

Supporting Climate Applications over Named Data Networking (NDN)

Christos Papadopoulos

Cathie Olschanowsky

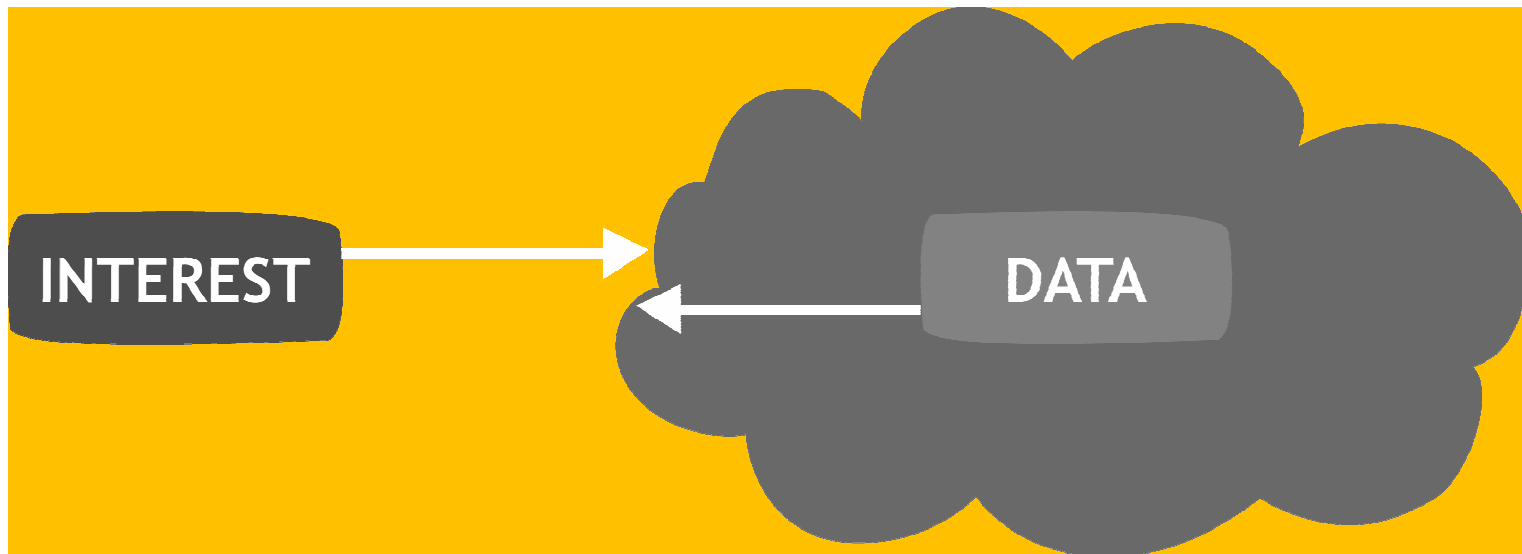
David Randall

Susmit Shannigrahi

CMMAP Team Meeting – Jan 8, 2014

#13410999: NSF CC-NIE Integration award

NDN: Focus on Data

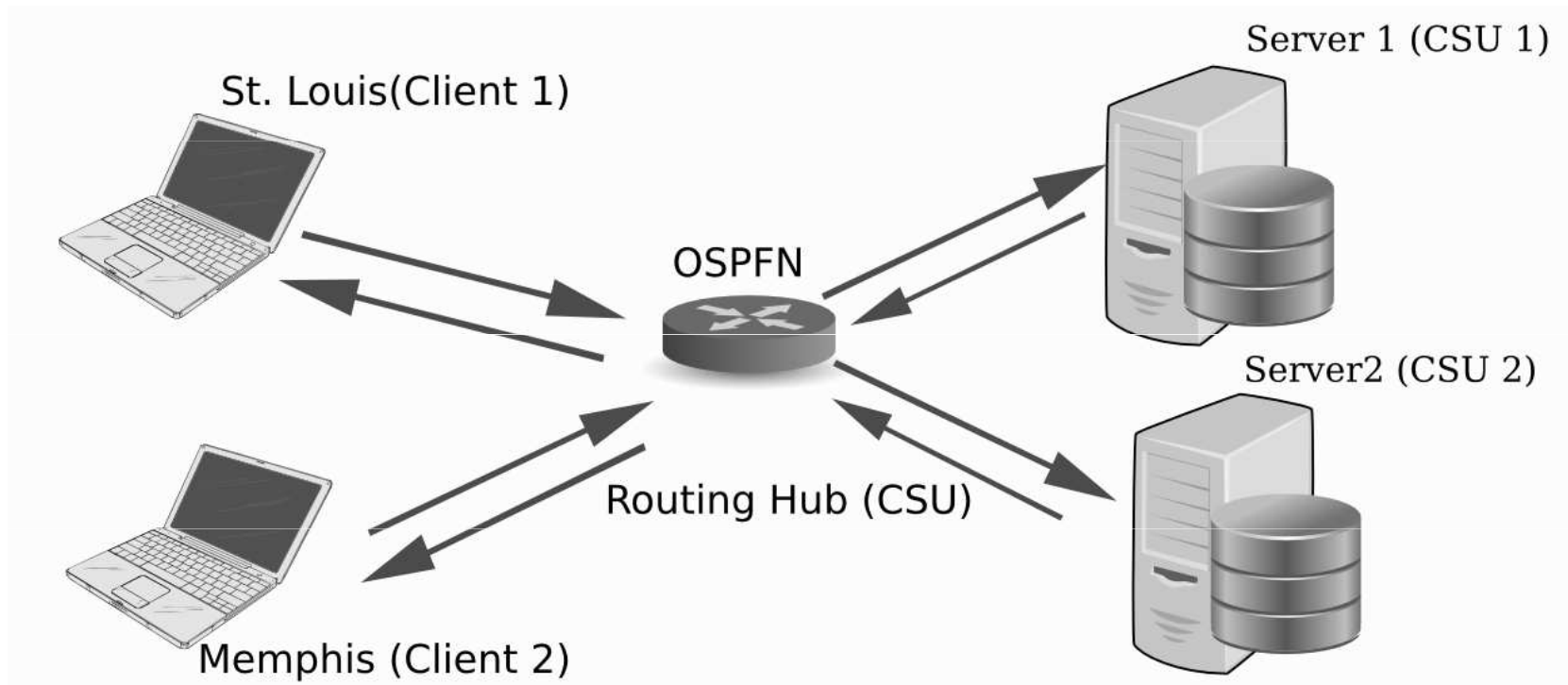


Abstracting away the notion of “host”

Superset of host-to-host communication model

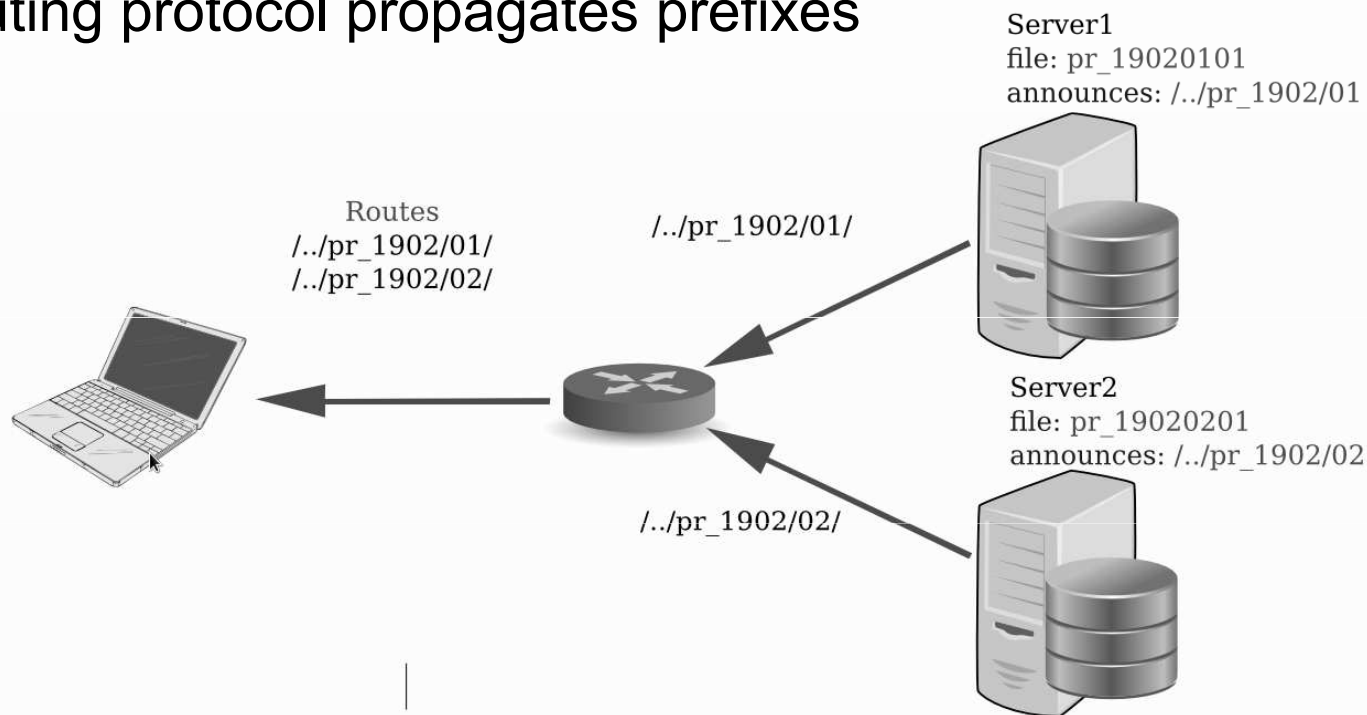
What Can NDN do for you

- Two servers and two clients
- Servers at CSU, clients at Memphis and St. Louis
- Nodes exchange name prefixes



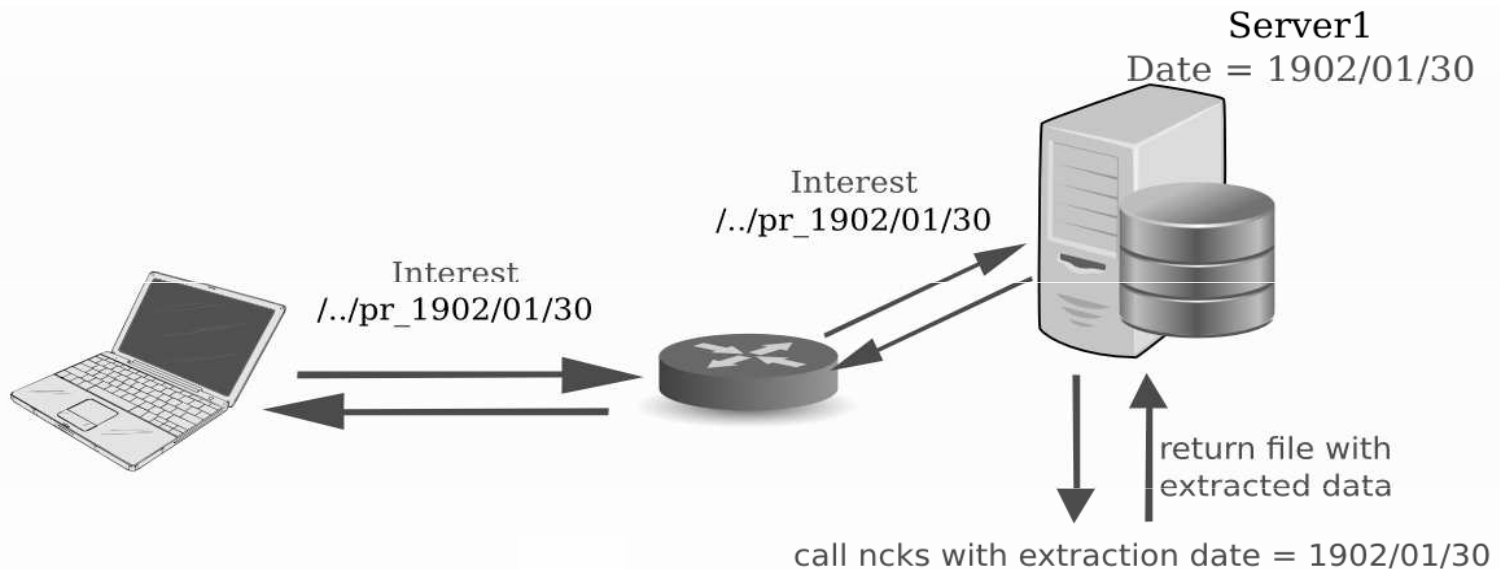
Announcements

- Servers have `.nc` files, each `.nc` file have one month's data
- Route announcements in network are based on filename
- Each server advertises one prefix for a file
 - Server having file `pr_19020101.nc` announces `../../pr_1902/01/`
- Routing protocol propagates prefixes



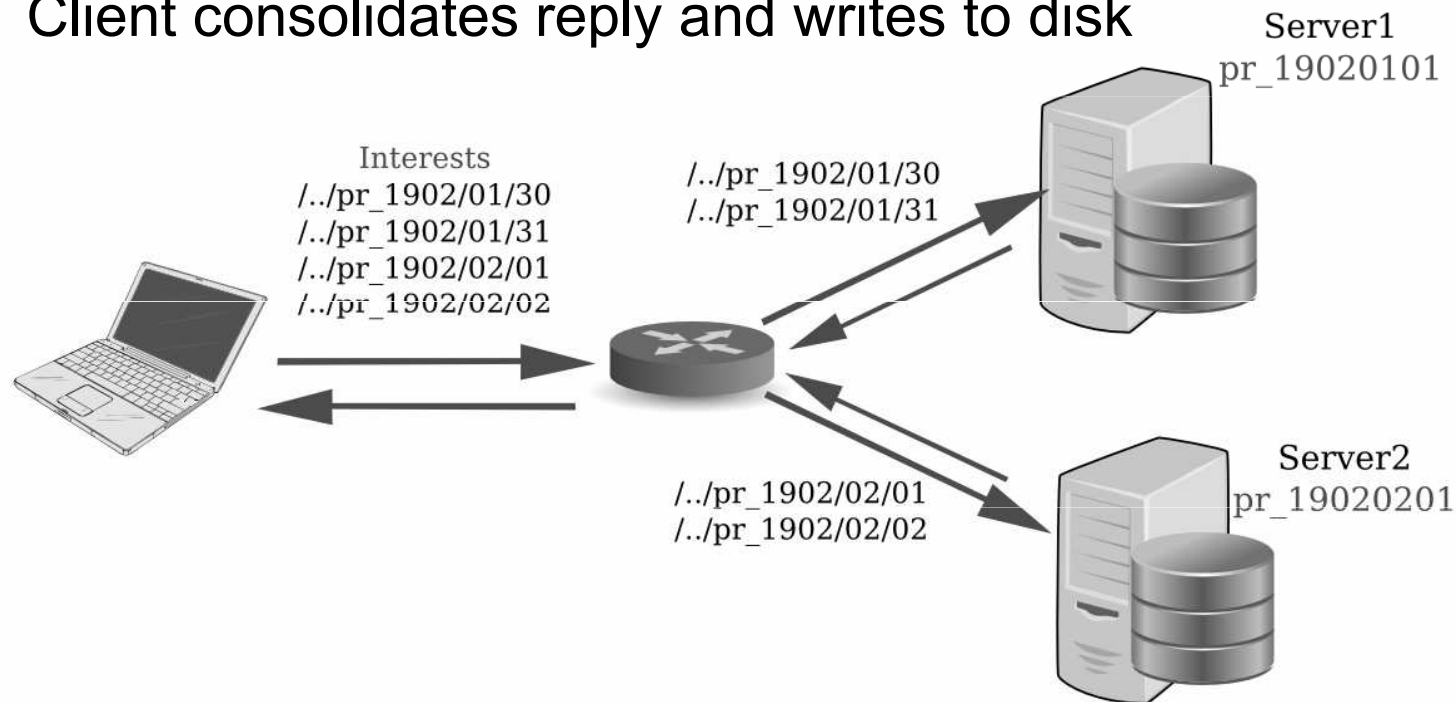
Dynamic Data Generation

- Servers parse interest names and find the date range
- Pass date range to ncks tool.
- ncks tool extracts data, writes to file and returns the filename to server
- Server sends back file



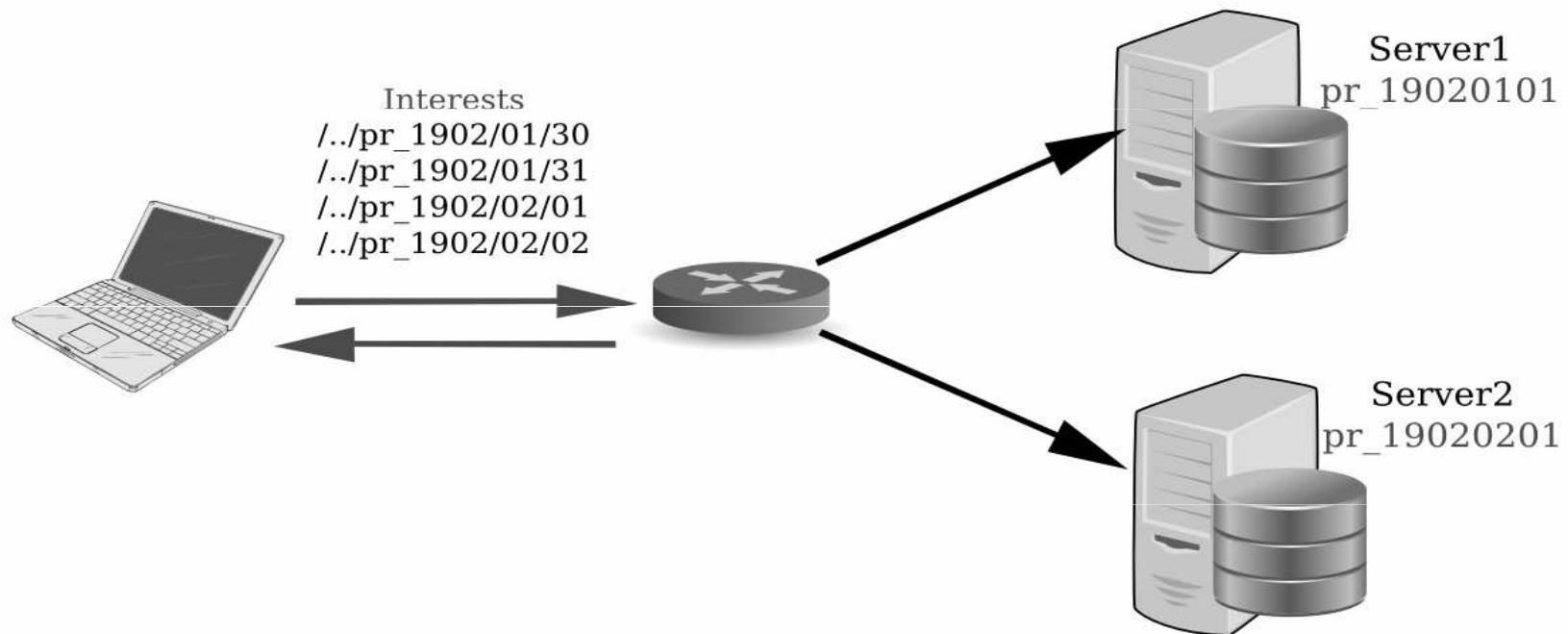
An Example Data Request

- Want data for Jan 30 – Feb 02
- Client expresses interests, one for each day
- Interests for Jan 30-31 go to server1
- Interests for Feb 01-02 go to server2
- Data is dynamically generated and sent back
- Client consolidates reply and writes to disk



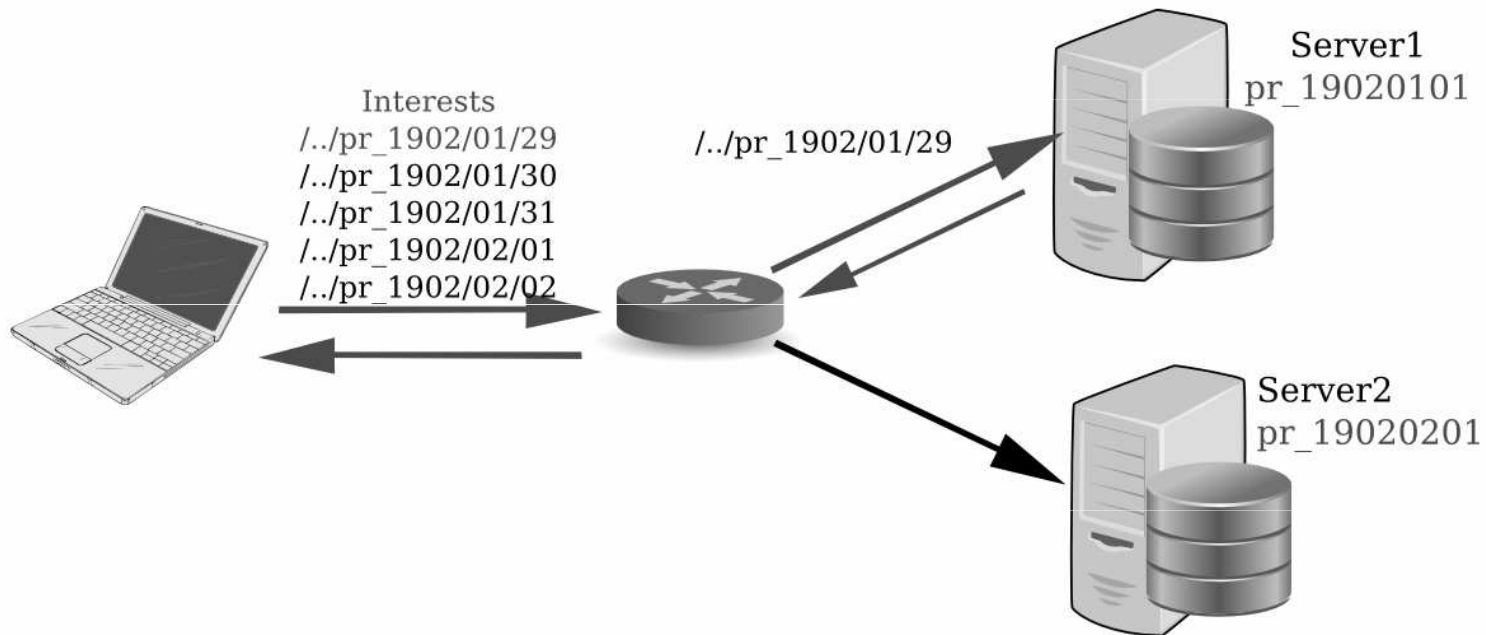
Repeat Requests and Cache

- If asked for same data, requests are answered from cache
- Saves transmission time, extraction time and transfer time



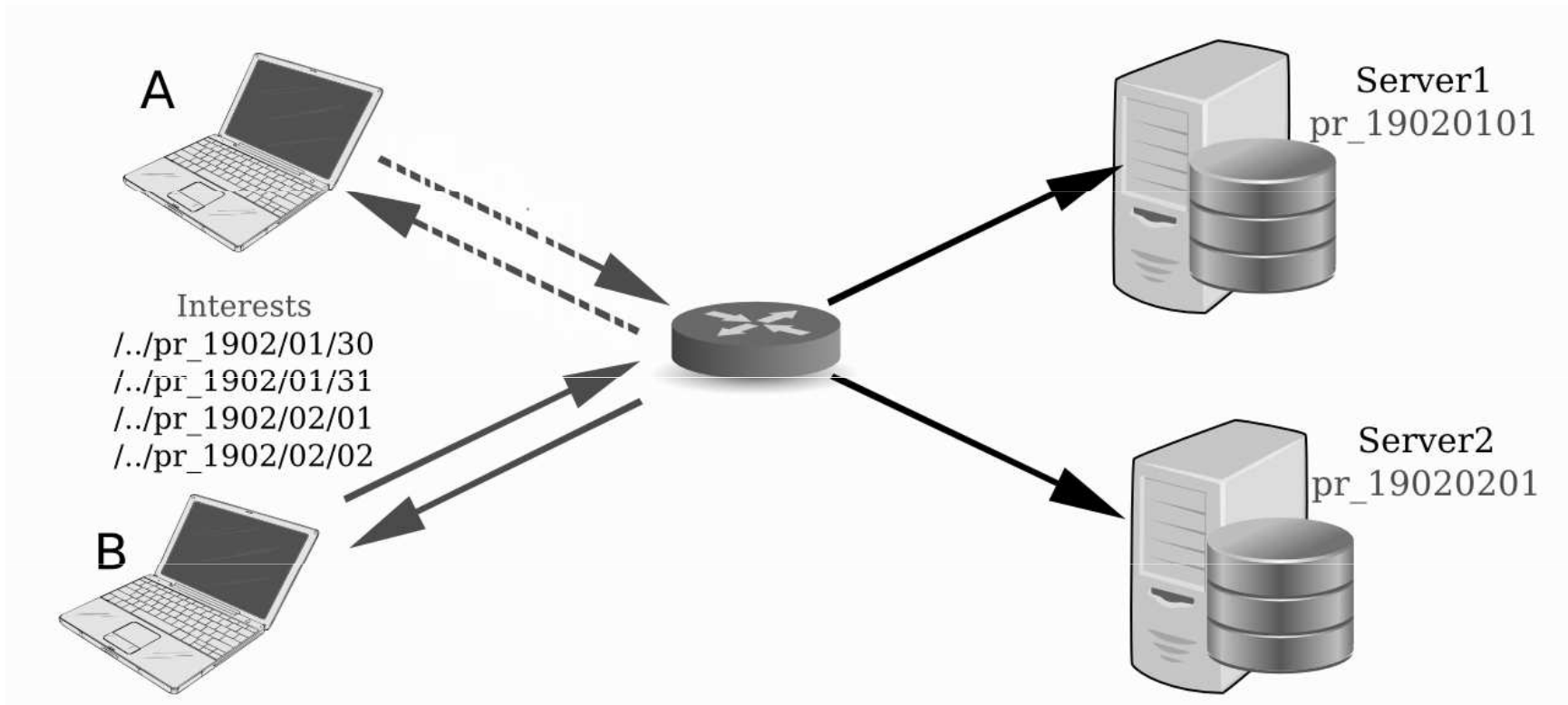
Partially Cached Data

- What happens if we ask for Jan 29 – Feb 2 ?
- Request for data not cached goes to server
- Rest is answered from cache



Collaborations

- A asked for data for Jan 30- Feb 2.
- B later asks for same data.
- B receives data from cache.



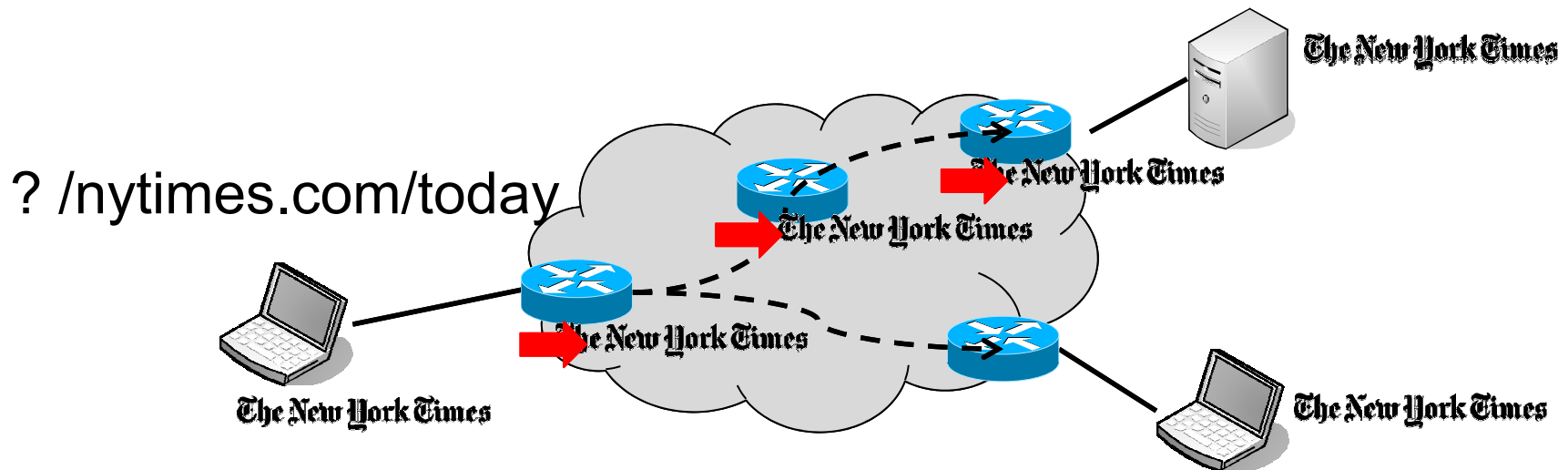
Two Problems with Current Internet

- Focus is on end-point communication
 - Artifact of original thinking: share resources, not content
 - Login to fast machine, access to the tape drive, the printer, etc.
- Security
 - To get data, you build a secure path
 - Once you authenticated with the server, you trust the content

Two Focal Points in NDN

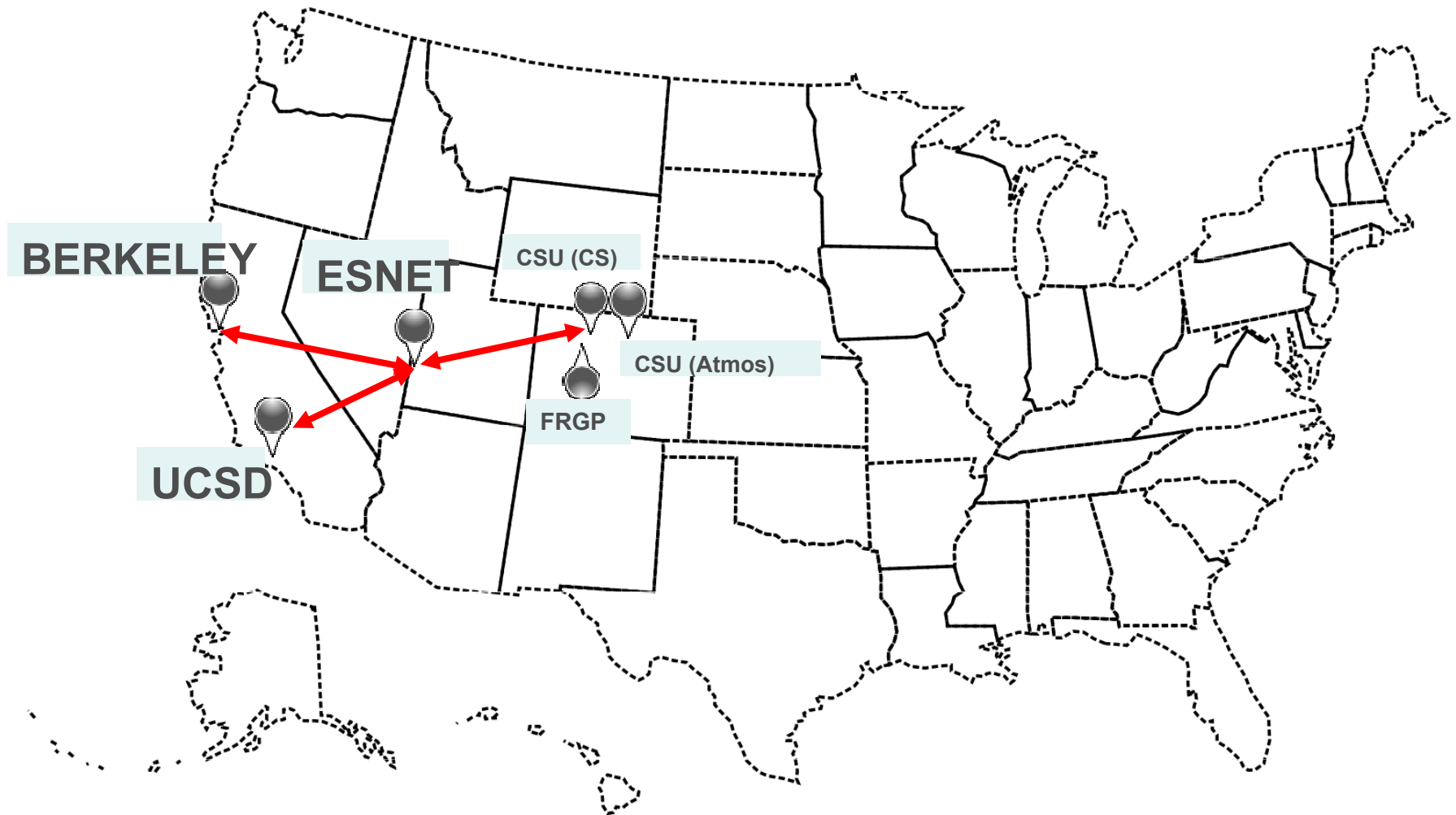
- Focus on the what not the where
- Secure the data not the container

Summary



- Route on content names
- Content from anywhere: not just the producer
- “Breadcrumbs” & de-duplication of requests
- Cache retrieved data in Content Store (CS)

Planned Climate Testbed



Naming In Climate Applications

- Climate community has recognized the importance of structured naming
- Reference documents are beginning to appear
- Mostly targeted to file and dataset naming
 - Including dataset publication
- Provide consistency across distributed archives

Example: CMIP5: Coupled Model Inter-comparison Project, phase 5

- “.. standard experimental protocol for studying the output of coupled atmosphere-ocean general circulation models.”
- International effort, started in 1995
- “CMIP5 is meant to provide a framework for coordinated climate change experiments.”

CMIP5 Data Reference Syntax (DRS) and Controlled Vocabularies

- “.. a common naming system to be used in files, directories, metadata, and URLs to identify datasets wherever they might be located within the distributed CMIP5 archive. It defines controlled vocabularies for many of the components comprising the data reference syntax (DRS).”
- Current version is 1.3.1, June 2012

DRS: Directory and File Structure

- **Directory encoding:**
 - /<activity>/<product>/<institute>/<model>/<experiment>/ <frequency>/<modeling realm>/ <variable name>/<ensemble member>
 - **Example:**
/CMIP5/output/MOHC/HadCM3/decadal1990/day/atmos/tas/r3i2p1/
- **Filename encoding:**
 - filename = <variable name>_<MIP table>_<model>_<experiment>_<ensemble member>[_<temporal subset>][_<geographical info>].nc
 - Example: tas_Amon_HADCM3_historical_r1i1p1_185001-200512.nc

Observations

- CMIP5 DRS is NDN-compliant
 - Hierarchical
 - Clearly defined name components
 - Well-defined vocabularies
 - Organizes components from less specific to more specific
- Utilities to translate into DRS (CMOR)
- Ports readily into NDN names

Potential Naming Components

- Identify a run
 - Model Name + Experiment Name
 - Where does activity belong?
- Identify a data file
 - Timestamp + Sample Frequency
- Identify variables within a data file or across data files
 - Data type + timestamp + sampling frequency

Moving Forward

- NDN offers two potential improvements to the scientific community:
 - Make data discovery far more convenient
 - Speed up data retrieval
- Data naming and discovery can be done now
- Speeding up data retrieval – forwarding, strategy layer, etc.
 - Real application to test congestion control

Current Plan

- Deploy hardware
- Design appropriate namespaces
- Implement translators
- Integrate with existing workflow
 - Discovery and retrieval
- Investigate access controls
- Evaluate
- Extend to other domains