Integration of European Grid Infrastructure with BOINC - latest steps and applications

Peter Kacsuk and Jozsef Kovacs
MTA SZTAKI – LPDS
{kacsuk, smith}@sztaki.hu

IDGF-SP is to be supported by the FP7 Capacities Programme under contract nr RI-312297.
Why BOINC is not taken by the Grid community?

- David told:
  - The grid community ignored BOINC
  - The take of BOINC is not as wide spread as he has expected

- Indeed BOINC is very matured, proved to be useful in large grand challenge projects with even million volunteers

- In fact, this is the only technology that really enables volunteer computing

- So why isn’t it as popular as we would expect?
Why BOINC is not taken by the Grid community?

The reasons:

– To port an application to BOINC requires significant effort
– To run a BOINC project you must become a BOINC expert
– Originally BOINC was not designed to accept many different kind of submitted jobs (and this is what grids supported)
– The response time of a volunteer grid is not as good as the one of a cluster grid (see the tail problem)

Conclusions:

– BOINC was designed to create long (even for years) running BOINC projects for a small number of grand challenge applications – and it is excellent for this
Goal of our work

- Learning the lessons from the previous analysis, the goal of our work was to enable the mass usage of BOINC:
  - To extend BOINC (without exchanging it) in order to enable
    - its dynamic, on-demand use (even only for days)
    - For a very large number of communities (even for individual researchers)
    - Without requiring any BOINC expertise
      - in porting applications
      - in set up and operate BOINC systems
Goal of our work

- To enable the use of BOINC as an **HTC infrastructure** (like Condor, like many grids)
- To use BOINC as a **volunteer co-infrastructure**:
  - To enable the **extension of existing grid** and **supercomputer** infrastructures with BOINC to enable the **collection of cheap resources**
    - parameter sweep jobs should be **transparently transferred** to the BOINC co-infrastructure
    - where their execution is **much cheaper** due to the help of volunteers
- To enable the use of BOINC systems **via science gateways**
Implementation of these goals

- **A series of EU projects** have been funded and successfully run
- Two technical projects (coordinator P. Kacsuk):
  - EDGeS: 2008-2010
  - EDGI: 2010-2012
- Two support action projects (coordinator R. Lovas):
  - DEGISCO: 2010-2012
  - IDGF-SP: 2012-2014
- **Result of these projects**: BOINC has been taken by many user communities for supporting the regular every day work of scientists and this approach getting more and more popular in Europe and elsewhere
Plans for EDGeS and EDGI

EDGeS scope
- for compute intensive applications
- for gLite

EDGI scope
- for both compute and data intensive applications

Extend Desktop Grids with Clouds
Use virtualization
Establishing the generic volunteer co-infrastructure concept and technical solution

 Gate way

 OSG

 XSEDE

 EGI

 Supercomputers

 Gate way

 3G Bridge (Generic Grid-Grid Bridge)

 Gate way

 BOINC

 Volunteers

 Gate way

 Gate way

 Gate way

 Gate way

 Gate way

 Gate way
Establishing the generic volunteer co-infrastructure concept and technical solution

The co-infrastructure concept works in the other direction, too.

As a result in DEGISCO, EGI grid resources supported newcomer BOINC projects as long as they were not able to collect large number of volunteer resources.
Concrete solution for using BOINC as co-infrastructure of EGI grids

Service Grid infrastructure

- gLite UI
- gLite SE
- gLite WMS
- gLite CEs

Service Grid extension

- EDGI
- gLite modified
- CREAM CE

Desktop Grid infrastructure

- Volunteer
- BOINC clients
- 3GBridge
- BOINC server
- Cloud infrastructure
- On-demand
- dedicated
- BOINC clients
- EDGeS@home

• MCE solutions for:
  - ARC, gLite, UNICORE
• 3GBridge is the same for every grid
EDGI solutions for show-stopping problems

- GBAC (Generic BOINC Application Client):
  - To avoid application porting
  - To run any applications on any type of client machines
  - By using virtualbox based virtualization
- Dedicated cloud resources as clients to solve the tail problem
- Result: EDGeS@home is now actively used by EGI user communities as
  - co-infrastructure or
  - HTC infrastructure
Applications at EDGeS@home for EGI scientists

- CNS [GBAC] (WeNMR)
- DIRAC [GBAC] (DIRAC)
- BBGC/MUSO (BIOVEL)
- Autodock (publicly available)
- Zeta-search (ELTE, Hungary)
- LinAlgOpt [GBAC] (Pannon, Hungary)
Overview of job submission alternatives for EDGeS@home

- gUSE portal
- EDGeS@home
- METAJOB PLUGIN
- 3GBridge
- Volunteering computers
- Cloud resources
- gLite grid
- gLite CREAM mCE
- DIRAC, WeNMR
- http://home.edges-grid.eu/home/
- cr2.edgi-grid.eu:8443/cream-pbs-homeboinc
Use case 1: submission through gLite (WeNMR/CNS, [DIRAC])

- Suggested for those who
- want to skip porting application for Desktop Grid
- prefer gLite interface instead of learning 3GBridge API
- submit jobs in the range of hundreds
Use case 2: submission through the gUSE portal (e.g. Autodock portal)

- Suggested for those who
  - prefer customised web interface and/or workflows with the easily customisable gUSE portal
  - require higher abstraction than jobs
  - submit jobs in the range of (tens of) thousands
Use case 3: direct 3GBridge submission (BIOVEL)

- Suggested for those who
- prefer low-level interface integration to the desktop grid server
- prefer hiding the single job submission with an own portal
- submits jobs in the range of (tens of) thousands
Use case 4: 3GBridge with Metajob (Pannon University, linear programming)

- Parameter-sweep Metajob file describes thousands of jobs
- Submitted in one step
- Can be combined with GBAC
- Batch level directives (e.g. stop at 90% for Monte Carlo type)
- Results are downloadable in one step

More details: [http://doc.desktopgrid.hu](http://doc.desktopgrid.hu)

LinAlgOpt at E@H: trying to find optimal values for system solvers of Linear Programming problems by doing parameter sweep of a large number of run time parameters.
Use case 5: Automatic workunit generation and submission (ELTE)

- A huge parameter space is defined
- **Parameter generator** is executed as part of the infrastructure
- Jobs are automatically generated and submitted
- Results are collected (preprocessed if needed) and **sent** to the application owner

**Zeta-search** at E@H and SZDG: locating many values where \( Z(t) \) (Riemann-Siegel formula) is large in order to get a better understanding of the behavior of the distribution of primes, scans the numbers towards infinite

- Suggested for those who
  - have easily programmable parameter sweep application **running for years**
  - generate jobs in the range of millions
Acceptance of the co-infrastructure concept by EGI: establishing IDGF (Regional) Operation Center

- IDGF OC has been established
- IDGF OC to collect Desktop Grid resources for EGI
- Each site represents a DG server
- So far two sites have been set-up
- EDGeS@home and SZDG
Creating BOINC system with or without a job submission gateway on-demand in a cloud

- User can deploy by one click a BOINC infrastructure that is extended with a science gateway
- The BOINC infrastructure running in the cloud can be extended with ordinary home computers as in ordinary BOINC projects.
- Recommended to user communities having no BOINC expertise

University Cloud

WS-PGRADE

3G Bridge (GBAC)

BOINC Server (GBAC)

BOINC Client (VirtualBox)

BOINC Client (VirtualBox)
Summary: IDGF-SP core production infrastructure

Supported VOs:
- xfered
- farmen
- glida
- hunggrid
- integrgrid
- edggrid, vvo.edggrid.eu
- chem.vul.hu
- edggrid.eu

Volunteer Resources:
- ABC@home
- AlmereGrid
- Charity Engine
- EDGeS@home
- EDGI
- Demonstration
- SZTAKI Desktop Grid
- Westminster Campus DG

Desktop Grid Servers:
- 3G BRIDGE
- Applications:
  - GBAC

Clouds:
- Amazon (on demand)
- Westminster Cloud
- SZTAKI Cloud
- L 1/P 0 Cloud

AutoDock Portal:
- GUSE

Application Repositories:
- EDGI
- IDGF
- egi
Summary

- The BOINC extension (called as SZTAKI Desktop grid) helped to make generally accept the co-infrastructure concept in Europe
- More and more user communities start to use the concept
- If this will generally be accepted then BOINC will be used by massive number of user communities as we planned
Thank you for your attention!

Project websites:
http://idgf-sp.eu
http://desktopgridfederation.eu
http://doc.desktopgrid.hu

Acknowledgement:

- IDGF-SP EU support project (RI- 312297)