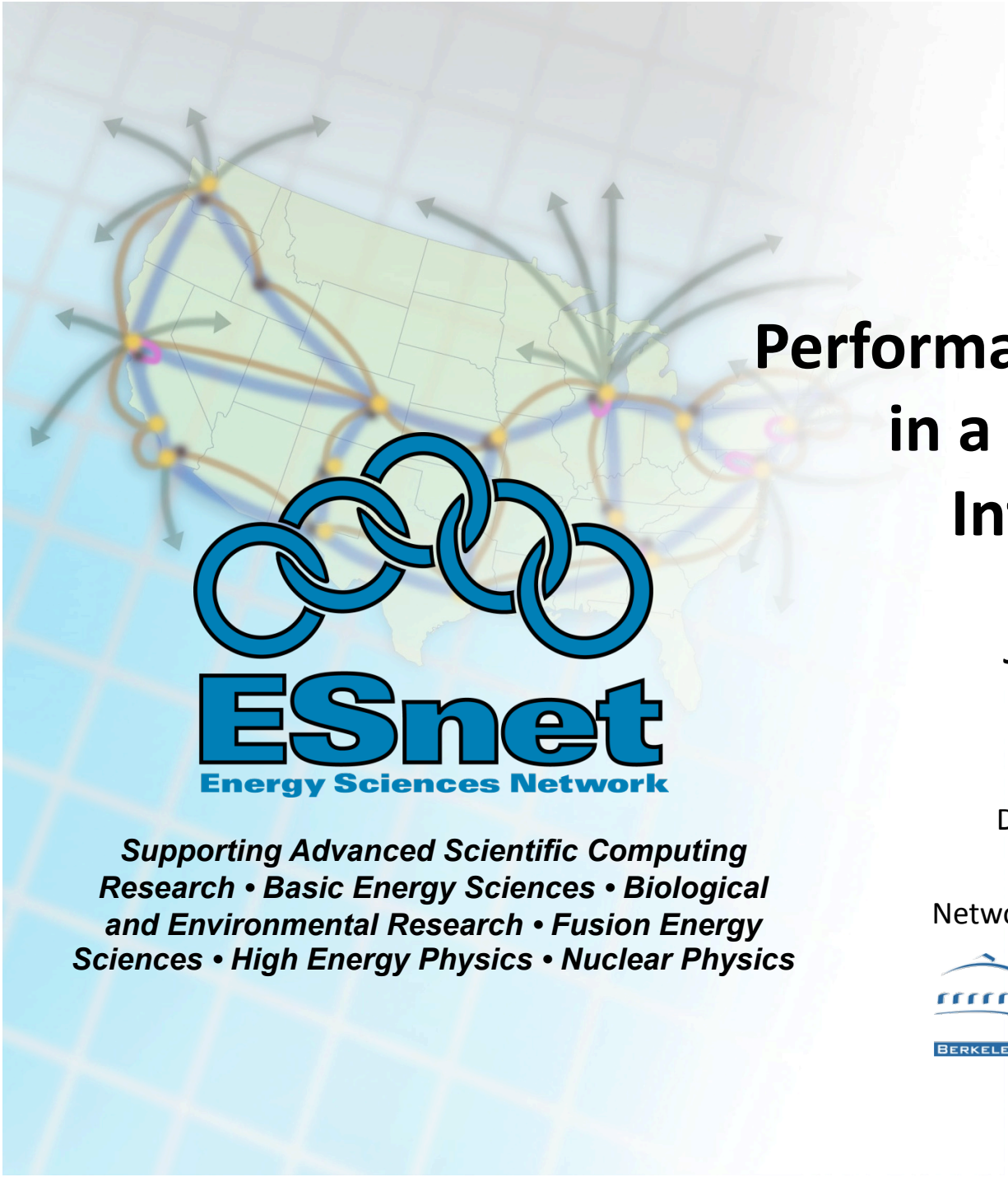


ESnet
Energy Sciences Network

Supporting Advanced Scientific Computing Research • Basic Energy Sciences • Biological and Environmental Research • Fusion Energy Sciences • High Energy Physics • Nuclear Physics

Performance Measurement in a Multi-Domain Infrastructure

June 21, 2009

Thomas Ndousse
DOE Program Manager
DOE Office of Science
Joe Metzger
Network Engineer at ESnet/LBNL

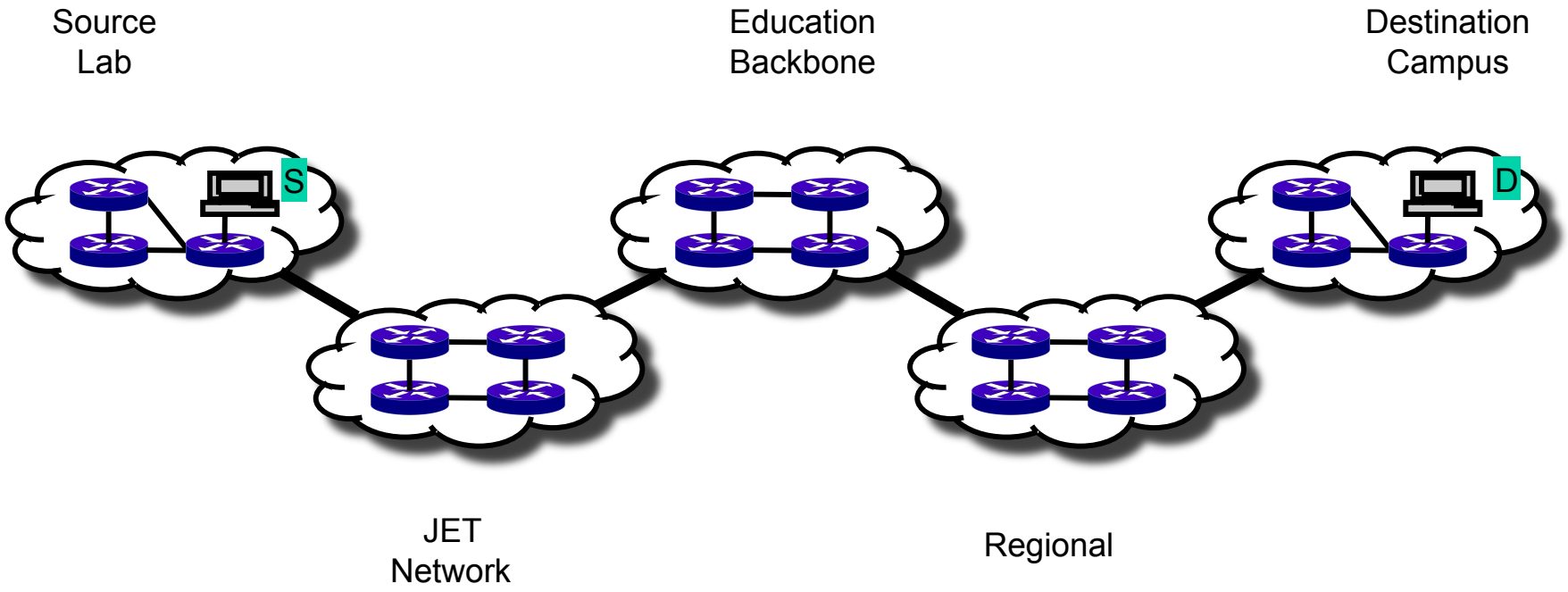
BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science

High Performance Networking

- The Department of Energy's Office of Science is one of the largest supporters of basic research in the physical sciences in the U.S.
  - Directly supports the research of some 15,000 scientists, postdocs and graduate students at DOE laboratories, universities, other Federal agencies, and industry worldwide
  - Operates major scientific facilities at DOE laboratories that that have participation by the US and international research and education (R&E) community
- Many users are not achieving the end to end performance they need.
  - Naive users should be able to easily get
    - 200 Mbps/per stream between properly maintained systems
    - 2 Gbps aggregate rates between significant computing resources
  - "There is widespread frustration involved in the transfer of large data sets between facilities, or from the data's facility of origin back to the researcher's home institution. " From the BES network requirements workshop: http://www.es.net/pub/esnet-doc/BES-Net-Req-Workshop-2007-Final-Report.pdf
- We can increase performance by measuring and reporting it!

Network Troubleshooting is a Multi-Domain Problem



Soft Network Failures

- Soft failures are where basic connectivity functions, but high performance is not possible.
- TCP was intentionally designed to hide all transmission errors from the user:
  - "As long as the TCPs continue to function properly and the internet system does not become completely partitioned, no transmission errors will affect the users." (From IEN 129, RFC 716)
- Soft failures are common at DMZ's between networks.
- Common network management systems do not detect many soft failures

Addressing the Problem: perfSONAR

- Developing an open web-services based framework for collecting, managing and sharing network measurements
- Deploying the framework across the science community
- Encouraging people to deploy *'known good'* measurement points near domain boundaries

perfSONAR Developers

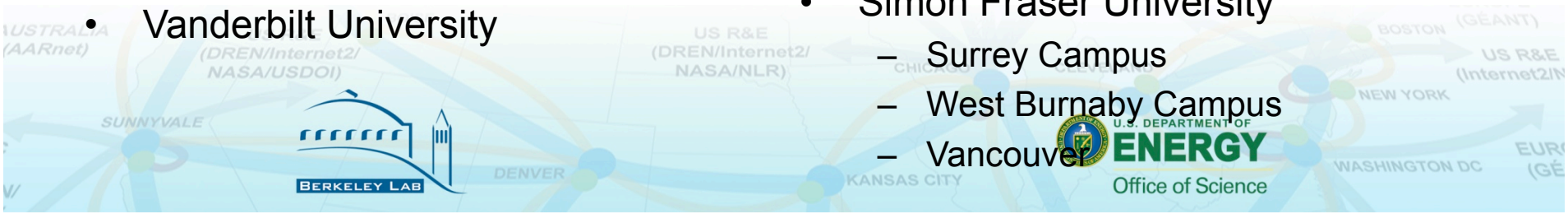
- ESnet
- GEANT
- Internet2
- RNP
- University of Delaware
- FERMI
- Georgia Tech
- SLAC
- ARIES
- BELNET
- CARNet
- CESNET
- DANTE
- DFN
- FCCN
- Consortium GARR
- GRNET
- IST
- POZNAN Supercomputing Center
- Red IRIS
- Renater
- SURFnet
- SWITCH
- UNINETT

perfSONAR Deployments

- Internet2
- University of Michigan, Ann Arbor
- Indiana University
- Boston University
- University of Texas Arlington
- Oklahoma University, Norman
- Michigan Information Technology Center
- William & Mary
- University of Wisconsin Madison
- Southern Methodist University, Dallas
- University of Texas Austin
- Vanderbilt University

- ESnet
- Argonne National Lab
- Brookhaven National Lab
- Fermilab
- National Energy Research Scientific Computing Center
- Pacific Northwest National Lab

- APAN
- GLORIAD
- JGN2PLUS
- KISTI Korea
- Monash University, Melbourne
- NCHC, HsinChu, Taiwan
- Simon Fraser University
  - Surrey Campus
  - West Burnaby Campus
  - Vancouver



perfSONAR Deployments (2)

- GEANT
- GARR
- HUNGARNET
- PIONEER
- SWITCH
- CCIN2P3
- CERN
- CNAF
- DE-KIT
- NIKHEF/SARA
- PIC
- RAL
- TRIUMF
- ASCC?

perfSONAR JET deployment

- The Joint Engineering Team is developing a perfSONAR deployment plan
  1. Reviewing the network measurement data each network is willing to share, or would like to access
  2. Reviewing the perfSONAR tools & monitoring functions to evaluate which networks will deploy which ones.
- First deployments in the nets with open science missions and exchange points

Example: US Atlas

- Tier 1 to Tier 2 Center Data transfer problem
  - Couldn't exceed 1 Gbps across a 10GE end to end path that included 5 administrative domains
  - Used perfSONAR tools to localize problem
  - Identified problem device
    - An unrelated domain had leaked a full routing table to the router for a short time causing FIB corruption. The routing problem was fuxed, but router started process switching some flows after that.
  - Fixed it
    - Rebooting device fixed the symptoms of the problem
    - Better BGP filters on that peer will prevent reoccurrence (of 1 cause of this particular class of soft faults)

Example: NERSC & OLCF

- Users were having problems moving data between supercomputer centers
  - One user was: “waiting more than an entire workday for a 33 GB input file”
- perfSONAR Measurement tools were installed
  - Regularly scheduled measurements were started
- Numerous choke points were identified & corrected
- Dedicate wide area transfer nodes were setup
  - Tuned for Wide Area Transfers
  - Now moving 40 TB in less than 3 days

How to Participate

- Deploy perfSONAR
- Use perfSONAR to find & correct the hidden performance problems in your networks.
- Attend the BOF during Lunch today for more info.

ESnet
Energy Sciences Network

Questions?