Integrating Geographical Information Systems and Grid Services for Earthquake Forecasting

Marlon Pierce
Community Grids Lab
Indiana University
May 4, 2005
A Big Picture for SERVOGrid

Figure 1: Science, Critical Infrastructure Protection (CIP) and Emergency Preparedness and Response (EPR) Grids built as a Grid of Web Service (WS) Grids
The Problem: Integrating Data, Applications, and Client Devices

- The key issue we try to solve is building the distributed computing infrastructure that can connect
  - Legacy data archives
  - Executable codes
  - Real time data sources
  - Collaboration services ([http://www.globalmmcs.org](http://www.globalmmcs.org))
  - Client tools for collaboration
    - Audio/Video systems, whiteboard annotators, etc

- Various application-specific grids can be built out of the **common infrastructure**
  - Science Grids (described here)
  - Emergency planning, crisis response

- We choose certain fixed points for our **foundations**
  - Web Service standards: SOAP and WSDL
  - Other standards where available: GIS standards
  - Universal messaging substrate for SOAP and other messages: [http://www.naradabrokering](http://www.naradabrokering)
SERVO Apps and Their Data

- As summarized below, many SERVO codes use observational data measurements as input and create geo-located results.

- **GeoFEST**: Three-dimensional viscoelastic finite element model for calculating nodal displacements and tractions. Allows for realistic fault geometry and characteristics, material properties, and body forces.
  - Relies upon fault models with geometric and material properties.

- **Virtual California**: Program to simulate interactions between vertical strike-slip faults using an elastic layer over a viscoelastic half-space.
  - Relies upon fault and fault friction models.

- **Pattern Informatics**: Calculates regions of enhanced probability for future seismic activity based on the seismic record of the region
  - Uses seismic data archives

- **RDAHMM**: Time series analysis program based on Hidden Markov Modeling. Produces feature vectors and probabilities for transitioning from one class to another.
  - Used to analyze GPS and seismic catalog archives.
  - Can be adapted to detect state change events in real time.
Geographical Information System Services as Data Grid Components

- We decided that the Data Grid components of SERVO are best implemented using standard GIS services.
  - Use Open Geospatial Consortium standards
  - Maximize reusability in future SERVO projects
  - Provide downloadable GIS software to the community as a side effect of SERVO research.

- We implemented two cornerstone standards
  - **Web Feature Service (WFS):** data service for storing abstract map features
    - Supports queries
    - Faults, GPS, seismic records
  - **Web Map Service (WMS):** generate interactive maps from WFS’s and other WMS’s.
    - Maps are overlays: we grab images from OnEarth, overlay our additional images generated from features.
    - Can also extract features (faults, seismic events, etc) from user GUIs to drive problems such as the PI code and (in near future) GeoFEST, VC.

- We have also recently completed initial GPS Sensor Grid services
Building the GIS Grid

- We built these as Web Services
  - “WS-I+” style Grid
  - WSDL and SOAP: programming interfaces and messaging formats
  - You can work with the data and map services through programming APIs as well as browser interfaces.
  - Running demos and downloadable code are available from www.crisisgrid.org.

- Recent/ongoing work
  - Improved WFS performance
  - Integrating WMS clients with more applications
  - WMS clients publicly available and downloadable (as portlets).
  - Implementing WMS as a streaming video server.
  - Implementing SensorML for streaming, real-time data.
Pattern Informatics (PI)

- PI is a technique developed at University of California, Davis for analyzing earthquake seismic records to forecast regions with high future seismic activity.
  - They have correctly forecasted the locations of 15 of last 16 earthquakes with magnitude > 5.0 in California.


- PI is being applied other regions of the world, and John has gotten a lot of press.
  - Google “John Rundle UC Davis Pattern Informatics”
Pattern Informatics in a Grid Environment

- PI in a Grid environment:
  - Hotspot forecasts are made using publicly available seismic records.
    - Southern California Earthquake Data Center
    - Advanced National Seismic System (ANSS) catalogs
  - Code location is unimportant, can be a service through remote execution
  - Results need to be stored, shared, modified
  - Grid/Web Services can provide these capabilities

- Problems:
  - How do we provide programming interfaces (not just user interfaces) to the above catalogs?
  - How do we connect remote data sources directly to the PI code.
  - How do we automate this for the entire planet?

- Solutions:
  - Use GIS services to provide the input data, plot the output data
    - WFS for data archives, WMS for generating maps
  - Use HPSearch tool to tie together and manage the distributed data sources and code.
Tying It All Together: HPSearch

- **HPSearch** is an engine for orchestrating distributed Web Service interactions
  - It uses an event system and supports both file transfers and data streams.
  - Legacy name

- HPSearch flows can be scripted with JavaScript
  - HPSearch engine binds the flow to a particular set of remote services and executes the script.

- HPSearch engines are Web Services, can be distributed interoperate for load balancing.
  - Boss/Worker model

- **ProxyWebService**: a wrapper class that adds notification and streaming support to a Web Service.
Data can be stored and retrieved from the 3rd party repository (Context Service).

**WMS**

**WS Context (Tambora)**

**HPSearch (TReX)**

**HPSearch (Danube)**

**Data Filter (Danube)**

**PI Code Runner (Danube)**

- Accumulate Data
- Run PI Code
- Create Graph
- Convert RAW -> GML

**GML (Danube)**

**WFS (Gridfarm001)**

NaradaBroker network: Used by HPSearch engines as well as for data transfer.

**Virtual Data flow**

**Actual Data flow**

HPSearch controls the Web services.

Final Output pulled by the WMS.

HPSearch Engines communicate using NB Messaging infrastructure.
Some Challenges

- **Performance:** Are GIS services suitable for non-trivial data transfers?
  - Entire California seismic record since 1932 is 12 MB.
  - Global records obviously larger
  - This is not really suitable for HTTP transport.
  - We more recently implemented streaming data transfers for higher performance.

- **Adoption:** We must get the tools and services to the point where science application developers want to use them *early* in the development process rather than *later*.
  - Web Service client libraries to remote GIS data
  - Develop codes to work with data streams rather than files.

- **Security:** A global version of this has interesting security requirements
  - Authentication, authorization, federation for different countries
  - Time/event dependent security for crisis response
Sensor Grid

- A flexible computing environment for coupling real-time data sources to High Performance Geographic Information Systems (GIS) applications.

- Basing this on Open Geospatial Consortium’s SensorML suite
  - SensorML provides metadata about sensors
    - In development
  - Observations and measurements extensions to GML.
    - Currently implemented

- Codes such as RDAHMM can analyze real-time data for state change detection in GPS and other time series data.
  - Individual GPS state changes can be monitored and aggregated to detect changes in GPS networks.

- We are implementing this in conjunction with Scripps and JPL
Support for Streaming Data

- We use NaradaBrokering messaging software to manage data streams and filters.
  - Open source, Java-based software from the Community Grids Lab
  - Based on topic-based publication/subscription for delivery of messages from/to multiple endpoints.
  - “Message” can be anything, including SOAP and binary data streams.
  - We use this for audio/video collaboration.
  - More recently using it to build Web Service messaging substrates
    - SOAP 1.2 routing model, WS-Reliability, WS-Eventing

- NB ensures **reliable delivery** of events in the case of broker or client failures and prolonged entity disconnects.
  - Also supports **replay**.

- Implements high-performance protocols (message transit time of 1 to 2 ms per hop)
SOPAC GPS Services

OCRTN: Orange County Real Time Network
RICRTN: Riverside/Imperial Counties Real Time Network
SDCRTN: San Diego County Real Time Network
SOPAC GPS Services

- As a case study we implemented services to provide real-time access to GPS position messages collected from several SOPAC networks.

- Next step is to couple data assimilation tools (such as RDAHMM) to real-time streaming GPS data.

- Next steps
  - Programming APIs: currently we assume the subscriber speaks NaradaBrokering Java APIs (either NB’s native API or Java Messaging Service).
    - Need to investigate appropriate Web Service standards and C/C++ bindings.
  - SOAP enveloping of the GML message stream.
  - A Sensor Collection Service will be implemented to provide metadata about GPS sensors in SensorML.
Position Messages

- SOPAC provides 1-2Hz real-time position messages from various GPS networks in a binary format called RYO.
- Position messages are broadcasted through RTD server ports.
- We have implemented tools to convert RYO messages into ASCII text and another that converts ASCII messages into GML.
Real-Time Access to Position Messages

- We have a Forwarder tool that connects to RTD server port to forward RYO messages to a NB topic.
- RYO to ASCII converter tool subscribes this topic to collect binary messages and converts them to ASCII. Then it publishes ASCII messages to another NB topic.
- ASCII to GML converter subscribes this topic and publishes GML messages to another topic.
Current implementation provides real-time access to GP messages to following stations in RYO, ASCII and GML formats:

<table>
<thead>
<tr>
<th>RTD Port No</th>
<th>7010</th>
<th>7011</th>
<th>7012</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPS Network</strong></td>
<td>Orange County</td>
<td>Riverside/Imperial Counties</td>
<td>San Diego County</td>
</tr>
<tr>
<td>AZRY</td>
<td>BLSA</td>
<td></td>
<td>DSME</td>
</tr>
<tr>
<td>COTD</td>
<td>CAT2</td>
<td></td>
<td>DVLW</td>
</tr>
<tr>
<td>CRRS</td>
<td>FVPK</td>
<td></td>
<td>OGHS</td>
</tr>
<tr>
<td>DHLG</td>
<td>MJPK</td>
<td></td>
<td>PMOB</td>
</tr>
<tr>
<td>DVLE</td>
<td>OEOC</td>
<td></td>
<td>RAAP</td>
</tr>
<tr>
<td>GLRS</td>
<td>SACY</td>
<td></td>
<td>SIO5</td>
</tr>
<tr>
<td>KYVW</td>
<td>SBCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIN1</td>
<td>SCMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIN2</td>
<td>TRAK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSAP</td>
<td>WHYT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIDC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NaradaBrokering topics

<table>
<thead>
<tr>
<th>Network</th>
<th>Format</th>
<th>NB Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCRTN</td>
<td>RYO</td>
<td>SOPAC/GPS/Positions/OCRTN/RYO</td>
</tr>
<tr>
<td></td>
<td>ASCII</td>
<td>SOPAC/GPS/Positions/OCRTN/ASCII</td>
</tr>
<tr>
<td></td>
<td>GML</td>
<td>SOPAC/GPS/Positions/OCRTN/GML</td>
</tr>
<tr>
<td>RICRTN</td>
<td>RYO</td>
<td>SOPAC/GPS/Positions/RICRTN/RYO</td>
</tr>
<tr>
<td></td>
<td>ASCII</td>
<td>SOPAC/GPS/Positions/RICRTN/ASCII</td>
</tr>
<tr>
<td></td>
<td>GML</td>
<td>SOPAC/GPS/Positions/RICRTN/GML</td>
</tr>
<tr>
<td>SDCRTN</td>
<td>RYO</td>
<td>SOPAC/GPS/Positions/SDCRTN/RYO</td>
</tr>
<tr>
<td></td>
<td>ASCII</td>
<td>SOPAC/GPS/Positions/SDCRTN/ASCII</td>
</tr>
<tr>
<td></td>
<td>GML</td>
<td>SOPAC/GPS/Positions/SDCRTN/GML</td>
</tr>
</tbody>
</table>
More Information

- Contact: mpierce@cs.indiana.edu
- GIS Work at CGL: www.crisisgrid.org
  - Software, demos, publications
  - Several recent manuscript submissions are/will be posted soon.
- HPSearch at CGL: www.hpsearch.org
- SERVOGrid Web Sites
  - Our fine parent project
  - http://servo.jpl.nasa.gov/
  - http://quakesim.jpl.nasa.gov/
Status and Software

- Web Feature Service 1.x software available now
  - www.crisisgrid.org

- Our SERVO WFS includes
  - Fault data
  - GPS records
  - Seismic records now for most areas of the globe
  - Note these are Web Services, so you can build your own clients to connect to our running services.

- Web Map Service
  - Client portlets (shown) available from www.collab-ogce.org.
  - Server software downloads available soon.

- HPSearch
  - Currently available, www.hpsearch.org

- WS-Context and Information Services Work
  - http://grids.ucs.indiana.edu/~maktas/fthpis/
Acknowledgements

- Prof. John Rundle and his group at UC-Davis, particularly James Holliday.
- Prof. Geoffrey Fox, director of the Community Grids Lab, and CGL grad students
  - Ahmet Sayar: Web Map Server development
  - Galip Aydin: Web Feature Service development and Sensor Grid
  - Mehmet Aktas: Information and Context Services
  - Harshawardhan Gadgil: HPSearch Workflow development
- Satellite images from NASA OnEarth WMS
- This work is supported by the NASA Advanced Information Systems Technology Program.
Backup Slides
Open Standards

- SensorGrid will combine Open GIS standards for data and services with Web Services specifications.
- GML and SensorML, OGC (Open Geospatial Consortium Inc.) specifications for encoding geospatial data and sensor metadata in XML will be adopted for universal compatibility with larger GIS community.
- WS-* specifications will be utilized to ensure access to these data via standard interfaces.
CGL Work on GIS Services

- Some example OGC services include
  - Web Feature Service (WFS): for retrieving GML encode features, like faults, roads, county boundaries, GPS station locations,...
  - Web Map Service (WMS): for creating maps out of Web Features

- Problems with current GIS services
  - Not (yet) Web Service compliant
    - Efforts underway to provide this within OGC.
    - But current specs are “pre” web service, no SOAP or WSDL
    - Use instead HTTP GET/POST conventions.
  - Often define general Web Service services as specialized standards
    - Information services
    - Notification services in sensor grids
  - Can’t use other Web Service standards for reliability, security, etc.

- CGL is developing Web Service versions of OGC standard services