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).



Thomas Prokosch RINGrid: Evaluation of Remote Instrumentation Infrastructures

Remote Instrumentation Research Approach

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 Research Approach

Remote Instrumentation (RI)

- 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 Research Approach

Remote Instrumentation (RI)

- 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 Research Approach

Remote Instrumentation (RI)

- 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 Research Approach

Remote Instrumentation (RI)

- 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

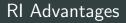


Remote Instrumentation Research Approach

Remote Instrumentation (RI)

- 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

Remote Instrumentation Research Approach



- 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



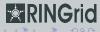
Remote Instrumentation Research Approach

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)



Remote Instrumentation Research Approach

Research Approach

Traditional approach:

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



Remote Instrumentation Research Approach

Research Approach

Traditional approach:

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



Classification Instrument Information Form User Identification Form Problems Identified

User Requirements

Outline of this section:

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



Classification Instrument Information Form User Identification Form 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

Classification Instrument Information Form User Identification Form Problems Identified

Instrument Information Form

Three details were asked:

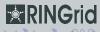
basic information about the instrument type, manufacturer, model, year of manuf., price, owner

2 technical parameters

concerning user access and experimental measurements

3 current and potential possibilities

of remote access to the instruments



Classification Instrument Information Form User Identification Form Problems Identified

User Identification Form

It consists of three main sections:

- **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



Classification Instrument Information Form User Identification Form **Problems Identified**

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

Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

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 Interactive Experiment Steering Data Transmission Data Storage 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 Interactive Experiment Steering Data Transmission Data Storage Collaboration

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

< ロ > < 同 > < 三 > < 三 >

Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Middleware Components for RI

Experiments conducted remotely require:

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Interactive Experiment Steering

utilize rapid turnaround try several variations of parameters \rightarrow better results



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Interactive Experiment Steering

utilize rapid turnaround try several variations of parameters \rightarrow better results

Requirements:

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Middleware Components for RI

Experiments conducted remotely require:

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Data Transmission

Transmission qualities:

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

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.

Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Data Transmission Gaps

We have already done a first classification of RI applications.

Missing:

Verification of classification by conducting experiments.



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Middleware Components for RI

Experiments conducted remotely require:

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration



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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration



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

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration



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

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

Data Storage

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

- 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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

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)

Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration

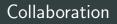
Middleware Components for RI

Experiments conducted remotely require:

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration



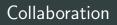
good thing: no hard technical restrictions

communication

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



Workflow Management Interactive Experiment Steering Data Transmission Data Storage Collaboration



good thing: no hard technical restrictions

communication

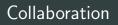
- instant messaging software like Skype, Gadu-Gadu
- custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)

common file system

 Storage Resource Broker (SRB by San Diego Supercomputer Center)



Workflow Management Interactive Experiment Steering Data Transmission Data Storage 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.

(日)



- 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.