GISolve

A Grid-based problem solving environment for computationally intensive geographic information analysis

Shaowen Wang

Department of Geography &
Grid Research & educatiOn group @ IoWa (GROW)
ITS Academic Technologies – Research Services
The University of Iowa
Iowa City, IA 52242, USA

July 24, 2005

THE UNIVERSITY OF IOWA

G.R.O.W. Grid Research and Education Group at Iowa

Collaborators

- Marc P. Armstrong, Professor, Department of Geography and Program in Applied Mathematical and Computational Sciences
- Jun Ni, Ph.D., Associate Research Scientist, ITS Academic Technologies – Research Services
- Yan Liu, Graduate Research Assistant, ITS Academic Technologies – Research Services

Purpose

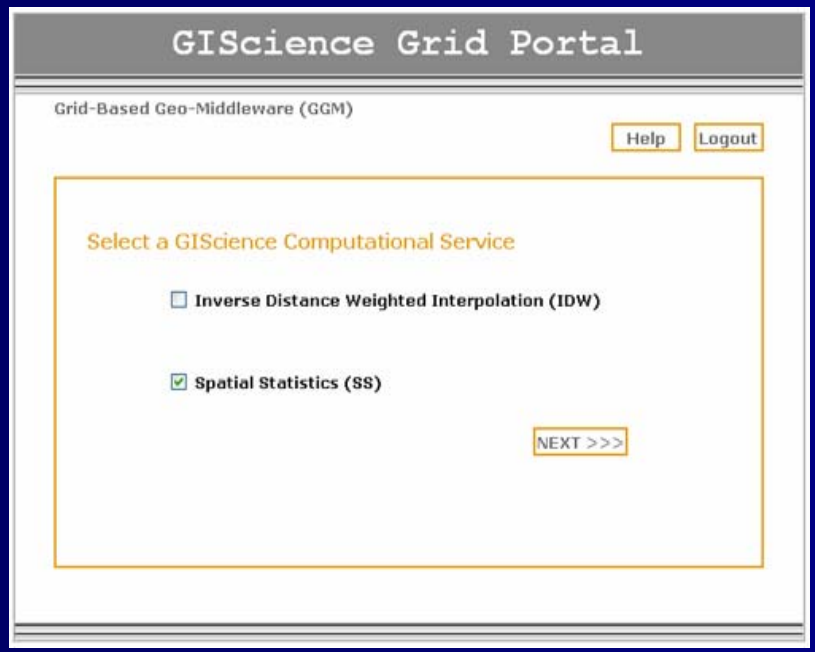
- To address why the Grid is important to geographic information analysis research
- To demonstrate the design and implementation of GISolve – a Grid-based problem solving environment for computationally intensive geographic information analysis

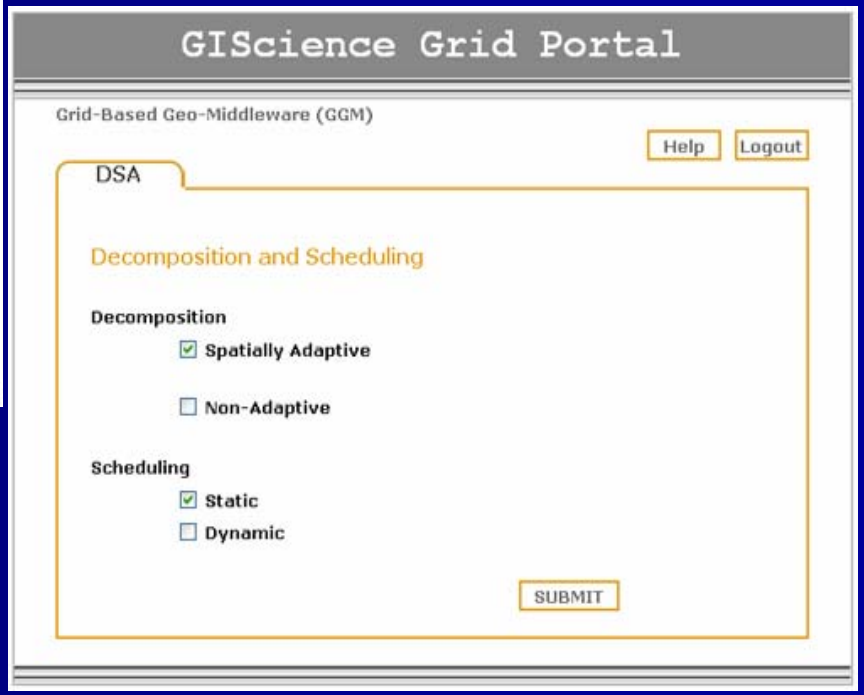
Background

- The amount of geographic information
  - Ever increasing
    - Application driven
  - GPS, LBS, RS
- Computationally intensive geographic analysis
  - Heuristic search
  - Simulation
  - Optimization
  - Statistical methods
- Grid computing
  - Cyberinfrastructure
  - Open Grid Service Architecture (OGSA)
    - Web services
- Problem solving environments
  - Grid portals

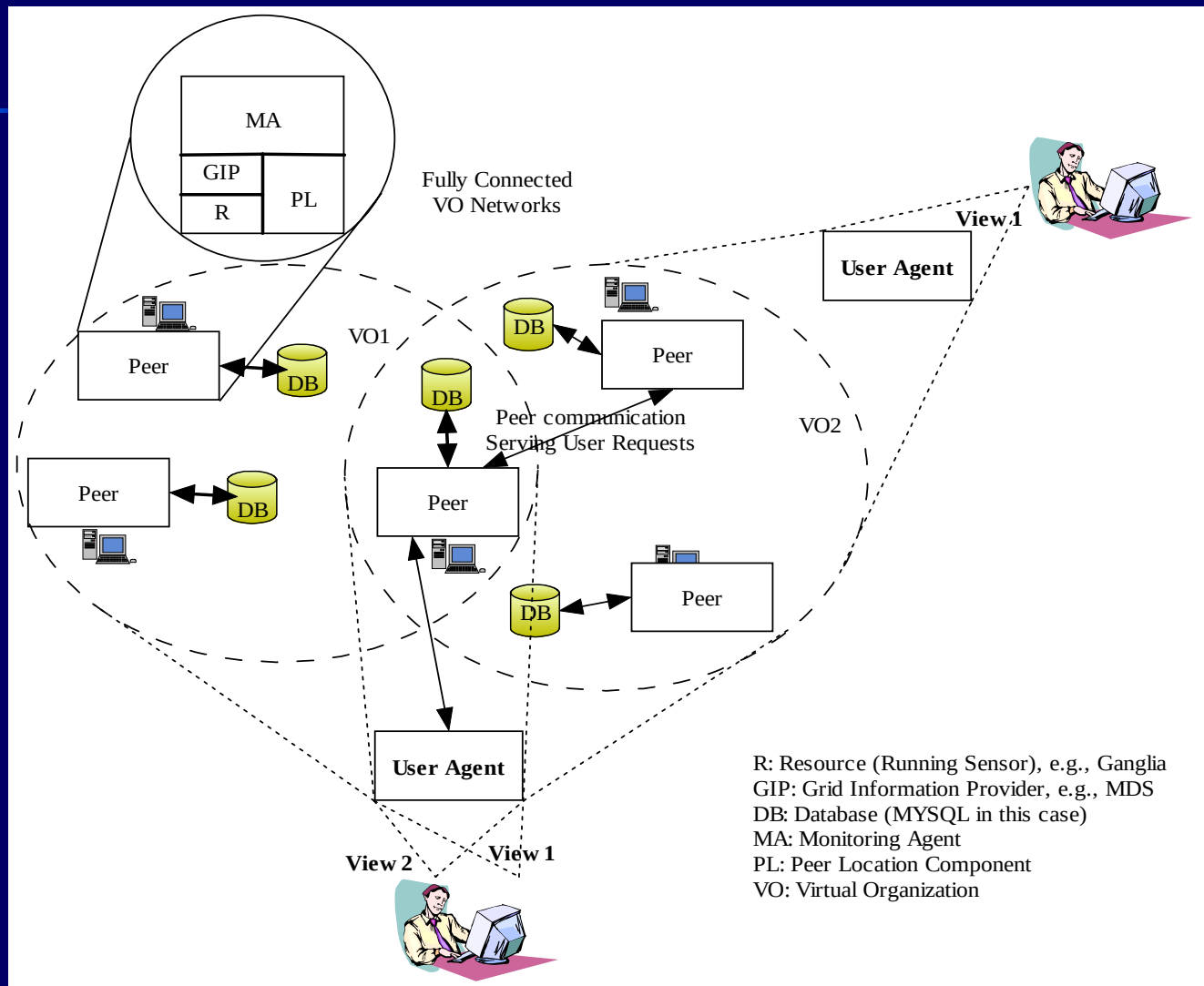
GIScience Grid Portal

- Purpose
- **Background**
- Design
- **Implementation**
- Case Study
- Evaluation
- Conclusions





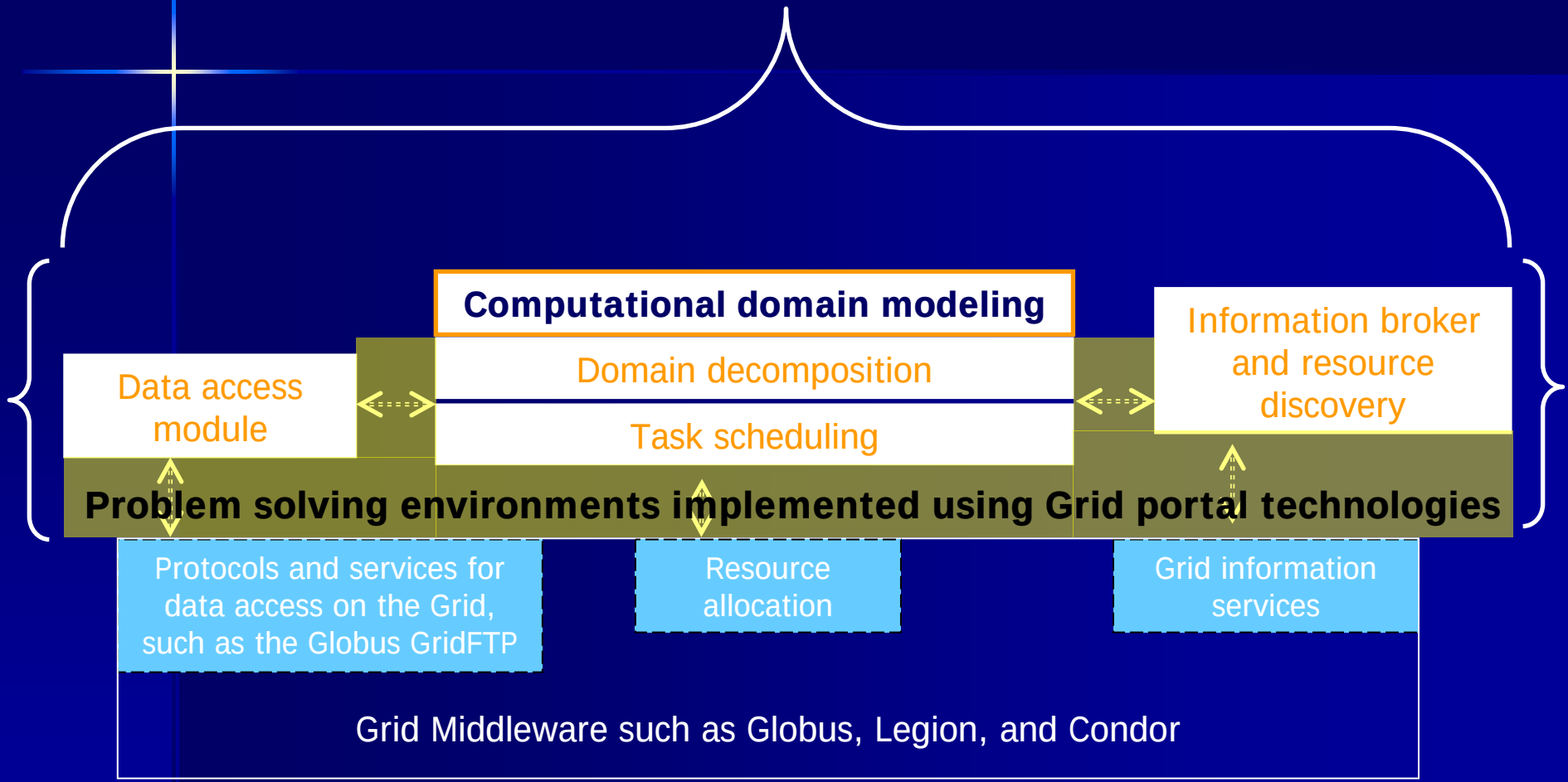
Grid Resources – A User View



Grid Complexity

- Grid middleware limitation
  - Provides a selected set of protocols and services
    - Not directly user- or application- oriented
- Heterogeneous
- Dynamic
- Administrated in different security domains
- Faults
  - Usually unpredictable

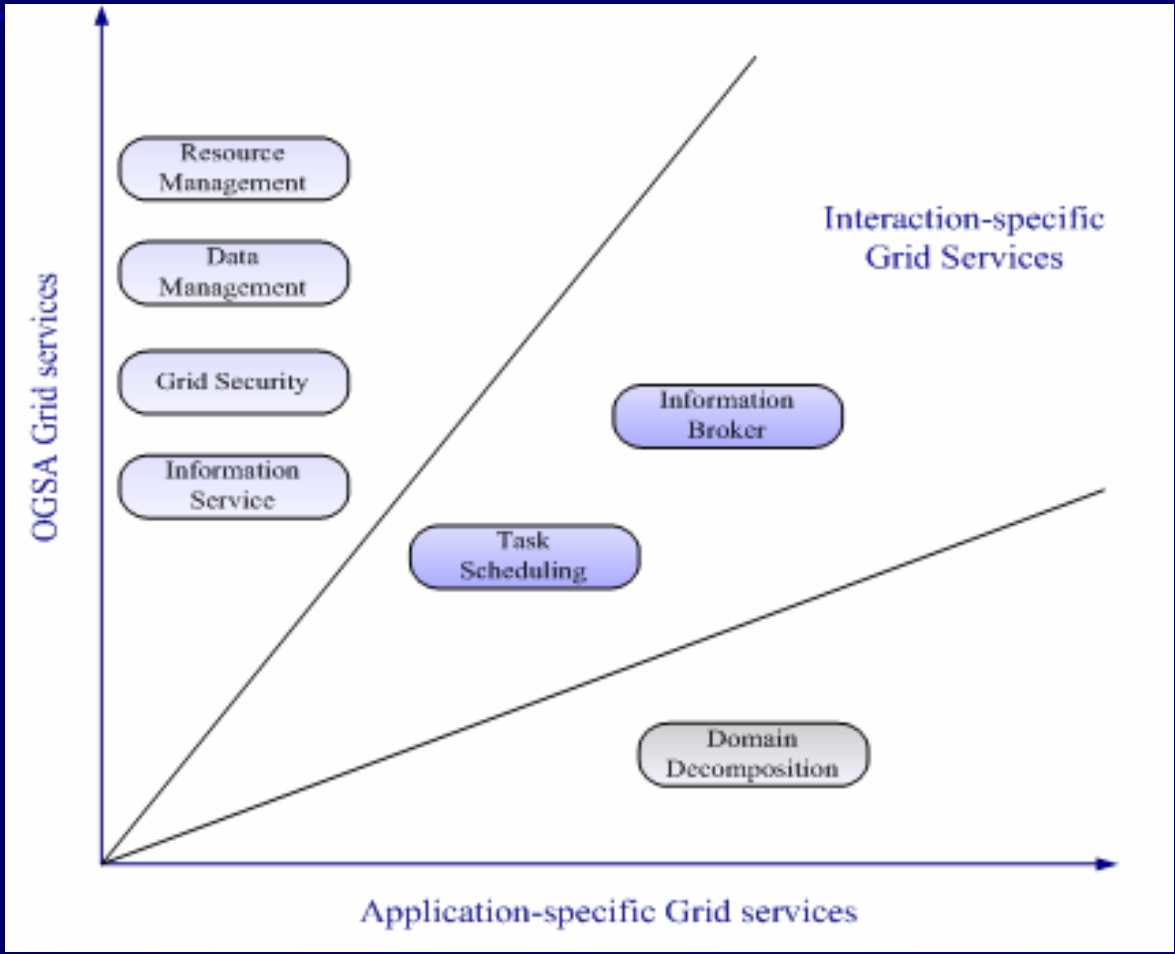
GISolve Architecture



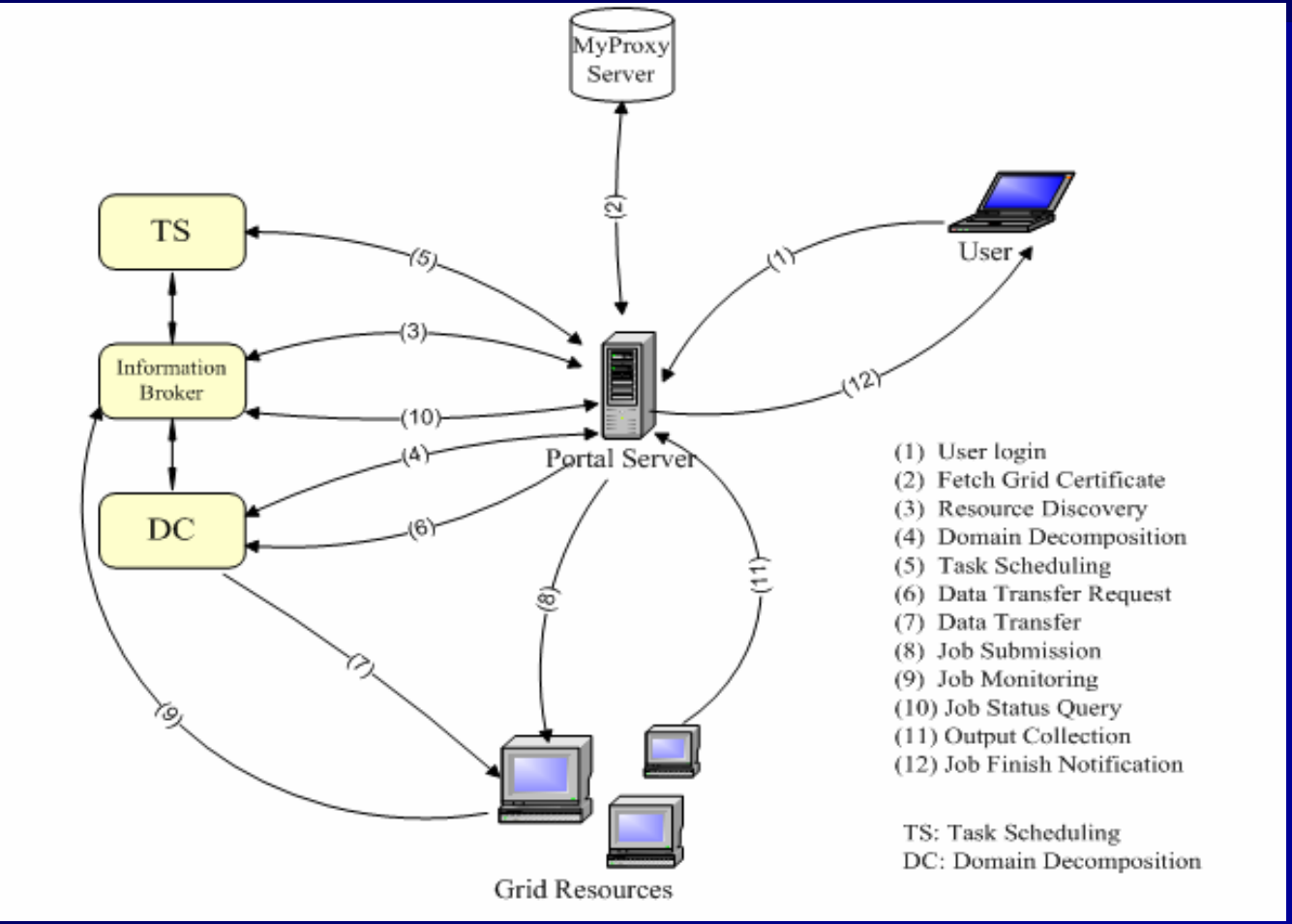
GISolve Services

- Security
- Decomposition and task scheduling
- Geographic data access
- Resource information brokering

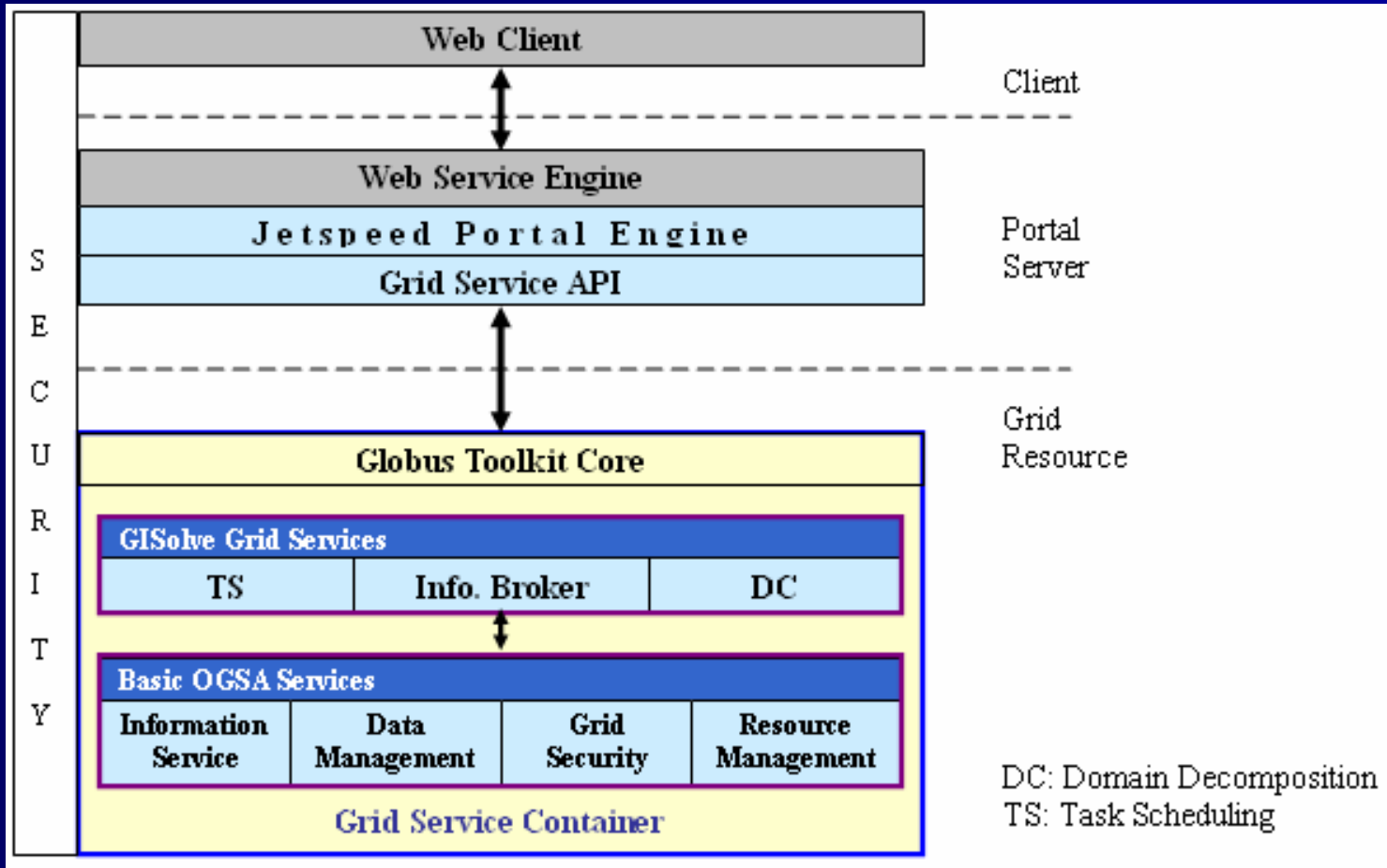
Classification of GISolve Services



GISolve Workflow



A Three-Layer Model of Grid Portal Technologies



Technology Specification

- Jetspeed portal server
  - Jetspeed: portal engine
  - Turbine: MVC (Model-View-Control) framework
  - Velocity: dynamic html generation
  - Tomcat: portal container
- Multi-user support
  - User registration and management
  - User-based state management
- Configurable portal interface
  - Customized layout
  - Display control of each portlet interface

APIs Implemented

- APIs developed
  - [org.gisolve.globus]: Globus Toolkit enhancement
  - [edu.uiowa.gisolve.ogsa]: Grid service (interface definition, user-level Grid service implementation)
  - [org.apache.jetspeed.modules.actions.gisolveportlets]: Jetspeed portlet action implementation
  - [gisolveportlets]: portal support for Grid service APIs (state management, Grid service client APIs)
- APIs integrated
  - Java COG (GT2), GT3/GT4 Core, OGSA (WS GRAM, RFT), MyProxy, LDAP, JNI, JPL

Defining GISolve Services

- A task scheduling service gWSDL:

```
<definitions name="TSService targetNamespace=http://uiowa.edu/grow/namespaces/2004/04/gisolve/TS ... />
<!-- import external types -->
<import location="../ogsi/ogsi.gwsdl“ .../>
<!-- self-defined types -->
<import location="gidTypes.xsd“ namespace="http://uiowa.edu/grow/namespaces/2004/04/gisolve/gidtypes"/>
<!-- local types schema -->
<types><xsd:schema .../></types>
<!-- message definition -->
<message name="SchedInputMessage">...</message>
<message name="SchedOutputMessage">...</message>
<!-- service operation definition -->
<gwsdl:portType name="TSPortType" extends="ogsi:GridService">
        <operation name="schedule">
                <input message="tns:SchedInputMessage"/>
                <output message="tns:SchedOutputMessage"/>
                <fault name="Fault" message="ogsi:FaultMessage"/>
        </operation>
        <sd:serviceData name="TSState“ ... />
</gwsdl:portType>
```

OGSI to WSRF

WSDL

```
<?xml  version="1.0" encoding="UTF-8"?>
<definitions name="DCService" ... ...
   xmlns:wsrp="http://docs.oasis-open.org/wsrf/2004/06/wsrf-WS-ResourceProperties-1.2-draft-01.xsd"
   xmlns:wsrpw="http://docs.oasis-open.org/wsrf/2004/06/wsrf-WS-ResourceProperties-1.2-draft-01.wsdl"
   xmlns:wsdlpp="http://www.globus.org/namespaces/2004/10/WSDLPreprocessor" >
   <wsdl:import namespace="http://docs.oasis-open.org/wsrf/2004/06/wsrf-WS-ResourceProperties....wsdl"
      location="../wsrf/properties/WS-ResourceProperties.wsdl"/>
   <types>... ...<xsd:element name="DCResourceProperties">... ...</xsd:element></types>
   <!-- WSRF Grid service PortType definition -->
   <portType name="DCPortType" wsdlpp:extends="wsrpw:GetResourceProperty"
      wsrp:ResourceProperties="tns:DCResourceProperties">
      <operation name="decompose">... ...</operation>
   </portType>
```

gWSDL

```
<?xml  version="1.0" encoding="UTF-8"?>
<definitions name="DCService“ ... ...
   xmlns:ogsi="http://www.gridforum.org/namespaces/2003/03/OGSI"
   xmlns:gwsdl="http://www.gridforum.org/namespaces/2003/03/gridWSDLExtensions"
   xmlns:sd="http://www.gridforum.org/namespaces/2003/03/serviceData">
   <!-- import ogsi Grid Service gwsdl. Location is under gt3/schema/ogsi/ -->
   <import location="../ogsi/ogsi.gwsdl" namespace="http://www.gridforum.org/namespaces/2003/03/OGSI"/>
   ...   ...
   <!-- Grid service interface definition -->
   <gwsdl:portType name="DCPortType" extends="ogsi:GridService">
      <operation name="decompose">... ...</operation>
      <sd:serviceData name="DCState" type="xsd:int" ... ... />
   </gwsdl:portType>
```

GISolve Services in Action

- Deployed GISolve services:

```
[globus4@rsgridportal globus4]$ globus-start-container
Starting SOAP server at: https://128.255.162.167:8443/wsrf/services/
With the following services:
 ... ...
[12]: https://128.255.162.167:8443/wsrf/services/gisolve/TSService
... ...
[47]: https://128.255.162.167:8443/wsrf/services/gisolve/DCService
```

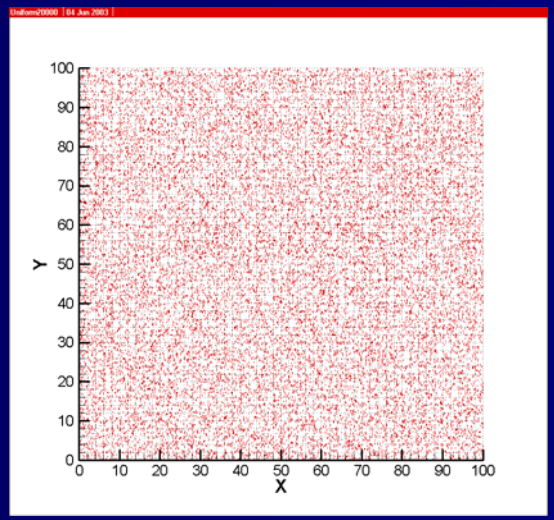
Case Study – $G_i^*(d)$ Statistic

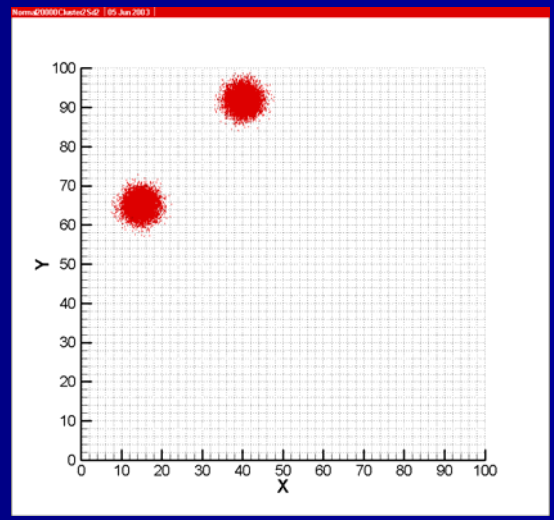
$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - W_i^*\bar{x}}{s\{[(nS_{1i}^*) - W_i^{*2}]/(n-1)\}^{1/2}}$$

- References
  - Getis, A., and Ord, J.K., 1992, The analysis of spatial association by use of distance statistics. *Geographical Analysis*, **24**(3): 189-206.
  - Ord, J. K., and Getis, A., 1995, Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, **27**(4): 286-306.

Synthetic Datasets

Purpose
Background
Design
Implementation
Case Study
Evaluation
Conclusions



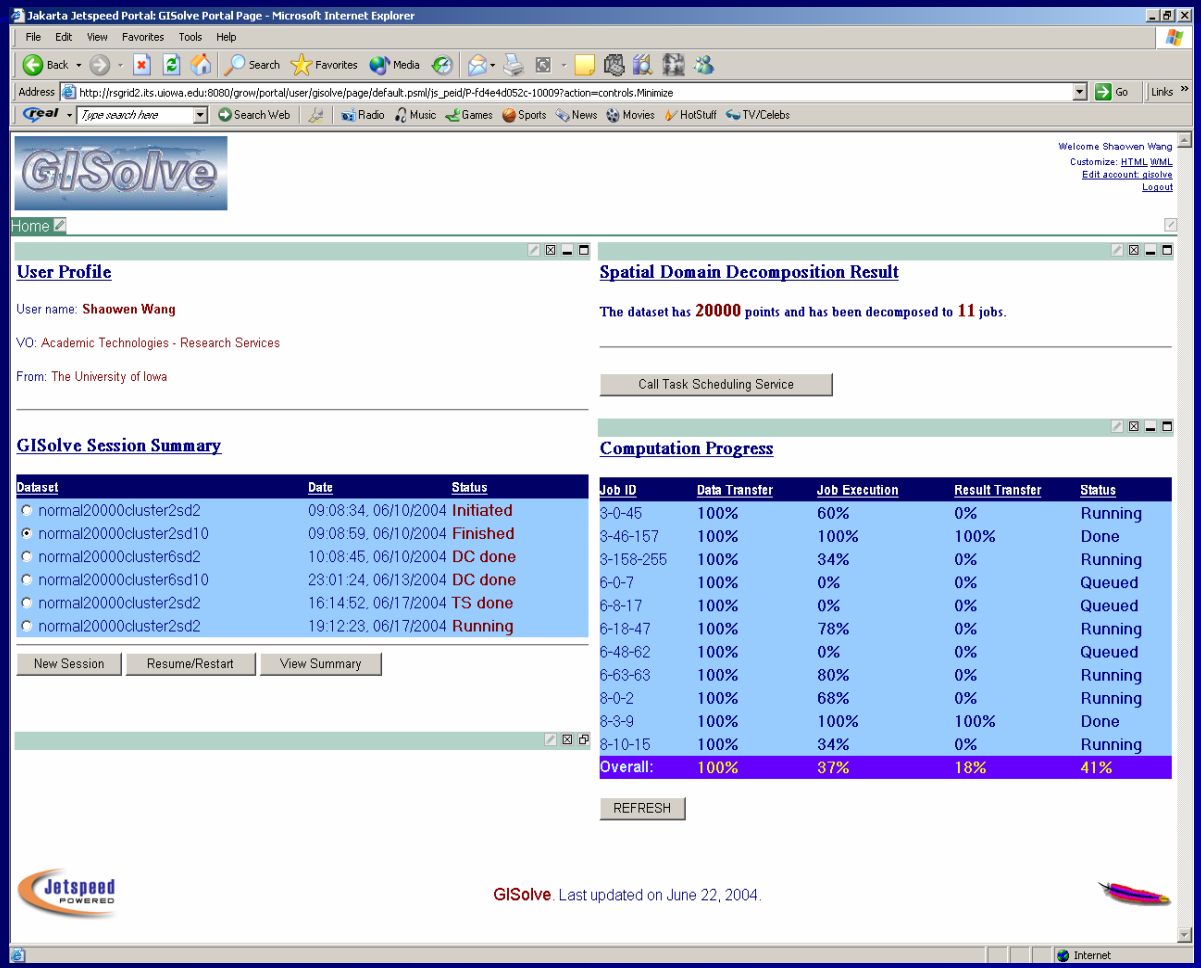


20000-point a uniform random distribution

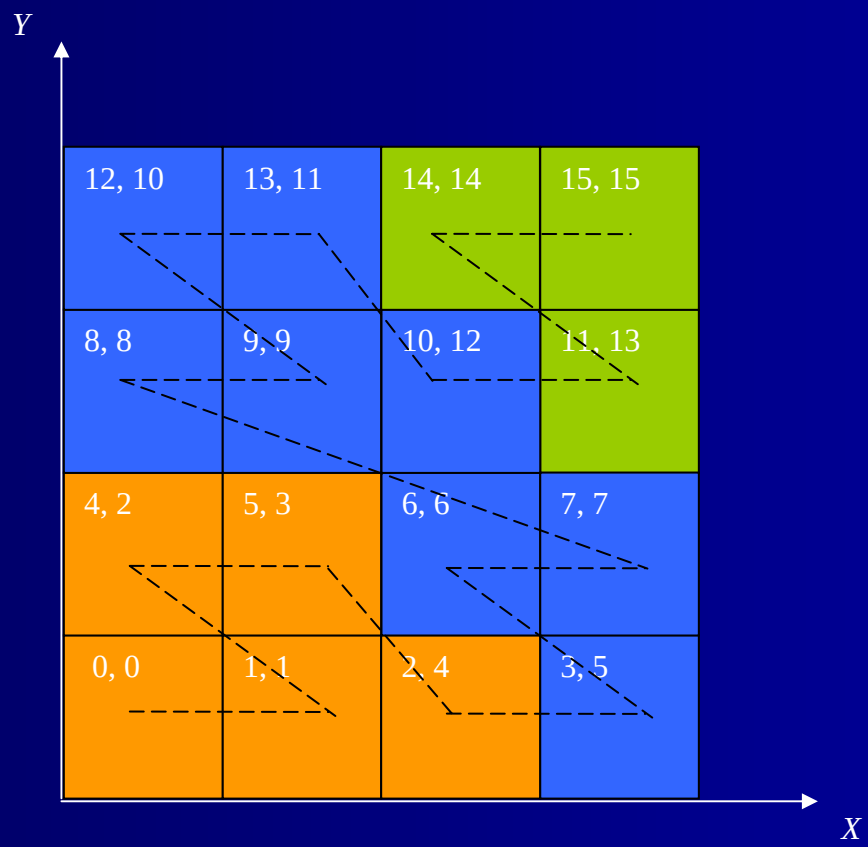
20000-point : two clusters, each of which has a normal distribution with a standard deviation of 2

User Interface

- Purpose
- Background
- Design
- **Implementation**
- **Case Study**
- Evaluation
- Conclusions



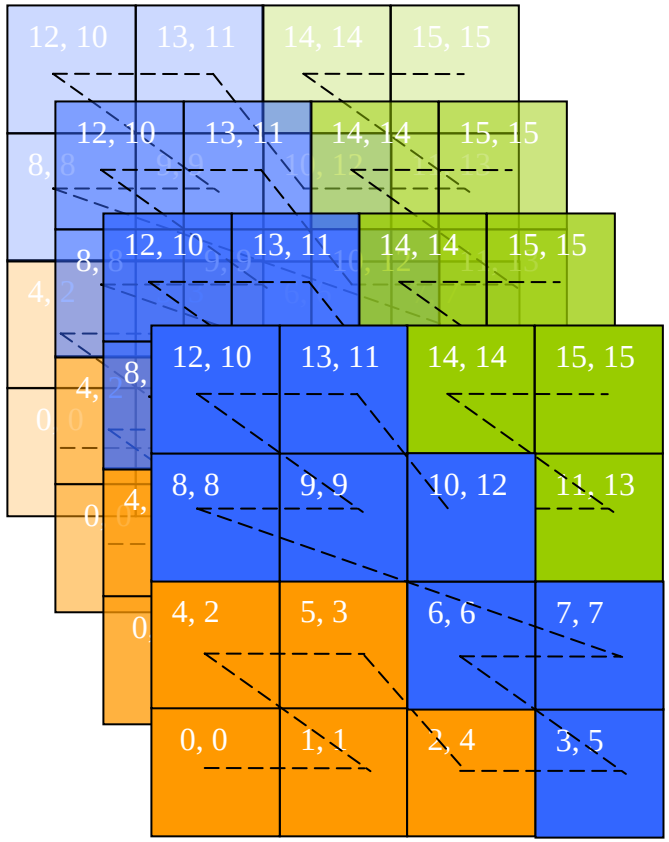
An Example of Spatial Computational Domain Decomposition

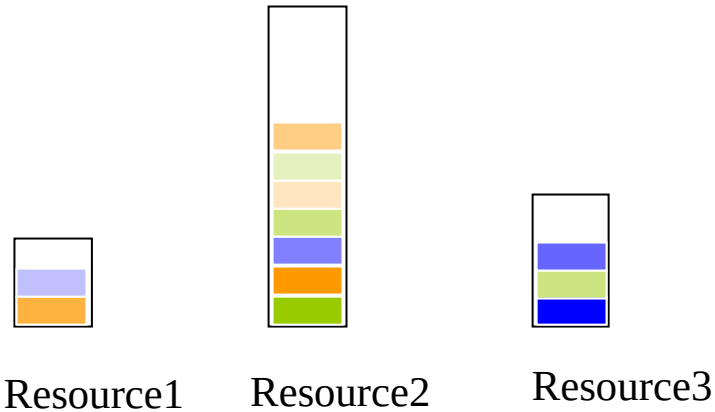


A Task Scheduling Service

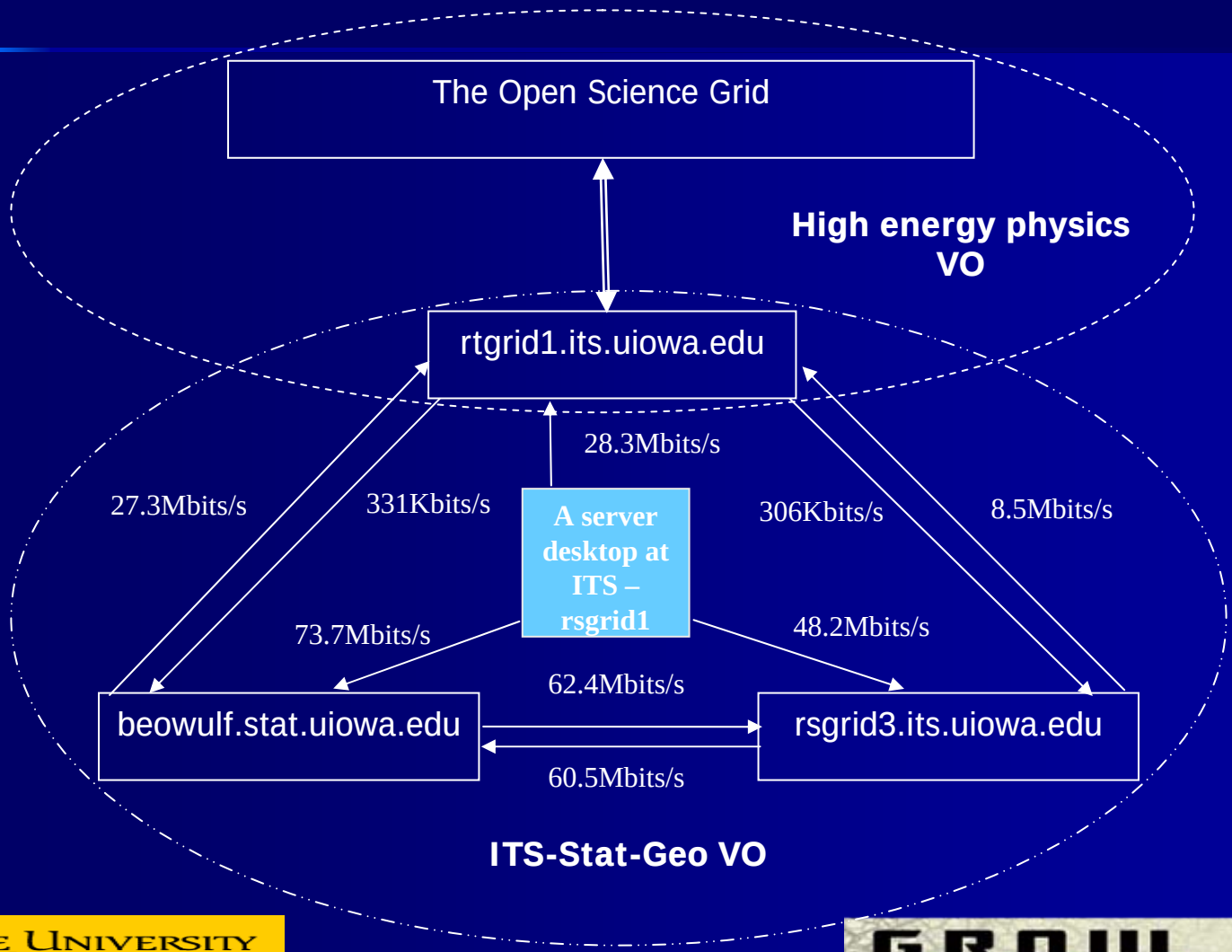
- Decoupled from domain decomposition services
- *NP*-complete problem
  - Max-min algorithm implemented
- Using the theoretical estimate to compute the Expected Time to Compute (ETC) matrix

A Diagrammatic Example of Static Task Scheduling





Grid Testbed Implementation - HawkGrid



Purpose

Background

Design

Implementation

Case Study

Evaluation

Conclusions

DEMO

Performance

- The problem cannot be solved by any single computer in our Grid testbed
  - 3.2G memory
- Solved within 10 minutes through GISolve

Conclusions

- GISolve demonstrates how the Grid can benefit research on computationally intensive geographic information analyses
- GISolve integrates OGSA-based Web services to support the computational aspects of GIServices

Ongoing Research

- Interoperability of GISolve services
- Adaptive domain decomposition services
- Evaluation of GISolve performance
- Extension of the types of geographic information analyses GISolve supports

Acknowledgement

- This research is partially supported by the HawkGrid project funded by the Office of Vice President for Research at The University of Iowa
- Computational resources used for experiments include a cluster supported by the U.S. NSF iVDGL (international Virtual Data Grid Laboratory) project

THE UNIVERSITY OF IOWA

G.R.O.W. Grid Research and Education Group at Iowa