NetSolve:
Grid Enabling Scientific Computing Environments
Keith Seymour, Asim Yarkhan, Sudesh Agrawal, and Jack Dongarra
CS Department, University of Tennessee
Presented by: Maoyuan Xie

Outline
- Introduction
- NetSolve
- GridRPC API
- GridSolve
- Applications
- Related works
- Conclusion

Introduction
NetSolve is developed as the middleware necessary for providing a seamless bridge between the simple, standard programming interfaces and desktop systems and the services supported by the emerging Grid architecture.

Unlike Globus or Condor, which are some toolkit for resource management or scheduling, NetSolve is only API for easy access to Grid resources.

NetSolve
1. Architecture
It “provides remote access to hardware and software resources from a variety of scientific problem solving environments such as Matlab, Mathematica, and Octave, as well as traditional programming languages like C and Fortran”.

To interface with other existing Grid services seamlessly, NetSolve supports servers which acts as a proxy to those services and also client by using GridRPC API.

NetSolve (Cont.)
2. Feature
a) Ease of use
NetSolve provides interfaces to interactive scientific computing environments (Matlab, Mathematica and Octave). In contrast to typical RPC systems, “the NetSolve client library determines the calling sequence at run-time by downloading a specification from the server” and this “is used to determine how to marshall the arguments to and from the server”, therefore, makes it a straightforward NetSolve interactive environment.

b) Pre-configured interfaces to a variety of numerical routines.
c) Agent as Resource Broker
Uses information from computation resources as well as clients to estimate fastest time to completion, and returns a list of qualified servers to the client.

d) Resource Aggregation
All services are treated as external library of NetSolve, which makes same problem solved by different services possible and transparent to client.

e) Transparent Fault Tolerance
NetSolve automatically resubmits failed job to the next server on the list initially sent to client until no server remaining in the list.

f) Request Sequencing
To avoid unnecessary data transfer, NetSolve performs a data flow analysis of input and output parameters of every request in the sequence to produce a DAG that illustrates tasks and their dependencies.

To exploit more parallelism and less communication overhead when no single server has all the software required to execute the entire sequence.

g) Task Farming in NetSolve
Special API provided for task farming, constructing arrays for different requests’ parameters and doing actual farming.

No interference from user until all tasks have been completed.

h) Distributed Storage Infrastructure
DSI helps the NetSolve to operate in a cache-like manner and provides access to commonly available storage service systems.

GridRPC API
A standardized, portable, and simple programming interface for remote procedure call over the Grid, and a layer on top of the normal API.

GridSolve
Three major concerns for future work (GridSolve):
1. Server-side ease of use
2. Interoperability
3. Scalability
GridSolve (cont.)

• Network Address Translators
  In order to address the problem of potential duplication of IP addresses, GridSolve introduces “a mechanism that allows a client to submit a problem, break the connection, and reconnect later at a more convenient time to retrieve the results”.

  To implement this, GridSolve uses unique component ID and Grid-Solve proxy, which acts on behalf of the component.

GridSolve (cont.)

• Scheduling Enhancements
  GridSolve will allow clients to filter the available resources according to their needs and servers to specify constraint on clients. In this case, agent will have to store as much information as possible, which will cause a performance bottleneck as number of resources increases and call for multi-agent architecture.

GridSolve (cont.)

• IDL Improvements
  A parallel service for interface definition; Re-examining periodically or when it receives the appropriate signal.

Applications

- Environmental Modeling
- Statistical Parametric Mapping
- Image Compression
- Vertex Cover and Clique Problems
- Genetic Algorithms

Application in Environmental Modeling

Along with using GIS map information, Across Tropic Level System Simulation (ATLSS) developed and integrated a set of models to compare the future effects of alternative hydrologic scenarios on the biotic components of the systems.

1) process models for lower trophic levels (including benthic insects, periphyton and zooplankton),
2) structured population models for five functional groups of fish and macroinvertebrates, and
3) individual-based models for large consumers (wood storks, great blue herons, white ibis, American alligators, white-tailed deer, and Florida panther).
Application in Environmental Modeling (cont.)

Panther telemetry: 1999
Recent panther locations
Panther locations over forested habitat
Habitat edge metric

Application in Statistical Parametric Mapping

SPM is an academic software toolkit for the analysis of functional imaging data, for users familiar with the underlying statistical, mathematical and image processing concepts.

Implementing NetSolve-to-IBP library to SPM allows sharing files between NetSolve client and server.

Application in Image Compression

Original Image
5 Iterations 10 Iterations 60 Iterations 100 Iterations

Application in Image Compression (cont.)

• The first 5 iterations gives the actual shape of the image, and takes up to 99.9% less storage space than the original one.
• The 60 iterations provides better approximation and requires 84% less storage space than the original one.
• The 100 iterations offers a near perfect image and requires 55% less storage space than the original one.

Application in Vertex Cover and Clique Problems

• Vertex Cover problem
• Clique problem
• Duality between these two problems

Application in Genetic Algorithms

• “Genetic Algorithm (GA) is an optimization algorithm that imitates the evolution of living creatures”.
• Apply GA by encoding design variables of individuals to optimization problems.
• Suitable for performing parallel processing.
Related works

- Ninf-G
- DIET
- NEOS

“NetSolve provides a complete solution for easy access to remote resources and software”.

Conclusion

- Ease of use
- Complexity transparent
- User intervention free