

Towards Secure Cyberinfrastructure for Sharing Border Information

Ann Gates¹, Vladik Kreinovich¹, Luc Longpré¹,
Paolo Pinheiro da Silva¹, G. Randy Keller²

¹Department of Computer Science

²Department of Geological Sciences

University of Texas at El Paso

500 W. University, El Paso, TX 79968, USA

agates@utep.edu, vladik@utep.edu, longpre@utep.edu,

paulo@utep.edu, keller@utep.edu

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Table of Contents

«

»

«

»

Page 1 of 17

Go Back

Full Screen

1. Outline

- In many border-related issues ranging from economic collaboration to border security, sharing information is *important*.
- Sharing is *difficult*: different countries use different information formats and data structures.
- *Desirable*: to facilitate information sharing.
- *Our experience*: geoinformatics.
- This experience can be applied to border collaboration.
- *Additional problem*: many security-related data are sensitive.

[General Approach: ...](#)[Interval Approach: ...](#)[Interval ...](#)[Interval ...](#)[Similar Situation: ...](#)[Let Us Use a Similar ...](#)[How to Represent Sets](#)[How to Propagate ...](#)[How to Propagate ...](#)[First Example: ...](#)[Second Example: ...](#)[How to Compute \$r_{ik}\$](#) [Distributivity: \$a \cdot \(b + \dots\$](#) [Distributivity: New ...](#)[Toy Example with ...](#)[Computation Time](#)[What Next?](#)[Probabilistic Case: In ...](#)[Acknowledgments](#)[Title Page](#)[◀◀](#)[▶▶](#)[◀](#)[▶](#)[Page 2 of 17](#)[Go Back](#)[Full Screen](#)

2. Practical Problem: Need to Combine Geographically Separate Computational Resources

- *Problem:*
 - In different domains, there is a large amount of data stored in different locations.
 - There are many software tools for processing this data, also implemented at different locations.
- Users may be interested in different information about this domain.
 - Sometimes, the information required by the user is already stored in *one of the databases*.
 - In other cases, different pieces of the information requested by the user are *stored at different locations*.
 - In many other situations, the appropriate answer to the user's request requires that we not only collect the relevant data, but that we also use some *data processing* algorithms to process this data.
- The need to combine computational resources (data and programs) located at different geographic locations seriously complicates research.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 3 of 17

Go Back

Full Screen

3. Centralization of Computational Resources – Traditional Approach to Combining Computational Resources; Its Advantages and Limitations

- *Traditional approach*: move all the resources to a *central location*.
- *Problem*: centralization requires a large amount of efforts:
 - data are presented in different formats,
 - the existing programs use specific formats, etc.
- To make the central data depository efficient, it is necessary:
 - to reformat all the data,
 - to rewrite all the data processing programs – so that they become fully compatible with the selected formats and with each other,
 - etc.
- *Conclusion*: the amount of work that is needed for this reformatting and rewriting is large.
- *Result*: none of these central depositories really succeeded in becoming an easy-to-use centralized database.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 4 of 17

Go Back

Full Screen

4. Cyberinfrastructure – A More Efficient Approach to Combining Computational Resources

- *Objective:*
 - provide the users with the efficient way to submit requests
 - without worrying about the geographic locations of different computational resources
 - and avoid centralization with its excessive workloads.
- *Main idea:* keep all (or at least most) computational resources
 - at their current locations,
 - in their current formats.
- *Specifics:* to expedite the use of these resources:
 - we supplement the local computational resources with the “metadata”, i.e., with the information about the formats, algorithms, etc.,
 - we “wrap up” the programs and databases with auxiliary programs that provide data compatibility into *web services*,
- *General description:* we provide a cyberinfrastructure that uses the metadata to automatically combine different computational resources.

5. Cyberinfrastructure: Example

- *User's request:* a user is interested in using the gravity data to uncover the geological structure of the Rio Grande region.
- *What the system will do:*
 - get the gravity data from the UTEP and USGS gravity databases,
 - convert them to a single format (if necessary),
 - forward this data to the program located at San Diego Supercomputer Center, and
 - move the results back to the user.
- *Comment:* this example is exactly what we are designing under the NSF-sponsored Cyberinfrastructure for the Geosciences (GEON) project.
- *General description:* this is similar to what other cyberinfrastructure projects are trying to achieve.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 6 of 17

Go Back

Full Screen

6. What Is Cyberinfrastructure: From the Official NSF Definition

- *Source:* NSF Blue Ribbon Advisory Panel on Cyberinfrastructure.
- *Motivation:* “a new age has dawned in scientific and engineering research, pushed by continuing progress in
 - computing,
 - information, and
 - communication technology,

and pulled by the

- expanding complexity,
- scope, and
- scale

of today’s challenges.”

- *Essence:* “The capacity of this technology has crossed thresholds that now make possible a comprehensive ‘cyberinfrastructure’ on which
 - to build new types of scientific and engineering knowledge environments and organizations and
 - to pursue research in new ways and with increased efficacy.”

General Approach: ...

Interval Approach: ...

Interval ...

Interval ...

Similar Situation: ...

Let Us Use a Similar ...

How to Represent Sets

How to Propagate ...

How to Propagate ...

First Example: ...

Second Example: ...

How to Compute r_{ik}

Distributivity: $a \cdot (b + \dots$

Distributivity: New ...

Toy Example with ...

Computation Time

What Next?

Probabilistic Case: In ...

Acknowledgments

Title Page

◀

▶

◀

▶

Page 7 of 17

Go Back

Full Screen

7. What Is Cyberinfrastructure: From the Official NSF Definition (Examples)

- “Such environments and organizations, enabled by cyberinfrastructure, are increasingly required to address national and global priorities, such as
 - understanding global climate change,
 - protecting our natural environment,
 - applying genomics-proteomics to human health,
 - maintaining national security,
 - mastering the world of nanotechnology, and
 - predicting and protecting against natural and human disasters,
- as well as to address some of our most fundamental intellectual questions such as
 - the formation of the universe and
 - the fundamental character of matter.”

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 8 of 17

Go Back

Full Screen

8. Geoinformatics: Cyberinfrastructure for the Geosciences

- *Problems:*
 - the chaotic distribution of available data sets,
 - lack of documentation about them, and
 - lack of easy-to-use access tools and computer modeling and analysis codes
- *Means to solve this problem:* recent advances in
 - computational methods,
 - visualization, and
 - database interoperability.
- *Vision:*
 - a highly interconnected data system populated with high quality, freely available data, as well as, a robust set of software for analysis, visualization, and modeling.
 - This system would feature rich and deep databases and convenient access.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 9 of 17

Go Back

Full Screen

9. Geoinformatics: Objectives

- *User's view*: data and information can be easily found and unexpected relationships can be discovered via queries in “Google-like” fashion.
- *Objectives*:
 - manage, preserve, and efficiently access the vast amounts of Earth Science data that exist now and the vast data flows that will be coming online as projects such as EarthScope (www.earthscope.org) get underway;
 - foster integrated scientific studies that are required to address the increasingly complex scientific problems that face our scientific community;
 - accelerate the pace of scientific discovery and facilitate innovation;
 - create an environment in which data and software developed with public funds are preserved and made available in a timely fashion; and
 - provide easy access to high-end computational power, visualization and open source software to researchers and students.

General Approach: ...
Interval Approach: ...
Interval ...
Interval ...
Similar Situation: ...
Let Us Use a Similar ...
How to Represent Sets
How to Propagate ...
How to Propagate ...
First Example: ...
Second Example: ...
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New ...
Toy Example with ...
Computation Time
What Next?
Probabilistic Case: In ...
Acknowledgments

Title Page



Page 10 of 17

Go Back

Full Screen

10. GEON Project: A First Step Towards Cyberinfrastructure in Geosciences

- *What is it:* a large ITR grant funds the GEON (GEOscience Network) project.
- *Focus:*
 - craft the many relatively raw data sets in the Earth Science community into mature databases that can grow and evolve;
 - interlink and share these multidisciplinary databases;
 - create a robust toolbox of open source software for analysis, modeling, and visualization; and
 - provide the information technology infrastructure to manage and explore a highly distributed and diverse network.
- *Participants:*
 - true partnership between Computer Science and Earth Science researchers;
 - working closely with IRIS, the U.S. Geological Survey, SCEC, and UN-AVCO, and with other IT efforts within the Earth Science community;
 - U.S. Geological Survey is a major partner.

General Approach: ...
Interval Approach: ...
Interval ...
Interval ...
Similar Situation: ...
Let Us Use a Similar ...
How to Represent Sets
How to Propagate ...
How to Propagate ...
First Example: ...
Second Example: ...
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New ...
Toy Example with ...
Computation Time
What Next?
Probabilistic Case: In ...
Acknowledgments

Title Page



Page 11 of 17

Go Back

Full Screen

11. Cyberinfrastructure for the Geosciences: Challenges

- *Challenges:*
 - extreme heterogeneity of geoscience data formats, storage and computing systems and, most importantly,
 - ubiquity of “hidden semantics” and differing conventions, terminologies, and ontological frameworks across disciplines.
- *First idea: unified language.* A Unified Geosciences Language System (UGLS) is being developed to enable semantic interoperability.
- *Example:* a stratigraphic layer of rock often changes names across state lines.
- *Second idea: portal.* Portal provides:
 - advanced query interfaces to distributed, semantically-integrated databases,
 - Web-enabled access to shared tools, and
 - seamless access to distributed computational, storage, and visualization resources and data archives.
- *Similar projects:* GriPhyN, NEESGrid, NBII, and BIRN
 - indicate the readiness of the Computer Science community to provide the necessary interoperable infrastructure, and
 - testify to the value of integration of Computer Science with major science and education initiatives.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 12 of 17

Go Back

Full Screen

12. GEON's Objective: Brief Summary

- *Goal:*
 - to advance the field of geoinformatics,
 - to prepare and train current and future generations of geoscience researchers, educators, and practitioners in the use of cyberinfrastructure to further their research, education, and professional goals.
- Geoinformatics will foster:
 - new interdisciplinary research, for example, the gravity modeling of 3D geological features, such as plutons;
 - study of active tectonics by integrating LiDAR data and geodynamics models; and
 - study of lithospheric structure and properties across diverse tectonic environments.
- GEON is based on a service-oriented architecture (SOA) with support for
 - “intelligent” search,
 - semantic data integration,
 - visualization of 4D scientific datasets, and
 - access to high performance computing platforms for data analysis and model execution – via the GEON Portal.

General Approach: ...
Interval Approach: ...
Interval ...
Interval ...
Similar Situation: ...
Let Us Use a Similar ...
How to Represent Sets
How to Propagate ...
How to Propagate ...
First Example: ...
Second Example: ...
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New ...
Toy Example with ...
Computation Time
What Next?
Probabilistic Case: In ...
Acknowledgments

Title Page



Page 13 of 17

Go Back

Full Screen

13. Our Experience Can Be Applied to Border Collaboration

- *Need for sharing information:* in many border-related issues:
 - economic collaboration,
 - environmental collaboration;
 - border security, etc.
- *Sharing is difficult:* different countries use different information formats and data structures.
- *Desirable:* design infrastructure to facilitate this information sharing.
- *Centralization is not feasible:*
 - first, transforming the large amounts of data into different formats would require a large amount of effort;
 - second, the agreement on a single format for storing common data may be politically difficult.
- *Our suggestion:* combine different border-related databases by an appropriate cyberinfrastructure, by “wrapping” up each computational resource into the corresponding web service.
- *Advantages:*
 - much smaller amount of effort,
 - no need to make complex political decisions.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 14 of 17

Go Back

Full Screen

14. How to Make Cyberinfrastructure Secure

- Many security-related border-related data are sensitive.
- *Traditional approach to security*: mainly about data security.
- *Example*: data exchange that uses RSA encryption is successfully implemented in different e-commerce applications such as amazon.com.
- *Challenge*:
 - to process data, we must decode it, and this is where the data becomes vulnerable;
 - so, if we use remote computational resources to process data, we must make sure that this data is not compromised by the computations.
- *What is known*: there exist protocols for computing without learning the underlying data (Naor, Pinkas, and Sumner).
- *Limitations*: mostly re auctions.
- *What we did*: extended these protocols to security and privacy protection on web services.
- *Possible application*: compute overall characteristics of a border area based on information that both sides do not necessarily want to share in detail:
 - due to security concerns or
 - due to the need to keep commercial secrets.
- *Warning*: to ensure privacy, we must perform additional computations.

General Approach: . . .
Interval Approach: . . .
Interval . . .
Interval . . .
Similar Situation: . . .
Let Us Use a Similar . . .
How to Represent Sets
How to Propagate . . .
How to Propagate . . .
First Example: . . .
Second Example: . . .
How to Compute r_{ik}
Distributivity: $a \cdot (b + \dots$
Distributivity: New . . .
Toy Example with . . .
Computation Time
What Next?
Probabilistic Case: In . . .
Acknowledgments

Title Page



Page 15 of 17

Go Back

Full Screen

15. Conclusions

- In many border-related issues ranging from economic collaboration to border security, it is extremely important that bordering countries share information.
- One reason why such sharing is difficult is that different countries use different information formats and data structures.
- It is therefore desirable to design infrastructure to facilitate this information sharing.
- UTEP is a lead institution in a similar NSF-sponsored multi-million geoinformatics project, whose goal is to combine diverse and complex geophysical and geographical data stored in different formats and data structures.
- Our experience in using and developing related web service techniques can be applied to border collaboration.
- In particular, we can make sure that the designed cyberinfrastructure provides secure sharing.

[General Approach: ...](#)[Interval Approach: ...](#)[Interval ...](#)[Interval ...](#)[Similar Situation: ...](#)[Let Us Use a Similar ...](#)[How to Represent Sets](#)[How to Propagate ...](#)[How to Propagate ...](#)[First Example: ...](#)[Second Example: ...](#)[How to Compute \$r_{ik}\$](#) [Distributivity: \$a \cdot \(b + \dots\$](#) [Distributivity: New ...](#)[Toy Example with ...](#)[Computation Time](#)[What Next?](#)[Probabilistic Case: In ...](#)[Acknowledgments](#)[Title Page](#)[◀◀](#)[▶▶](#)[◀](#)[▶](#)[Page 16 of 17](#)[Go Back](#)[Full Screen](#)

16. Acknowledgments

This work was supported in part:

- by NASA under cooperative agreement NCC5-209,
- by NSF grant EAR-0225670,
- by NIH grant 3T34GM008048-20S1,
- by Army Research Lab grant DATM-05-02-C-0046,
- by Star Award from the University of Texas System,
- and by Texas Department of Transportation grant No. 0-5453.

General Approach: ...

Interval Approach: ...

Interval ...

Interval ...

Similar Situation: ...

Let Us Use a Similar ...

How to Represent Sets

How to Propagate ...

How to Propagate ...

First Example: ...

Second Example: ...

How to Compute r_{ik}

Distributivity: $a \cdot (b + \dots$

Distributivity: New ...

Toy Example with ...

Computation Time

What Next?

Probabilistic Case: In ...

Acknowledgments

Title Page

◀◀

▶▶

◀

▶

Page 17 of 17

Go Back

Full Screen