

NAMED DATA NETWORKING: AN INTERNET ARCHITECTURE FOR THE FUTURE

Christos Papadopoulos
Colorado State University

SwitchOn Workshop, Oct 16, 2015, Sao Paulo, Brazil
Work supported by NSF #1345236 and #13410999





The NDN Project

- Part of the NSF Future Internet Architecture FIA initiative
- Goal: design the next generation Internet Architecture
- NDN is one of four multi-institution teams funded in 2010-13, and 2014-16, ~\$15M

NDN Institutions



<http://named-data.net>

<http://github.com/named-data>

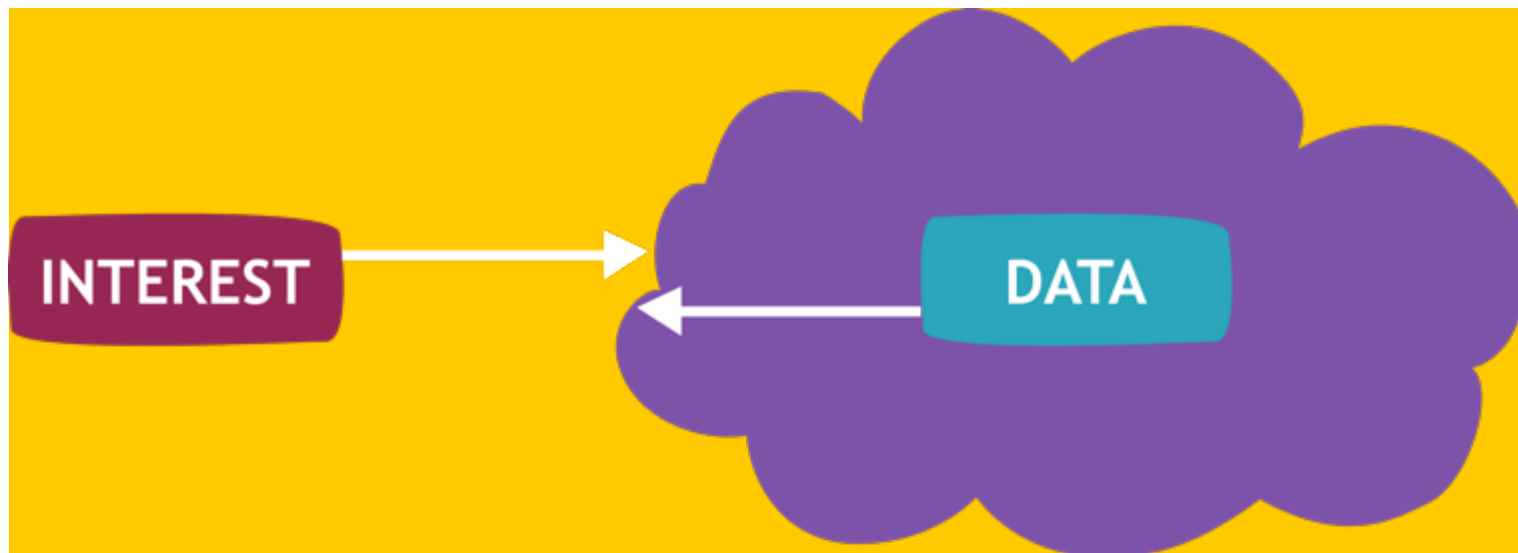
Colorado State University

Today's Internet Names Hosts

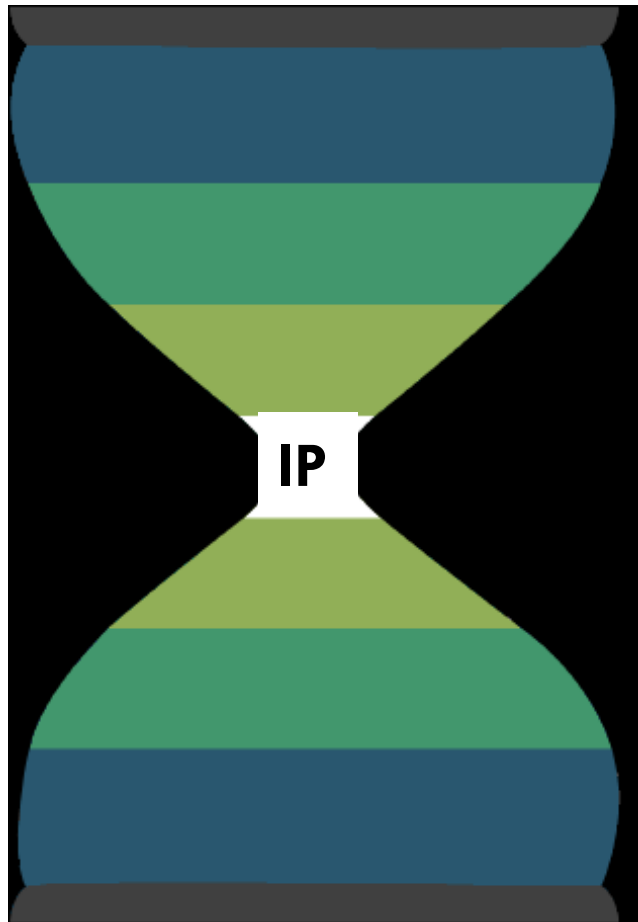
- To find content in the network
- ..you have to learn where the content is
- ..and then ask the network to take you there
- ..so you can tell the server what you want
- But no-one cares about the servers anymore..
- ..we care about the Data!
- **Service model mismatch**

Named Data Network (NDN)

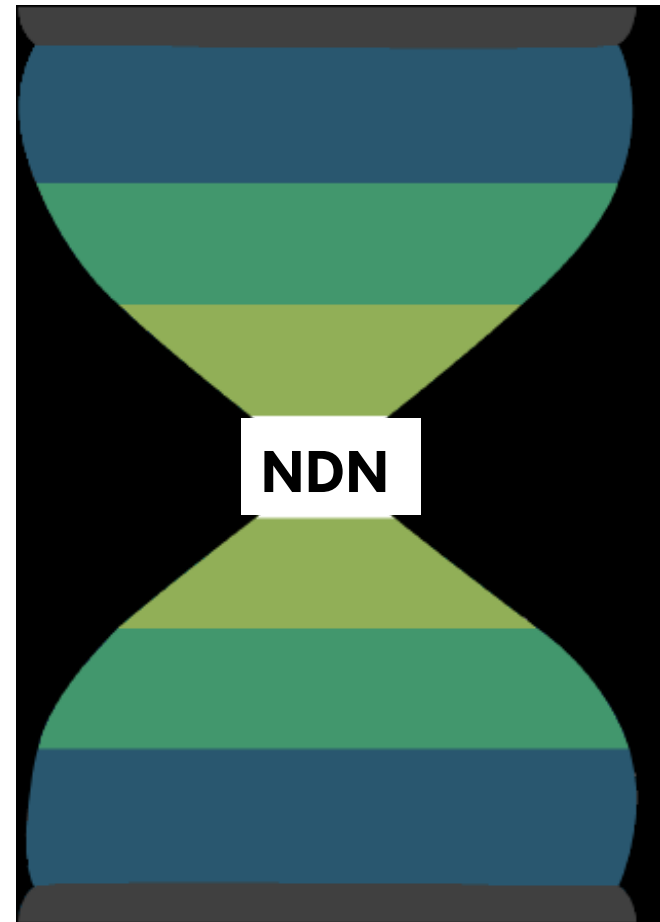
- The main idea: **Name the data, not the hosts!**
- ..so you just tell the network what you want..
- ..and let the network find it for you



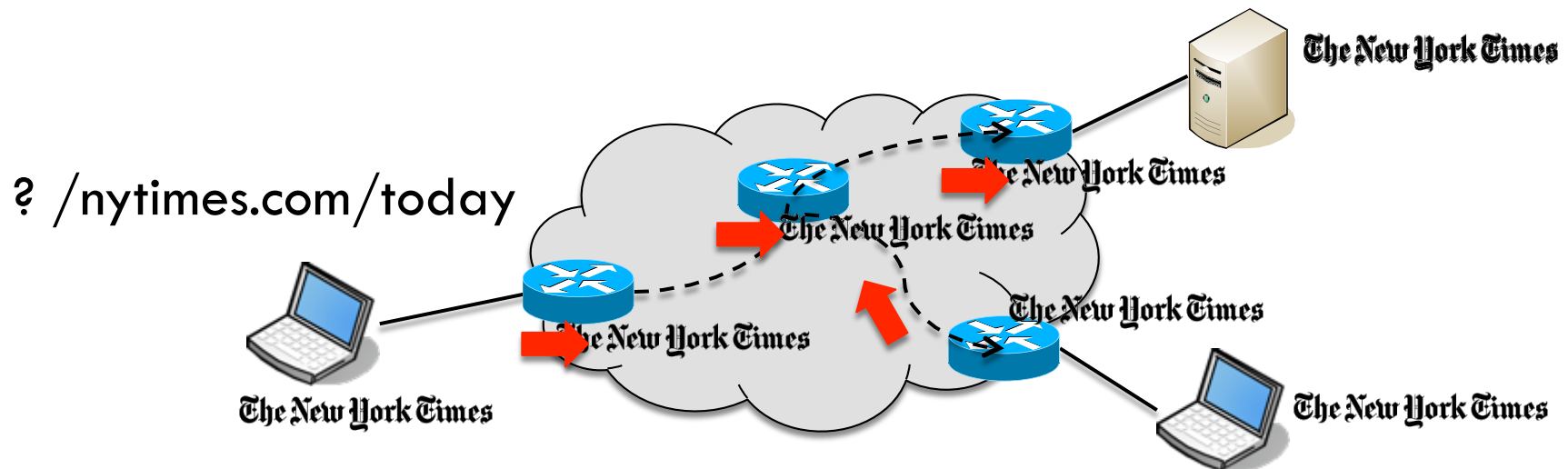
Host-centric addressing



Data-centric addressing

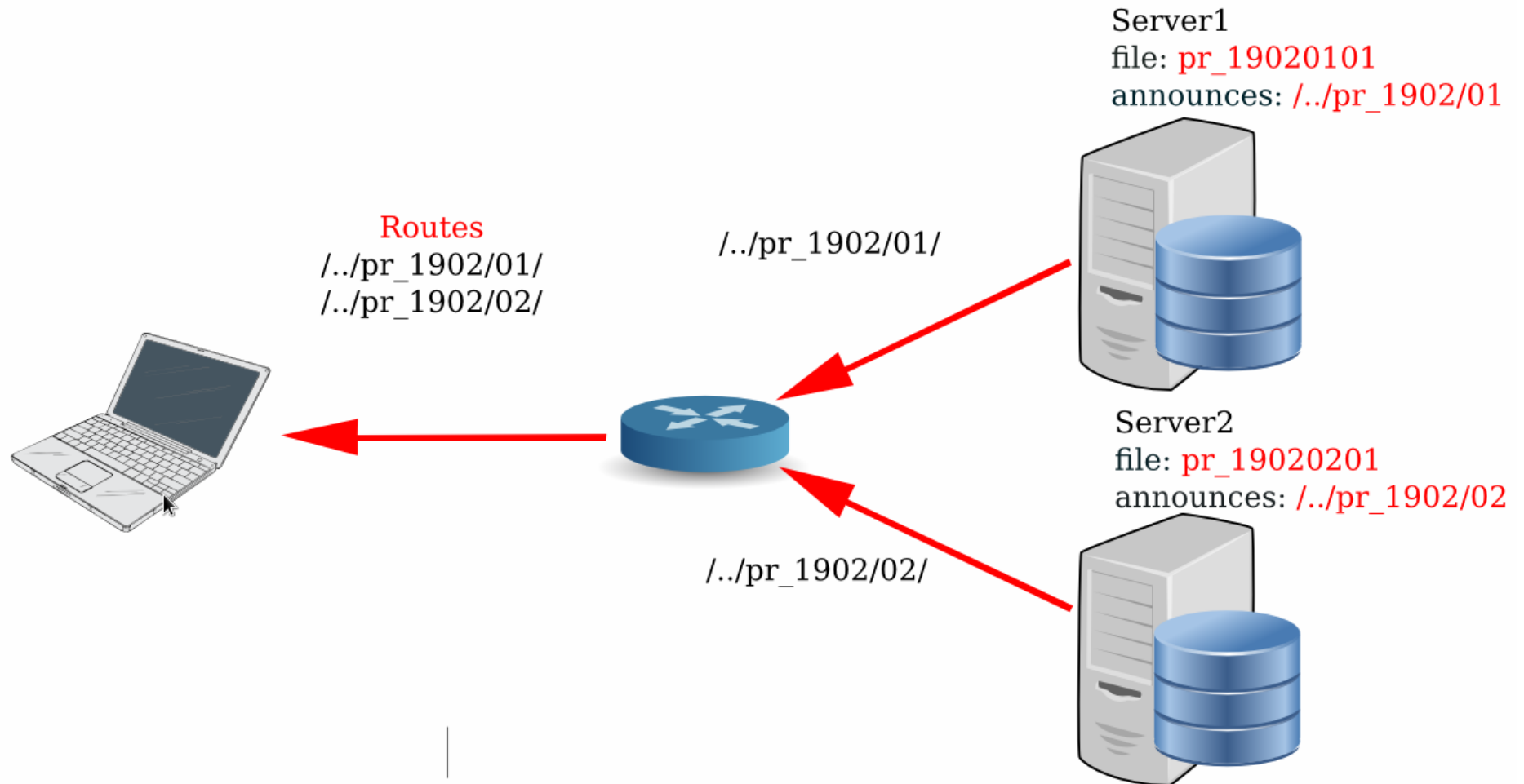


NDN Operation



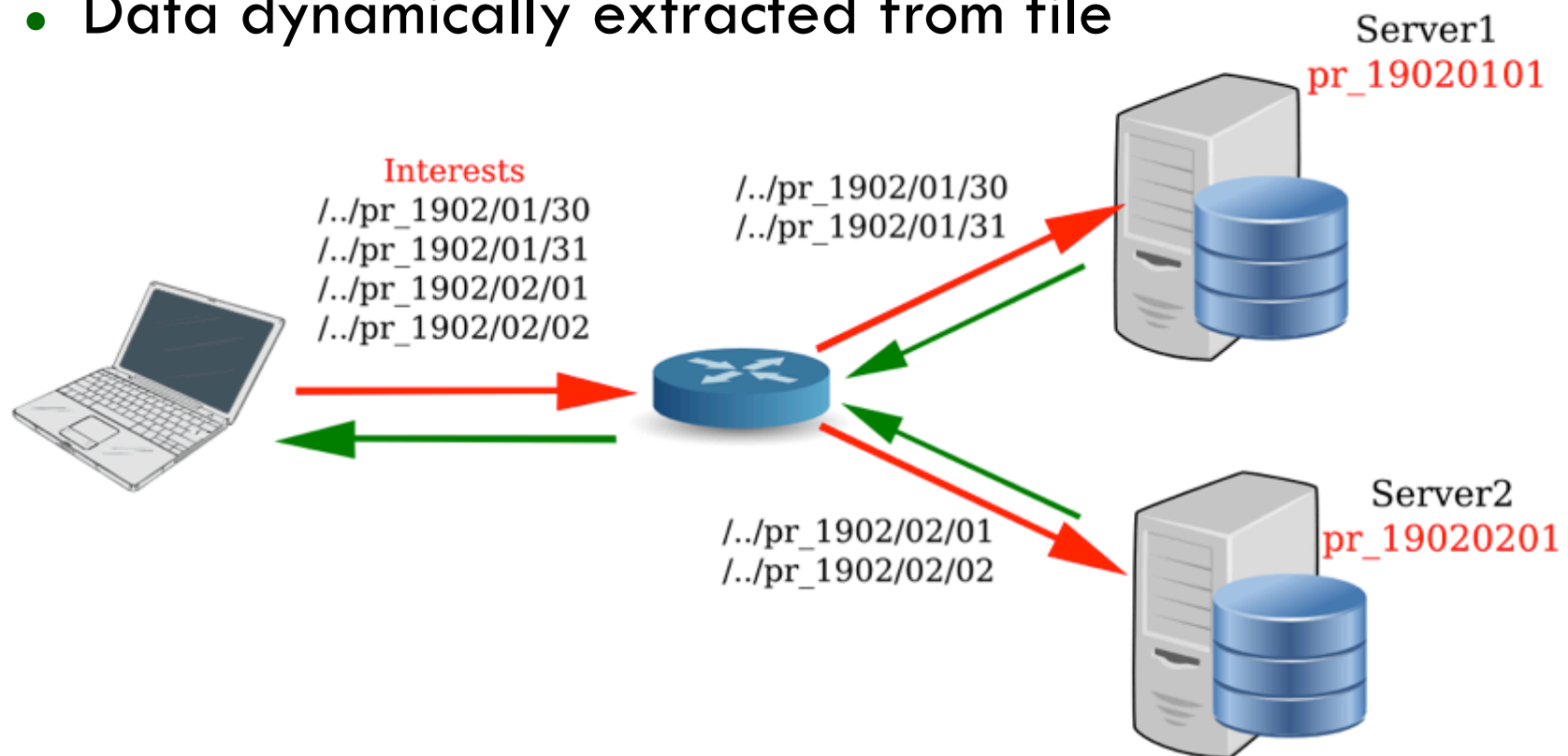
- ❑ Publishers push **hierarchical** name prefixes into the network
- ❑ Users send **Interests** follow path to published prefix
- ❑ “Breadcrumbs” direct **data** back to the user
- ❑ Data is **cached** into the network

Content Publishing



Data Request

- Interests for Jan 30-31 go to server1
- Interests for Feb 01-02 go to server2
- Data dynamically extracted from file



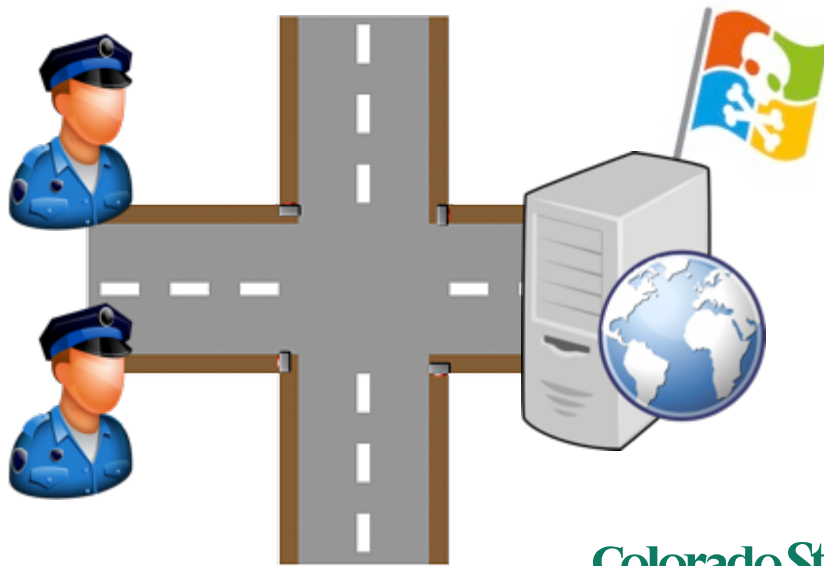
This Sounds Awfully Complex..

It's actually quite simple:

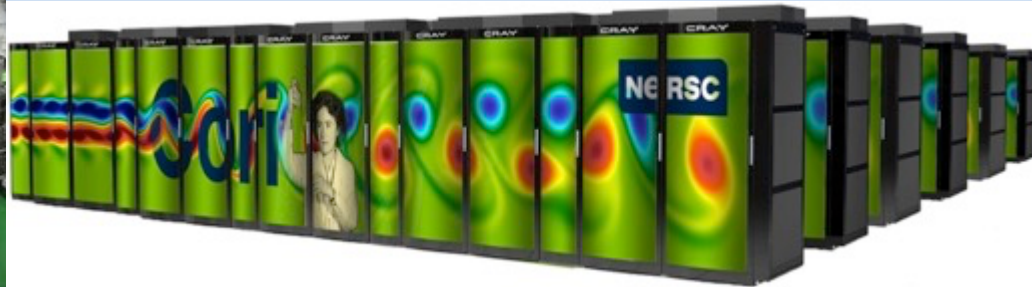
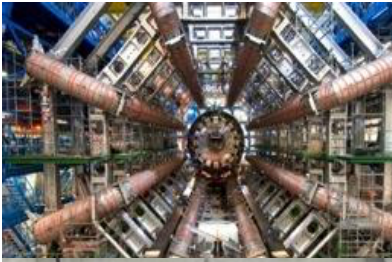
- First, name your datasets with a hierarchical, community agreed name:
 - `/store/mc/fall13/BprimeBprime_M_3000/GEN-SIM/POSTLS162_v1-v2/10000 /<UUID.root>`
- Then, advertise the *prefix* to the network:
 - I can answer any questions starting with:
 - `/store/mc/fall13/BprimeBprime_M_3000/GEN-SIM/POSTLS162_v1-v2/*`
- Finally, let users issue interests with the appropriate name or name prefix

Named Data is Easy to Secure

- In the Internet you secure your path..
- ..but the server may still be hacked!
- In NDN you **sign** the data with a **digital signature**..
- ..so the users know when they get bad data!



Experimental and observational science deals with big and small instruments, and a lot of data!



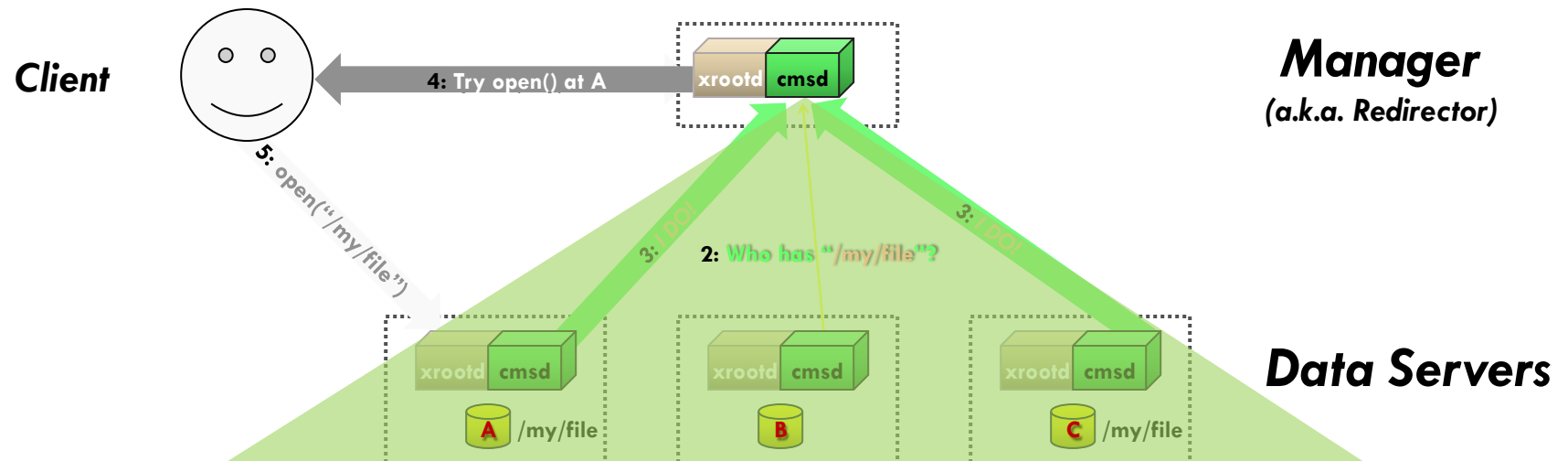
- Data volumes are increasing faster than Moore's Law
- New algorithms and methods for analyzing data
- Infeasible to put a supercomputing center at every experimental facility

Inder Monga, ESnet



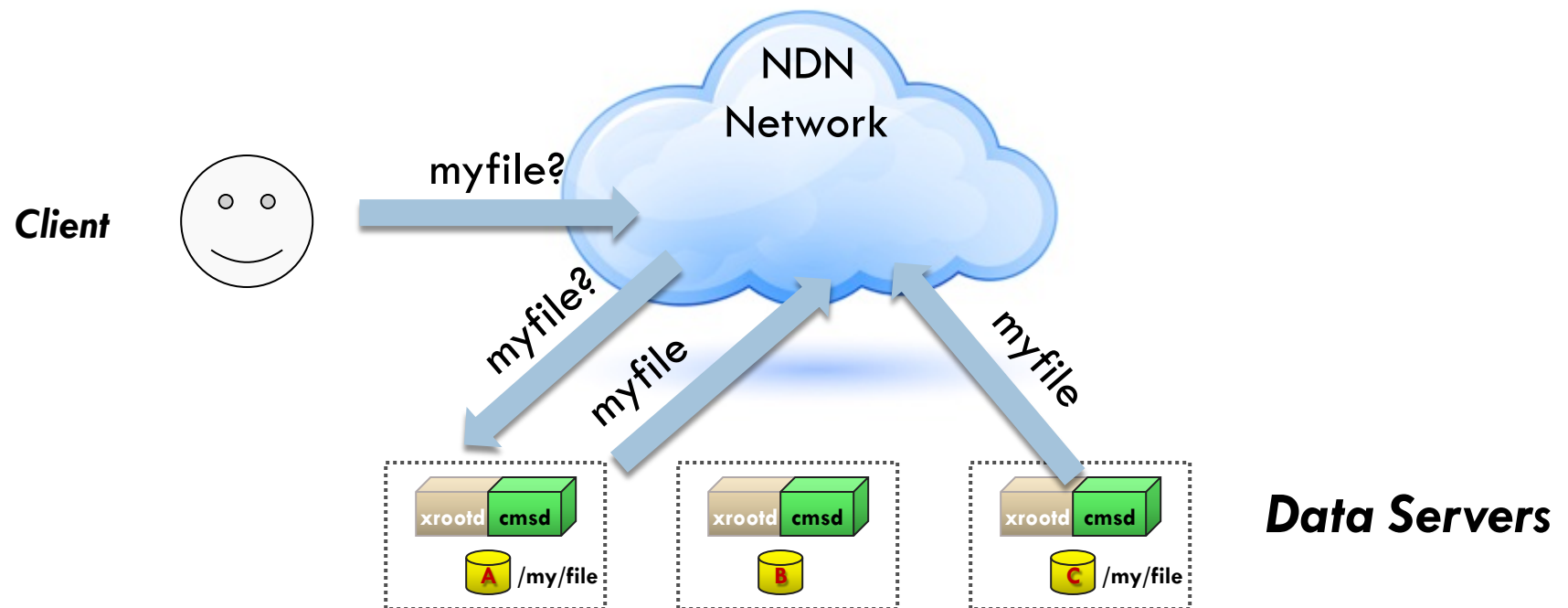
Simplifying a Complex System: xrootd Cluster

Here is how xrootd works today:



xrootd under NDN

No manager, fewer steps, more robust



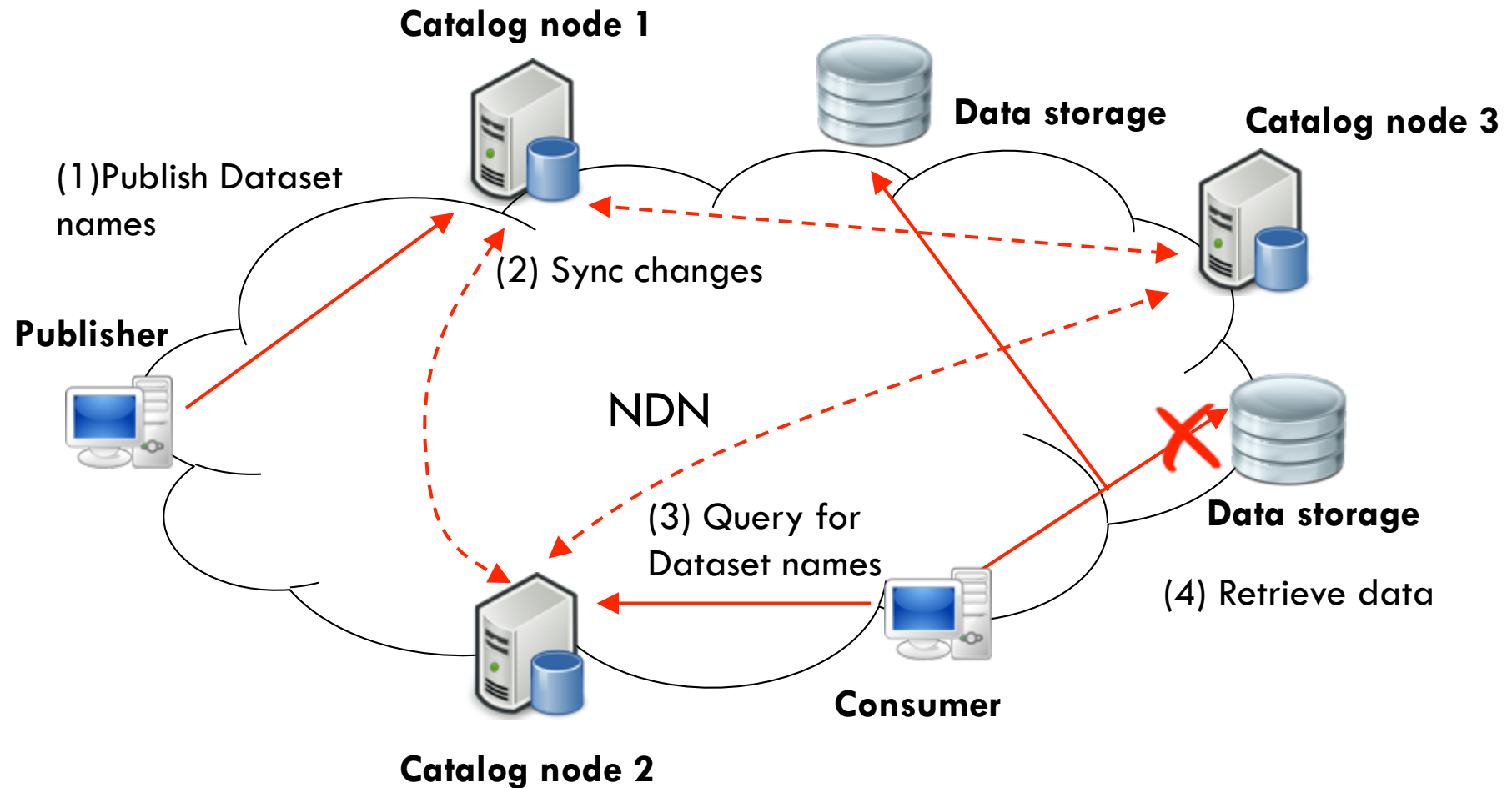
Supporting Science Applications

- Scientific apps generate tremendous amounts of data and face challenging management issues
 - Climate science CMIP5 dataset: 3.5 PB
 - High Energy Physics (HEP): 1 PB/s raw data, ATLAS project filters to 4 PB/yr
 - Data distributed to various local repositories
 - Variety of data naming schemes
 - E.g. different units and user defined parameters
 - Data provenance
- Existing, mature, software for dataset discovery, publishing, and retrieval
 - E.g. ESGF, xrootd, etc.
 - Lots of effort to overcome fragility of IP's host-centric paradigm

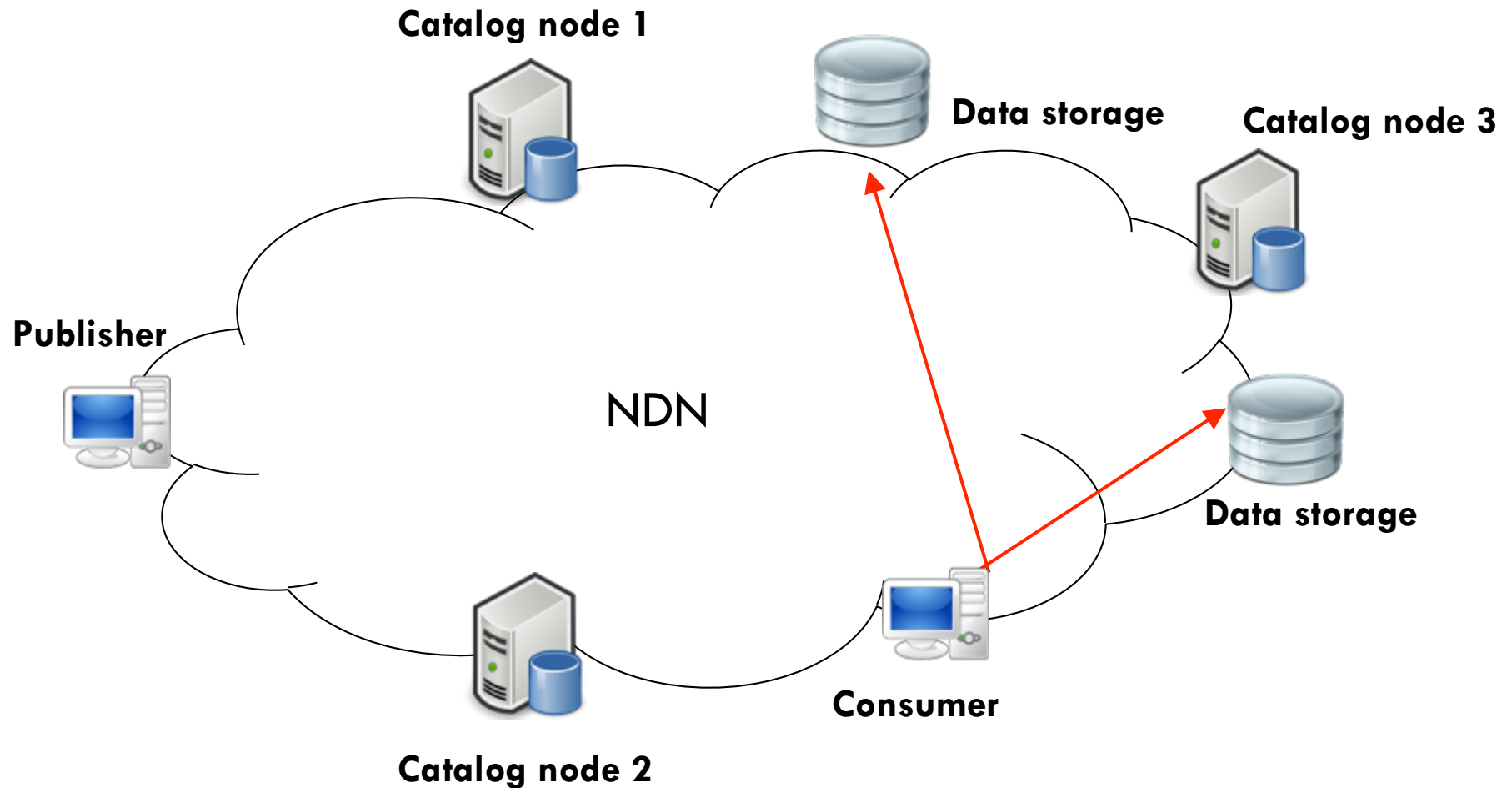
First Step – Build a Catalog

- Create a **shared resource** – a distributed, synchronized catalog of names over NDN
 - Provide common operations such as publishing, discovery, access control
 - Catalog only deals with name management, not dataset retrieval
 - Platform for further research and experimentation
- Research questions:
 - Namespace construction, distributed publishing, key management, UI design, failover, etc.
 - Functional services such as subsetting
 - Mapping of name-based routing to tunneling services (VPN, OSCARS, MPLS)

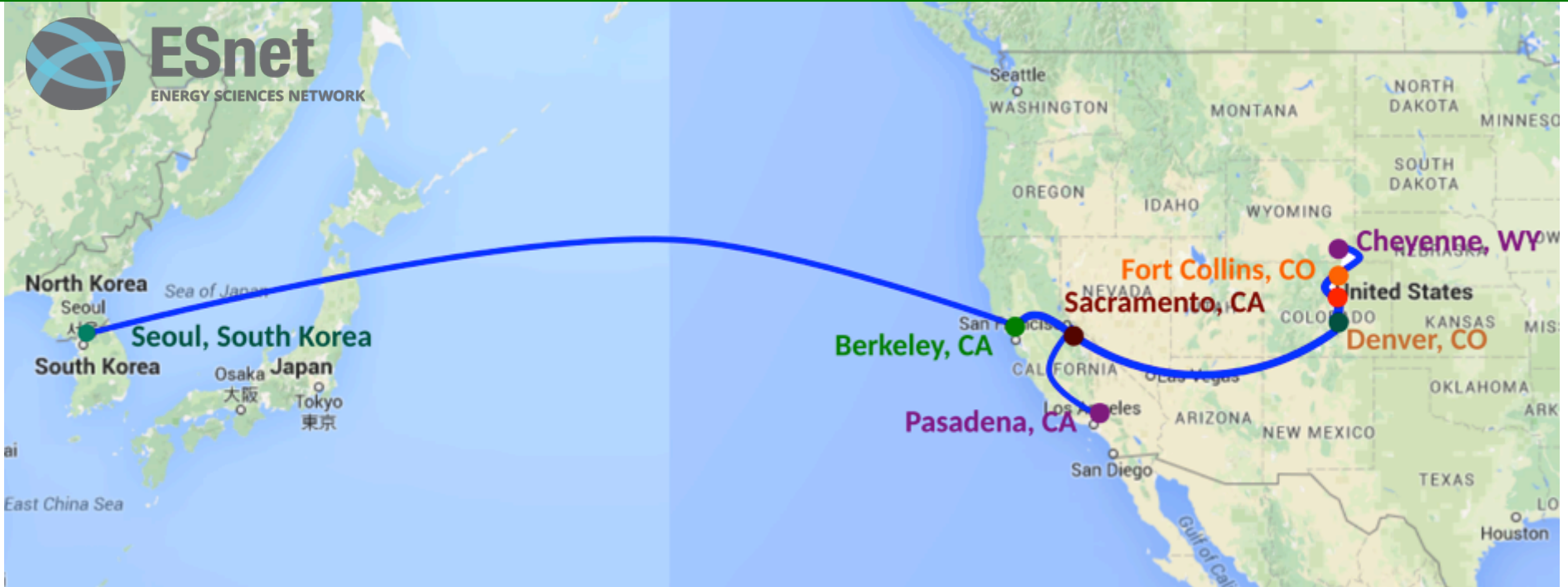
NDN Catalog



Forwarding Strategies



Science NDN Testbed



- NSF CC-NIE campus infrastructure award
 - 10G testbed (courtesy of ESnet, UCAR, and CSU Research LAN)
- Currently ~50TB of CMIP5, ~20TB of HEP data

General NDN Testbed

<http://ndnmap.arl.wustl.edu/>

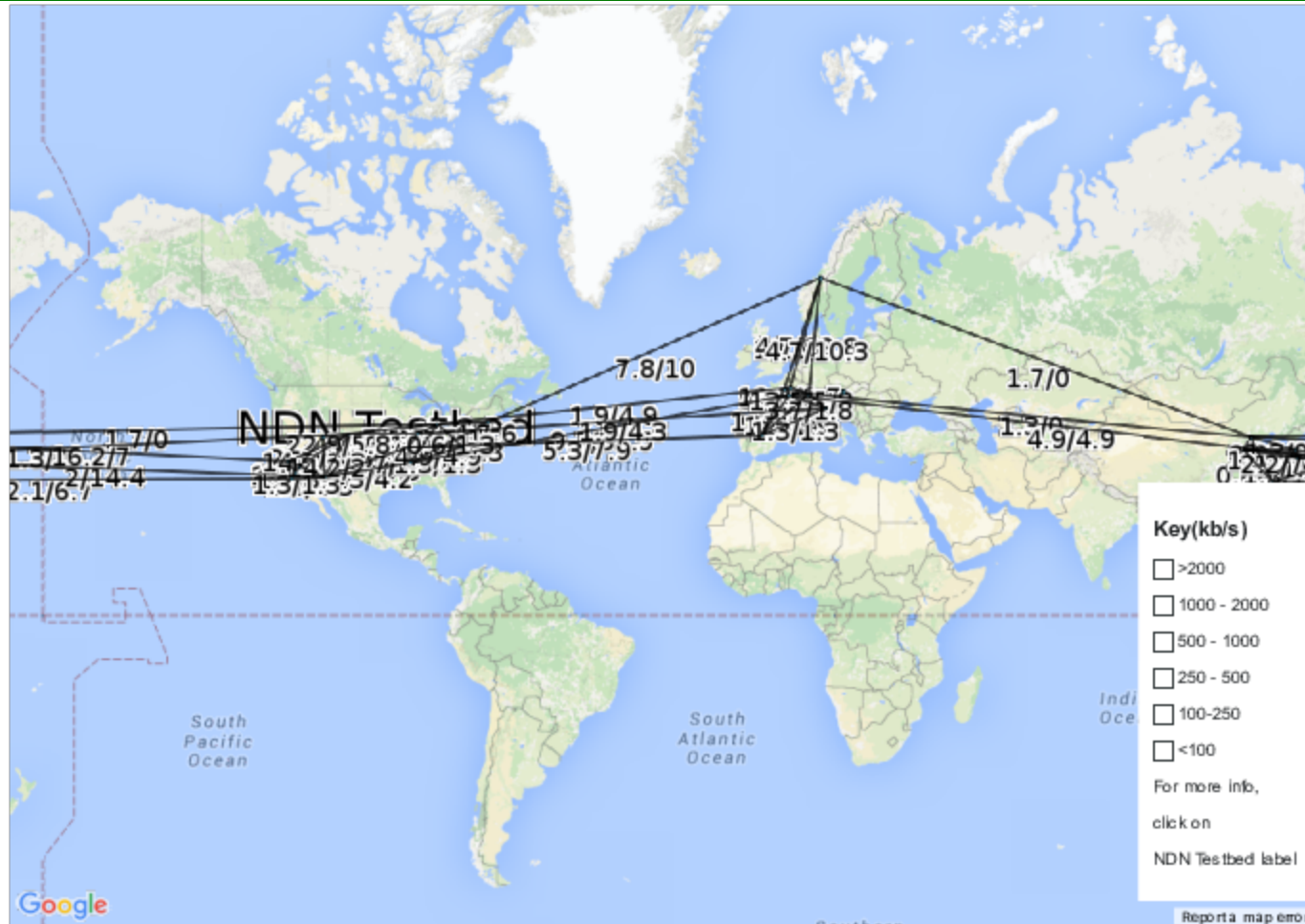
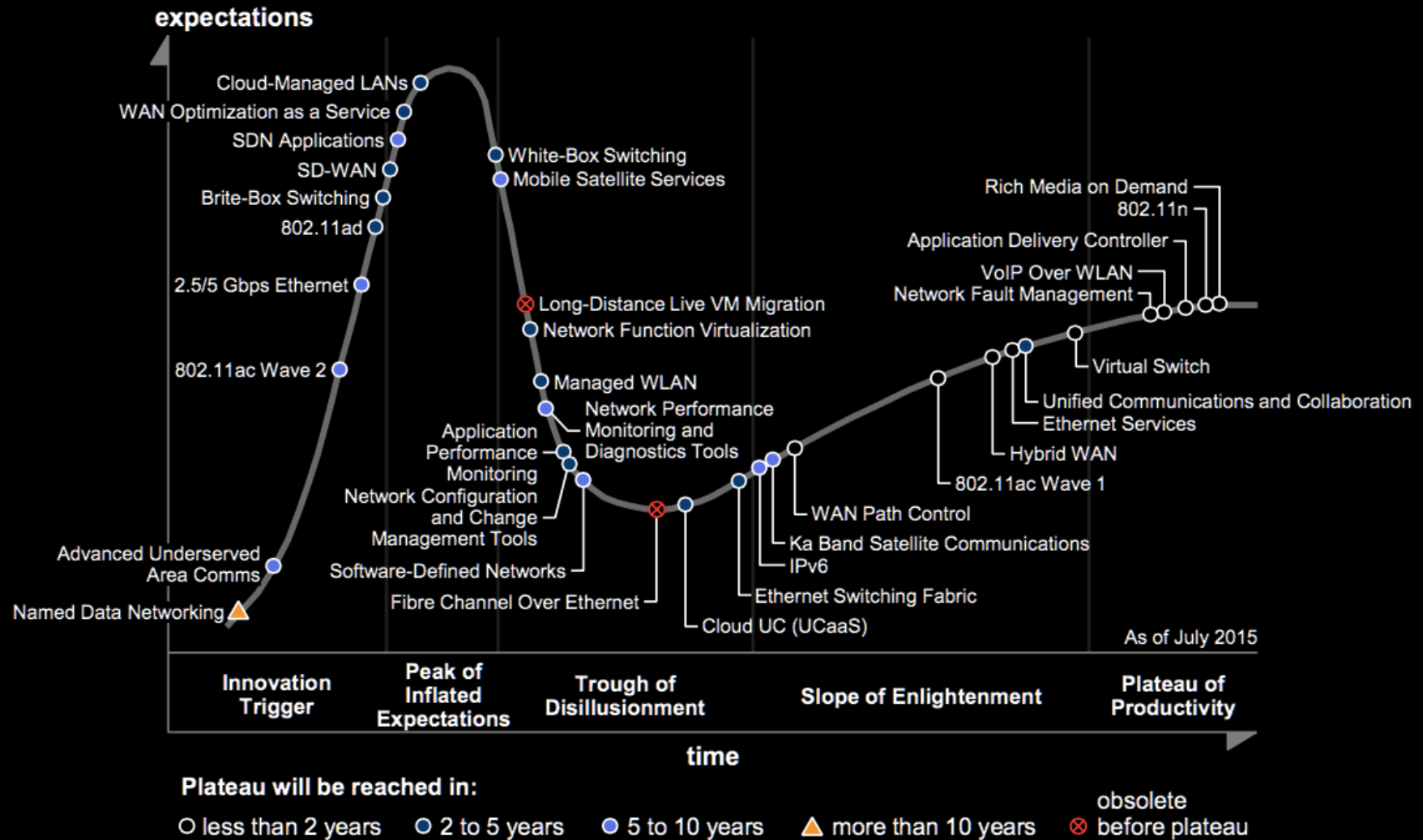


Figure 1. Hype Cycle for Networking and Communications, 2015



Source: Gartner (July 2015)

Collaboration Opportunities

- ❑ Big Science
- ❑ Joining Testbeds
- ❑ Hierarchical naming still challenging, not fully expressive, but compromise between expressiveness, performance, security
- ❑ Caching as an economic problem
- ❑ Privacy, encryption, advertising models
- ❑ Forwarding strategies
- ❑ Android, WRT implementations, IoT (light bulbs), Arduino devices
- ❑ Creative application development

Conclusions

- NDN encourages common **data** access methods where IP encourages common **host** access methods
 - NDN encourages interoperability at the content level
- Many playgrounds for you to play
 - Science, multimedia, IoT, Android, wireless and more
 - Ready-to-try catalog, supports a variety of applications
 - UI for data search and retrieval.

For More Info

christos@colostate.edu

<http://named-data.net>

<http://github.com/named-data>