dramatically wild application scenarios.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Grid computing; GridLab; GAT; Testbed

1. Introduction and summary

Grid computing is an exciting buzzword in the computing world today. Here we define it to mean the exploitation of a varied set of networked computing resources, including large or small computers, PDAs, file servers and graphic devices. The networks could be anything from high speed ATM to wireless or modem connections. Exploiting these connected resources could, for example, enable large scale simulations not possible on a single supercomputer, aid computational work of geographically distributed collaborations, simplify remote use of machines, and

enable the new dynamic application scenarios we propose.

While computational Grids are becoming increasingly common, promising ultimately to be ubiquitous and thereby change the way global resources are accessed and used, presently there is a dearth of real Grid users. This is partly because the whole concept is new, but also because there are few applications that can exploit Grid resources. Although some application developers are interested in writing Grid-enabled applications, there are few user level tools, while application developer tools are nonexistent and well understood Grid usage scenarios are lacking.

It is therefore imperative to attract real users into the Grid community (for example the Global Grid Forum, GGF) and ultimately onto the Grid. The Applications Research Group (ARG) of the former European Grid Forum (EGrid) has been addressing questions about user requirements, problems, and usage scenarios for several years now. In November

* Corresponding author. Tel.: +48-61-858-2072;
fax: +48-61-852-5954.

E-mail addresses: eseidel@aei-potsdam.mpg.de (E. Seidel), allen@aei-potsdam.mpg.de (G. Allen), merzky@zib.de (A. Merzky), naber@man.poznan.pl (J. Nabrzyski).

2000 the group established a pan-European testbed [1], based on the Globus toolkit, for prototyping and experimenting with various application scenarios. These testbed experiences inspired some members of the research group to submit a successful proposal for an application orientated project, called *GridLab* to the European Commission. This paper describes the motivation and architecture of the GridLab project.

The primary aim of GridLab is to provide users and application developers with a simple and robust environment enabling them to produce applications that can exploit the full power and possibilities of the Grid. The GridLab project brings together computer scientists with computational scientists from various application areas to design and implement a grid application toolkit (GAT), together with a set of Grid services, in a production grid environment. The GAT will provide functionality through a carefully constructed set of generic high-level APIs, through which an application will be able to call the grid services laying below. The project will demonstrate the benefits of the GAT by developing and implementing real application scenarios, illustrating wild, exciting, new uses of the Grid.

GridLab is a balanced program with co-development of a range of Grid applications alongside infrastructure development, working on transatlantic testbeds of varied supercomputers and clusters. This practical approach ensures that the developed software truly enables easy and efficient use of Grid resources in a real user environment, tested by several closely related user communities. We will maintain and upgrade the EU testbeds through deployment of new infrastructure and application technologies as they are developed. All deliverables are immediately prototyped and field tested under realistic conditions by existing application communities, on large scale applications, throughout the project.

We will develop core capabilities for simulation and visualization codes to be self-aware of the changing Grid environment, and to be able to fully exploit dynamic resources for fundamentally new and innovative applications scenarios. The applications themselves will possess the capability to migrate from site to site during the execution, both in whole or in part, to spawn related tasks, and to acquire additional resources as needed, according to both the changing availabilities of various resources in the grid, and the needs of the applications themselves. We will design and develop

the GAT to provide core, easy to use functionality through a carefully constructed set of generic APIs for both simulation codes and Grid software. This toolkit will contain independent modules for handling many different aspects of Grid programming, including simulation, performance and grid monitoring, resource brokering and selecting, simulation, performance prediction, interacting with information servers, security, notification, collaborative working, data handling, remote visualization, and remote application steering. We will also enhance real applications for the Grid, implementing new dynamic simulations with the GAT.

Two important aspects of Grid technology, which have been largely ignored, form the basis of our *GridLab* project, which aims to build components for Grid applications (as MatLab does for mathematics), and realistic testbeds for their development:

1. *Co-development of infrastructure and applications.* We propose a balanced program with *co-development* of a range of Grid applications (based on *Cactus* [3], the leading, widely used Grid-enabled open source application framework, and *Triana* [22], a dataflow framework used in gravitational wave research) *alongside* infrastructure development, working on transatlantic testbeds of varied supercomputers and clusters. This practical approach ensures that the developed software truly enables easy and efficient *use* of Grid resources in a real environment. We will maintain and upgrade the testbeds through deployment of new infrastructure and large scale application technologies as they are developed. All deliverables will be immediately prototyped and continuously field tested by several user communities [2,10,18]. Our focus on specific application frameworks allows us *immediately* to create *working Grid applications* to gain experience for more generic components developed during the project. Although *Cactus* and *Triana* are powerful example application frameworks, other applications will be able to take advantage of GridLab technologies as well.
2. *Dynamic grid computing.* We will develop capabilities for simulation and visualization codes to be self-aware of the *changing* Grid environment, and to be able to fully exploit dynamic resources

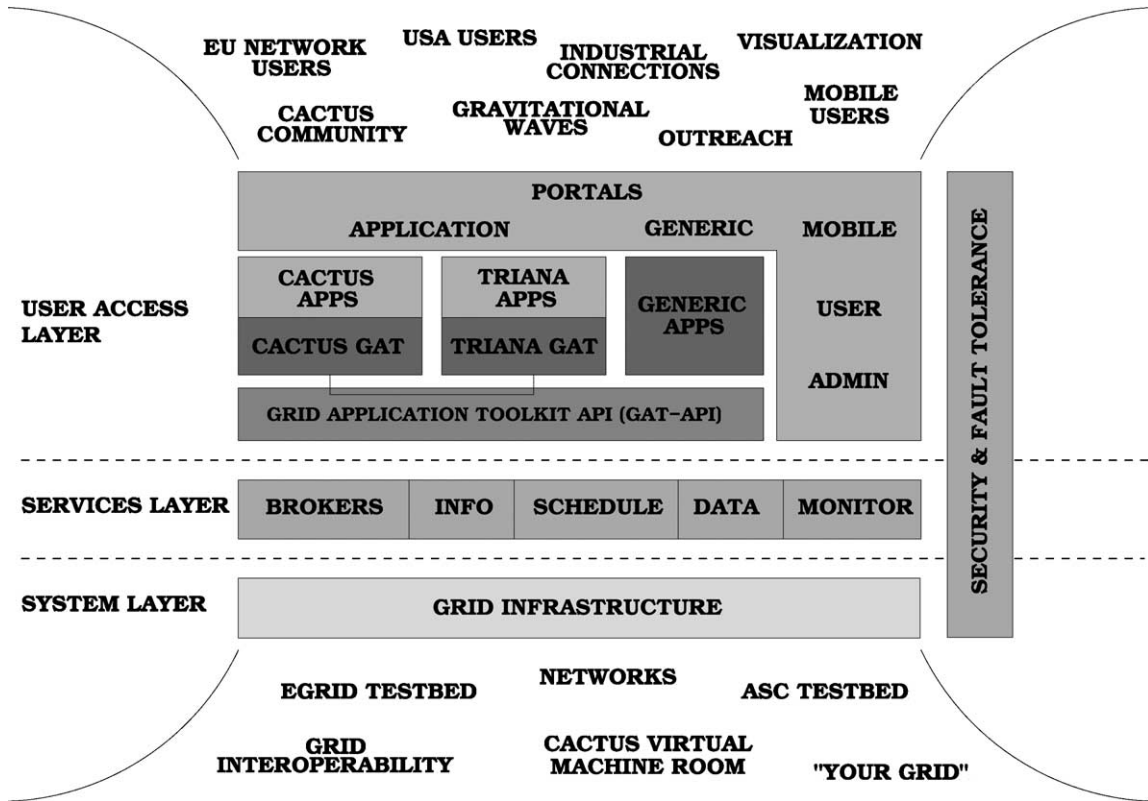


Fig. 1. The GridLab architecture provides key components needed to support Grid applications, uniting diverse application user and development communities, computational scientists and Grid infrastructure developers, and Grid resource providers, from testbeds to productive computing environments in USA and Europe.

for fundamentally new and innovative applications scenarios. For example, the applications themselves will possess the capability to migrate from site to site during the execution, both in whole or in part, to spawn related tasks, and acquire/release additional resources demanded by both the changing availabilities of Grid resources, and the needs of the applications themselves.

We will unify these elements in developing innovative, practical, Grid computing technologies, which will then be quickly and easily adopted and exploited by applications from many different research and engineering fields, as shown in Fig. 1. Specific key objectives of our project are given below.

- *Design and develop a GAT*, to provide core, easy to use functionality through a carefully constructed set of generic APIs for both simulation codes and

Grid software. The GAT will contain independent modules for handling many different aspects of Grid programming, including simulation, performance and grid monitoring, resource brokering and selecting, performance prediction, interaction with information servers, security, notification, collaboration, data handling, remote visualization, and remote application steering.

- *Simultaneously enhance real applications for the Grid*, implementing new dynamic simulation scenarios using the GAT. Both Cactus and Triana and other applications will be extended to integrate and exploit GAT elements, making Grid computing easily exploitable by a wide range of applications. Our simulation driven, compute intensive applications complement the highly data driven applications in many other Grid projects (e.g. DataGrid [5], GriPhyN [17], EuroGrid [9]).

- *Develop and test Grid infrastructure/applications on real testbeds*, constructed by linking heterogeneous collections of supercomputers and other resources spanning Europe and USA, using and extending existing testbeds [1,6]. Interoperability with different testbeds will be ensured by also using production testbeds in USA [2,13,24], driving international high speed network connectivity. Testing will be carried out by the project and several large, closely related user communities, including an EU Astrophysics Network [10], and various multidisciplinary US funded collaborations.

1.1. User Scenarios

The end technology developed through this project will enable scenarios, such as the following hypothetical examples, to become reality.

1.1.1. Gravitational wave detection and analysis

The gravitational wave detector network, including GEO600 in Germany, collects a TByte of data each day, which must be searched using different algorithms for possible events such as black hole or neutron star collisions, or pulsar signals.

Routine realtime analysis of gravitational wave data from the Hannover detector identifies a burst event, but this standard analysis reveals no information about the burst location. To obtain the location, desperately required by astrophysicists for turning their telescopes to view the event before it fades, a large series of templates must be cross-correlated against the detector data. An Italian astrophysicist accesses the **GEO600 portal**, and using the **performance** tool finds that three TFlops/s is needed to analyze the 100 GB of raw data in the required hour. Local resources are insufficient, so using the **brokering** tool, she locates the fastest available machines around the world. She selects five suitable machines, and with **scheduling** and **data management** tools, data is moved, executables created and the analysis starts. In an Amsterdam bar 20 min later, an SMS message from the portal's **notification** tool, informs her that one machine is overloaded, breaking the **runtime contract**. She connects with her PDA to the portal, and instructs the **migration** tool to move this part of the analysis to a different machine. Within the specified hour, a second SMS message tells her analysis is finished, and the resulting data is now on her local machine. Using this location data, observatories are able to find and view an exceptionally strong gamma-ray burst, characteristic of a collision of neutron stars.

1.1.2. Numerical Relativity

A single simulation of an astrophysical event, e.g. black hole or neutron star collisions, ideally requires over TByte and TFlop resources, not yet available on a single machine.

The GridLab project is made up of computer scientists, astrophysicists and other scientists from various

application areas. The project participants are listed in [Table 1](#).

2. Innovation

Grids promise—but have not yet delivered—application users the ability to harness global computing resources, providing easy access to significantly more resources, improved collaborative environments, dynamic interaction with simulations, and global mobile access. Our GridLab project will significantly advance the current state-of-the-art by developing (i) key components necessary for application oriented Grid computing (resource estimators and brokers, platform independent portals accessible even from mobile devices, security infrastructure, monitoring tools, etc.), (ii) interfaces to functionally similar components developed by others, (iii) a GAT, for both infrastructure and applications, enabling new generations of Grid-enabled applications, and (iv) innovative new Grid computing scenarios to dramatically increase the scale or throughput of possible applications. All components will be tightly integrated and built on present state-of-the-art infrastructure and

application tools, coordinated and co-developed with other leading international Grid projects and industrial partners, field tested by real application communities on inter-continental testbeds and production environments. The overall architecture of our project is summarized in [Fig. 1](#).

Learning more about the detected burst requires cross-correlating detector data with custom wave templates from full-scale neutron star simulations. Sufficiently accurate templates require running large scale simulations too big to fit on any current supercomputer. German members of an international numerical relativity collaboration are tasked with creating collision templates for 10 different neutron star mass combinations. They access the web-based **Simulation Portal**, selecting required code modules and building parameter files with the **code composition** tool. The **performance prediction** tool estimates that each simulation requires 1024 GB of memory and 10^{14} Flops, with an additional 50^{14} Flops required for processing data to create signal templates. The **brokering** tool finds that no single machine in the **simulation testbed** can supply enough memory, but locates two machines which can be connected to form a large enough virtual supercomputer, the dynamic grid **monitoring** tool indicates an acceptable bandwidth between them. The **scheduling** tool stages the five runs to appropriate queues on the machines, and the first simulation starts. The time-consuming task of creating templates is handled by **spawning simulations** to smaller machines dynamically located by the **broker**, at each time step data is streamed to a series of networked computers for analysis, creating a **simulation vector** using available machines on the grid. Collaborators around the world connect to the portal using networked workstations, home PCs and modems, as well as the latest wireless PDAs and mobile phones. They are able to use various **remote access** tools to **visualize** data, **monitor** performance and simulation properties, and **interactively steer** the simulation.

Table 1
List of participants

| Participant no. | Participant name | Participant short name | Country |
|-----------------|---|------------------------|-----------------|
| 1 | Poznan Supercomputing and Networking Center | PSNC | Poland |
| 2 | Max-Planck-Gesellschaft zur Foerderung der Wissenschaften e.V. | MPG | Germany |
| 3 | Zuse Institut Berlin | ZIB | Germany |
| 4 | Masarykova univerzita v Brne (Masaryk University Brno) | MU | Czech Republic |
| 5 | Computer and Automation Research Institute, Hungarian Academy of Sciences | MTA SZTAKI | Hungary |
| 6 | Vrije Universiteit Amsterdam | VU | The Netherlands |
| 7 | Universita' Degli Studi di Lecce | ISUFI | Italy |
| 8 | University of Wales, Cardiff | CARDIFF | UK |
| 9 | Sun Microsystems Gridware GmbH | GRIDWARE | Germany |
| 10 | Compaq Computer EMEA BV | COMPAQ | France |
| 11 | Institute of Communication and Computer Systems (ICCS)/National Technical University of Athens (NTUA) | ICSS/NTUA | Greece |
| 12 | Argonne National Laboratory | ANL | USA |
| 13 | Information Sciences Institute | ISI | USA |
| 14 | University of Wisconsin | UoW | USA |

2.1. Current state-of-the-art

Much Grid research was previously focused in USA where it has been developed over the last decade. Leading Grid infrastructure systems such as the Globus Toolkit [12] and Condor [4] are being extended and used in research projects such as GrADS [13] and GriPhyN [17]. The ASC project [2] is developing a leading edge portal [19] for Grid access. Companies

such as Entropia [8] are beginning to exploit unused cycles for both commercial and humanitarian ends.

Grid computing in Europe was recently boosted with the creation of the EGrid [1,7], which last year became part of the newly formed GGF [11]. Grid projects in Europe include UNICORE/UNICORE PLUS [23], the recent DATAGRID [5] and EUROGRID [9], the DFN-Verein TIKSL/GriKSL projects [16,21], along with commercial ventures such as GRIDWARE (now

Sun's Grid Engine [14]). The Cactus computational toolkit [3] is a collaborative simulation framework used in various application communities worldwide, and in many Grid development projects in US and Europe in prototyping new generations of Grid applications. The projects are described below.

However, Grid computing still has to arrive for most users. Problems, such as lacking standards, security issues and difficulty of use, have prevented most applications from utilizing the Grid, leading to a growing community of Grid developers without input from the community they seek to enable. GridLab addresses this by providing a functional development and testing environment, uniting the Grid community with application users and developers through collaborative Grid technologies like Cactus.

2.2. GridLab extends existing technologies

We seek to develop generic and innovative solutions for Grid computing by starting with, and extending, the most advanced and widely used Grid infrastructure and application toolkits. The Globus Toolkit is the standard grid middleware, providing tools for, e.g. security, job submission and information services, and is associated with many large collaborative projects; The Cactus code and computational toolkit is a modular, collaborative framework for developing parallel scientific and engineering applications which is already widely used for Grid computing.

Both Globus and Cactus emerged from experiments during the I-Way at SC95 [20], a pivotal moment in Grid history, and have driven Grid development for half a decade. We aim to *more tightly couple and extend* these projects in ways that *enable applications to be self-aware, of both their own varying computational needs and the changing Grid environment, and respond and adapt appropriately*. As shown in Fig. 1, we will achieve this by using, extending, and building components, such as brokers, information servers, monitoring systems, that work with the applications themselves through a GAT, connecting to basic Grid infrastructure software and hardware underneath. Application communities will have portals to access application and Grid information from any device.

In summary, the aims of this project are to develop exciting new and innovative uses of the Grid which directly address the needs of the application commu-

nity and, importantly, because we begin with specific applications already in use, to make them immediately usable by the community, by:

- Developing a Grid Laboratory for (i) developers to grid enable their applications, employing the best available tools through a high level, flexible GAT; (ii) users to deploy these applications through a simulation portal on different resources and testbeds, visualizing, interacting with, and analyzing their simulations both in the work and mobile environments; (iii) developers to easily include new Grid tools for testing and production use.
- Enhancing a variety of existing, resource intensive, applications with the capabilities of the GAT, including Triana and applications already using Cactus. The applications will be tested and deployed on different testbeds to ensure grid interoperability.
- Building communities through collaborative Grid technology. Cactus was designed from the beginning to enable collaborative work between application and computer science communities. Our portal will enable geographically distributed communities to jointly interact with and analyze results of their Grid simulations. Deployment across different transatlantic testbeds encourages joint development work crossing different project, discipline, and national boundaries. These technologies/testbeds will act as catalysts new organizational and collaborative practices in these communities.
- Disseminating results and distributing software through (i) our application communities, (ii) the gluing together of different US and EU research projects, (iii) our active participation in the GGF, (iv) our close association with leading computing centers worldwide, and (v) our partnerships with leading computer vendors, who will bring this technology directly to their industrial customers.

2.3. Specific issues addressed by GridLab

2.3.1. Co-development of infrastructure and applications

Much Grid development has focused on infrastructure with scant attention given to the actual scientific or engineering simulation applications which should use it. Testbed experience shows that existing Grid infrastructure components may, or may not, work cor-

rectly for many real-world applications. Difficulties occur in maintaining a coherent testbed, implementing workable security, interacting efficiently with information servers, migrating authentication rights, etc. Application communities are largely uninvolved in Grid computing, despite demanding larger machines and storage systems, shorter queues, and easier working environments, all of which could be provided by a working Grid. Application codes are generally Grid-unaware. Of the few Grid-aware applications, most are experimental, never used in production environments. Most application developers and users are unaware of new Grid-based alternatives to the traditional time-consuming methods for using supercomputers.

We will address these problems by developing the GAT, introducing it to research communities, and deploying it across testbeds in EU and US. Enhancing both the Cactus framework and Triana for Grid computing will help ensure maximum generality, and enable a broad range of applications to exploit global resources. We will also develop, implement, and deploy fundamentally new Grid application scenarios, and show application communities how to take advantage of them. Our project architecture is summarized in Fig. 1. The primary components of our solutions are:

The centerpiece of GridLab is a generic GAT integrating components for resource management, resource, application and performance monitoring, information services, and data management. The work packages are designed to develop working and extensible examples for each of these components, and just as importantly, generic APIs that work not only with GridLab components, but with components from other projects. GridLab will be tightly connected to other projects, and GGF, to ensure compatibility. API's will be developed for all GAT components and layers, including portals, applications and the underlying infrastructure (e.g. Globus services).

The Cactus and Triana specific GAT components will provide application specific functionality, e.g. runtime choice of components such as resource brokers. This will be of prime importance for comparing the performance and functionality of different implementations of Grid infrastructure with real-world applications; conversely, the same Grid components can be easily tested with different applications.

The success of the promises of global Grid computing will be measured by its effective use by application communities. We will establish and maintain a working European testbed for developing and deploying GridLab technologies, and making them easily available to application developers and users. The testbed will provide the necessary environment for developing and testing software which needs to work on the very different machines and operating systems (Cray T3E, IBM SP2, Hitachi SR2000, Origin 2000, Compaq Alpha clusters, Sun Enterprise, Windows NT and Linux clusters) interacting with local policies, file-systems, security measures, schedulers and other variations.

As our toolkits will be used by scientists working in different large scale, geographically distributed collaborations, we require solutions which enable use and interoperability across different testbeds. Our EU testbed will be linked to additional environments in US, ensuring interoperability, and will include various existing large scale collaborative application teams, such as an already existing EU Network of 10 institutions using Cactus for astrophysics, and many participants of the GGF.

2.3.2. Construction of dynamic grid computing capabilities and scenarios

The Grid will fundamentally change not only the way scientists and engineers work but also the way they *think* about computing. Fundamentally new approaches and capabilities will be not only possible but also required to solve the above dynamic problem scenarios, with the complexity, scale, and efficiency needed by the future generations of computational scientists and engineers.

The Grid is inherently dynamic; machine loads and availability, network latencies and bandwidths, change continually. Available resources, and application requirements, can be dramatically different from 1 h to the next. Simulation technologies using the Grid must be able to adapt to such dynamic environment, requiring many new adaptive technologies at both the application and infrastructure level, *and* new thinking about how problems and algorithms.

We will develop new and innovative Grid computing scenarios to treat the Grid as a single computer, implement them with our GAT and the Cactus and Triana toolkits, and change the way application communities think about their problems. Our simulation

Table 2

Network demands: inferior networking conditions may also be covered, but result in severe usability restrictions

| Scheme | Min throughput (MB/s) | Max latency (ms) | Max jitter (%) |
|----------------------|-----------------------|------------------|----------------|
| Remote checkpointing | 50.0 | ×000 | 100 |
| Raw data movement | 10.0 | ×000 | 100 |
| Remote visualization | 0.5 | 300 | 10 |
| Remote steering | 0.1 | 50 | 10 |

and visualization codes will be aware of the Grid environment, and *dynamically adapt to best exploit it*. These capabilities will allow fundamentally new uses of global computational resources. For example, spawning independent tasks, migrating entire or parts of running applications, or recovering from Grid failures. Applications themselves are now becoming increasingly dynamic. Techniques such as adaptive mesh refinement, interactive steering and the realtime addition of code modules change the resource requirements *during* a simulation. Many such scenarios will be developed and realized in practical applications.

3. Network infrastructure for GridLab

The primary goal of GridLab is the development of a Grid computing software infrastructure (called the GAT) and to make this toolkit publicly available on an open source basis. The direct involvement of user groups in the development of GridLab's software ensures the usefulness and applicability of the GAT. However, the purpose of the GAT is its widespread use by application communities (within and outside the GridLab consortium) rather than performing certain computations during the project lifetime.

The GridLab project anticipates Grid computing experiments in order to evaluate, improve, and demonstrate the software developed within the project. For the development of the GAT software infrastructure, a network testbed installation is vital. This testbed installation should be a realistic test case for the GAT. In the sections below the networking demands of the testbed applications and the GAT are described.

3.1. GridLab networking demands

While it is beneficial to have high-performance links between the individual sites in the GridLab testbed, the whole system needs to be open (i.e. not

built using private dedicated links) in order to allow easy incorporation of new sites, changes in topology, etc. Also, user access to the Grid system should not use any dedicated private links, as this severely limits the number of potential users (to only those with such special arrangements) as well as user mobility. Thus, the software developed by GridLab should be able to run on any Grid platform, even without specific network provision.

Various data movement schemes will be exploited by the GridLab scenarios. They vary in their demands on latency and throughput, as shown in Table 2. In fact, the application toolkit (GAT) developed as part of the GridLab project will provide support for adaptive behavior, in order to dynamically react to changes in the underlying network performance. Also, research and development in the area of mobile computing will emphasize the adaptability to the different network characteristics. This means that instead of relying on guaranteed network characteristics (like a particular throughput or latency), the GridLab solution is to provide a software infrastructure that is able to adapt to changing network conditions.

From the many different Grid use scenarios considered within the GridLab project, the following two will stress the network the most:

1. *Worm-type applications.* Such applications migrate from site to site within a Grid environment. For example, migration might become necessary when the resources (e.g. available compute time) are exceeded at the computation site that is currently used at a certain moment of the application run. For this purpose, a worm-type application performs application-level checkpointing that transfers the entire status (execution state and data) to another computation site. In this event, large data volumes (up to hundreds of gigabytes) are transferred from one site to another. However, as the computation does not continue during the migration itself, the

actual link quality between the sending and receiving nodes is not critical, as long as the data arrive in reasonable time.

Meanwhile, means for efficient transfer of large amounts of data across networks are currently studied and developed within other projects (e.g. GriPhyN, DataGrid, etc.) and will be available to the wider scientific community. Also, such huge data movement requirements will not be necessary before the second half of the project (where more “production-like” runs will be possible).

2. *Remote visualization.* Remotely visualizing data sets or running simulations can create data streams each of which typically requires 5 MB/s or more. Algorithms that are able to adapt to different network bandwidths will be developed to reduce the actual requirements. Also, visualization may only be required during the parts of the project time, leaving opportunity for cooperation with network operators offering managed bandwidth services that may be asked for the visualizations only.

With remote visualization, low latency is less important than small amounts of jitter (the latency variation), which should be kept as low as possible. However, current IP networks are not well suited to guarantee such a behavior; this is a topic of ongoing network research worldwide.

Accompanying visualization, remote steering and remote data exploration make greater demands on latency and jitter, while transferring only small amounts of data. Software adaptive to varying network quality will offer great benefits to such scenarios.

3.2. *Physical network infrastructure provision for GridLab*

The GridLab testbed will consist of the facilities of the project partners: PSNC, MPG, ZIB, MU, SZTAKI, VU, ISUFI, and NTUA. Via subcontracting, we expect Paderborn University, GMD Bonn, and the University of Vienna to provide additional testbed sites. In USA, we anticipate to include the GrADS testbed and facilities of ISI California, Argonne, and NCSA.

The GridLab testbed will be built on top of the existing, primary academic networking infrastructure. The tests and demonstrations performed in the past by some of the GridLab collaborators had shown

that this infrastructure does have suitable capacity for demonstration and testing purposes. The infrastructure is continuously improving, both at the national level (e.g. the Polish project Pionier, intercity academic gigabit networks in Germany, Czech Republic, etc.) and internationally via the project Geant and related activities (e.g. the upcoming project DataTag). All these activities not only increase the totally available bandwidth, but they are also accompanied by spin-off projects that target the management of available bandwidth (the bandwidth management service to be offered by Dante within the Geant project, large scale DiffServ experiments, etc.).

PSNC, the coordinator of GridLab, is also a national operator for the Polish academic high speed network. Masaryk University and its GridLab team has strong links to CESNET, the Czech national academic high speed network operator. PSNC and CESNET are the members of Dante consortium and collaborators in the Geant project. Both these GridLab partners are prepared and committed to provide necessary support for negotiating specific conditions with national academic operators and with Dante for pan-European connectivity with plans to use the managed bandwidth service for large scale demonstrations. Through Vrije Universiteit, another GridLab partner, we are prepared to approach the Dutch SURFnet, to get even more support in these issues if necessary. A rather optimistic view on available network capacities is supported by the current Geant topology that will provide 10 GB/s links to UK, Germany, Italy and Czech Republic and 2.5 GB/s links to the Netherlands, Poland and Hungary.

We expect that it will be possible to perform tests and demonstrations within the GridLab project without additional financial requirements. Support by the respective academic network providers should be given in form of special, temporary arrangements within the existing provision of network infrastructure.

With respect to the transatlantic connectivity, we plan to mostly use the existing network infrastructure. This will be available both via the Geant project as well as through direct national links to USA (and particularly to StarTap, the Internet2 major international hub). Several other projects to provide high speed transatlantic connectivity are also under way, e.g. the StarLight project which will connect SURFnet (Amsterdam) with StarTap (Chicago) with a 2.5 GB/s

link from September 2001. Among others, this link will connect the DAS clusters of Vrije Universiteit with machines of ANL, both are already anticipated as nodes of the GridLab testbed.

Moreover, project participants of GridLab have already been invited by Tom Defanti from StarTap to give a GridLab demo at the iGrid event in September 2002 where the showcase of the 10 Gigabit applications will take place. The transatlantic networking infrastructure will be given to GridLab for free for several days. Previous iGrid events showcased advanced scientific applications being enabled by StarTap; for details see iGrid 98 <http://www.startap.net/igrid98> and iGrid 2000 <http://www.startap.net/igrid2000>. Previous iGrids also provided researchers with a global testbed on which to develop and test application advancements, middleware innovations and new networking technologies. The challenges of participating in iGrid testbeds motivated participants—application scientists, computer scientists, artists, networking engineers, commercial vendors—to collaborate on a global scale to advance the state-of-the-art in high-performance computing and communications. The iGrid demo in September 2002 will allow us to gain further experience by using a 2.5 GB/s (possibly 10 or even 40 GB/s) global testbed.

We will closely cooperate with the national operators and with other Grid-related projects on the issues concerning networking, most notably with DataGrid with its own working group on networks, and also with new emerging projects like DataTag. For large scale experiments we will use direct support of the national operators (PSNC, CESNET), including, if necessary, their indirect financial contribution (e.g. providing part of the international and especially transatlantic capacity for the time of the experiment).

The basic idea behind the GridLab project is the use of commodity (or more precisely) already available systems including the available network infrastructure and allow their Grid-wide use through the development of new software toolkits and services. Adaptability, one of the GridLab focus areas, is our answer to the impossibility of guaranteeing precise service contracts within the current internet. Close relationship with national academic backbone operators will provide fresh updates on the current state of the European-wide networking infrastructure and its features to be used within the GridLab project.

4. Summary

We have presented the GridLab project, which is the newly funded project under the Fifth Framework Programme of the European Commission. We have briefly showed the main goals of the project and emphasized the innovations which GridLab brings to the Grid Community. For more information about the GridLab project we refer the readers to the GridLab web pages [15].

References

- [1] G. Allen, T. Damlitsch, T. Goodale, G. Lanfermann, T. Radke, E. Seidel, T. Kielmann, K. Verstoep, Z. Balaton, P. Kacsuk, F. Szalai, J. Gehring, A. Keller, A. Streit, L. Matyska, M. Ruda, A. Krenek, H. Frese, H. Knipp, A. Merzky, A. Reinefeld, F. Schintke, B. Ludwiczak, J. Nabrzyski, J. Pukacki, H.-P. Kersken, M. Russell, Early experiences with the egrid testbed, in: Proceedings of the IEEE International Symposium on Cluster Computing and the Grid, 2001. http://www.cactuscode.org/Papers/CCGrid_2001.pdf.gz.
- [2] Astrophysics Simulation Collaboratory (ASC) home page. <http://www.ascportal.org/>.
- [3] Cactus Computational Toolkit home page. <http://www.cactuscode.org/>.
- [4] Condor High Throughput Computing home page. <http://www.cs.wisc.edu/condor/>.
- [5] DataGrid project home page. <http://www.eu-datagrid.org/>.
- [6] EGrid Testbed and Applications Working Group Demonstrations at SC2000. <http://www.cactuscode.org/Development/Egrid.html>.
- [7] EGrid home page. <http://www.egrid.org/>.
- [8] Entropia home page. <http://www.entropia.com>.
- [9] EuroGrid: Application Testbed for European GRID Computing home page. <http://www.eurogrid.org/>.
- [10] EU Network for Theoretical Foundations of Sources for Gravitational Wave Astronomy of the Next Century: Synergy between Supercomputer Simulations and Approximation Techniques home page. http://www.aei.mpg.de/research/astro/eu_network/index.html.
- [11] Global Grid Forum home page. <http://www.gridforum.org/>.
- [12] Globus home page. <http://www.globus.org/>.
- [13] Grid Application Development Software (GrADS) project home page. <http://www.isi.edu/grads/>.
- [14] Sun Grid Engine Software home page. <http://www.sun.com/software/gridware/>.
- [15] GridLab home page. <http://www.gridlab.org/>.
- [16] DFN Gigabit Project Development of Grid Based Simulation and Visualization Techniques (GriKSL) home page. <http://www.griksl.org/>.
- [17] GriPhyN home page. <http://www.griphyn.org>.
- [18] Albert Einstein Institute Numerical Relativity Group home page. <http://jean-luc.aei-potsdam.mpg.de/>.

- [19] M. Russell, G. Allen, G. Daues, I. Foster, T. Goodale, E. Seidel, J. Novotny, J. Shalf, W.M. Suen, G. Von Laszewski, The astrophysics simulation collaboratory: a science portal enabling community software development, in: Proceedings of the Tenth IEEE International Symposium on High-performance Distributed Computing, HPDC-10, San Francisco, 2001. http://www.cactuscode.org/Papers/HPDC10_2001_Portal.ps.gz.
- [20] Supercomputing'95 Distributed Computing Experiments. <http://jean-luc.aei.mpg.de/Projects/SC95/>.
- [21] DFN Gigabit Project “Tele-immersion: Collision of Black Holes” (TIKSL) home page. <http://www.zib.de/Visual/projects/TIKSL/>.
- [22] Triana Graphical Programming Environment home page. <http://www.triana.co.uk/>.
- [23] Uniform Interface to Computing Resources (UNICORE) home page. <http://www.unicore.de/>.
- [24] Alliance Virtual Machine Room home page. <http://www.ncsa.uiuc.edu/SCD/Alliance/VMR/>.



Jarek Nabrzyski received his MSc and PhD degrees in Computer Science from Poznan University of Technology. Currently he occupies a research position at the Poznan Supercomputing and Networking Center, where he heads the Applications Department. His research interests over the last 10 years have focused on knowledge-based multiobjective project scheduling, and resource management for parallel and distributed computing. For the last 5 years he has been working on tools and middleware technologies for Computational Grids. He currently leads several Grid-related projects at PSNC: MCRM Portal—multicriteria resource management architecture for the Grids with Portal Access, GridPred—job runtime and queue wait time prediction system. Jarek Nabrzyski is a co-founder of the European Grid Forum and the Global Grid Forum. He is a Member of the Global Grid Forum Steering Committee and he actively promotes Grid ideas and concepts in Europe. He was the *Program Committee Chair* of the First Global Grid Forum meeting held in Amsterdam in March 2001.



Gabrielle Allen, who holds a PhD in Astrophysics from Cardiff University, has led the Cactus team for the past 3 years. Previous to this she spent 10 years as a Computational Physicist in the field of numerical astrophysics, and so has a deep understanding of user requirements for computational software and infrastructure. Allen is involved in collaborations with other advanced grid projects, including

Globus, the Global Grid Forum, the Grid Adaptive Development Software (GrADs) project, the EGrid Applications Working Group, the Grid Portal Collaboration and the Astrophysics Simulation Collaboratory. Allen is a member of the Steering Committee for the proposed EU project “*Computational and information infrastructure in the astronomical DATAGRID: towards virtual astronomy*”, and a PI for the DFN project “*GriKSL: development of grid-based simulation and visualization techniques*”. She is also a Member of the EU funded network *Sources of Gravitational Wave Astronomy* and holds a Visitor position at the Argonne National Laboratory in the USA.



Ed Seidel has led both the Computer Science and the Numerical Relativity Research groups at the AEI for the past 6 years, with over 120 publications spanning both disciplines, and has a PhD from Yale University. Seidel holds additional positions as Senior Research Scientist at NCSA and as Adjunct Professor at the University of Illinois, and a Visitor position at ANL. He has active collaborations and connections with leading centers and both computational science and astrophysics based research projects in the US, strengthening the cross-Atlantic nature of this project. Seidel is the Chair of the GGF Applications Working Group. In the US, he is a PI for the NSF funded KDI project to establish an *Astrophysics Simulation Collaboratory*, and a NASA Grand Challenge Project. In Europe, he led the DFN project “*TIKSL: teleimmersion of black holes*”, for which a follow-on proposal *GriKSL* has just been approved, and is the coordinator of the EU funded network for *Astrophysical Sources of Gravitational Wave Astronomy*. All these projects provide both technologies and user communities for testing Grid components in this project.



André Merzky received his degree in Particle Physics in 1998 and joined the Scientific Visualization Department at ZIB. He has worked in Grid-related topics concerning data management (see *GriKSL*-project with AEI), was the Chair of the EGrid Working Group *Data Storage and Management*, and is one of the Co-Chairs of the *Data Area* of the Global Grid Forum.