First Prototype of the CrossGrid Testbed

J.Gomes[1], M.David[1], J.Martins[1], L.Bernardo[1], J.Marco[2], R.Marco[2],
D.Rodríguez[2], J.Salt[3], S.Gonzalez[3], J.Sánchez[3], A.Fuentes[4], M.Hardt[5],
A.García[5], P.Nyczyk[6], A.Ozieblo[6], P.Wolniewicz[7], M.Bluj[8], K.Nawrocki[9],
A.Padee[8,9,10], W.Wislicki[8,9], C.Fernández[11], J.Fontán[11], A.Gómez[11],
I.López[11], Y.Cotronis[12], E.Floros[12], G.Tsouloupas[13], W.Xing[13],
M.Dikaiakos[13], J.Astalos[14], B.Coghlan[15], E.Heymann[16], M.Senar[16],
G.Merino[17], C.Kanellopoulos[18], and G.D.van Albada[19]

[1] Laboratório de Instrumentacão e Física de Partículas, Lisbon, Portugal
[2] Instituto de Física de Cantabria (CSIC),Santander, Spain
[3] Instituto de Física Corpuscular(CSIC),Valencia, Spain
[4] RedIris(CSIC),Madrid, Spain
[5] Forschungszentrum Karlsruhe GMBH, Germany
[6] Akademickie Centrum Komputerowe CYFRONET, Krakow, Poland
[7] Poznan Supercomputing and Networking Center, Poznan, Poland
[8] A.Soltan Institute for Nuclear Studies, Warsaw, Poland
[9] Interdisciplinary Centre for Mathematical and Computational Modelling,
University of Warsaw,Poland
[10] Instytut Radioelektroniki PW, Warsaw, Poland
[11] CESGA, Centro de Supercomputacion de Galicia, Santiago de Compostela, Spain
[12] National Center for Scientific Research "Demokritos", National and Kapodistrian
University of Athens, Dep. of Informatics and Telecommunications, Greece
[13] University of Cyprus, Cyprus
[14] Ustav Informatiky Slovenska Akademia Vied, Bratislava, Slovakia
[15] Trinity College Dublin, Ireland
[16] Universitat Autonoma de Barcelona, Spain
[17] Institut de Fisica d'Altes Energies, Barcelona, Spain
[18] Aristotle University of Thessaloniki, Greece
[19] Universiteit van Amsterdam, Netherlands

Abstract. The CrossGrid project is developing new grid middleware
components, tools and applications with a special focus on parallel and
interactive computing. In order to support the development effort and
provide a test infrastructure, an international grid testbed has been de-
ployed across 9 countries. Through the deployment of the testbed and its
supporting services, CrossGrid is also contributing to another important
project objective, the expansion of the grid coverage in Europe. This
paper describes the status of the CrossGrid testbed.

1 Introduction

The CrossGrid international distributed testbed will share resources across six-
teen European sites and this is one of the challenging points of the CrossGrid
project.

The sites list range from relatively small computing facilities in universities, to large computing centers, offering an ideal mixture to test the possibilities of the Grid framework.

National research networks and the high-performance European network, Géant, will assure the interconnectivity between all sites. The network includes usually three steps: the local step (typically inside a University or Research Center, via Fast or Gigabit Ethernet), the jump via the national network provider (at speeds that will range from 34 Mbits/s to 622 Mbits/s or even Gigabit) to the national node, and finally the link to the Géant network (155 Mbits/s to 2.5 Gbits/s).

The figure 1 shows the geographical map for the different nodes, including the major "network" links.
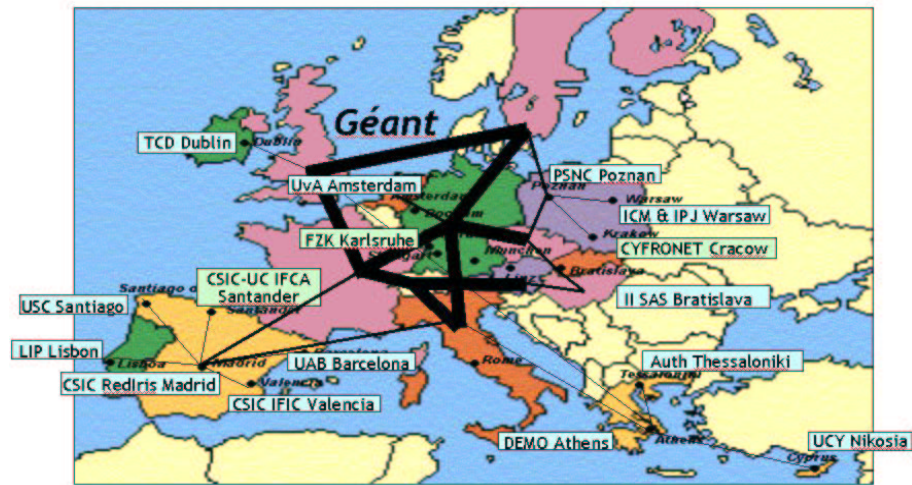

Fig. 1. CrossGrid testbed map.

The CrossGrid testbed largely benefits from the DataGrid [1] experience on testbed setup: since the CrossGrid middleware is being developed it was decided that initially the testbed middleware would have to be based entirely in Data-Grid (EDG) and Globus middleware distributions. The efforts to establish an integrated CrossGrid testbed started with the release of EDG 1.2.0, however several problems where found in the first deployed sites. The EDG release 1.2.2 with improved stability has allowed more sites to join in spite of some serious middleware limitations. Currently EDG 1.2.2 and 1.2.3 are deployed in the production testbed, while the version 1.4.3 is being tested at several validation sites; it is expected that it will overcome many of the major limitations of the previous versions allowing the interconnection of both CrossGrid and DataGrid testbeds.

2   Testbed status

At least sixteen sites are foreseen to be included in the CrossGrid testbed[2], this number may be increased with the addition of more sites interested in the technologies and applications being developed by CrossGrid, and with the addition of more sites belonging to CrossGrid partners. These 16 sites are being deployed and registered into the CrossGrid Resource Broker (RB).

The table 1 shows the testbed deployment status in the end of February 2003.

Site	Location	Version	RB registration
CYFRONET	Cracow	1.2.2	Production
ICM	Warsaw	1.2.2	Production
INS	Warsaw	1.2.2	Production
UvA	Amsterdam	1.2.2	NOT registered
FZK	Karlsruhe	1.4.3	Validation
IISAS	Bratislava	1.2.3	Production
PSNC	Poznan	1.2.2	Production
UCY	Nikosia	1.2.2	NOT registered
TCD	Dublin	1.2.3	Production
IFIC	Valencia	1.2.3	Production
IFCA	Santander	1.2.2	NOT registered
UAB	Barcelona	1.2.2	Production (under test)
USC/CESGA	Santiago	1.2.2	Production
Demokritos	Athens	1.4.3	Validation
AUTH	Thessaloniki	1.2.2	Production
LIP	Lisbon	1.2.3 1.4.3	Production, Validation (two clusters)

Table 1. Testbed site status.

3   Testbeds

According with the CrossGrid testbed architecture and minimum hardware requirements[3] each site must have at least five system components:

Gatekeeper: is the system that provides the gateway through which Jobs are submitted to local farm nodes. The gatekeeper is the interface through which grid enabled systems can use the local resources.

Worker Node (WN): is a local farm computing node where jobs are actually executed. Jobs received by the Gatekeeper are sent to the WN through the local batch scheduling system. Many worker nodes can exist behind a single Gatekeeper. The combination of a Gatekeeper with their worker nodes is usually called a Computing Element (CE).

Storage Element (SE): is the generic name used for any storage resource that includes a Grid interface ranging from large Hierarchical Storage Management Systems to disk pools.

User Interface (UI): is the system used by end-users to submit jobs to the grid Computing Elements. The job submission is usually performed through a resource broker.

LCFG installation server: is used to install, configure and maintain the above systems from a single management system. The installation server is required for the proper installation and ease of maintenance of the EDG middleware.

Although we talk about a generic CrossGrid testbed, in fact several "testbeds" will co-exist to support different efforts, from pure development, to production after validation. The following testbeds are foreseen [4]:

Development testbed: new middleware or application software development will require a dedicated small but well maintained testbed, allowing its test in a Grid aware framework. The testbed setup will include a dedicated minimal configuration, and the flexibility to adapt to the required changes in the external software packages (like for example migrating from Globus 2 to Globus 3, or from RedHat 6.2 to RedHat 7.3). The development testbed will also provide the environment to integrate the developed components into software releases.

Test and Validation testbed: each new middleware release will require a complete validation process that will include extensive tests. These tests must be performed in a tight controlled environment without disturbing the production and development testbeds. The dedicated validation testbed supports this effort, and is being offered by LIP, FZK, CSIC(IFIC), USC(CESGA) and DEMO.

Finally the **Production testbed** will be used to run the applications, showing the Grid potential. It will be extended over all testbed sites: CYFRONET, ICM, INS, UvA, IISAS, FZK, PSNC, UCY, TCD, IFCA, UAB, USC, DEMO, AUTh, IFIC and LIP.

The first CrossGrid testbed prototype was born from the initial "test and validation" testbed, managed by the LIP group. This effort has been very successful: external site resources have been included in the corresponding Resource Broker and users joined the Virtual Organizations. Jobs have been successfully submitted through User Interface nodes at each testbed site.

Basic tests covering the Globus and EDG middleware functionalities have been performed using the testbed. These tests cover the job submission using Globus and the EDG RB, file transfer with GSI ftp, file replication with GDMP, file replication with the Replica Manager, the VO server and the MDS information system.

Currently the initial "test and validation" testbed running EDG 1.2.2 and 1.2.3 was moved into production, and a new "test and validation" testbed running EDG 1.4.3 is now deployed and running within a limited set of sites (LIP, FZK and Demokritos). This new "test and validation" testbed will be used to test the first CrossGrid software release that will include: monitoring tools, development tools, a remote access server, portals and a prototype of the parallel resource broker.

Tables 2 and 3 show the resources available in the production and validation testbeds respectively.

Computing Elements	15
Worker Nodes	69
CPUs	115
Storage Elements	14
Storage Capacity	2.7TB

Table 2. "Production" testbed resources.

Computing Elements	3
Worker Nodes	4
CPUs	5
Storage Elements	3
Storage Capacity	1.2TB

Table 3. "Test and validation" testbed resources.

The resource tables for the production testbed already show a considerable amount of computing and storage resources. The amount of resources will grow as the sites commit more nodes. Namely it is expected that once RedHat 7.3 is made available in the testbed several existing production cluster will be added to the grid. A method for adding existing clusters running different Linux flavours from the one supported by EDG, is being studied. If a suitable solution is found, then it will be possible to deploy existing Linux clusters as worker nodes without interfering with the installed software.

Although the "test and validation" testbed has fewer resources it has been very successful in testing the EDG releases 1.4.x. Their resources are currently providing the platform for the integration of the first CrossGrid software release.

4   Tests and applications

Three CrossGrid sites (FZK, IFIC and LIP) have participated at the IST 2002 demonstration event showing the interoperation of the CrossGrid sites with other testbeds. During the demonstration several jobs have been successfully executed in the three participating CrossGrid sites.

In addition the first tests with MPI in the testbed have been performed. Since the EDG RB doesn't support parallel applications the program executions have been performed using the Globus job submission capabilities directly. The MPI package used for the tests was MPICH-G2 a variant of the MPICH implementation using Globus IO for inter-process communication and Globus DUROC for resource co-allocation. In the same context tests were performed on how to start the parallel applications. Although MPICH provides its own program start-up command named mpirun, the MPICH-G2 mpirun command has some limitations. Therefore other methods to submit jobs to selected resources have been tested.

At the same time, the first CrossGrid software prototypes are being developed. A prototype of the HEP application has been released, requiring MPI and very low latency. Tests aiming to validate the prototype architecture and the testbed middleware with a full CrossGrid application are being performed in the testbed. The tests of the HEP application have shown the feasibility of running full applications across testbed sites with MPICH-G2.

5   Collaborative tools

The basic collaborative tools are the WP4 web pages (that can be found at http://grid.ifca.unican.es/crossgrid/wp4) and the videoconference system VRVS (http://vrvs.org): its usage allows that, coordination meetings take place regularly with the participation of all sites (14 meetings in 9 months). VRVS has proved to be both inexpensive and flexible since it uses the Internet as communications medium and supports a wide range of audio and video devices. Many operating systems are supported and most portable computers can be used with VRVS.

Web pages containing information about the central testbed services are available at the LIP web server. The main link for the pages related with the production testbed is http://www.lip.pt/computing/cg-services. These web pages also contain pointers to usage statistics and information on how to configure the testbed sites. Useful information on how to diagnose problems and a list of common configuration problems is also provided. The pages are both a central place to find updated information about the testbed and a first level user support tool where users can cross check their site configurations and obtain helpful hints on how to solve common problems. A second web page containing the same information for the test and validation testbed is available at http://www.lip.pt/computing/cg-tv-services.

6   Central services

A set of central services[2] is required for proper testbed operation. These services are essential for user authentication, authorization, workload management, monitoring and location of file replicas.

One of the most important central services is the Virtual Organizations server that contains the CrossGrid VO. To use the CrossGrid VO for authentication, configuration changes have to be made in all testbed systems. The support for VO site configuration is provided through the central services web pages at http://www.lip.pt/computing/cg-services. CrossGrid users wishing to join the CrossGrid VO should send their request by Email with an explanation of their participation in the project and containing the X.509 user certificate. This procedure will be used until a web interface for VO enrollment,which is under development is made available.

The Resource Broker (RB) is the central service responsible for the management of the testbed workload. The RB receives user job requests submitted

through User Interface nodes. Access control is performed using a local authorization file rebuilt from the CrossGrid VO server. Once a user is authenticated and authorized to access the RB the job description is transferred from the UI to the RB and a matchmaking operation is started to find computing resources matching the job requirements. In the matchmaking the RB uses information about the existing computing resources in the testbed obtained from the Globus MDS information system. For this purpose the RB consults a MDS Information Index (II) server containing pointers to the MDS information servers (published by every CE and SE). When the job description contains references to logical files the RB must also query the central Replica Catalogue to obtain the location of the corresponding physical replicas in order to select the CE that has a near SE containing replicas of the requested files. Finally the RB uses the central MyProxy server to renew the job credentials when needed.

The MyProxy server is an authentication credential proxy used by the RB to renew the proxy certificates of long-lived jobs. MyProxy stores certificates on behalf of other subjects and can issue or renew short-term proxy certificates based on stored certificates.

The system hosting the RB also hosts the Logging and Bookkeeping (LB) database. The LB is used to record information about job status and can be queried to obtain the status of a job or historic data about previously submitted jobs. The logging information is stored in a MySQL relational database.

Two RBs have been deployed, one for the production testbed and a second for the "test and validation" testbed. However in the production testbed the II server is hosted in the RB while in the "test and validation" it is hosted in a separate system.

The last central service is the Replica Catalogue (RC). Currently the RC is an LDAP server containing logical file names. For each logical file name the server contains the location of the physical file replicas. The physical file replicas can be stored in multiple testbed SEs. The replica catalogue is a key piece in the data access optimization since when combined with the RB it allows jobs to be executed at sites where the required data files are already present, hence minimizing data transfers and reducing data access time and network bandwidth usage. File replication software such as the Replica Manager and GDMP make use of the Replica Catalogue to register the location of the replicated files. Again two RCs have been deployed one for each testbed.

All CrossGrid central services are being hosted at the LIP Computer Centre in Lisbon.

7   Certification authorities

The authentication of users and systems is performed through a public key infrastructure based on X.509 certificates. Since certificates must be issued by trusted certification authorities (CAs) CrossGrid choose to trust the national CAs already established by DataGrid, and to coordinate the deployment of new national CA's where necessary[2]. New grid certification authorities have been

established in the countries where they were not previously available namely: Poland, Germany, Slovakia, Greece and Cyprus. A strong effort was made to make the new CAs recognized by DataGrid creating new opportunities for sharing resources between CrossGrid and DataGrid and therefore extending the grid coverage in Europe.

8   Testbed monitoring

Users and site administrators can obtain information about the testbed status [2] from the mapcenter web pages available at http://mapcenter.lip.pt. Mapcenter is an important tool for testbed support since it provides a quick global view of the testbed status and historical data that can be used to identify the origin of problems occurred in the past. The mapcenter web pages are extremely important for site administrators since they allow them to check the site services connectivity. Mapcenter has been improved with the addition of links for usage statistics covering the RBs and CEs usage.

The statistics module for the RB collects information from the MySQL logging and bookkeeping database containing historical data about the job submissions performed through the RB. The current RB doesn't support parallel applications therefore all MPI application tests involving communication across clusters don't appear in the RB statistics. CrossGrid is working on extending the EDG RB to support MPI parallel applications. The tables 4 and 5 show the usage of both the production and validation RBs. The meaning of each table row is explained below:

- **Total users:** number of different user certificates that have been used to submit jobs through the RB.
- **Total jobs sent to the RB:** number of job submissions that have been attempted from user interfaces.
- **Total jobs accepted by the RB:** number of job submissions that have been authenticated and accepted by the RB.
- **Total jobs with good match:** number of jobs submitted for which the RB has found testbed resources matching the request. Matchmaking failures can be caused by jobs requesting unavailable or even inexistent resources.
- **Total jobs submitted by JSS:** jobs submitted by the RB to grid computing resources (Gatekeepers) using the Globus GRAM service.
- **Total jobs run:** number of jobs successfully submitted to grid resources.
- **Total jobs done:** number of jobs that have finished successfully.

The CE statistics module collects information from the gatekeeper log file in every testbed CE. The log file is first retrieved and then parsed to reconstruct the job submission records then web pages with a summary and statistics indexed per user and IP address are produced for each site.

Table 6 shows the total aggregated usage of the "production" CEs while table 7 shows the total aggregated usage of the "test and validation" CEs. The meaning of the CE statistics fields are the following:

Total users	35
Total jobs sent to the RB	2313
Total jobs accepted by the RB	2168
Total jobs with good match	2010
Total jobs submitted by JSS	1988
Total jobs run	1755
Total jobs done	1207

Table 4. Production RB usage.

Total users	9
Total jobs sent to the RB	5407
Total jobs accepted by the RB	5396
Total jobs with good match	5178
Total jobs submitted by JSS	5167
Total jobs run	5022
Total jobs done	4976

Table 5. Validation RB usage.

– **Total job submissions attempted:** number of connections attempted to the gatekeeper TCP port. Unfortunately due to TCP scans this number can be higher than the actual real job submission attempts.
– **Total ping jobs submitted:** number of globus GRAM pings submitted. These aren't real jobs, instead they are used to verify whether a specific globus job-manager is available.
– **Total jobs successfully submitted:** number of jobs submitted, accepted and for which a job-manager was started.

The failed jobs are divided into the following four fields:

– **LCAS failed jobs:** jobs submissions failed due to authorization issues.
– **GSS failed jobs:** jobs submissions failed the GSS authentication protocol. The port scans are included in this error type.
– **CRL failed jobs:** job submissions failed due to outdated CRLs.
– **Jobman failed jobs:** jobs submissions failed due to job-manager related problems such as an inexistent job-manager.

Another important source of information for site administrators is the "host check" web site at http://www.lip.pt/computing/cg-services/site_check. The "host check" web pages are produced by a diagnostics tool that verifies the installation and configuration of the CE and SE systems in the testbed by running several tests covering: file transfer, job submission, accuracy of the information published through MDS and the correctness of the relevant system configuration files. The "host check" web pages are being used to support the deployment of

Total job submissions attempted	39168
Total ping jobs submitted	1496
Total jobs successfully submitted	14353
LCAS failed jobs	411
GSS failed jobs	22071
CRL failed jobs	254
Jobman failed jobs	583

Table 6. "Production" CEs aggregated usage.

Total job submissions attempted	86913
Total ping jobs submitted	2383
Total jobs successfully submitted	79836
LCAS failed jobs	63
GSS failed jobs	4557
CRL failed jobs	18
Jobman failed jobs	56

Table 7. "Test and Validation" CEs aggregated usage.

new sites by providing an automated site verification tool that covers the majority of the configuration issues. This tool has also proved to be extremely valuable in the testbed quality assurance helping on the monitoring and certification of new sites.

9 Software repository

Regarding the software repository[5, 6], the Savannah software package has been installed on http://gridportal.fzk.de. Savannah is a web-portal-application based on Sourceforge. It is based on a MySQL database and php/perl scripts and was adapted to the CrossGrid specific needs.

The portal provides the central software repository for CrossGrid. Furthermore Mailinglist-Forums, BugTrackers and CVS services are offered for developer and user communication. The software repository is the official site for the distribution of CrossGrid developed software. The repository also contains other user information such as the installation profiles for the several sites and documentation.

10 Helpdesk

The User Help Desk [5] is the main tool of the User Support Team, this specific development has been carried taking into account the guidelines for the

unification of the DataGrid and the CrossGrid Helpdesk. The user Help Desk infrastructure allows all CrossGrid testbed users to get support for encountered problems or questions and access the CrossGrid user documentation. A user could be a scientist using the CrossGrid or a local system administrator of a CrossGrid testbed site. Users can ask all kind of questions related to the Cross-Grid Testbed, covering issues such as certificates usage, installation and config-uration, job submission, network security, resource availability, etc. The Help Desk Database administrator takes care of the utility and will control the ef-ficiency of the method trying to improve it whenever possible. The CrossGrid HelpDesk is a Web based helpdesk system incorporating PHP, Javascript and MySQL, customized from the OneOrZero initiative. The help desk is oriented not only to end users but also to developers and system administrators.

11   Acknowledgements

We would like to thank the support from the following projects and institutions:

References

1. For a description of the European Data Grid project testbed see for example *Eval-uation of Testbed Operation, EU-DataGrid IST-2000-25182, Feb 2002*, and ref-erences there in. Available as project deliverable D6.4 from Workpackage 6 at http://www.eu-datagrid.org
2. J.Gomes. *Testbed Extension and Site Status, EU-CrossGrid IST-2001-32243, Feb 2003*. Available as project deliverable D4.4 from Workpackage 4 at http://www.eu-crossgrid.org
3. J.Marco, R.Marco, J.Gomes, M.David, M.Hardt, J.Salt. *Detailed Planning for Testbed Setup, EU-CrossGrid IST-2001-32243, May 2002*. Available as project de-liverable D4.1 from Workpackage 4 at http://www.eu-crossgrid.org
4. J.Gomes, M.David. *Test and Validation Testbed Architecture, EU-CrossGrid IST-2001-32243, Sep 2002* Available as project deliverable D4.2 from Workpackage 4 at http://www.eu-crossgrid.org

5. J.Marco, R.Marco, J.Gomes, M.David, M.Hardt, J.Salt. *Setup of First Testbed in Selected Sites, EU-CrossGrid IST-2001-32243, Sep 2002* Available as project deliverable D4.2 from Workpackage 4 at http://www.eu-crossgrid.org
6. J.Marco, R.Marco. *Testbed Prototype Release, EU-CrossGrid IST-2001-32243, Feb 2003* Available as project deliverable D4.4 from Workpackage 4 at http://www.eu-crossgrid.org