Supporting scientific and Web 2.0 communities by desktop grids

P. Kacsuk
MTA SZTAKI
(Univ. of Westminster)
The strength of communities

- **Web 2.0 services** provide frameworks for communities to put together contents

- **Volunteer computing (VC)** is a community activity to provide desktop resources for typically scientific communities -> Desktop Grids (DGs)

- Our goal is to bring together the two communities:
  - Web 2.0 communities can help scientific projects to advertise their projects and attract volunteer donors
  - Web 2.0 communities can extend the scope of their applications by using VC technology provided for science
  - Web 2.0 service providers can increase their services without further resource investment
Goals of the EDGI and Web2Grid projects

- To support both scientific and Web 2.0 communities by desktop grids
- EDGI: enables scientific communities to extend the existing service grids (SGs) with DG resources
- Web2Grid enables Web 2.0 communities to create large computing infrastructure as community effort
Possible options to extend the infrastructure for communities

Scientific Community Users

U1 U2

Un

Grid

Cloud

D1 D2

Donors Dj

Faster, more reliable but more expensive

Web 2.0 Community Users

U1 U2

Um

Web 2.0 Community Service

Cloud

D1 D2

Donors Dk

Less expensive since uses the free cycles of existing PCs
EDGI: Extending the grid infrastructure for scientific communities

EDGeS scope only for compute intensive applications for EGEE (gLite)

EDGI scope for both compute and data intensive applications for EMI/EGI (gLite, ARC, Unicore)

Extend Desktop Grids with Clouds for QoS
Extending the grid infrastructure for scientific communities

- U1, U2, Un: Scientific Community Users
- Cloud: Provides reliable donors from dedicated cloud resources
- Faster, more reliable but more expensive

Grid:
- DG: Donors
- D1, D2, Dj: Donors

Diagram explanation:
- The grid infrastructure is extended to include cloud resources.
- Scientific community users (U1, U2, Un) access the grid through dedicated cloud resources.
- The cloud provides faster, more reliable, but more expensive donors (D1, D2, Dj).

Diagram notes:
- The diagram visualizes the integration of grid and cloud resources to enhance data availability and reliability for scientific communities.
Extending the grid infrastructure for scientific communities

Grid

U1 U2 Un
Scientific Community Users

DG

D1 D2 Dj
Donors

Cloud
Provides reliable donors from dedicated cloud resources
Steps of the technical solution

The starting state
Security issues

• **SG -> DG direction**
  - In DG projects
    • Applications are trusted
    • DG server is trusted
    • Clients are not trusted
  - Therefore any SG application that is intended to run in a DG should be **validated** to make it trusted
  - Therefore EDGI provides:
    • Application validation service
    • **Application repository** (AR) where trusted applications are stored
    • EDGI bridge transfers only those applications to DGs that are stored in the AR
Step 1: Providing application repository

- metadata templates describe applications and their implementations to support browse and search operations
- access without registration for all users to browse and search the repository
- API to enable repo access for modified CEs
Step 2: Submit job from gLite to DG project

- Download
  - User IF
  - AR IF
  - Bridge IF

- Submit
  - CREAM MCE
    - AR
    - attic
    - monitor

- gLite grid

- 3GBridge
- DG Project

- DG client
  - Volunteer/Institutional Resources
Generic Grid-Grid (3G) Bridge to integrate SGs and DGs

3G Bridge

Job Database

Queue Manager

Job Handler Interface

GridHandler Interface

DC-API Plugin

WS interface

SG

BOINC

XtremWeb

XtremWe b Plugin
Step 3: Submit job from ARC to DG project

ARC grid

ARC MCE
- AR
- attic
- monitor

ARC grid

CREAM MCE
- AR
- attic
- monitor

3GBridge

DG Project

DG client

Volunteer/Institutional Resources

User IF

AR

Bridge IF

submit

download

submit
Step 4: Support for Data-intensive applications
Usage of Attic P2P File System in EDGI
Step 5: QoS Support by Clouds

ARC grid
- ARC MCE
  - AR
  - attic
  - monitor

ARC grid
- gLite grid
  - CREAM MCE
  - AR
  - attic
  - monitor

gLite grid
- upload

OpenNebula/Amazon
- DG client
  - attic

3GBridge
- Attic FS
  - DG Project
  - DG client
  - attic

Volunteer/Institutional Resources
- DG client
  - attic

User IF
- AR
- Bridge IF

submit

download

3G Bridge Extension for supporting Clouds

1. Job is submitted from MCE to 3G Bridge.
2. 3G Bridge submits the job to a BOINC DG using the BOINC Plugin.
3. The Scheduler keeps track of the number of jobs in the BOINC queue (Queue 1) and of the number of the running Cloud Resources (workers).
4. If the BOINC DG is overloaded, the Scheduler starts new workers by submitting a job to the queue of the Cloud Plugin (Queue 2).
5. Each job in Queue 2 launches a new BOINC Worker in the Cloud.
6. If the cluster is underutilized, the Scheduler stops some workers (cloud resources) by sending cancel jobs into Queue 2.
Step 6: Monitoring

ARC grid
ARC MCE
ARC
attic
monitor

OpenNebula/
Amazon
DG client
attic

User IF
AR
Bridge IF

3GBridge
attic
cloud
monitor

DG Project

CREAM MCE
AR
attic
monitor
gLite grid

Monitor
UI

Upload

Volunteer/
Institutional Resources

Submit

Download

Inspect
**EDGI business model**

**Grid**

- SC Users can ask QoS support (cloud resources) if they have collected credits

**Institutes of SC users can donate DG resources and can collect credits (by BOINC mechanism)**

**Students and public can support universities by offering their credits**

**EDGI provides reliable donors from dedicated cloud resources**

**After EDGI, IDGF members or cloud providers can provide cloud resources**
Universities can create local DGs (e.g. Univ. of Westminster, Univ. of Portsmouth, Univ. of Szeged, etc.)

University level DG could be connected to the EDGI DG and collect credit for the univ. researchers
WEB2GRID Project

Building comm.
Collect resources
Applications
Services for comm.

Existing community appl1
Existing community appl2
Watermark appl.
Media appl.
New community apps
Facebook interface for watermarking photos

- To prevent the illegal use of unprotected photos
- Use of steganographic methods, embedding hidden watermarks into the photos
- Protected images can be found by a search-bot
- Cryptography (timestamp and digital signature) is used to show the ownership rights
Example Web2Grid application: watermarking photos

- Job 1
- Job 2
- Job 3

- Grid applications
  - Watermark Workflow
  - Media Workflow
  - ...

- 3G Bridge

- Desktop Grid
  - Watermark App
  - Media App
  - ...

- Digital signature

- Micropayment (Abaqoos)

- "Credit"

Example watermarking photos
Possible business models

• **Service provider provides the whole framework**

• **Supplier model:**
  – Service provider invites trusted individuals to connect clients
  – Pays for the collected credits

• **Volunteer model:**
  – Web 2.0 community members provide the clients
  – They collect credits
  – Credits are transferred to real money and offered to charity organizations
Support those NGIs that want to extend their VO with volunteer or local DGs

Support those universities, institutes, companies who want to set up local DG

Run dissemination campaign to attract more volunteers to donate their computer resources

Federation is open for organisations and individuals in research and industry. Current members:

- >100 individuals
- > 30 organizations (including companies)
Summary

- Volunteer DG resources can be provided for scientific and Web 2.0 communities as an inexpensive alternative to cloud resources
- EDGI and Web2Grid projects have developed the required technical solutions
- They also show possible business models
- Future task: to implement the business models
Thank you for your attention ...

For more information please visit the EDGeS and EDGI Websites:

http://www.edges-grid.eu/
http://edgi-project.eu

and/or send e-mail to me:

kacsuk@sztaki.hu