RINGrid
Evaluation of Remote Instrumentation Infrastructures

Thomas Prokosch
prokosch@gup.uni-linz.ac.at

GUP, Joh. Kepler University Linz, Austria

OGF 21: Seattle, WA, US

Funded by the EC under 6FP
(contract number 031891).

RINGrid

Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil

Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs

Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling

Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling
  - experiments may fail

Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling
  - experiments may fail
  - limited expertise on instrument

Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling
  - experiments may fail
  - limited expertise on instrument
- idea: use these instruments over the Internet

RINGrid

RI Advantages

- resources are shared
- **instruments are better utilized**
- expertise is increased in LA countries
- new communities formed
- expenses are justified
  better instruments will be bought
- new opportunities created

Problem Definition

What does an experiment look like?

- data retrieval (instrument as grid component)
- raw data (storage on the grid)
- data processing (on the grid)
- collaboration (needs grid support, timezones)
- filtered data (results, stored on the grid)

workflow management (in need of grid support)

Research Approach

Traditional approach:

- What do user require?
- What is there?
- What is missing?
- Develop an architecture.
- Verification: Test the architecture.

Research Approach

Traditional approach:

- What do user require?
- What is there?
- What is missing?
- Develop an architecture.
- Verification: Test the architecture.

User Requirements

Outline of this section:

1. classification of instruments
2. instrument information form
3. user identification form
4. problems identified

Classification

RINGrid's WP2 identified user requirements with the help of an **instrument information** and **user identification forms**.
When talking about instruments, 3 classes are important:

- **local**
  widely available, low cost, limited interest for remote access
- **regional**
  some interest for remote access, mostly within the same country/county
- **global**
  great interest in remote access, mostly supported by international science collaborations

Instrument Information Form

Three details were asked:

1. **basic information about the instrument**
   type, manufacturer, model, year of manuf., price, owner
2. **technical parameters**
   concerning user access and experimental measurements
3. **current and potential possibilities**
   of remote access to the instruments

User Identification Form

It consists of three main sections:

1. **basic information about the user**
   position, affiliation, contact information
2. **experience with remote access equipment**
   research field, years of experience, conventional equipment used
3. **feedback**
   everything else the user wants to say

Problems Identified

It is not hard to see what users require.
More interesting are the problems user face. They are our working areas.

- video/display problems
- lack of information which is available in the control room
- tunnel stability
- (network speed)
- real time control
- firewalls
- closed source software

Middleware Components for RI

Experiments conducted remotely require:

1. workflow management
2. interactive experiment steering
3. data transmission
4. data storage
5. collaboration

Workflow Management

"Where can data for the next step in the experiment be retrieved?"

- dependencies inherent within an experimental setup
- for automating processes, i.e. experiments

data from instruments have to be **stored near** the instrument
data needs to be **forwarded** to other hosts for processing
need to **keep track** of data

Workflow Management Solutions

We had a look at:

- **g-Eclipse**
  workflow capability cannot be used independently
- **Yet Another Workflow Language (YAWL)**
  execution engine tied to graphical editor
- **XML Process Definition Language (XPDL)**
  create diagrams based on the XML files
- **Business Process Execution Language (BPEL)**
  long-running applications
  span multiple organizational entitites

Middleware Components for RI

Experiments conducted remotely require:

1. workflow management
2. interactive experiment steering
3. data transmission
4. data storage
5. collaboration

RINGrid

Interactive Experiment Steering

utilize rapid turnaround

try several variations of parameters $\longrightarrow$ better results

Interactive Experiment Steering

utilize rapid turnaround
try several variations of parameters $\longrightarrow$ better results

Requirements:

- **display of** (preliminary) **experimental data** (simple/complex)
  data rate, evtl. latency
- **show control elements for instrument**
  latency

Visualization Components

glogin

- bi-directional channels
- forward data securely (X11, TCP, VPNs, shell)
- no account needed

GVid

video rendering on any node, complex tasks split

- transmission of video data to user's desktop
- interaction elements communicated back

missing: adapt components for RI (latency, etc.)
encoding of instrument controls

Middleware Components for RI

Experiments conducted remotely require:

1. workflow management
2. interactive experiment steering
3. data transmission
4. data storage
5. collaboration

Data Transmission

Transmission qualities:

- high connection bandwidth
- reasonable round trip time
- small jitter
- no packet loss

Data Transmission

Transmission qualities:

- high connection bandwidth
- reasonable round trip time
- small jitter
- no packet loss

Luckily, not all applications demand all these qualities.

Classify applications by the requirements and choose underlying network connectivity appropriately.

Data Transmission Gaps

We have already done a first classification of RI applications.

Missing:
Verification of classification by conducting experiments.

Middleware Components for RI

Experiments conducted remotely require:

1. workflow management
2. interactive experiment steering
3. data transmission
4. data storage
5. collaboration

Data Storage

huge amounts of data (think LHC)
not a problem for a single experiment
data archive/digital library

Data Storage

huge amounts of data (think LHC)
not a problem for a single experiment
data archive/digital library

1. reduce data to a reasonable amount
   data reduction, compression, preliminary analysis

Data Storage

huge amounts of data (think LHC)
not a problem for a single experiment
data archive/digital library

1. reduce data to a reasonable amount
   data reduction, compression, preliminary analysis
2. distribute data among several grid nodes

Data Storage

huge amounts of data (think LHC)
not a problem for a single experiment
data archive/digital library

1. reduce data to a reasonable amount
   data reduction, compression, preliminary analysis
2. distribute data among several grid nodes
3. retrieval of data with data **management service**

Data Storage: Pandora's Box

- **data movement**
  done: GridFTP/RFT (Reliable File Transfer)
- **data replication and access**
  good management of data metadata
- **data consistency**
  data locking, lazy-copy (consistency problems)
- **movement planning**
  access patterns, statistical methods
- **replica management**
  finding an optimal place for replicas is NP-complete (Wolfson, Milo)

RINGrid

Middleware Components for RI

Experiments conducted remotely require:

1. workflow management
2. interactive experiment steering
3. data transmission
4. data storage
5. collaboration

Collaboration

good thing: no hard technical restrictions

- **communication**
  - instant messaging software like Skype, Gadu-Gadu
  - custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)

Collaboration

good thing: no hard technical restrictions

- **communication**
  - instant messaging software like Skype, Gadu-Gadu
  - custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)
- **common file system**
  - Storage Resource Broker (SRB by San Diego Supercomputer Center)

Collaboration

good thing: no hard technical restrictions

- **communication**
  - instant messaging software like Skype, Gadu-Gadu
  - custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)
- **common file system**
  - Storage Resource Broker (SRB by San Diego Supercomputer Center)

Problem: Missing integration.

Summary

1. We need to identify user requirements.
   - Especially address the problems mentioned.
2. We need to choose middleware standards.
3. Middleware needs to be adopted.
   - adaptation of a workflow management solution for RI
   - adaptation of visualization components
   - verify application classification in terms of data transmission
   - data storage: replica management
4. An architecture needs to be created.

RINGrid