ESnet Activities & Plans

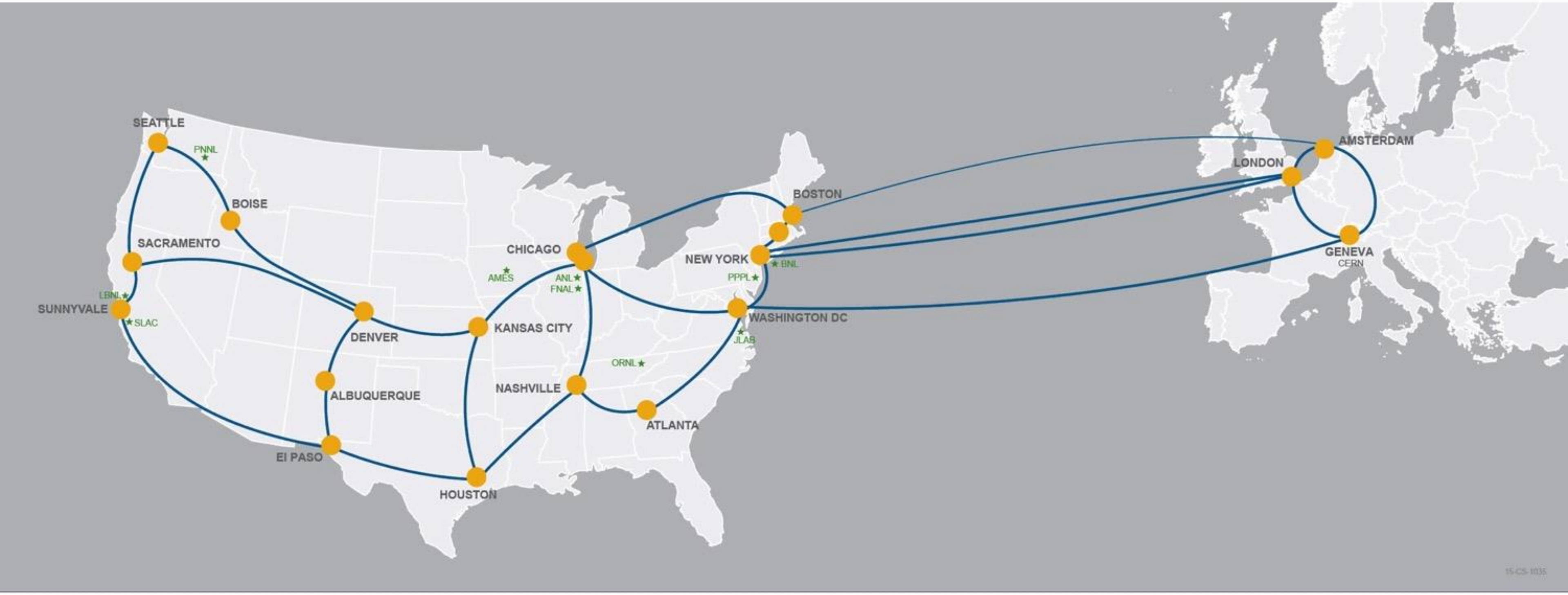
Michael O'Connor moc@es.net
Network Engineering
Energy Sciences Network
Lawrence Berkeley National Lab

LHCOPN-LHCONE Meeting
Taipei Taiwan
March 13, 2016

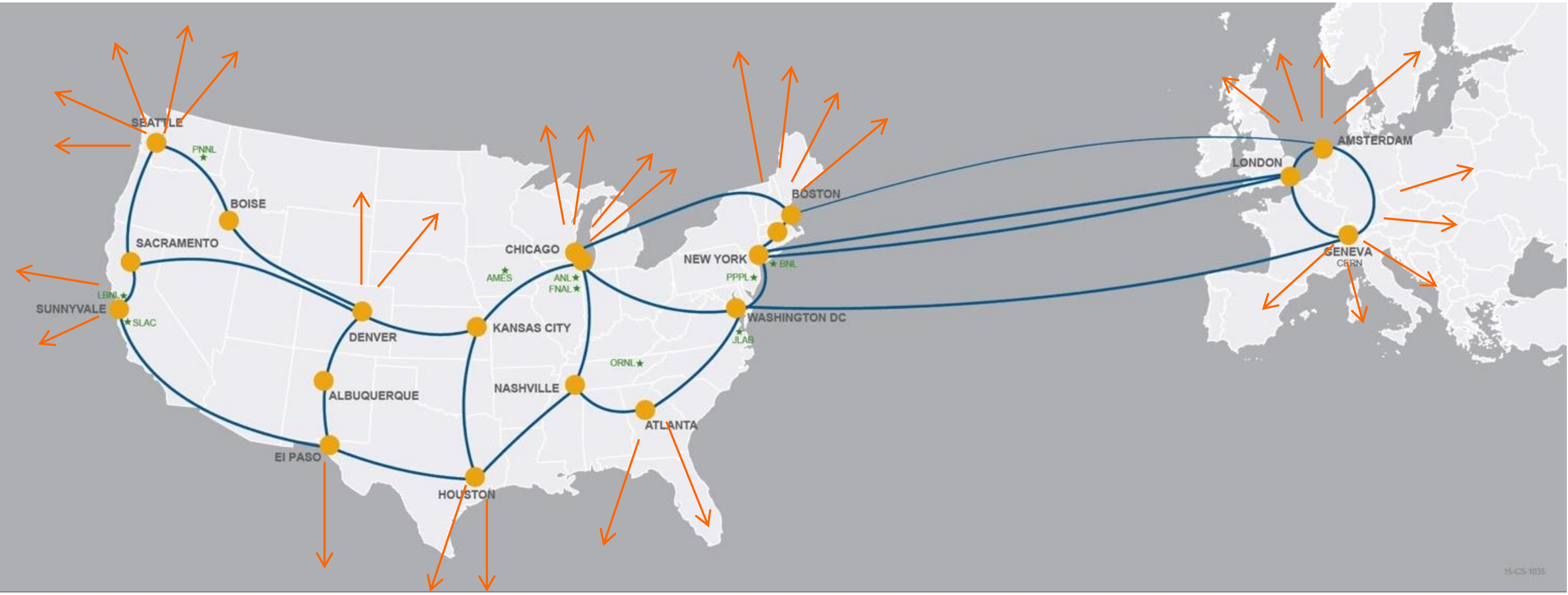
ESnet vision:

Scientific progress will be **completely unconstrained** by the physical location of instruments, people, computational resources, or data.

DOE’s Energy Sciences Network (ESnet):

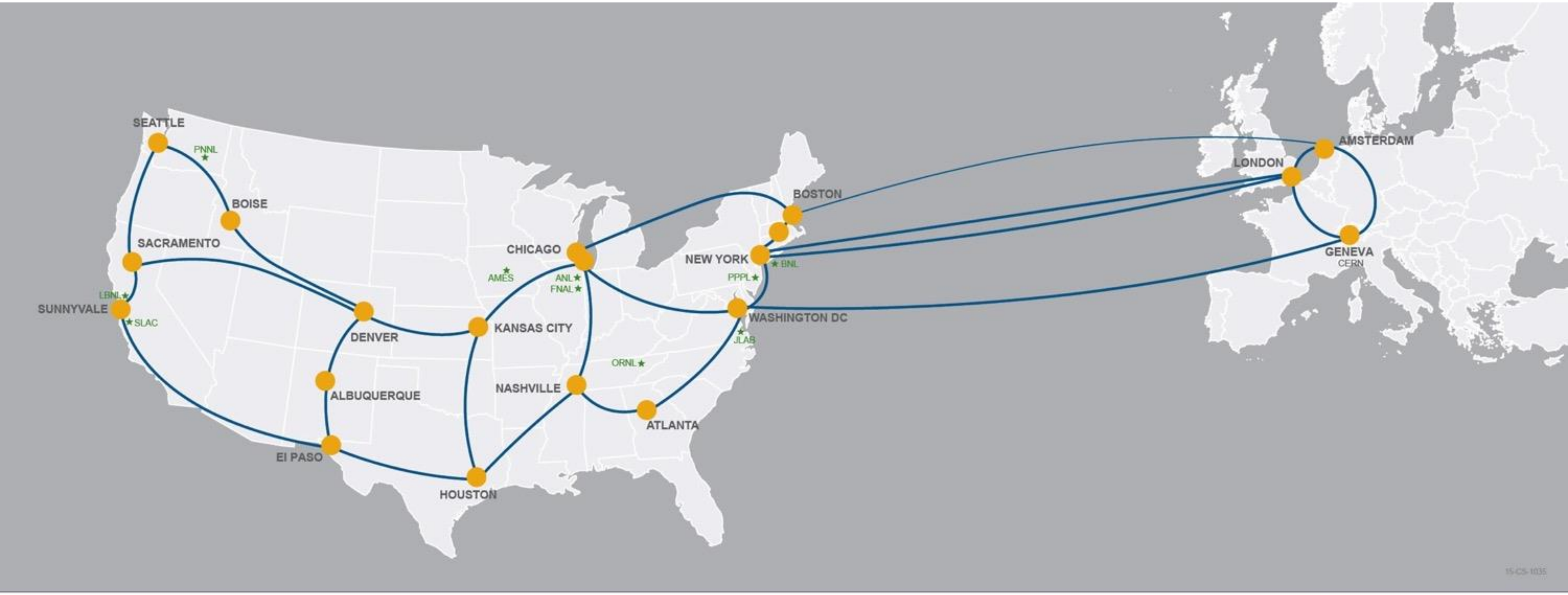


NRENs share fate. No research network can succeed in isolation.

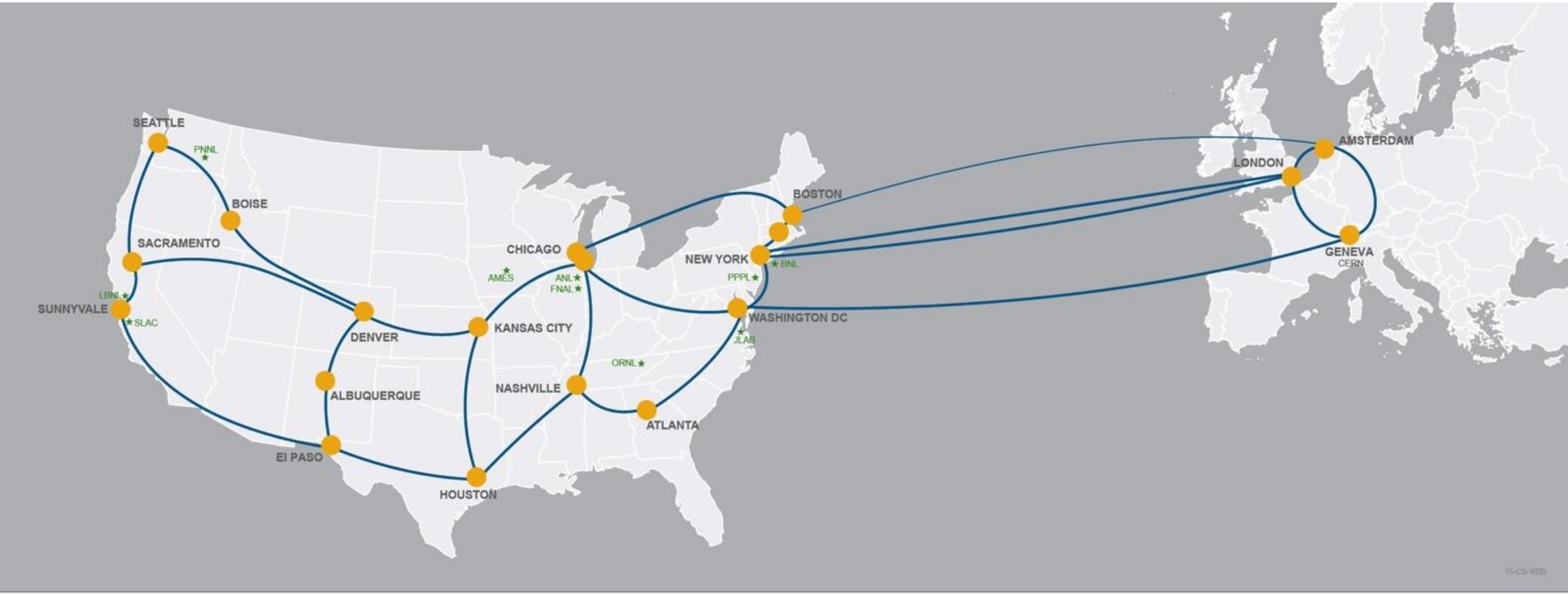


80% of ESnet traffic originates or terminates outside the DOE complex.

The most important thing to know about NRENs: they are not ISPs.

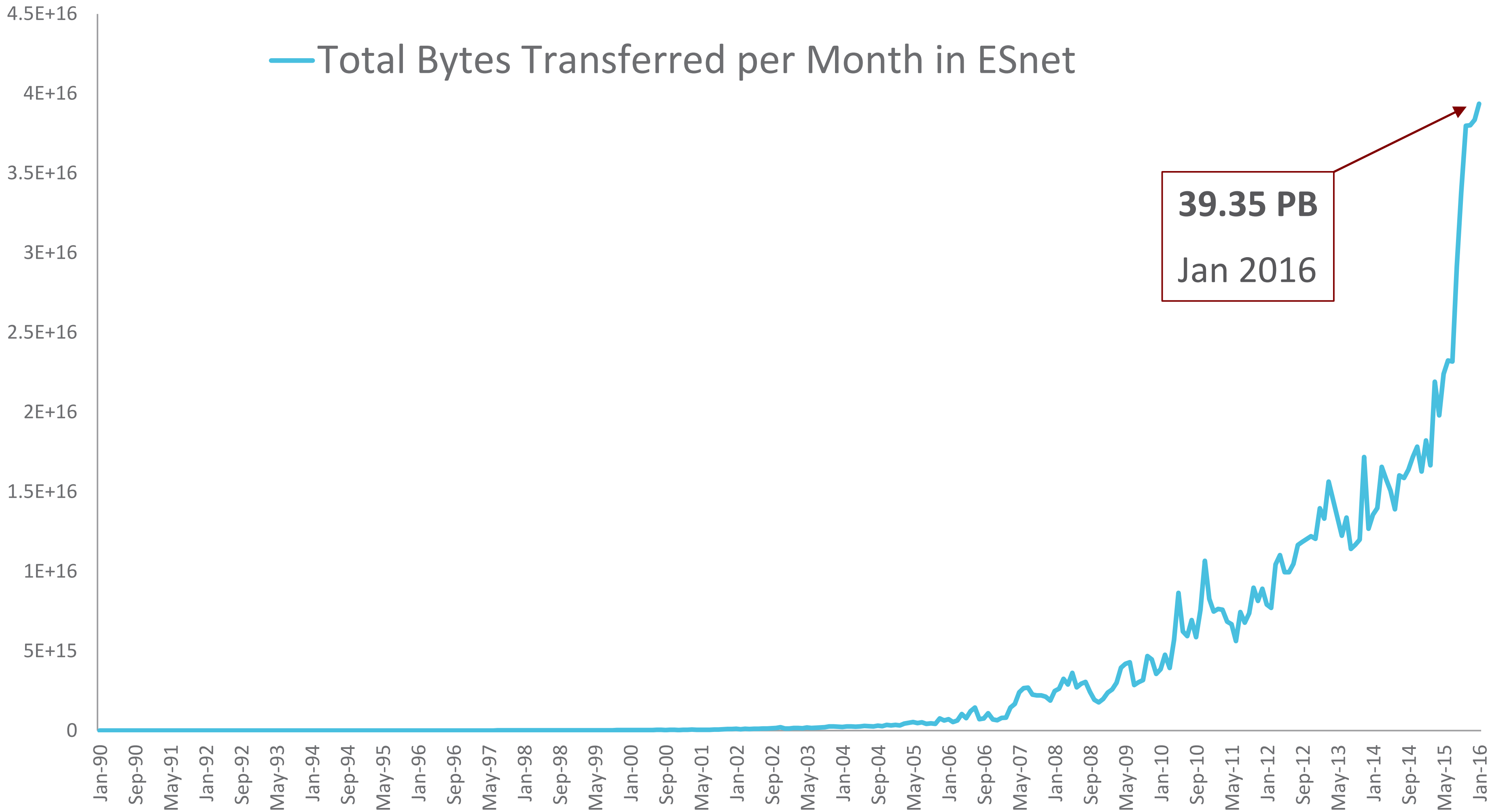


They are instruments for discovery designed to overcome the constraints of geography.



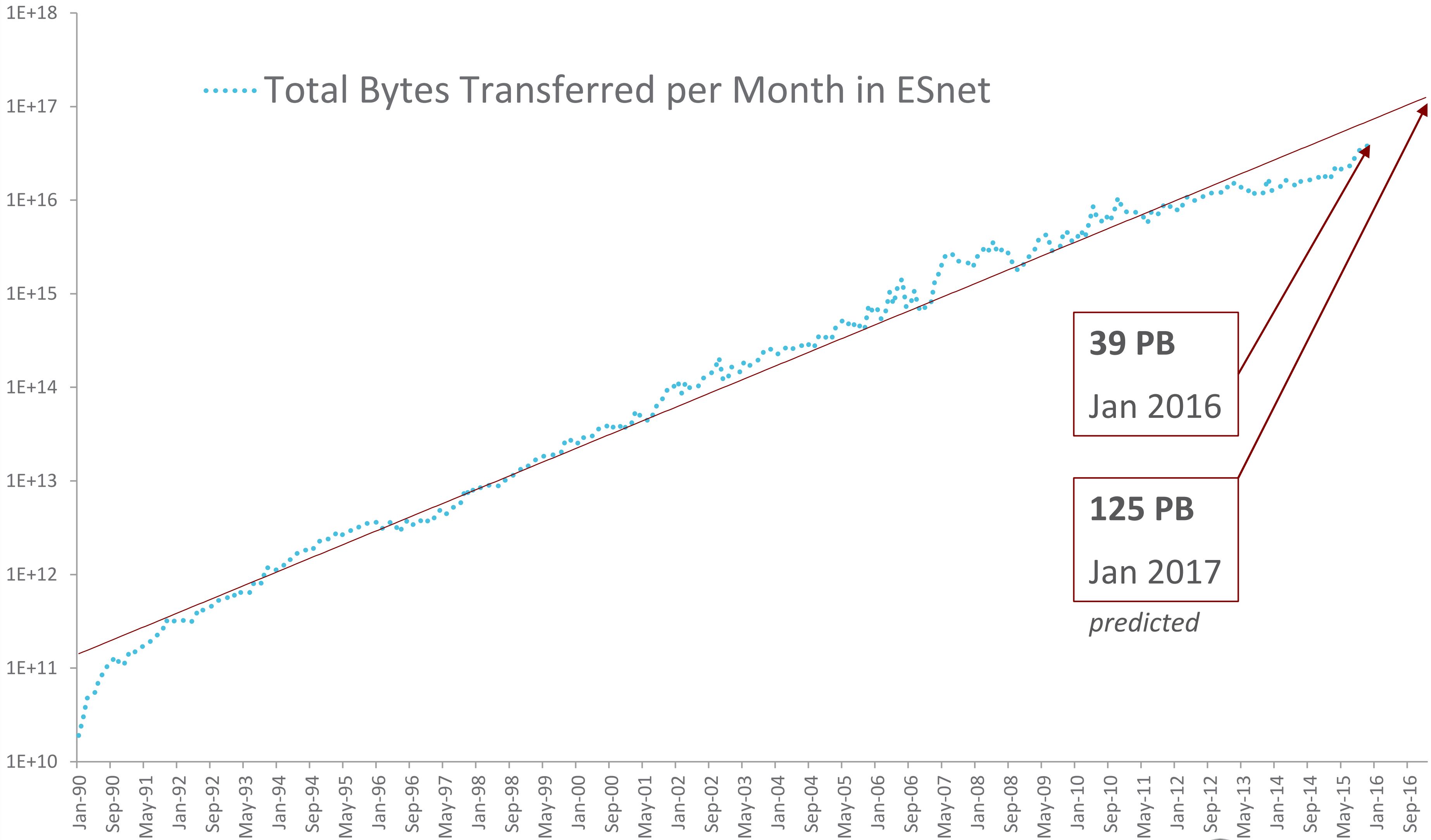
Offering unique capabilities – and optimized for data acquisition, data placement, data sharing, data mobility.

What does success look like?



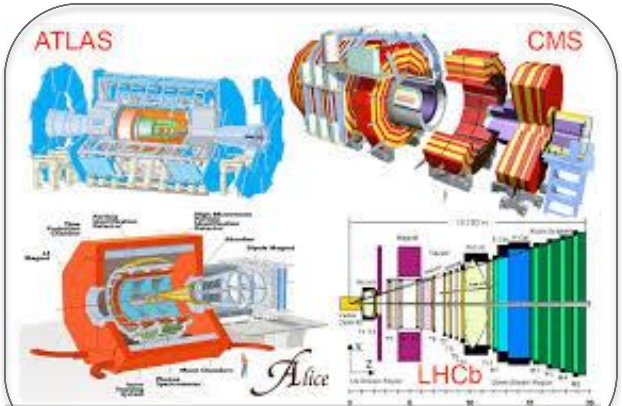
ESnet

Planning for growth

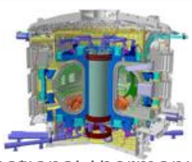


ESnet

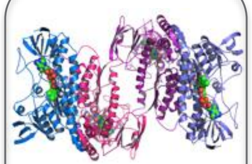
Design Pattern #1: Protect your *Elephant* Flows



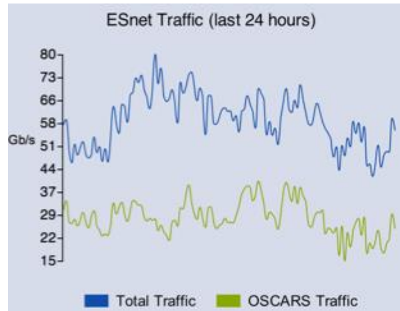
Large Hadron Collider (LHC)



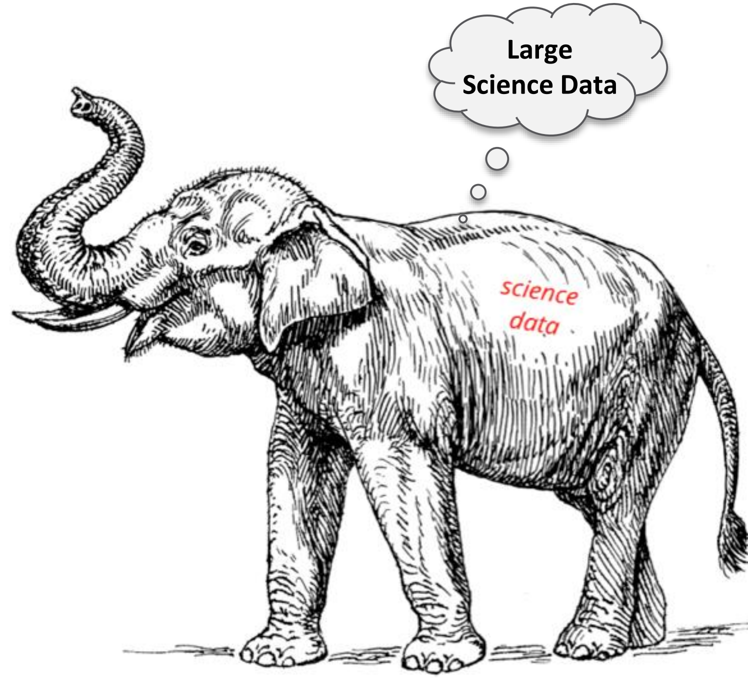
International Thermonuclear Experimental Reactor (ITER)

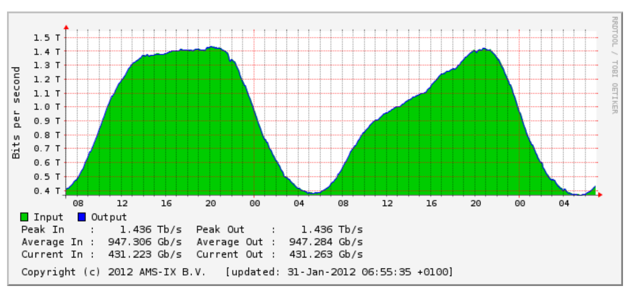


Genomics



ESnet

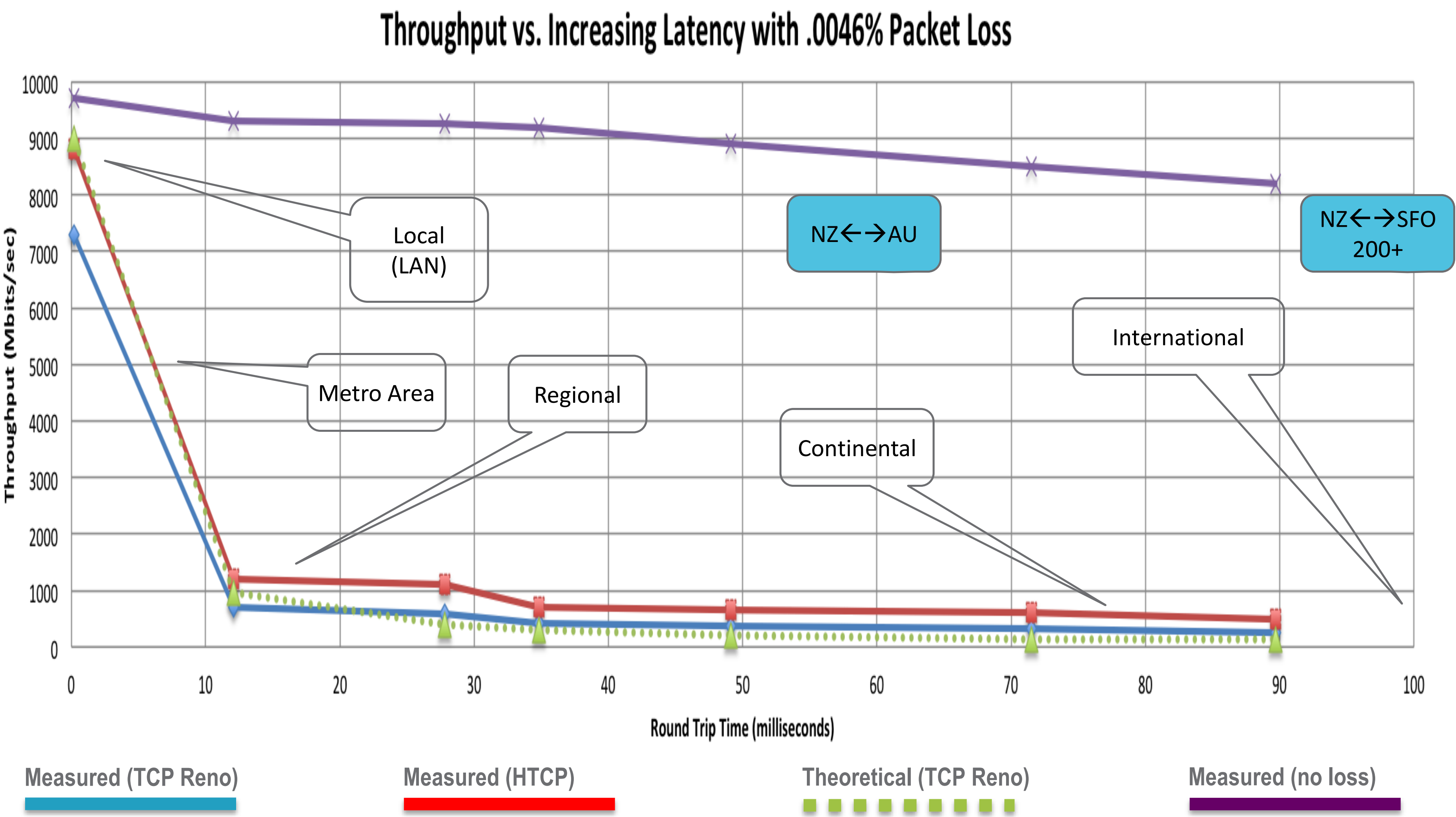




ESnet

General Internet

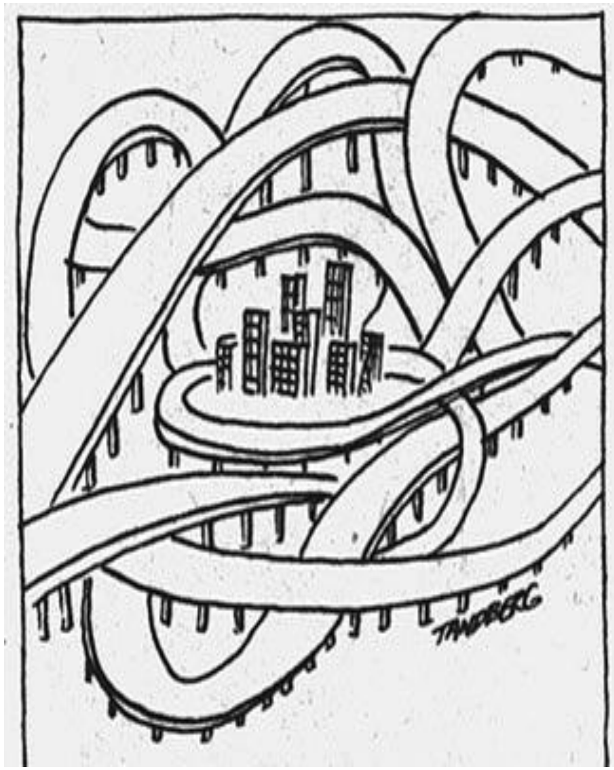
Elephant flows require almost *lossless* networks.



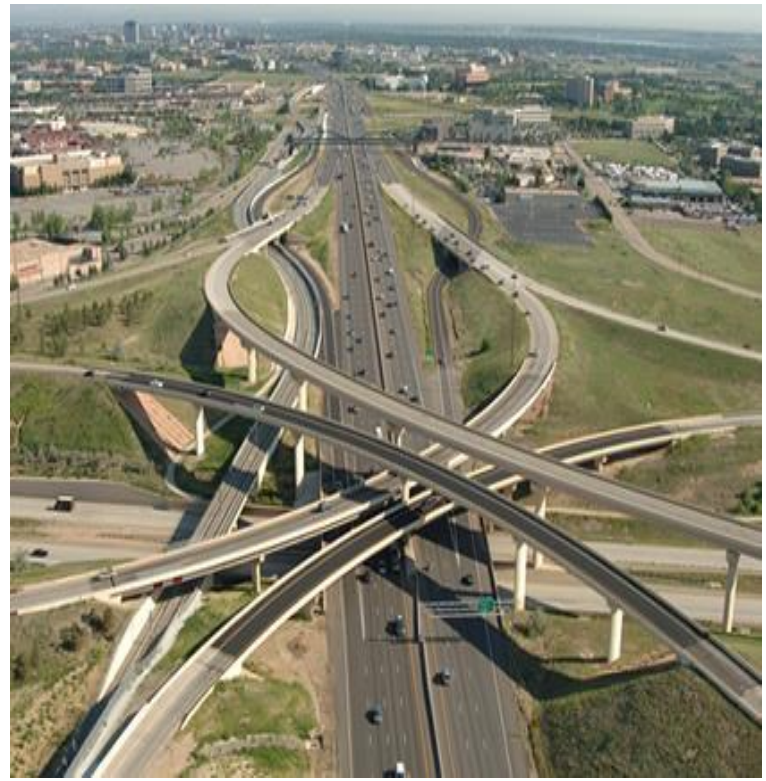
See Eli Dart, Lauren Rotman, Brian Tierney, Mary Hester, and Jason Zurawski. The Science DMZ: A Network Design Pattern for Data-Intensive Science. In *Proceedings of the IEEE/ACM Annual SuperComputing Conference (SC13)*, Denver CO, 2013.

Design Pattern #2: Unclog your data taps

Problem and Solution explained illustratively



Big-Data assets **not optimized** for **high-bandwidth access** because of **convoluted campus network and security design**



Science DMZ is a **deliberate, well-designed architecture** to simplify and **effectively on-ramp 'data-intensive' science** to a capable WAN

Data Set Mobility Timeframes - Theoretical

Data set size		1 Minute	5 Minutes	20 Minutes	1 Hour
10PB		1,333.33 Tbps	266.67 Tbps	66.67 Tbps	22.22 Tbps
1PB		133.33 Tbps	26.67 Tbps	6.67 Tbps	2.22 Tbps
100TB		13.33 Tbps	2.67 Tbps	666.67 Gbps	222.22 Gbps
10TB	> 100Gbps	1.33 Tbps	266.67 Gbps	66.67 Gbps	22.22 Gbps
1TB		133.33 Gbps	26.67 Gbps	6.67 Gbps	2.22 Gbps
100GB	100Gbps	13.33 Gbps	2.67 Gbps	666.67 Mbps	222.22 Mbps
10GB	< 10Gbps	1.33 Gbps	266.67 Mbps	66.67 Mbps	22.22 Mbps
1GB		133.33 Mbps	26.67 Mbps	6.67 Mbps	2.22 Mbps
100MB	< 100Mbps	13.33 Mbps	2.67 Mbps	0.67 Mbps	0.22 Mbps

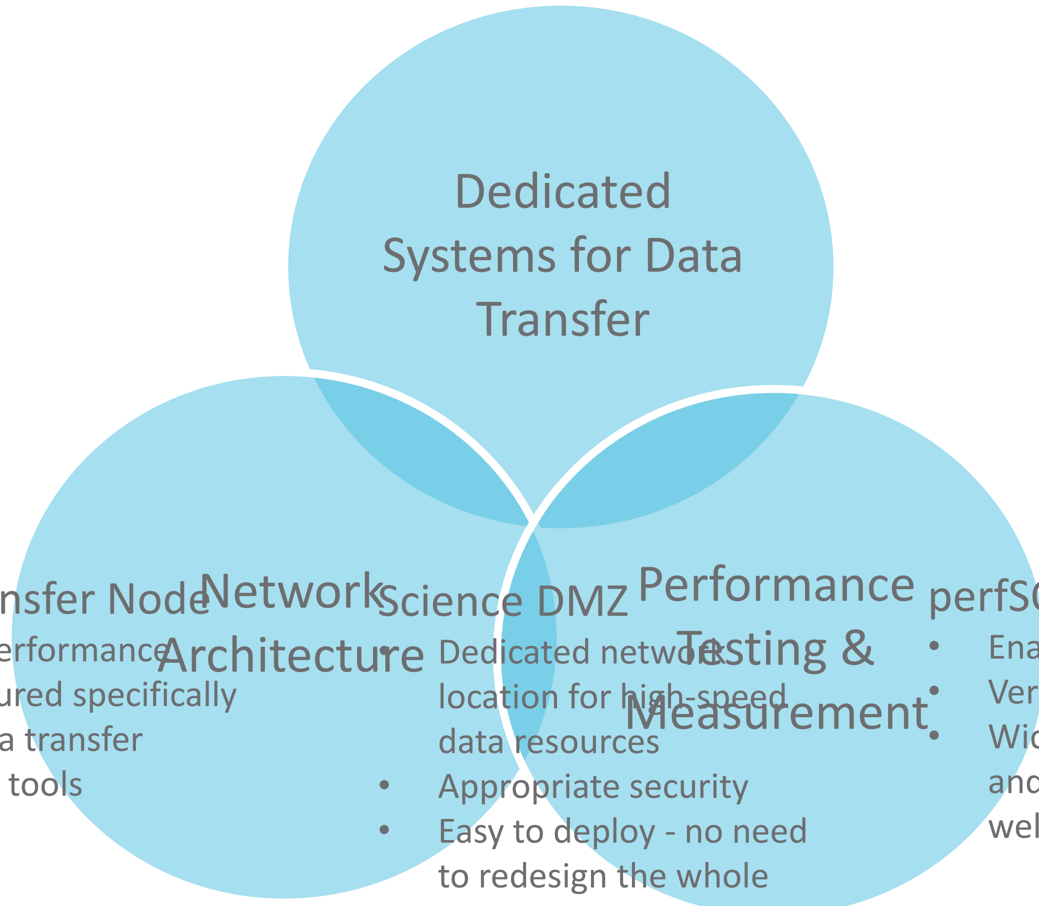
Time to transfer

This table available at:

http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/

ESnet

The Science DMZ Design Pattern



Data Transfer Node

- High performance
- Configured specifically for data transfer
- Proper tools

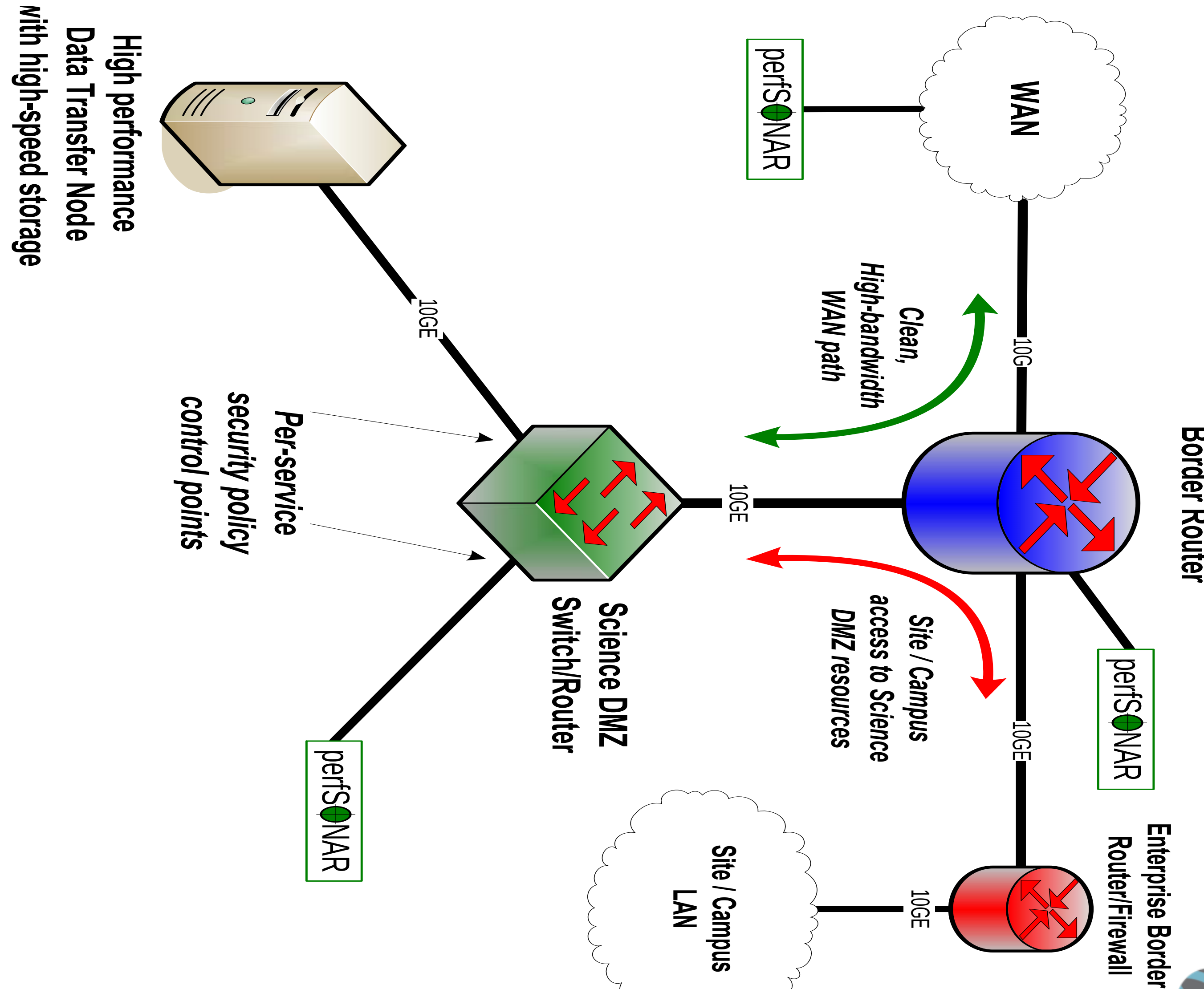
Science DMZ

- Dedicated network location for high-speed data resources
- Appropriate security
- Easy to deploy - no need to redesign the whole network

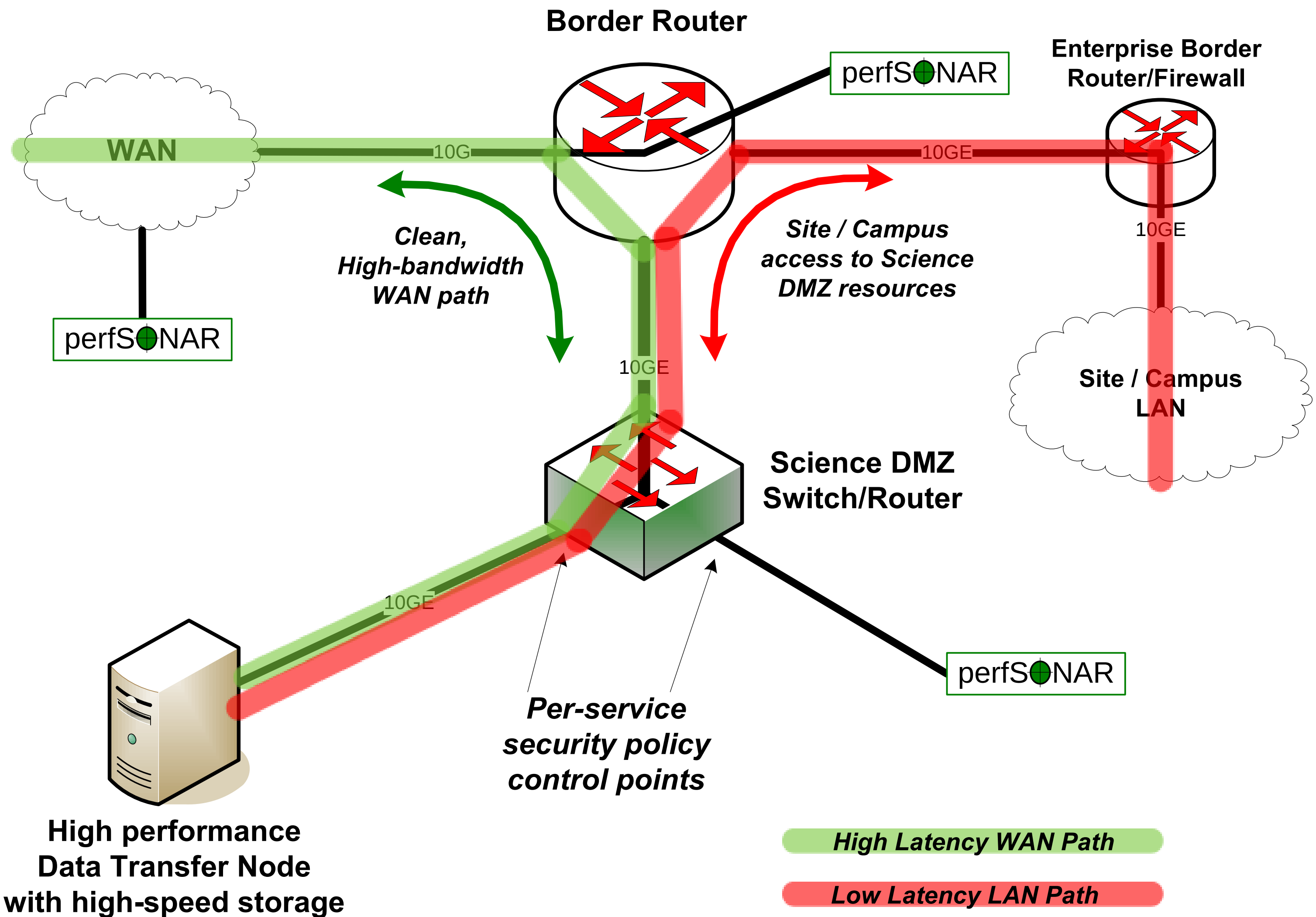
perfSONAR

- Enables fault isolation
- Verify correct operation
- Widely deployed in ESnet and other networks, as well as sites and facilities

Science DMZ Design Pattern (Abstract)

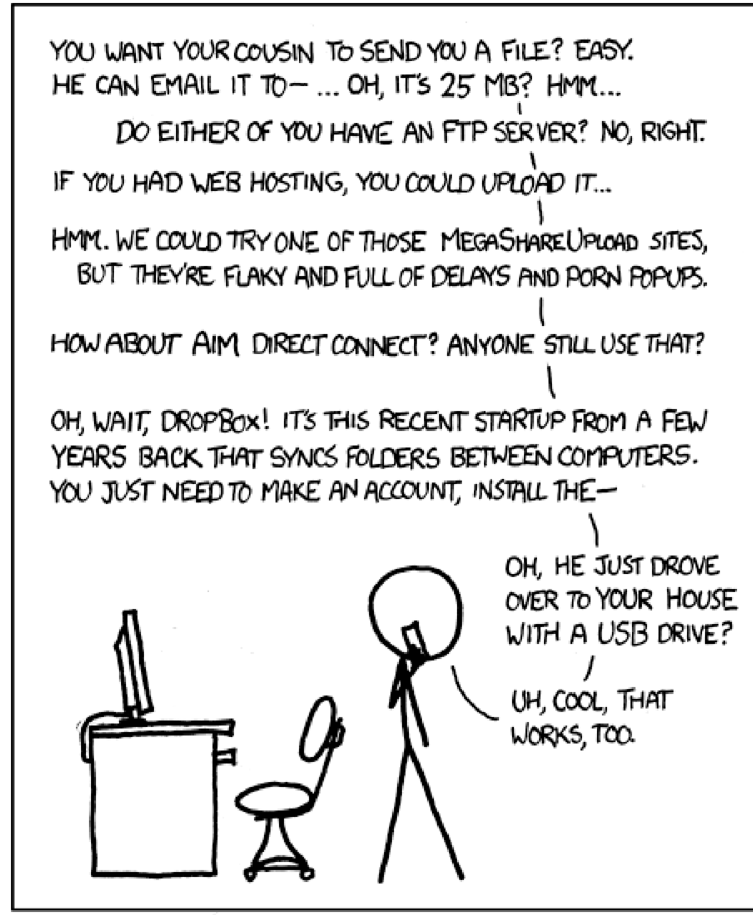


Local And Wide Area Data Flows



Dedicated Systems – Data Transfer Node

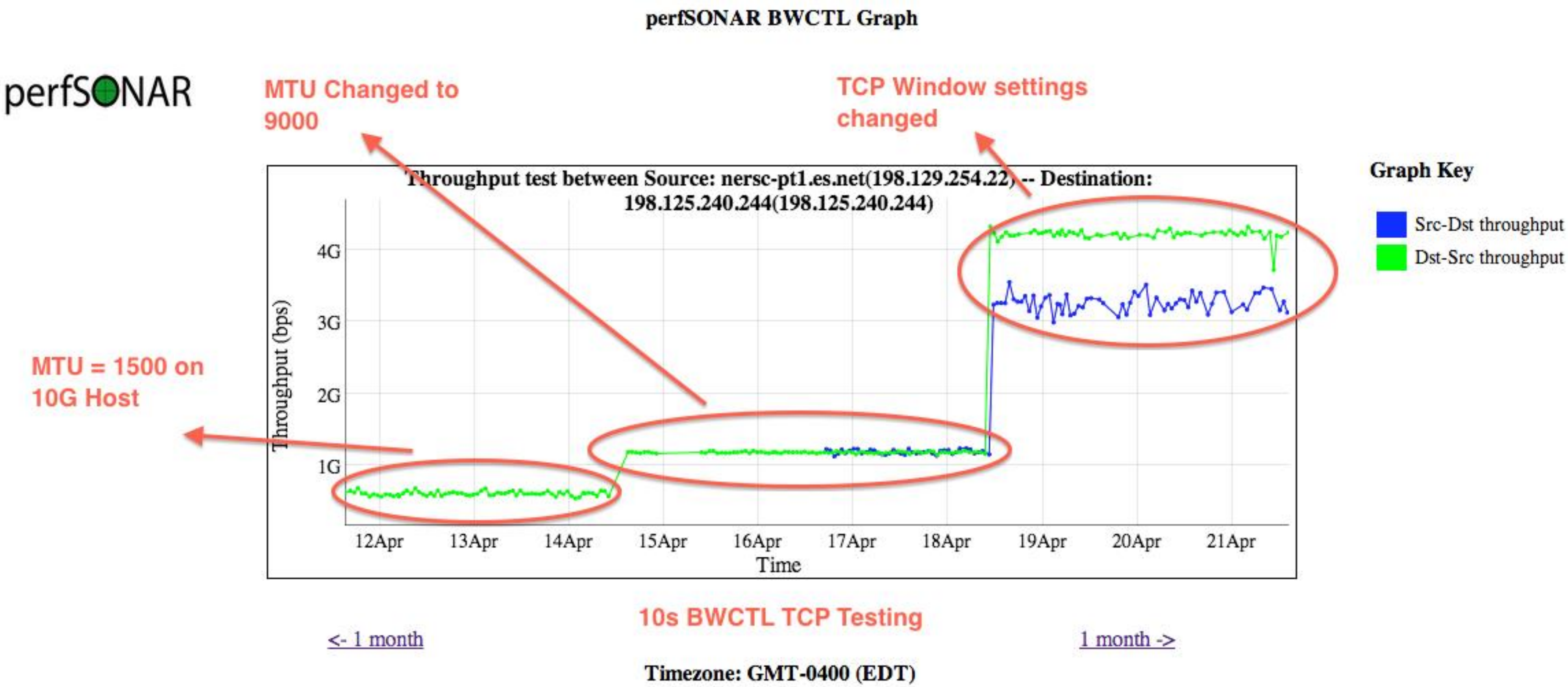
- Set up ***specifically*** for high-performance data movement
  - System internals (BIOS, firmware, interrupts, etc.)
  - Network stack
  - Storage (global filesystem, Fibrechannel, local RAID, etc.)
  - High performance tools
  - No extraneous software
- **Limitation of scope and function is powerful**
  - No conflicts with configuration for other tasks
  - Small application set makes cybersecurity easier



I LIKE HOW WE'VE HAD THE INTERNET FOR DECADES, YET "SENDING FILES" IS SOMETHING EARLY ADOPTERS ARE STILL FIGURING OUT HOW TO DO.

ESnet

Example of perfSONAR monitoring



Improving things, when you don't know what you are doing, is a random walk. Sharing and educating the local community is important

Emerging global consensus around Science DMZ architecture.

>120 universities in the US have deployed this ESnet architecture.

NSF has invested >>$60M to accelerate adoption.

Australian, Canadian universities following suit.

http://fasterdata.es.net/science-dmz/

ESnet

ESnet 6 Planning Process

2018 – 100G ESnet 5 architecture will exceed 6 years in production

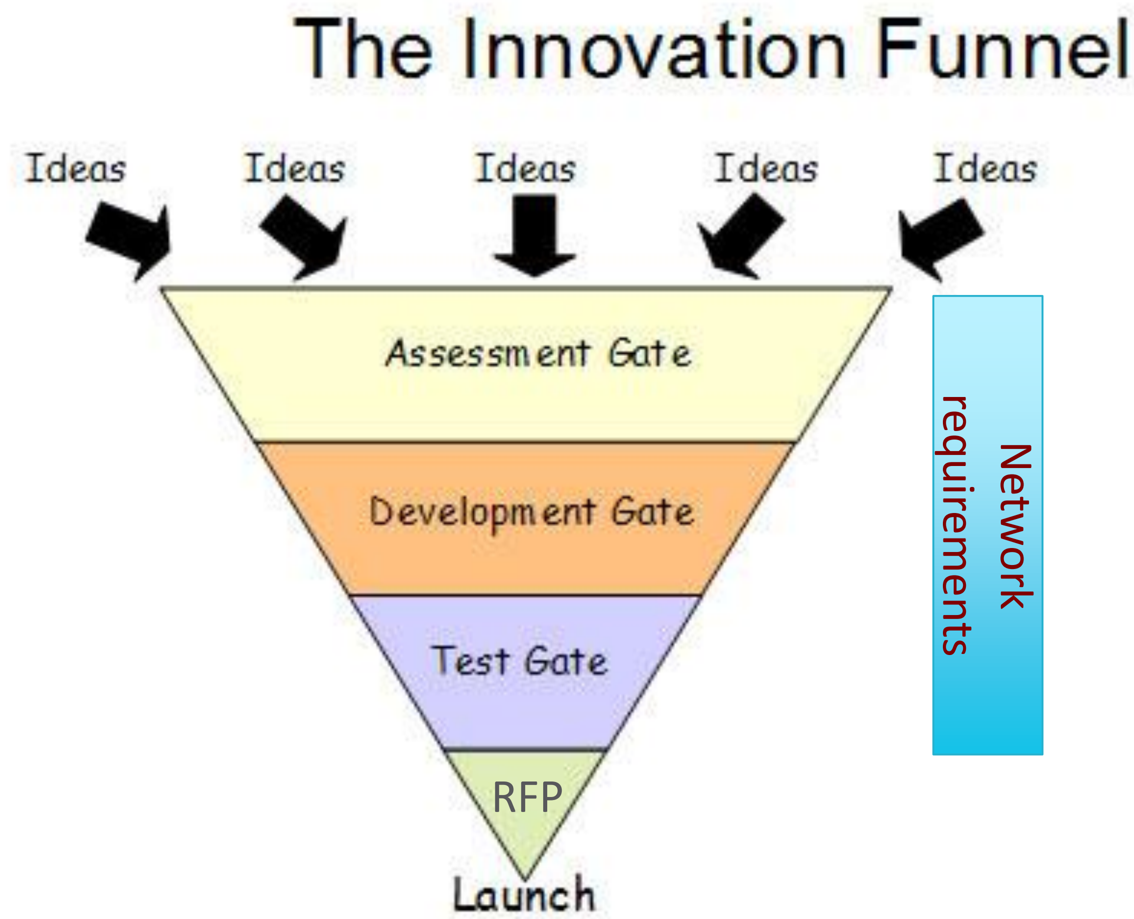
ESnet has begun the planning process for ESnet 6

Architectural Choices

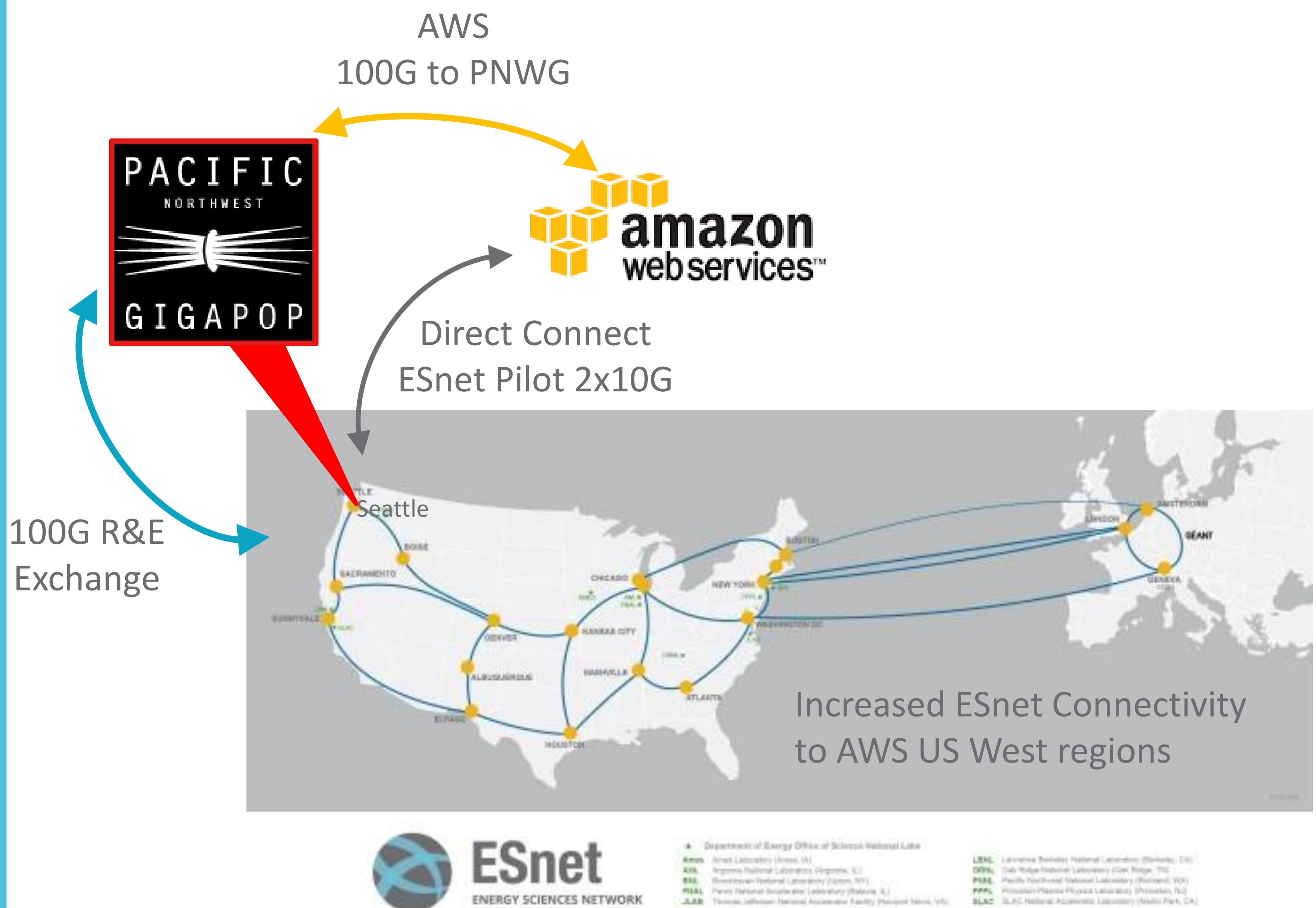
- SDN
  - Hybrid SDN/Traditional Routing
  - Full Control and Data Plane separation
- Packet Optical Convergence
  - Packet services in Optical Transport Networks
  - OTN Transceivers in classical routers
- Business as Usual
  - Disruptive price/performance advances?

ESnet 6 investigation should align with Next Gen service development

ESnet6 Architecture Investigation



2015 ESnet AWS Pilot



The AWS Cloud
Changing the rules of the Network Game

The Amazon Web Service cloud bends a number or longstanding Internet Architecture rules.

- AWS ASN 16509 is not contiguous as Autonomous Systems are intended to be. This "Cloud" is a set of regional Data Centers.
- BGP with AWS establishes connectivity with only the geographically close region(s), **NOT** the entire AS. Establishing BGP with AWS, ASN 16509 in one region doesn't establish connectivity to other regions, continents etc.
- Since AWS does not advertise remote region routes, routing to those remote regions will NOT use your established BGP with them, choosing the general public Internet instead.
- Similarly, the reverse path will ignore AS Path rules and prefer commodity public Internet over the direct peering with AWS.

This is a challenge for globally collaborative computing, but it's not necessarily a bad thing

ESnet

Amazon is not in the "Networking" business

- Amazon Web Services (AWS) offers an extensive portfolio of computing resources, but they are not in the "networking" business.
- AWS recharges customers for egress traffic out of the cloud. Some researchers have described these fees as "holding their data hostage".
- AWS must pay for upstream transit to the Internet and so they pass this on to their customers.
- Their inter-region long haul circuits are used for internal control and management rather than customer transport.
- AWS offers a service that will migrate data between regions, but this is strictly controlled and scheduled by AWS, not customers.
- Since AWS peers for free with R&E networks, they offer customers of these networks an "Egress Traffic Fee Waiver".

Research & Education Networks have an opportunity to scale "Cloud" beyond the local region for their customers.

ESnet

Routing Exercise

How do European NRENs connect to AWS

Prefixes from several AWS regions

- `us-east-1    50.19.0.0/16`
- `us-west-1    50.18.0.0/16`
- `us-west-2    50.112.0.0/16`
- `eu-central-1 52.28.0.0/16`
- `eu-west-1    52.18.0.0/15`

Try Looking up these BGP routes using a couple of these European NREN looking glass services

- **Check the AS path to the US regions**
- **Check the AS path to the European regions**

What is interesting about these paths?

European NREN Looking Glasses

- GEANT - https://tools.geant.net/portal/links/lg/
- DFN Germany – https://www.noc.dfn.de/lg/
- GARR Italy - https://gins.garr.it/LG/
- RENATER France - https://portail.noc.renater.fr/lookingglass/v2/lg_index.pl

The "On-Ramp" to R&E Networking
Virtual Private Cloud

The public cloud is being used privately, however the private cloud enhances data sharing over R&E networks.

- Currently, cloud service usage within the ESnet customer-base, employs a simple "Single Consumer Model", where each site purchases services that they consume directly.
- Their connection is a standard routed path through their upstream provider ESnet, no virtual circuits or additional routing protocol setup, simple.
- As global collaborations shift computing resources to the "Cloud", this simple single consumer model will give way to one that better supports data sharing, the "Collaborative Cloud".
- VPC enables a customer to control the routing of their data by providing a means of mapping customer IP allocations into the cloud.
- The principal assumption: When geographically dispersed collaborations begin to transit large datasets directly out of the cloud, VPC services will be essential in order to control routing between those collaborating institutions.

Global collaborations need to control transport of large data sets

ESnet

Proposal: Collaborative Cloud Services (CCS)

SDN site router - A virtual router that is provisioned to BGP peer with the cloud provider using the customer ASN.

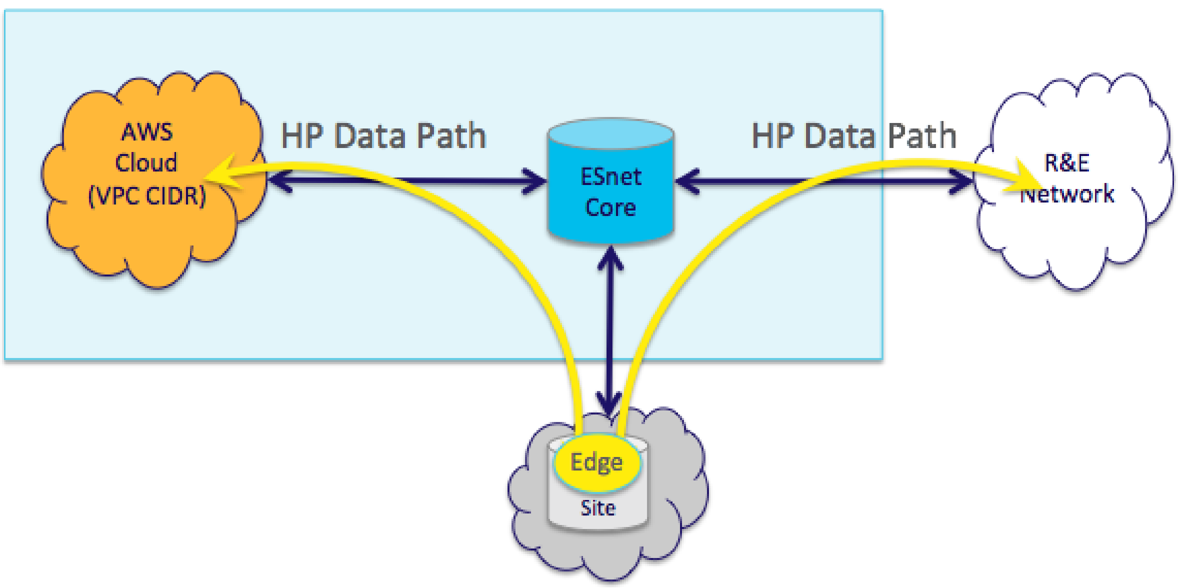
- Multiple virtual routers per hardware device.
- IPsec will connect the SDN Site router to the cloud in order to support VPC.
- Collaborator transit paths take the best path to the cloud and back.

The potential:

- Become an important application for SDN and a new ESnet service.
- Transform the ESnet site connection architecture by alleviating upgrade pressure on local loop circuits.

Proposed: Collaborative Cloud Services (CCS)

Standard VPC Third Party Data Path

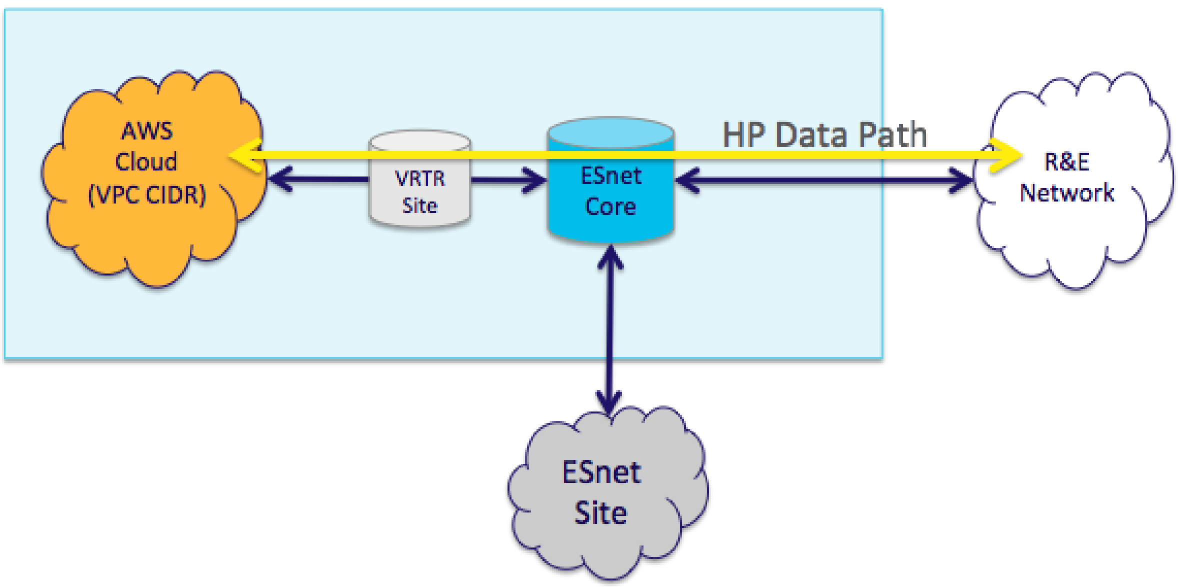


Standard VPC implementation controls routing as an **"On-Ramp"** for R&E Network Transport.

Virtual "Site Router" (VRTR) Service

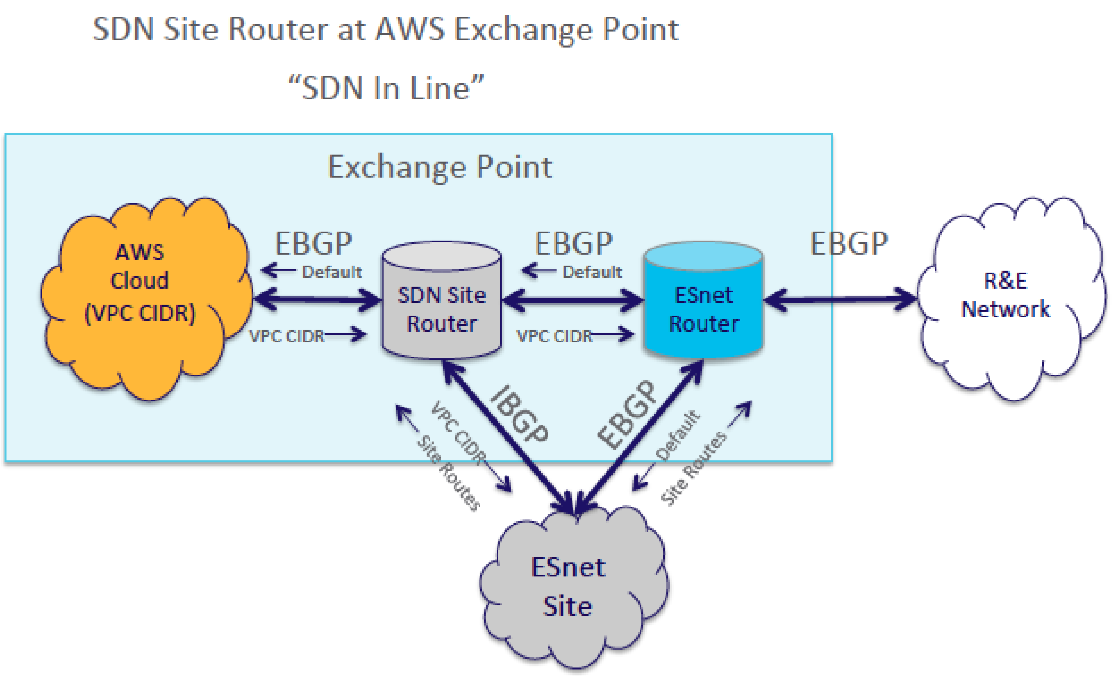
At the edge of the cloud

Virtual Site Router at AWS Exchange Point



Virtual "Site Router" improves path efficiency and takes pressure off of the site local-loop.

CCS Protocol Description



The CCS service would reside on ESnet HW, but serve multiple customer site router instance in software.

In conclusion:

- Global science networks are not ISPs – rather, extensions of science discovery instruments.
- Design patterns, architectures, workflows and challenges from science are now crossing over to domains.
- Packet Optical convergence, SDN and Cloud all present challenges and opportunities in network design.
- Cloud computing services offer a compelling business case for research institutions along with challenges and opportunities for NRENs.

ESnet

Thank You

Michael O’Connor moc@es.net
Network Engineering
Energy Sciences Network
Lawrence Berkeley National Lab

LHCOPN-LHCONE Meeting
Taipei Taiwan
March 13, 2016