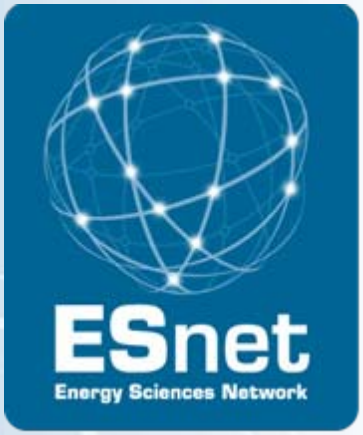


ESnet's Advanced Networking Initiative and Testbed

Steve Cotter, Lawrence Berkeley National Lab

ESnet Department Head

GLIF Workshop 2010

Geneva, Switzerland

October 13, 2010

The Energy Sciences Network

Mission: provide the critical network infrastructure that supports the Department of Energy's Office of Science missions

- Sharing of massive amounts of data
- Thousands of collaborators world-wide
- Distributed data processing
- Distributed data management
- Distributed simulation, visualization, and computational steering

ESnet directly supports the research of ~15,000 scientists, postdocs and graduate students at DOE laboratories, universities, other federal agencies, and industry worldwide

Since 1986, providing the reliable connections, science-driven innovation and user focus that enables scientists to collaborate, manage, and exchange data

ESnet Network Facts

One of two largest R&E networks in the US by footprint

- Transports massive quantities of scientific data from Office of Science facilities to the associated community of science collaborators
  - Including all of CERN's Large Hadron Collider data in the US
- Also transports Labs' operational traffic

Rich connectivity with R&E and commercial networks

- >140 peerings with commercial & research networks around the world
- >85% of all traffic either originates or terminates off-net

Recent budget

- 2010: $29.8M + $62.4M in ARRA funding
- 2011: $30.0M

ESnet4

4th Generation of the Energy Sciences Network

Uses wavelengths on a dedicated optical infrastructure managed by Level3

Hybrid Network Architecture

- IP: for regular flows
- Science Data Network: for huge data flows

OSCARS

- Guaranteed, agile multi-domain virtual circuits

perfSONAR

- Performance measurements, archival, testing and debugging
  http://www.perfsonar.net

http://weathermap.es.net

Strategic Imperatives for ESnet

Higher bandwidth for lower cost
- Pursuing dark fiber as a means to affordably scale to terabit capacities

Richer services set
- Optimized network services in a layer-aware, multi-technology, multi-vendor environment

Effective operational environment
- Real-time analysis of network traffic trends for predictive provisioning and re-configuration

Security
- Develop real-time "global" view of network attacks individual Labs would not see

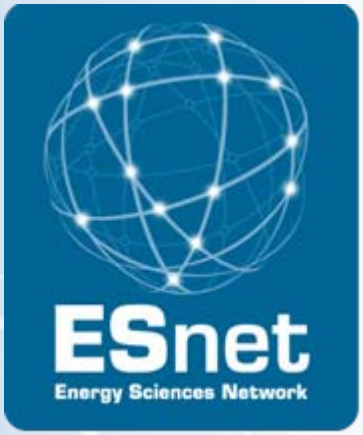
Energy efficiency
- Efficiency of the network – lowest layer, eliminate layers, devices
- Using the network to improve energy efficiency of science

Network research and experimentation
- Investing in innovations and developing advanced capabilities

Developing a Sustainable Research Ecosystem

Four main elements to an effective network research ecosystem:

- Participate in research and software development to enhance our end-user's experience
  - Advanced services development like OSCARS and perfSONAR
  - Current DOE funded research: E-Center and ARCHSTONE projects
- Collaborative peek into the future with scientists
  - Requirements workshops, PI meetings, etc.
- Guidance and actively following longer-term research
  - Professors, researchers, R&E networks, SBIRs
  - 20+ letters of support, review of proposals and guidance
- Attract external researchers to help solve relevant problems
  - Through ANI Testbed capabilities and proposal process
  - Interaction with NSF GENI programs and other research initiatives
  - Joint faculty appointments (Dr. Ben Yoo)



100 Gbps Prototype Network, Testbed Facility

Advanced Networking Initiative

Advanced Networking Initiative

Project has 3 goals:

1)Build an end-to-end 100G prototype network between the three DOE supercomputing facilities and NYC international exchange point
   a) Accelerate the deployment of 100G technologies in uncertain economic times
   b) First step toward DOE's vision of 1 Tbps links between Exascale supercomputers

2)Build an experimental network research environment for researchers and industry at sufficient scale to usefully test experimental approaches to next generation networks

3)Utilize one-time funding to build a persistent, cost-effective infrastructure for science

DOE is funding an additional $5M in network research using the testbed facility with the goal of near-term technology transfer to the production ESnet network

ANI 100G Prototype Network Timeline

RFP issued June 15, asking for:

- 100G Service between MANLAN, ORNL, ANL, and NERSC
- 20-year dark fiber IRU for nationwide footprint
  - Including metro fiber in SanFran Bay area & Chicago area

Responses received Aug 23

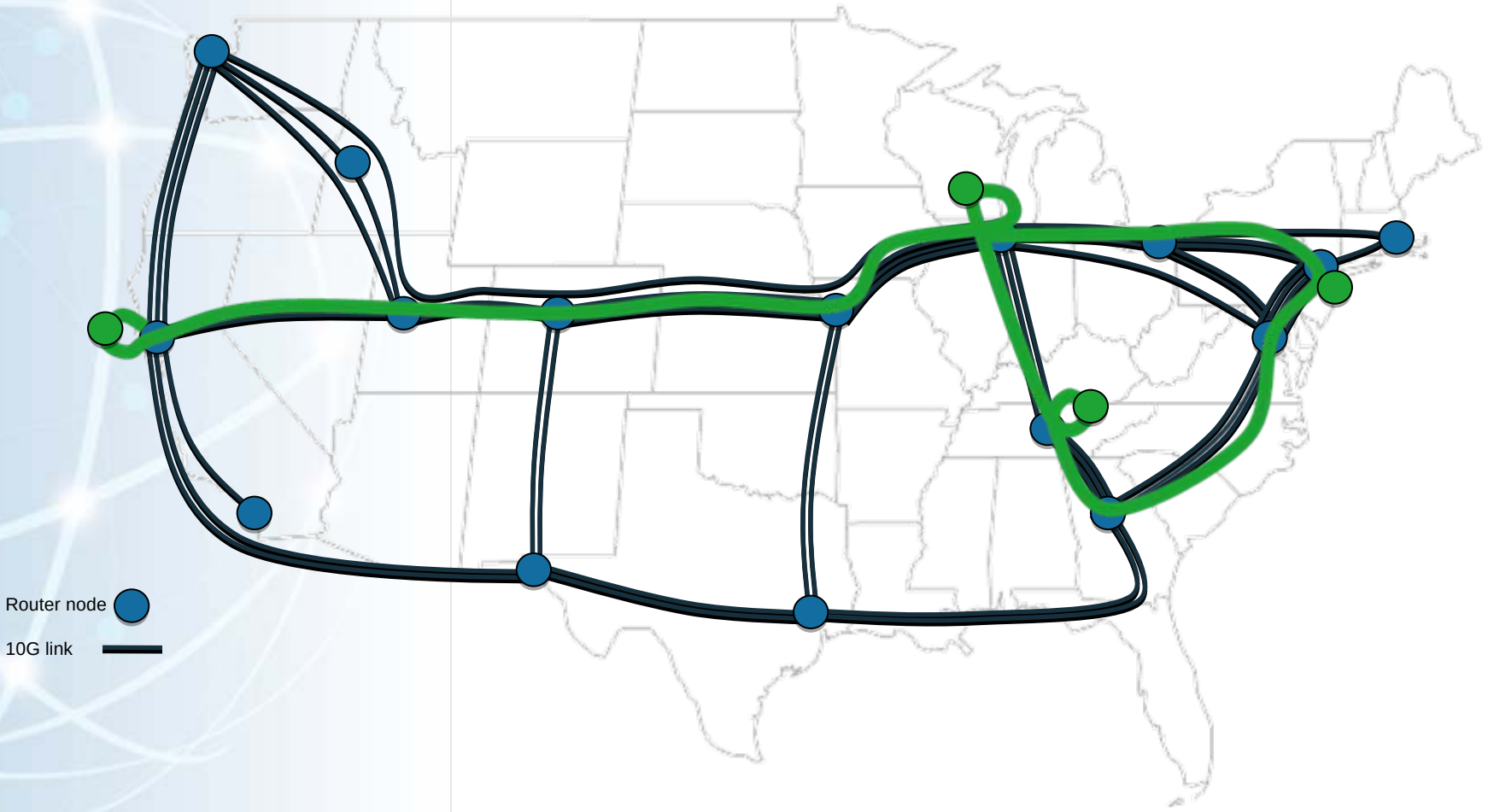
- Proposals are currently being evaluated, negotiations set to start

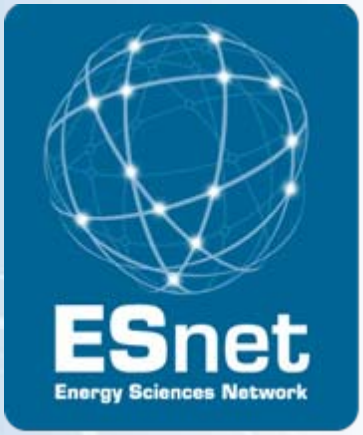
Started testing 100G routers in our lab in July

Tentative schedule:

- Decision on 100G RFP by end of Oct
- Contract negotiation and DOE approval: Nov-Dec
- Contract awarded: December
- 100G router RFP will be issued in January

100 Gbps Prototype Topology





Overview, Configuration, Research

ANI Testbed Facility

Testbed Overview

Progression

- Starting out as a tabletop testbed
- Move to Long Island MAN when dark fiber build is complete (Nov)
- Extend to WAN when 100G prototype network available

Capabilities

- Ability to support end-to-end networking, middleware and application experiments, including interoperability testing of multi-vendor 100 Gbps network components
- Researchers get “root” access to all devices
- Use Virtual Machine technology to support custom environments
- Detailed monitoring so researchers will have access to all possible monitoring data

Network Testbed Components

Tabletop network testbed consists of:

- 6 DWDM devices (Layer 0-1)
- 4 Layer 2 switches supporting Openflow
- 2 Layer 3 routers
- Test and measurement hosts
  - Virtual Machine based test environment
  - 4x10G test hosts initially
  - Eventually 40G and 100G from Acadia 100G NIC project

This configuration will evolve over time

Why Openflow?

Allows researchers to manipulate flows

- A software programmable network

Enables multi-layer experimentation

- For example: divert flows to Infinera's or routers

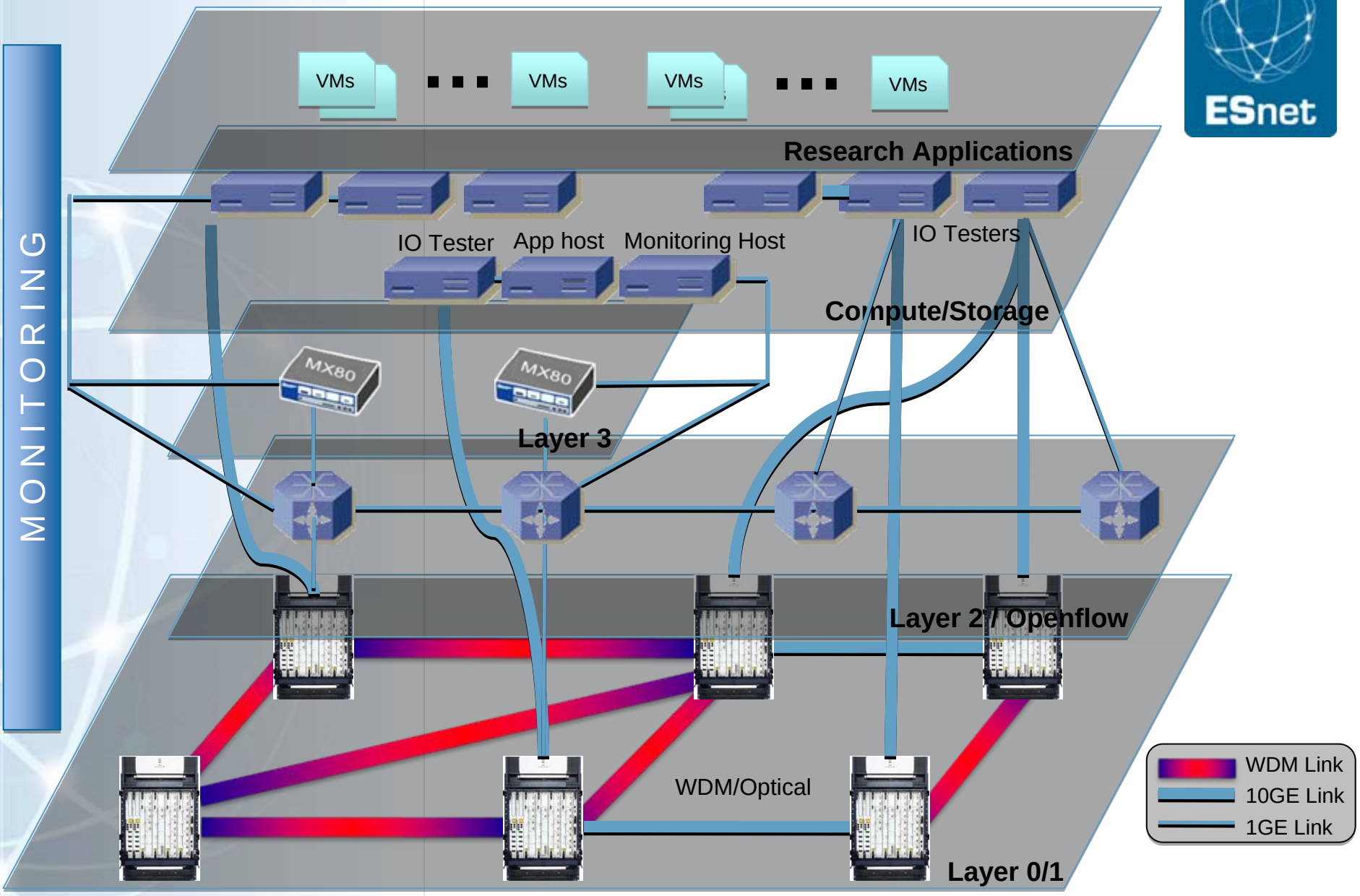
Create your own flow routing paradigm

- Router flows based on your NOX application rather than well known protocols.

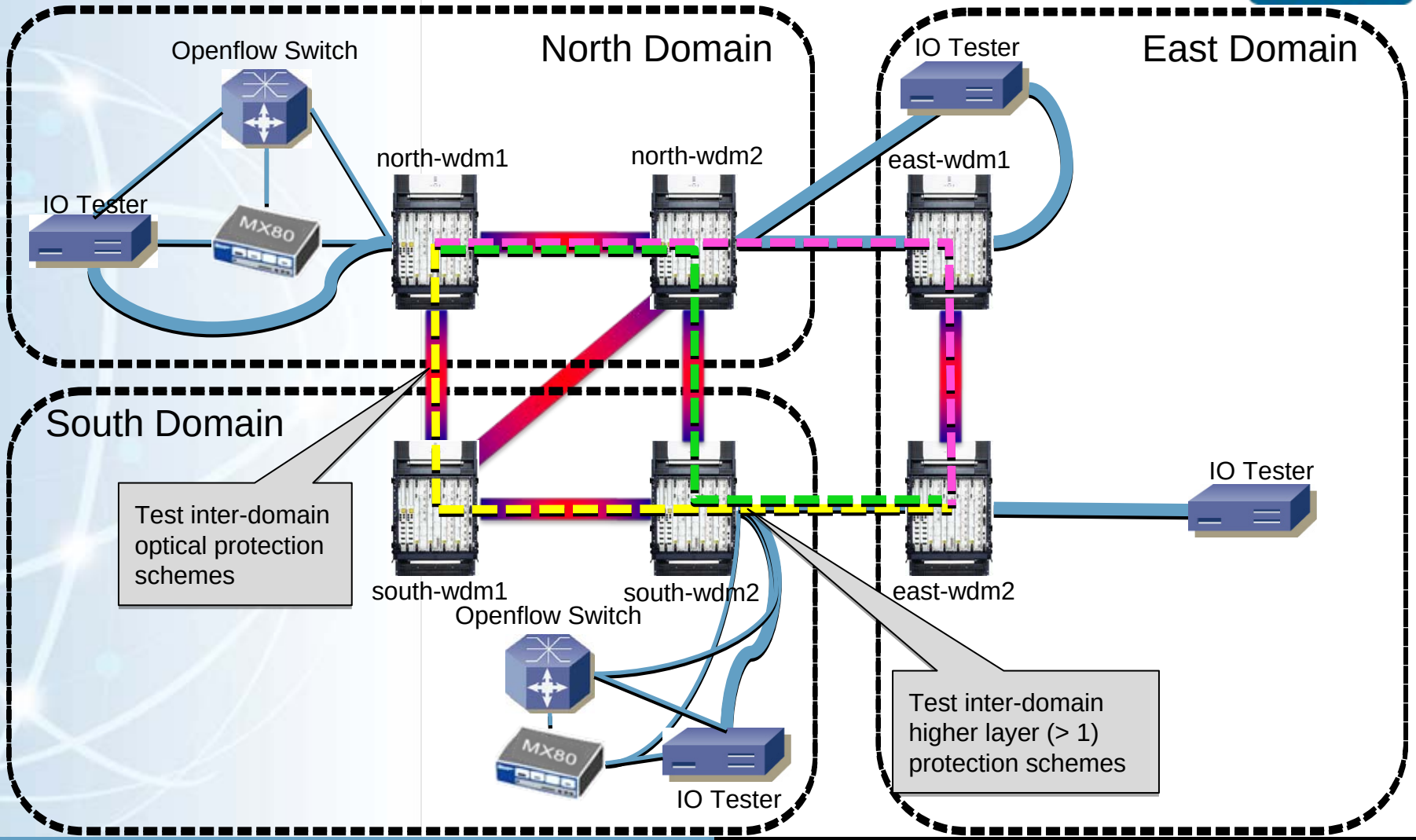
Limitations due to flow record and vendor hardware capabilities

- Few matches/actions supported (like instruction set of a processor)

Tabletop: A layered view



Sample Configuration: Multi-Domain Multi-Layer Protection Testing



Testbed Status

Tabletop testbed available for log in since June

- Researchers are logging in, configuring VMs, running tests, etc.
- Reserving testbed components using Google calendar
- Project PIs asked to submit monthly reports (Google spreadsheet survey form)

User documentation mostly complete:

- https://sites.google.com/a/lbl.gov/ani-testbed-user-guide/

Per-project monitoring is set up

- https://tb-webdav-1.es.net/ganglia/

Testbed-support@es.net email list is quite active

A few remaining tasks to be done: e.g. web interface to claim reserved resources

Current Projects Using the Testbed

Project	Summary	Expected Results
Archstone PI: Tom Lehman, ISI	To dynamically create “slices” of resources across multiple network layers in a vertically integrated manner, so as to generate virtual network topologies. This requires a highly-advanced path computation element which extends the concept of simple path computation to multi-layer, multi-dimensional topologies.	This project will utilize the ANI Testbed to determine design requirements, test alternatives, and evaluate performance of the developed technologies.
FlowBench PI: Prasad Calyam, Ohio Supercomputer Center	To set up different physical topologies in testbed using resources such as NEC Openflow switches, App Hosts, and Monitoring hosts. On these topologies, we will experiment with Openflow and benchmark performance of GridFTP file transfers with enhanced TCP/UDP variants.	The Testbed will be used to confirm that our developed technologies will operate as desired with production network equipment, topologies, and configurations.
HNTES PI: Malathi Veeraraghavan, University of Virginia	Hybrid Network Traffic Engineering Software: The purpose of HNTES is to leverage both an IP datagram network and a high-speed optical dynamic circuit network to best serve users' data communication needs.	Experiments on the testbed will be conducted to determine whether flows can be redirected on-the-fly to newly established optical circuits without impacting TCP behavior, and user-perceived performance.

Projects Using the Testbed

Project	Summary	Expected Results
Climate 100 PI: Alex Sim, LBL	The Climate100 project integrates massive climate datasets, emerging 100 Gbps networks, and state-of-the-art data transport and management technologies. The goal of this project is to improve the understanding and use of network technologies and transition the climate community to a 100 Gbps network for production and research.	The testbed will be used to test the direct memory access over the network and new data transfers/management algorithms including the use of the 100G transfer protocol.
	Projects waiting for 100G prototype Network	
100G FTP PI: Dantong Yu, BNL	This project will design and develop an ultra high speed end-to-end file transfer protocol and tool to move science data at a speed of 100 gigabit per second (Gbps) across the national scale 100Gbps data network interconnecting research centers.	The testbed will be used to verify that this tool scales to 100Gbps on a single wavelength or multiple modulated wavelengths.
100G NIC PI: Jesse Wen, Acadia	This project will develop network interface controller (NIC) hardware and device-driver/protocol-specific software for host and gateway systems operating at 40 and 100 Gb/s.	The testbed will be used to investigate issues that do not arise in initial back-to-back testing. Such issues include interoperability with core-network gear and the effect of long-haul physical impairments.

Testbed Access

Proposal process to gain access described at:

https://sites.google.com/a/lbl.gov/ani-testbed/

First round of proposals were due Oct 1

- Accepted proposals announced Dec 10, 2011

Received 6 proposals:

1: TCP congestion control algorithm experiments

3: Control plane experiments

1: Energy efficient networking

1: High performance data movement middleware

Testbed Access

Proposal review committee members:

- DOE Lab: Phil DeMar (FNAL); Les Cottrell (SLAC)
- University: Ben Yoo (UC Davis)
- Industry: Bikash Koley (Google); David Richardson (Amazon); Steve Wolff (Cisco); Wes Doonan (ADVA)
- International: Cees De Laat (U Amsterdam); Mauro Camponelli (GARR); Tomohiro Kudoh (AIST)
- Other: Jerry Sobiesky (NORDUnet); Kevin Thompson, (DHS)

Testbed Support

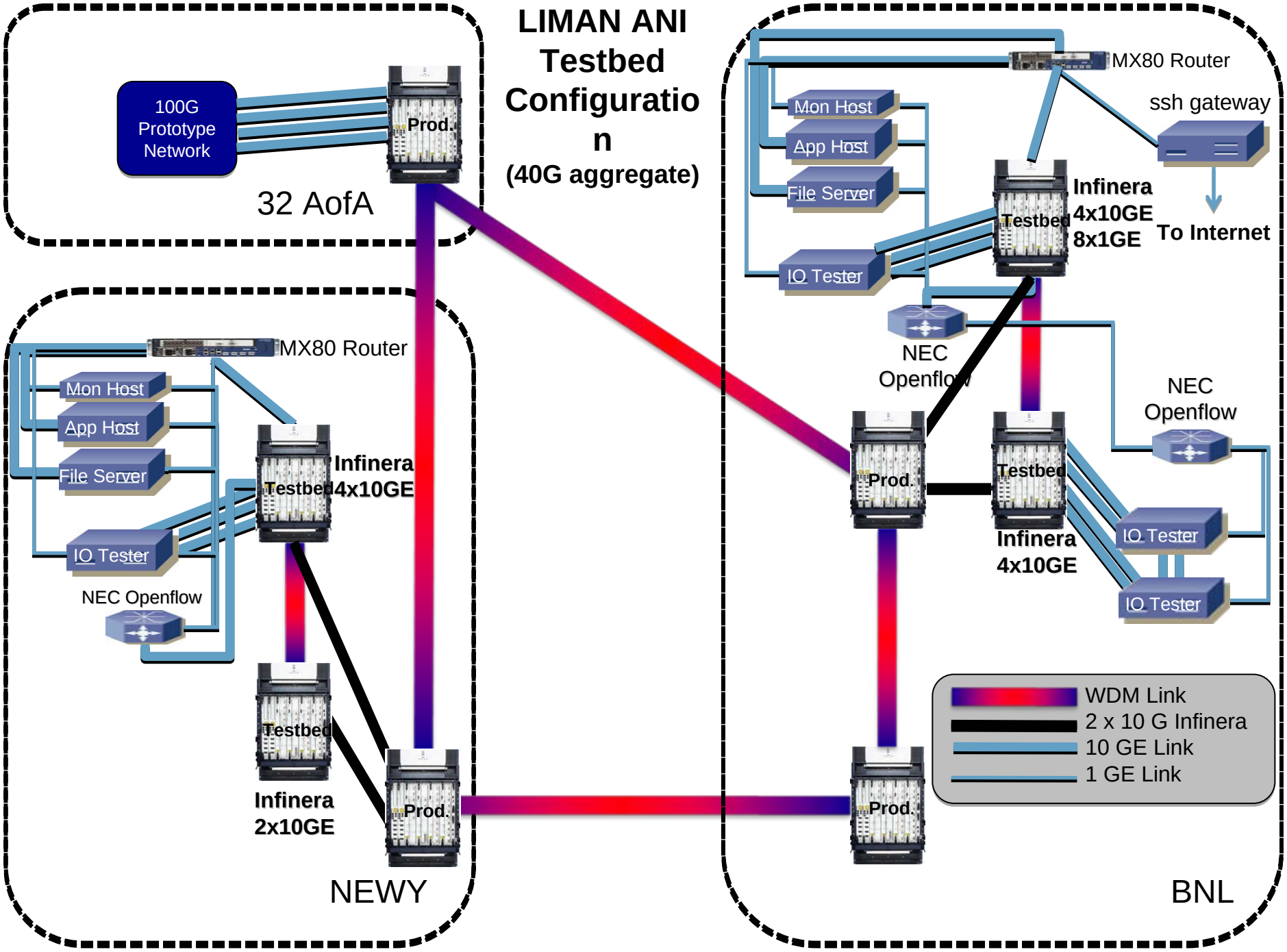
Hired 2 FTEs dedicated to the testbed and user support

10-12 ESnet staff members have helped build the testbed

Great learning opportunity for staff!

- Learned about many new technologies, several of which will likely become part of ESnet's production services
  - Infinera optical platform
  - Openflow
  - VM technologies (XEN)
  - Cloud software (Eucalyptus)
  - IPMI for remote host management
  - VPN server software

Phase 2: Move to Long Island MAN (Nov)


LIMAN ANI Testbed Configuration (40G aggregate)
100G Prototype Network
Prod.
32 AofA
MX80 Router
Mon Host
App Host
File Server
IO Tester
NEC Openflow
Testbed
Infinera 4x10GE
Infinera 2x10GE
NEWY
MX80 Router
ssh gateway
Mon Host
App Host
File Server
Testbed
Infinera 4x10GE 8x1GE
To Internet
IO Tester
NEC Openflow
Prod.
Testbed
Infinera 4x10GE
NEC Openflow
IO Tester
IO Tester
Prod.
BNL
WDM Link
2 x 10 G Infinera
10 GE Link
1 GE Link

OpenDevNet: R&D Proving-ground

ESnet has built a VM-based environment for testing distributed software

- Accessible to any ESnet collaborator (separate security enclave)
- Used by perfSONAR and OSCARS project for testing
- Used by ANI testbed project for testing software
- Also used by ESnet staff for a variety of testing

Available to the community for training / demos

- 3 application hosts, each capable of running 4-8 VMs
- 2 IO tester hosts, capable of 40 Gbps memory to memory, 15 Gbps disk to disk
- Openflow switch (coming soon)

Thank you

Email: steve@es.net

Follow us:

http://esnetupdates.wordpress.com
http://www.twitter.com/esnetupdates