Integrating BOINC-based DGs with EGEE

Peter Kacsuk
MTA SZTAKI
Parallel and Distributed Systems Lab

The EDGeS project receives Community research funding
The problem we address

- **We have service grids (SG):**
  - Guaranteed service based on clusters but
  - Relatively small number of processors (1K - 50K)
  - Many users with many applications but misuse the resources
    - Most of the applications are parameter sweep applications that better fit to desktop grids
    - Take away the cluster resources from MPI type applications

- **We have desktop (volunteer) grids (DG):**
  - No guaranteed service but
  - Large number of processors (100K – 1M)
  - Few applications: 1-3 applications per desktop grid
EGEE Infrastructure

> 200 sites in 40 countries
  ~ 60 000 CPUs
  ~ 10 PB storage
  ~ 100k jobs/day
> 200 Virtual Organizations
⇒The world’s largest multi-disciplinary service Grid
Enabling Desktop Grids for e-Science (EDGeS)

- New FP7 project started on the 1st of January 2008
- Goals of the project:
  - To integrate Service Grids and Desktop Grids to attract new scientific communities that needs very large number of computing resources (100K-1M processors)
  - To enable the seamless usage of the integrated SG-DG infrastructure, i.e. the user does not need to be aware if the application is executed in SG or DG
  - To provide a Grid application development environment
  - To provide application repository and broker for the seamless execution of applications in the SG-DG system
## Project partners

<table>
<thead>
<tr>
<th>Partic. No.</th>
<th>Participant name</th>
<th>Participant short name</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Computer and Automation Research Institute of the Hungarian Academy of Sciences</td>
<td>MTA SZTAKI</td>
<td>Hungary</td>
</tr>
<tr>
<td>2</td>
<td>Centro de Investigaciones Energéticas Medio Ambientales y Tecnológicas</td>
<td>CIEMAT</td>
<td>Spain</td>
</tr>
<tr>
<td>3</td>
<td>Foundation for the Development of Science and Technology in Extremadura</td>
<td>Fundecyt</td>
<td>Spain</td>
</tr>
<tr>
<td>4</td>
<td>The French National Institute for Research in Computer Science and Control</td>
<td>INRIA</td>
<td>France</td>
</tr>
<tr>
<td>5</td>
<td>University of Westminster</td>
<td>UoW</td>
<td>UK</td>
</tr>
<tr>
<td>6</td>
<td>Cardiff University</td>
<td>CU</td>
<td>UK</td>
</tr>
<tr>
<td>7</td>
<td>Faculty of Sciences and Technology of the University of Coimbra</td>
<td>FCTUC</td>
<td>Portugal</td>
</tr>
<tr>
<td>8</td>
<td>Stichting AlmereGrid</td>
<td>AlmereGrid</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>9</td>
<td>Centre National de la Recherche Scientifique - Institut National de Physique Nucleaire et de Physique des Hautes Energies</td>
<td>IN2P3</td>
<td>France</td>
</tr>
</tbody>
</table>
The EDGeS production infrastructure

- **BOINC based DGs**
  - Public DG EGEE@home
    - Planned 10,000 PCs
  - Public DG SZDG
    - 30,000 PCs
  - Public DG Extremadura Grid
    - 70,000 PCs

- **XtremWeb based DGs**
  - Public DG EGEE->XtremWeb
    - 1,000 PCs
  - Public DG IN2P3 Grid
    - 300 PCs
  - Public DG AlemereGrid
    - 3,000 PCs

- **Local DG**
  - UoW Grid
    - 1,500 PCs
  - IN2P3 Grid
    - 300 PCs
  - EGEE@home
    - Planned 10,000 PCs

- **Sevice Grid (EGEE)**

- **Network**
  - Connections between different DGs and grids.
The proposed DG-SG architecture

Grid 1
  3GB source Client

Grid 2
  3GB source Client

Grid 3
  3GB source Client

Generic Grid-Grid Bridge (3G Bridge)

Grid 4

Grid 5

Grid 6

Grid 7
BOINC->EGEE bridge based on 3G Bridge

Modified BOINC client transforms a WU into an EGEE job. It collects, schedules and transfers jobs. The 3G Bridge appears to BOINC as a large multi-processor resource, transferring jobs to EGEE in an optimized way.
BOINC to EGEE direction

- EGEE supports the connected DGs whenever it has available spare resources
- No changes at the BOINC servers
- The solution has already been tested by SETI@home and SZTAKI Desktop Grid
- Production version by the end of September
EGEE User access to various grids

EGEE

EDGeS

University DG

LocalDEG

GlobalDEG

Volunteer DG

Service Grid

EGEE

Volunteer DG

Repository

EDGeS

Appl.

Repository
EGEE VO → BOINC system

- EGEE BDII
- EGEE WMS
- EGEE LB
- EGEE UI

EGEE UI

Report resources and performance
Submit job
Log events
Send output
Submit job
Watch Get output

EGEE VOMS

Check EXE
X509 proxy
Get EXE

EGEE Application Repository

BOINC CE
Info provider
Add job
Watch job

BOINC-LRMS

3G Bridge
BOINC1 (SZDG)
BOINC2 (UoW)
BOINC3 (EGEE@home)

3G Bridge
BOINC1 (SZDG)

DC-API plugin

3G Bridge
BOINC2 (UoW)

DC-API plugin

3G Bridge
BOINC3 (EGEE@home)

DC-API plugin

3G Bridge
BOINC1 (SZDG)

DC-API plugin

3G Bridge
BOINC2 (UoW)

DC-API plugin

3G Bridge
BOINC3 (EGEE@home)

DC-API plugin
User access to various grids at workflow level

EDGeS
Appl. Repository

WS-PGRADE
gUSE

Service Grid EGEE

University DG

LocalDEG

GlobalDEG

Volunteer DG
An example: CancerGrid workflow

N = 30K, M = 100 => about 0.5 year execution time

Generator job

N = 30K
User communities

• Another goal of EDGeS
  – To identify user communities that need such large computational infrastructure
  – To support these communities in porting their application to EDGeS

• 14 identified communities:
  – BioInfoGrid, e-Health, Fusion, computational chemists, etc.
  – 4 new partners

• Establishing EDGeS User Forum

• 5 identified companies
  – 3 new partners

• Establishing EDGeS Industry Forum
Conclusions

• BOINC -> EGEE bridge
  – Is prototyped
  – Production version by end of September

• EGEE -> BOINC bridge
  – Is under development
  – Prototype by end of September
  – Production version by end of December

• We are looking for BOINC projects/DGs that would like to join the EDGeS infrastructure

• BOINC projects’ advantages by joining:
  – Extra resources from EGEE
  – Larger publicity by EGEE

• More information on EDGeS:
  http://www.edges-grid.eu/