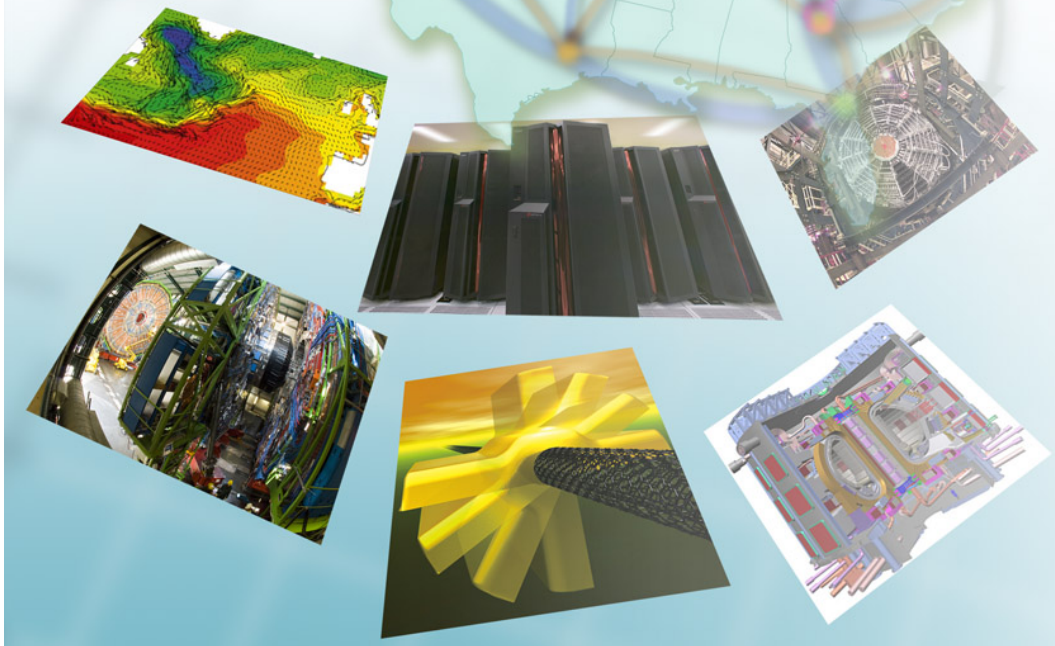
ESnet Status Update

ESCC
July 18, 2007

William E. Johnston
ESnet Department Head and Senior Scientist

Energy Sciences Network
Lawrence Berkeley National Laboratory

wej@es.net, www.es.net
This talk is available at www.es.net/ESnet4



Networking for the Future of Science

DOE Office of Science and ESnet – the ESnet Mission

- **ESnet's primary mission is to enable the large-scale science that is the mission of the Office of Science (SC) and that depends on:**
  - Sharing of massive amounts of data
  - Supporting thousands of collaborators world-wide
  - Distributed data processing
  - Distributed data management
  - Distributed simulation, visualization, and computational steering
  - Collaboration with the US and International Research and Education community
- ESnet provides network and collaboration services to Office of Science laboratories and many other DOE programs in order to accomplish its mission

Talk Outline

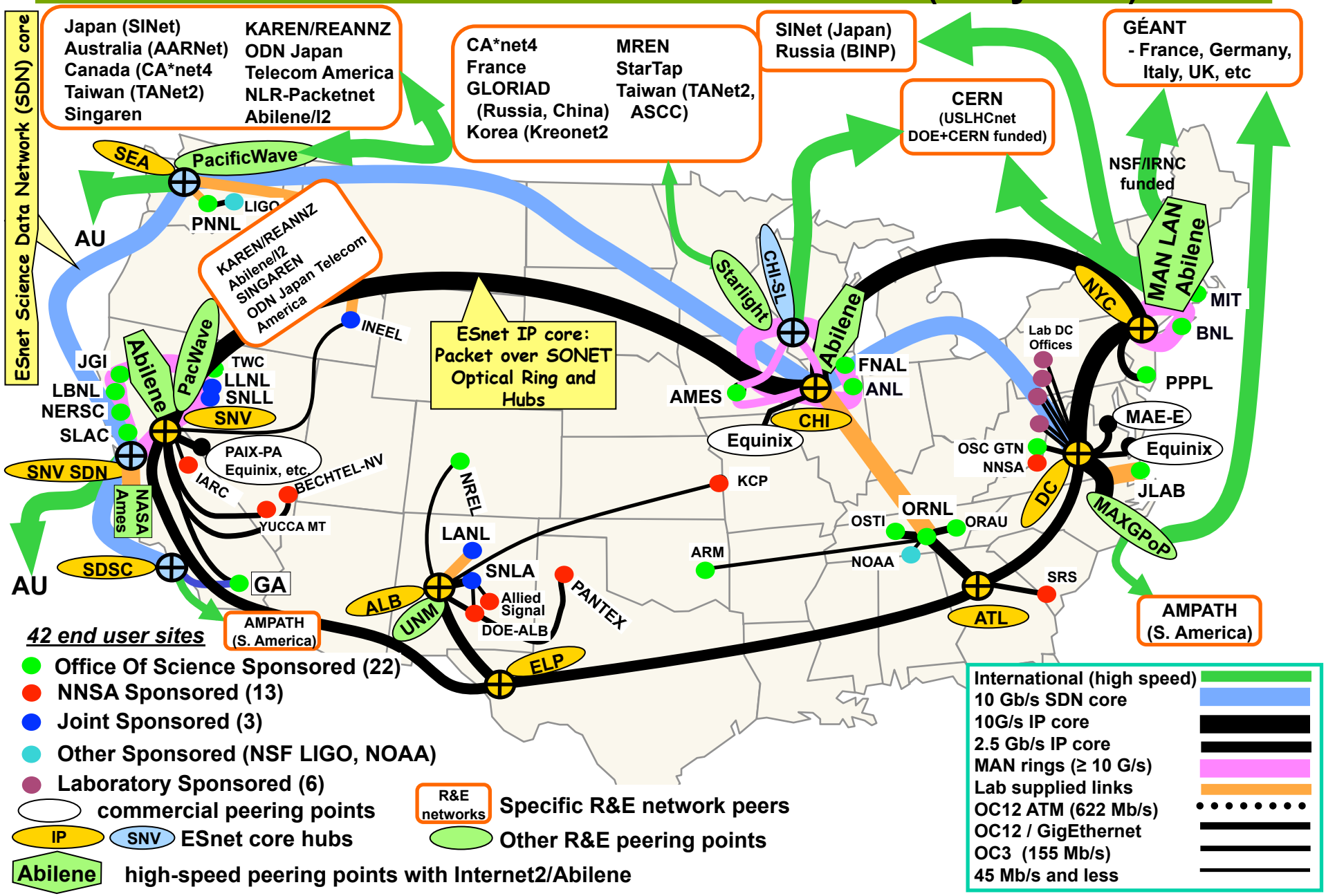
I. Current Network Status

II. Planning and Building the Future Network - ESnet4

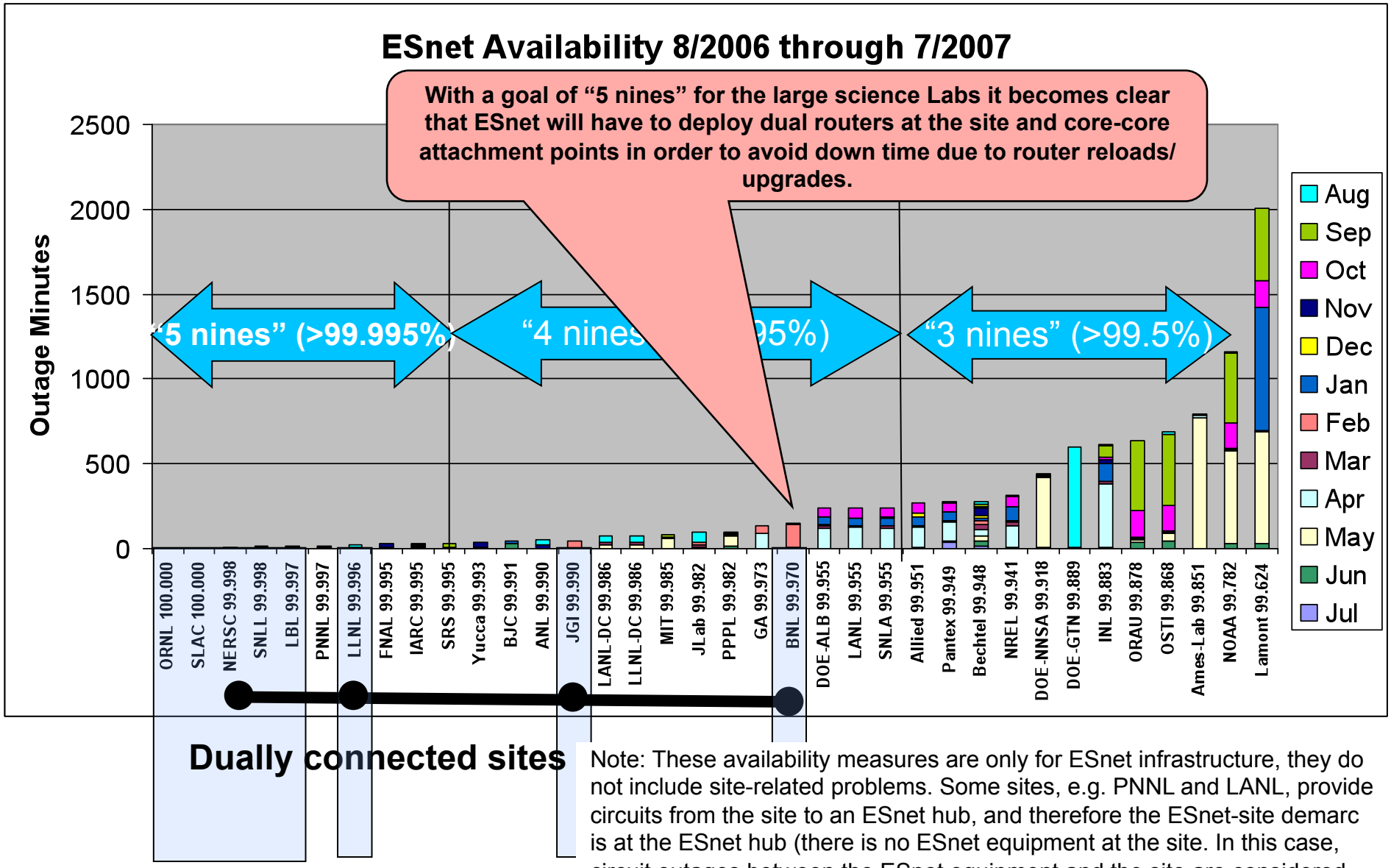
III. Science Collaboration Services - 1. Federated Trust

IV. Science Collaboration Services - 2. Audio, Video, Data Teleconferencing

I. ESnet3 Today Provides Global High-Speed Internet Connectivity for DOE Facilities and Collaborators (Early 2007)



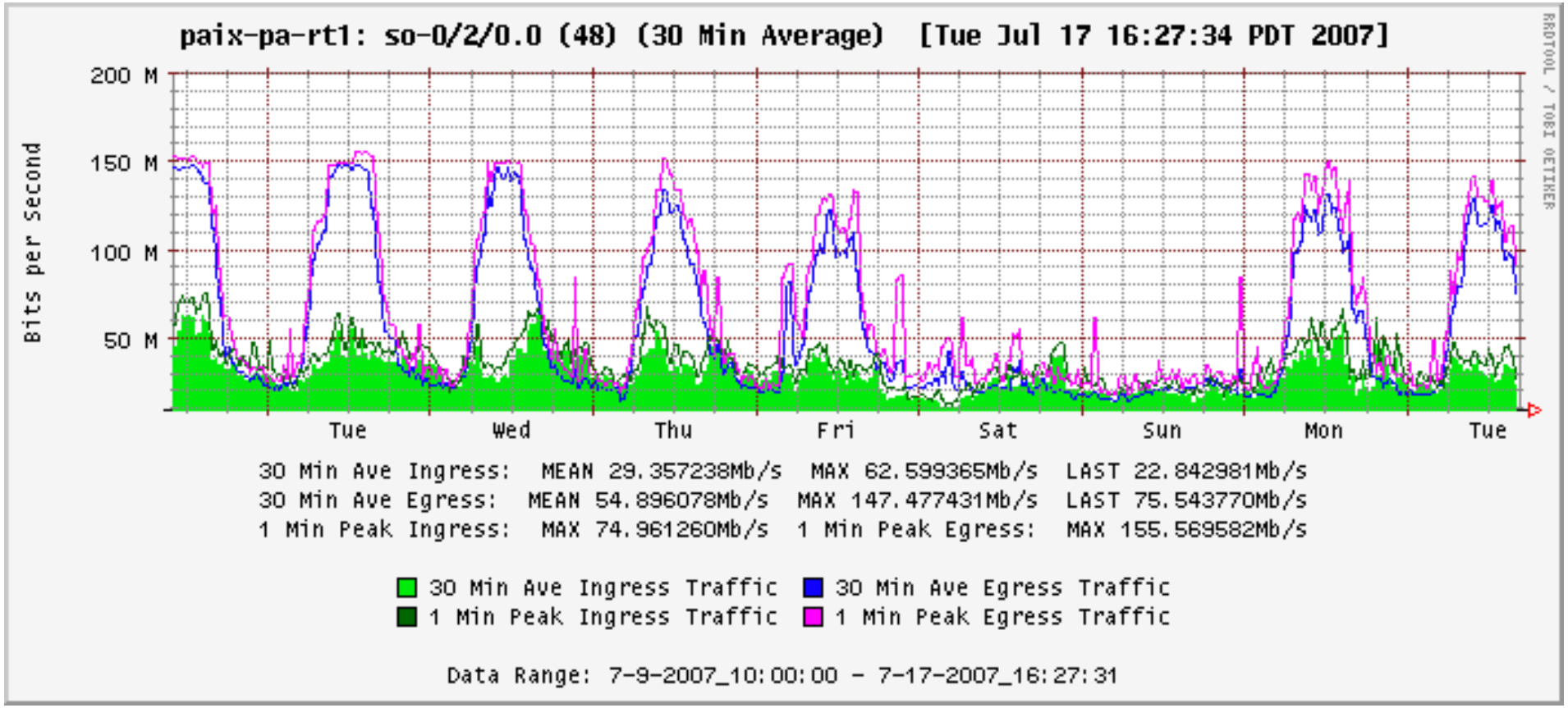
ESnet Availability



Note: These availability measures are only for ESnet infrastructure, they do not include site-related problems. Some sites, e.g. PNNL and LANL, provide circuits from the site to an ESnet hub, and therefore the ESnet-site demarc is at the ESnet hub (there is no ESnet equipment at the site. In this case, circuit outages between the ESnet equipment and the site are considered site issues and are not included in the ESnet availability metric.

Peering Issues

- ESnet has experienced congestion at both the West Coast and mid-West Equinix commercial peering exchanges



Commercial Peers Congestion Issues: Temporary changes Long-term Fixes

- The OC3 connection between paix-pa-rt1 and snv-rt1 was very congested, with peaks clipped for most of the day.
  - Temporary mitigation
    - Temporarily forcing West coast Level3 traffic to eqx-chicago - Traffic is now only clipped (if at all) at the peak of the day
  - Long term solution
    - Establish new Level3 peering at eqx-chicago (7/11/07)
    - Working on establishing a second peering with Global Crossing
    - Upgrade current loop (OC3) and fabric (100Mpbs) to 1Gbps
- Congestion to AT&T
  - Long term solution
    - Upgraded AT&T peering at eqx-sanjose from OC3 to OC12 (3/15/07)
  - Established OC12 peering with AT&T at eqx-ashburn (1/29/07) and eqx-chicago (07/11/07)
- The Equinix shared fabric at eqx-ashburn is congested
  - Long term solution
    - New Level3 peering at eqx-chicago has helped to relieve congestion
  - Additional mitigation
    - Third peering with Google at eqx-chicago, third peering with Yahoo at eqx-chicago
  - Future mitigation
    - Establish a second peering with Global Crossing at eqx-chicago
    - Upgrade equinix-sanjose and equinix-ashburn fabrics connections from 100Mb/s to 1Gbps

II. Planning and Building the Future Network - ESnet4

- Requirements are primary drivers for ESnet – science focused
- Sources of Requirements
  1. Office of Science (SC) Program Managers
     - The Program Offices Requirements Workshops
       - BES completed
       - BER in July, 2007
       - Others to follow at the rate of 3 a year
  2. Direct gathering through interaction with science users of the network
     - Example case studies (updated 2005/2006)
       - Magnetic Fusion
       - Large Hadron Collider (LHC)
       - Climate Modeling
       - Spallation Neutron Source
  3. Observation of the network
- Requirements aggregation
  - Convergence on a complete set of network requirements

1. Basic Energy Sciences (BES) Network Requirements Workshop

- Input from BES facilities, science programs and sites
  - Light Sources
  - SNS at ORNL, Neutron Science program
  - Nanoscience Centers
  - Combustion Research
  - Computational Chemistry
  - Other existing facilities (e.g. National Center for Electron Microscopy at LBL)
  - Facilities currently undergoing construction (e.g. LCLS at SLAC)

Workshop Process

- Three inputs
  - Discussion of Program Office – goals, future projects, and science portfolio
  - Discussions with representatives of individual programs and facilities
  - Group discussion about common issues, future technologies (e.g. detector upgrades), etc.
- Additional discussion – ESnet4
  - Architecture
  - Deployment schedule
  - Future services

BES Workshop Findings (1)

- BES facilities are unlikely to provide the magnitude of load that we expect from the LHC
  - However, significant detector upgrades are coming in the next 3 years
  - LCLS may provide significant load
  - SNS data repositories may provide significant load
  - Theory and simulation efforts may provide significant load
- Broad user base
  - Makes it difficult to model facilities as anything other than point sources of traffic load
  - Requires wide connectivity
- Most facilities and disciplines expect significant increases in PKI service usage

BES Workshop Findings (2)

- Significant difficulty and frustration with moving data sets
  - Problems deal with moving data sets that are small by HEP's standards
  - Currently many users ship hard disks or stacks of DVDs
- Solutions
  - HEP model of assigning a group of skilled computer people to address the data transfer problem does not map well onto BES for several reasons
    - BES is more heterogeneous in science and in funding
    - User base for BES facilities is very heterogeneous and this results in a large number of sites that must be involved in data transfers
    - It appears that this is likely to be true of the other Program Offices

(1A)

- ESnet action item – build a central web page for disseminating information about data transfer tools and techniques
- Users also expressed interest in a blueprint for a site-local BWCTL/ PerfSONAR service

2. Case Studies For Requirements

- Advanced Scientific Computing Research (ASCR)
  - NERSC
  - NLCF
- Basic Energy Sciences
  - Advanced Light Source
    - Macromolecular Crystallography
  - Chemistry/Combustion
  - Spallation Neutron Source
- Biological and Environmental
  - Bioinformatics/Genomics
  - Climate Science
- Fusion Energy Sciences
  - Magnetic Fusion Energy/ITER
- High Energy Physics
  - LHC
- Nuclear Physics
  - RHIC

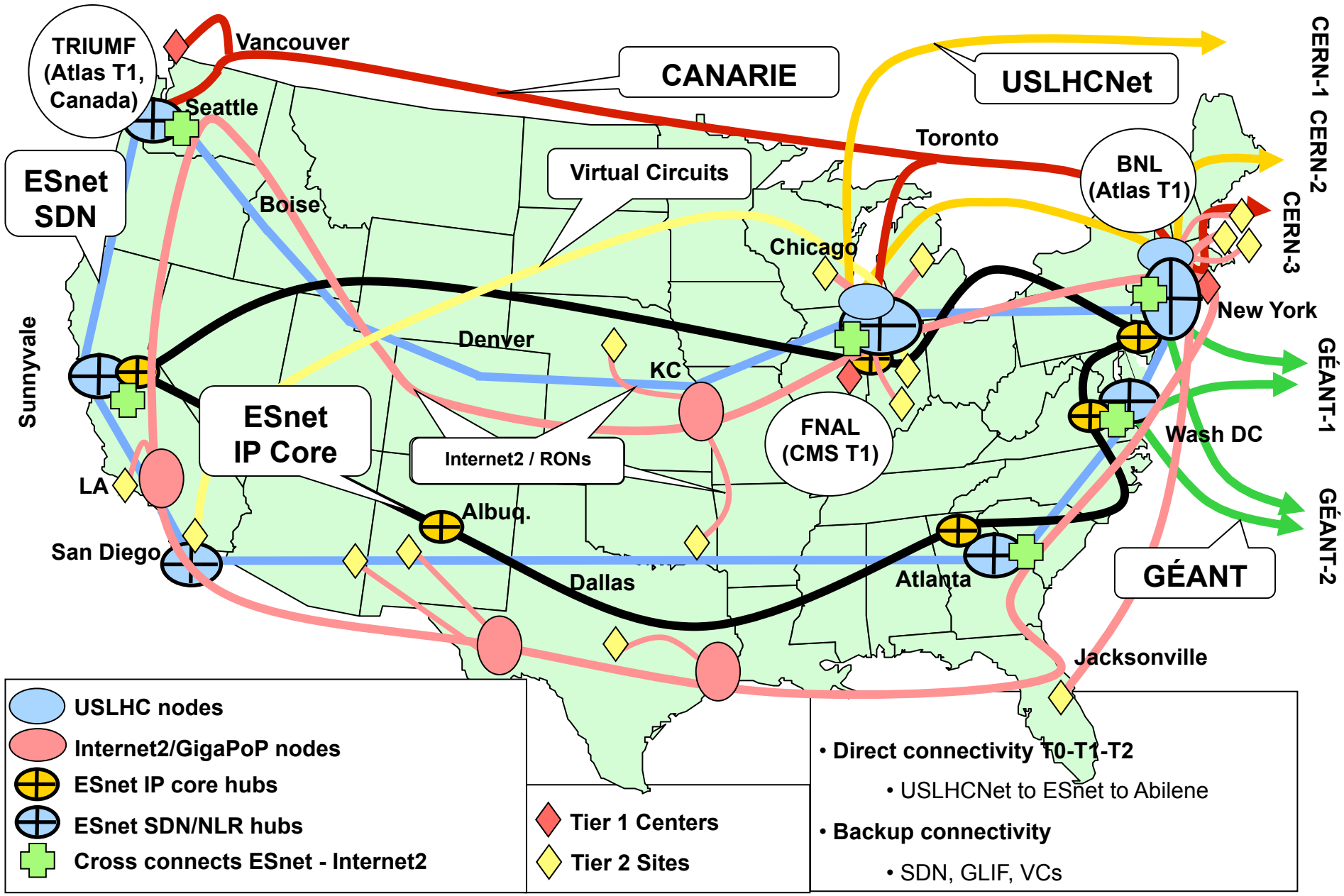
(2A) Science Networking Requirements Aggregation Summary

Science Drivers Science Areas / Facilities	End2End Reliability	Connectivity	Today End2End Band width	5 years End2End Band width	Traffic Characteristics	Network Services
Magnetic Fusion Energy	99.999% (Impossible without full redundancy)	• DOE sites • US Universities • Industry	200+ Mbps	1 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • Guaranteed QoS • Deadline scheduling
NERSC and ACLF	-	• DOE sites • US Universities • International • Other ASCR supercomputers	10 Gbps	20 to 40 Gbps	• Bulk data • Remote control • Remote file system sharing	• Guaranteed bandwidth • Guaranteed QoS • Deadline Scheduling • PKI / Grid
NLCF	-	• DOE sites • US Universities • Industry • International	Backbone Band width parity	Backbone band width parity	• Bulk data • Remote file system sharing	
Nuclear Physics (RHIC)	-	• DOE sites • US Universities • International	12 Gbps	70 Gbps	• Bulk data	• Guaranteed bandwidth • PKI / Grid
Spallation Neutron Source	High (24x7 operation)	• DOE sites	640 Mbps	2 Gbps	• Bulk data	

Science Network Requirements Aggregation Summary

Science Drivers Science Areas / Facilities	End2End Reliability	Connectivity	Today End2End Band width	5 years End2End Band width	Traffic Characteristics	Network Services
Advanced Light Source	-	• DOE sites • US Universities • Industry	1 TB/day 300 Mbps	5 TB/day 1.5 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • PKI / Grid
Bioinformatics	-	• DOE sites • US Universities	625 Mbps 12.5 Gbps in two years	250 Gbps	• Bulk data • Remote control • Point-to-multipoint	• Guaranteed bandwidth • High-speed multicast
Chemistry / Combustion	-	• DOE sites • US Universities • Industry	-	10s of Gigabits per second	• Bulk data	• Guaranteed bandwidth • PKI / Grid
Climate Science	-	• DOE sites • US Universities • International	-	5 PB per year 5 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • PKI / Grid
		Immediate Requirements and Drivers				
High Energy Physics (LHC)	99.95+% (Less than 4 hrs/year)	• US Tier1 (FNAL, BNL) • US Tier2 (Universities) • International (Europe, Canada)	10 Gbps	60 to 80 Gbps (30-40 Gbps per US Tier1)	• Bulk data • Coupled data analysis processes	• Guaranteed bandwidth • Traffic isolation • PKI / Grid

(2B) The Next Level of Detail: LHC Tier 0, 1, and 2 Connectivity Requirements Summary



(2C) The Next Level of Detail: LHC ATLAS Bandwidth Matrix as of April 2007

Site A	Site Z	ESnet A	ESnet Z	A-Z 2007 Bandwidth	A-Z 2010 Bandwidth
CERN	BNL	AofA (NYC)	BNL	10Gbps	20-40Gbps
BNL	U. of Michigan (Calibration)	BNL (LIMAN)	Starlight (CHIMAN)	3Gbps	10Gbps
BNL	Boston University	BNL (LIMAN)	Internet2 / NLR Peerings	3Gbps (Northeastern Tier2 Center)	10Gbps (Northeastern Tier2 Center)
BNL	Harvard University				
BNL	Indiana U. at Bloomington	BNL (LIMAN)	Internet2 / NLR Peerings	3Gbps (Midwestern Tier2 Center)	10Gbps (Midwestern Tier2 Center)
BNL	U. of Chicago				
BNL	Langston University	BNL (LIMAN)	Internet2 / NLR Peerings	3Gbps (Southwestern Tier2 Center)	10Gbps (Southwestern Tier2 Center)
BNL	U. Oklahoma Norman				
BNL	U. of Texas Arlington				
BNL	Tier3 Aggregate	BNL (LIMAN)	Internet2 / NLR Peerings	5Gbps	20Gbps
BNL	TRIUMF (Canadian ATLAS Tier1)	BNL (LIMAN)	Seattle	1Gbps	5Gbps

LHC CMS Bandwidth Matrix as of April 2007

Site A	Site Z	ESnet A	ESnet Z	A-Z 2007 Bandwidth	A-Z 2010 Bandwidth
CERN	FNAL	Starlight (CHIMAN)	FNAL (CHIMAN)	10Gbps	20-40Gbps
FNAL	U. of Michigan (Calibration)	FNAL (CHIMAN)	Starlight (CHIMAN)	3Gbps	10Gbps
FNAL	Caltech	FNAL (CHIMAN)	Starlight (CHIMAN)	3Gbps	10Gbps
FNAL	MIT	FNAL (CHIMAN)	AofA (NYC)/ Boston	3Gbps	10Gbps
FNAL	Purdue University	FNAL (CHIMAN)	Starlight (CHIMAN)	3Gbps	10Gbps
FNAL	U. of California at San Diego	FNAL (CHIMAN)	San Diego	3Gbps	10Gbps
FNAL	U. of Florida at Gainesville	FNAL (CHIMAN)	SOX / Ultralight at Starlight	3Gbps	10Gbps
FNAL	U. of Nebraska at Lincoln	FNAL (CHIMAN)	Starlight (CHIMAN)	3Gbps	10Gbps
FNAL	U. of Wisconsin at Madison	FNAL (CHIMAN)	Starlight (CHIMAN)	3Gbps	10Gbps
FNAL	Tier3 Aggregate	FNAL (CHIMAN)	Internet2 / NLR Peerings	5Gbps	20Gbps

Large-Scale Data Analysis Systems (Typified by the LHC) have Several Characteristics that Result in Requirements for the Network and its Services

- The ***systems are data intensive and high-performance***, typically moving terabytes a day for months at a time
- The ***system are high duty-cycle***, operating most of the day for months at a time in order to meet the requirements for data movement
- The ***systems are widely distributed*** – typically spread over continental or inter-continental distances
- Such ***systems depend on network performance and availability***, but these characteristics cannot be taken for granted, even in well run networks, when the multi-domain network path is considered
- The applications ***must be able to get guarantees from the network*** that there is adequate bandwidth to accomplish the task at hand
- The applications ***must be able to get information from the network*** that allows graceful failure and auto-recovery and adaptation to unexpected network conditions that are short of outright failure

This slide drawn from [ICFA SCIC]

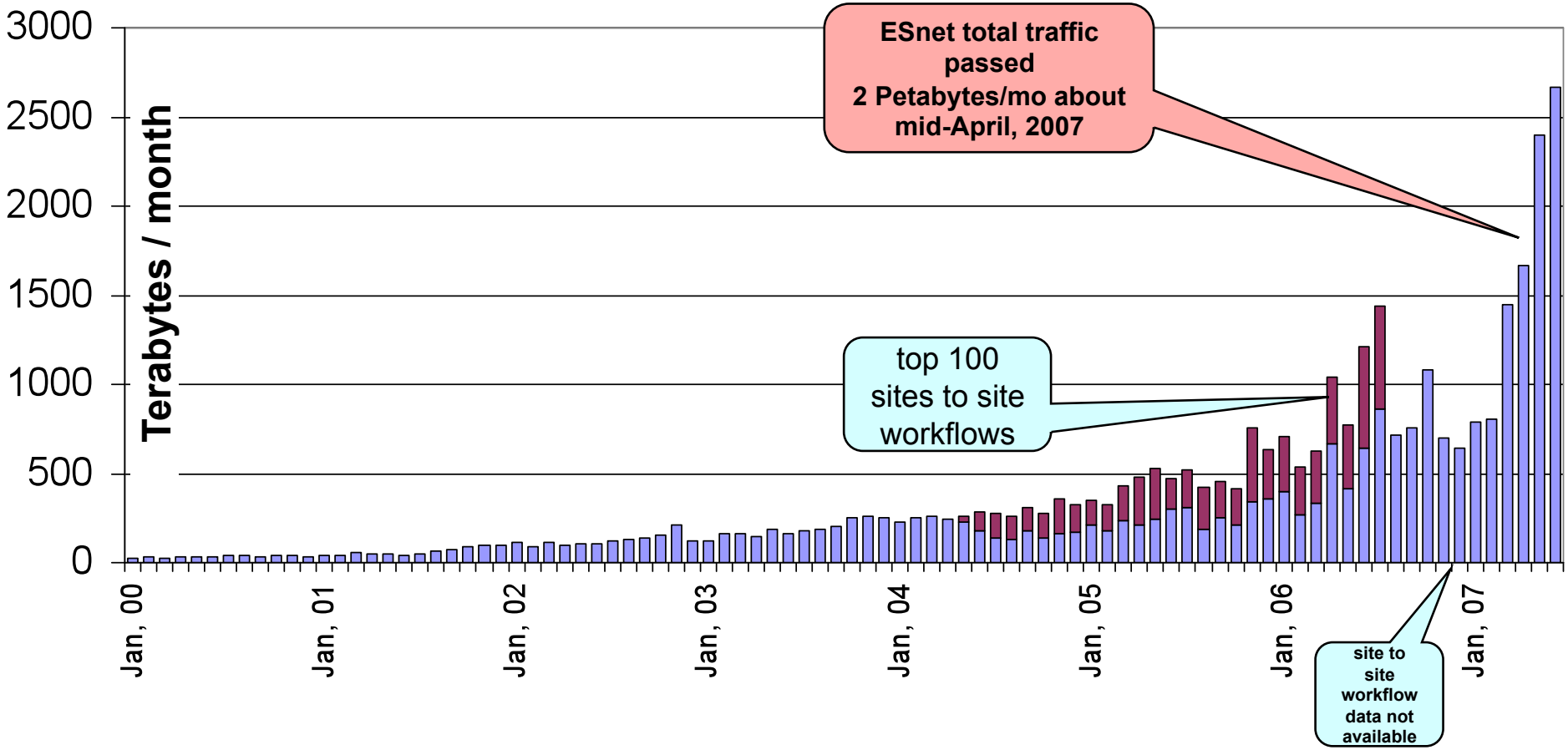
Enabling Large-Scale Science

- These requirements are generally true for systems with widely distributed components to be reliable and consistent in performing the sustained, complex tasks of large-scale science

- Networks must provide communication capability that is service-oriented: configurable, schedulable, predictable, reliable, and informative – and the network and its services must be scalable

(2D)

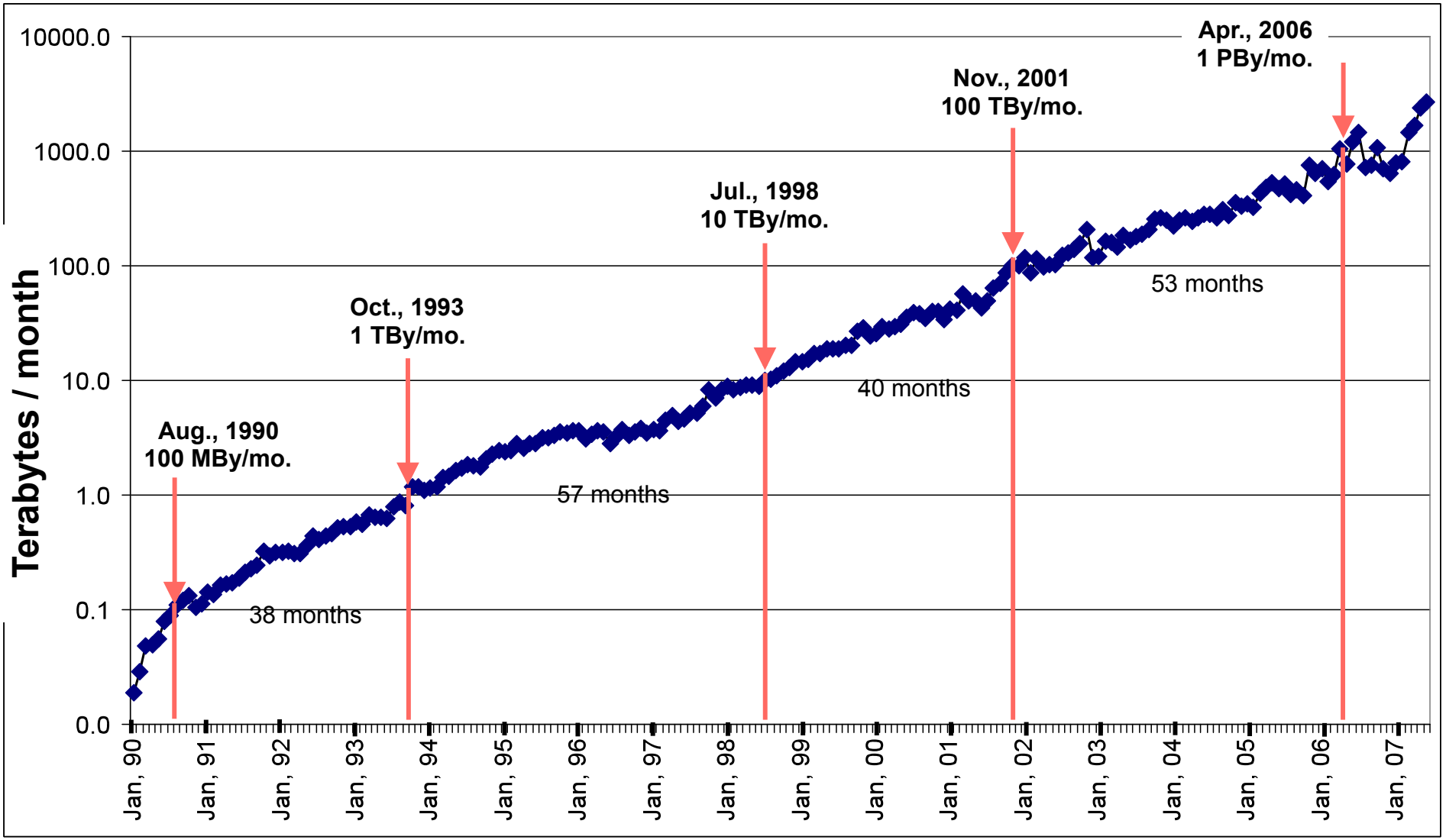
3. Observed Evolution of Historical ESnet Traffic Patterns



ESnet Monthly Accepted Traffic, January, 2000 – June, 2007

- ESnet is currently transporting more than1 petabyte (1000 terabytes) per month
- More than 50% of the traffic is now generated by the top 100 sites ⇒ large-scale science dominates all ESnet traffic

ESnet Traffic has Increased by 10X Every 47 Months, on Average, Since 1990



Log Plot of ESnet Monthly Accepted Traffic, January, 1990 – June, 2007

Requirements from Network Utilization Observation

- In 4 years, we can expect a 10x increase in traffic over current levels *without the addition of production LHC traffic*
  - Nominal average load on busiest backbone links is ~1.5 Gbps today
  - In 4 years that figure will be ~15 Gbps based on current trends

(3A)

- Measurements of this type are science-agnostic
  - It doesn't matter who the users are, the traffic load is increasing exponentially
  - Predictions based on this sort of forward projection tend to be conservative estimates of future requirements because they cannot predict new uses

Requirements from Traffic Flow Observations

- Most of ESnet science traffic has a source or sink outside of ESnet

(3B)

  - Drives requirement for high-bandwidth peering
  - Reliability and bandwidth requirements demand that peering be redundant
  - Multiple 10 Gbps peerings today, must be able to add more bandwidth flexibly and cost-effectively
  - Bandwidth and service guarantees must traverse R&E peerings
    - Collaboration with other R&E networks on a common framework is critical
    - Seamless fabric

- Large-scale science is now the dominant user of the network

(3C)

  - Satisfying the demands of large-scale science traffic into the future will require a purpose-built, scalable architecture
  - Traffic patterns are different than commodity Internet

Summary of All Requirements To-Date

Requirements from SC Programs:

1A) Provide “consulting” on system / application network tuning

Requirements from science case studies:

2A) Build the ESnet core up to 100 Gb/s within 5 years

2B) Deploy network to accommodate LHC collaborator footprint

2C) Implement network to provide for LHC data path loadings

2D) Provide the network as a service-oriented capability

Requirements from observing traffic growth and change trends in the network:

3A) Provide 15 Gb/s core within four years and 150 Gb/s core within eight years

3B) Provide a rich diversity and high bandwidth for R&E peerings

3C) Economically accommodate a very large volume of circuit-like traffic

➢ ESnet4 - The Response to the Requirements

I) A new network architecture and implementation strategy

- Provide two networks: IP and circuit-oriented Science Data Netework
  - Reduces cost of handling high bandwidth data flows
    - Highly capable routers are not necessary when every packet goes to the same place
    - Use lower cost (factor of 5x) switches to relatively route the packets
- Rich and diverse network topology for flexible management and high reliability
- Dual connectivity at every level for all large-scale science sources and sinks
- A partnership with the US research and education community to build a shared, large-scale, R&E managed optical infrastructure
  - a scalable approach to adding bandwidth to the network
  - dynamic allocation and management of optical circuits

II) Development and deployment of a virtual circuit service

- Develop the service cooperatively with the networks that are intermediate between DOE Labs and major collaborators to ensure and-to-end interoperability

III) Develop and deploy service-oriented, user accessable network monitoring systems

IV) Provide "consulting" on system / application network performance tuning

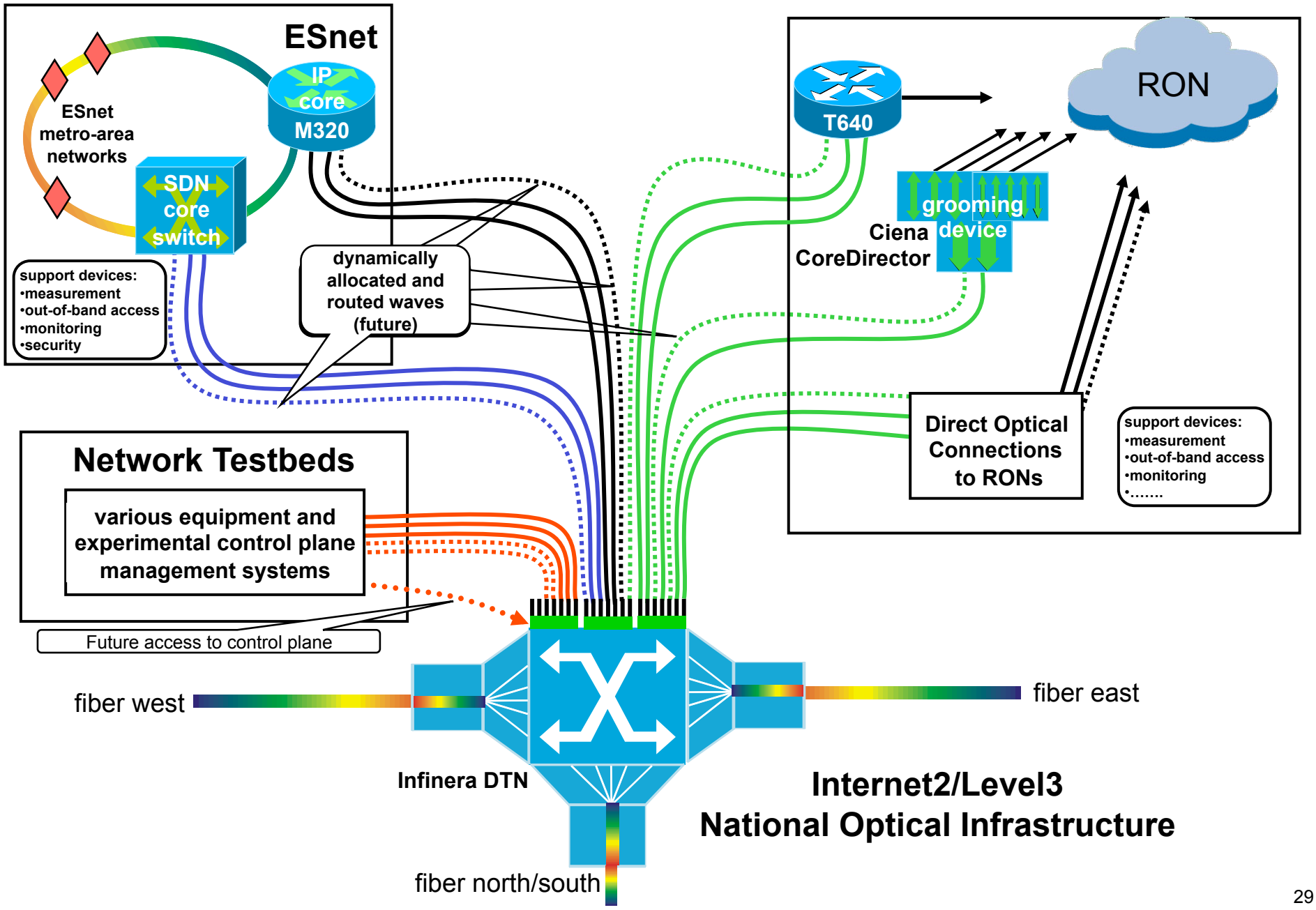
ESnet4

- Internet2 has partnered with Level 3 Communications Co. and Infinera Corp. for a dedicated optical fiber infrastructure with a national footprint and a rich topology - the “Internet2 Network”
  - The fiber will be provisioned with Infinera Dense Wave Division Multiplexing equipment that uses an advanced, integrated optical-electrical design
  - Level 3 will maintain the fiber and the DWDM equipment
  - The DWDM equipment will initially be provisioned to provide10 optical circuits (lambdas - λs) across the entire fiber footprint (80 λs is max.)
- ESnet has partnered with Internet2 to:
  - Share the optical infrastructure
  - Develop new circuit-oriented network services
  - Explore mechanisms that could be used for the ESnet Network Operations Center (NOC) and the Internet2/Indiana University NOC to back each other up for disaster recovery purposes

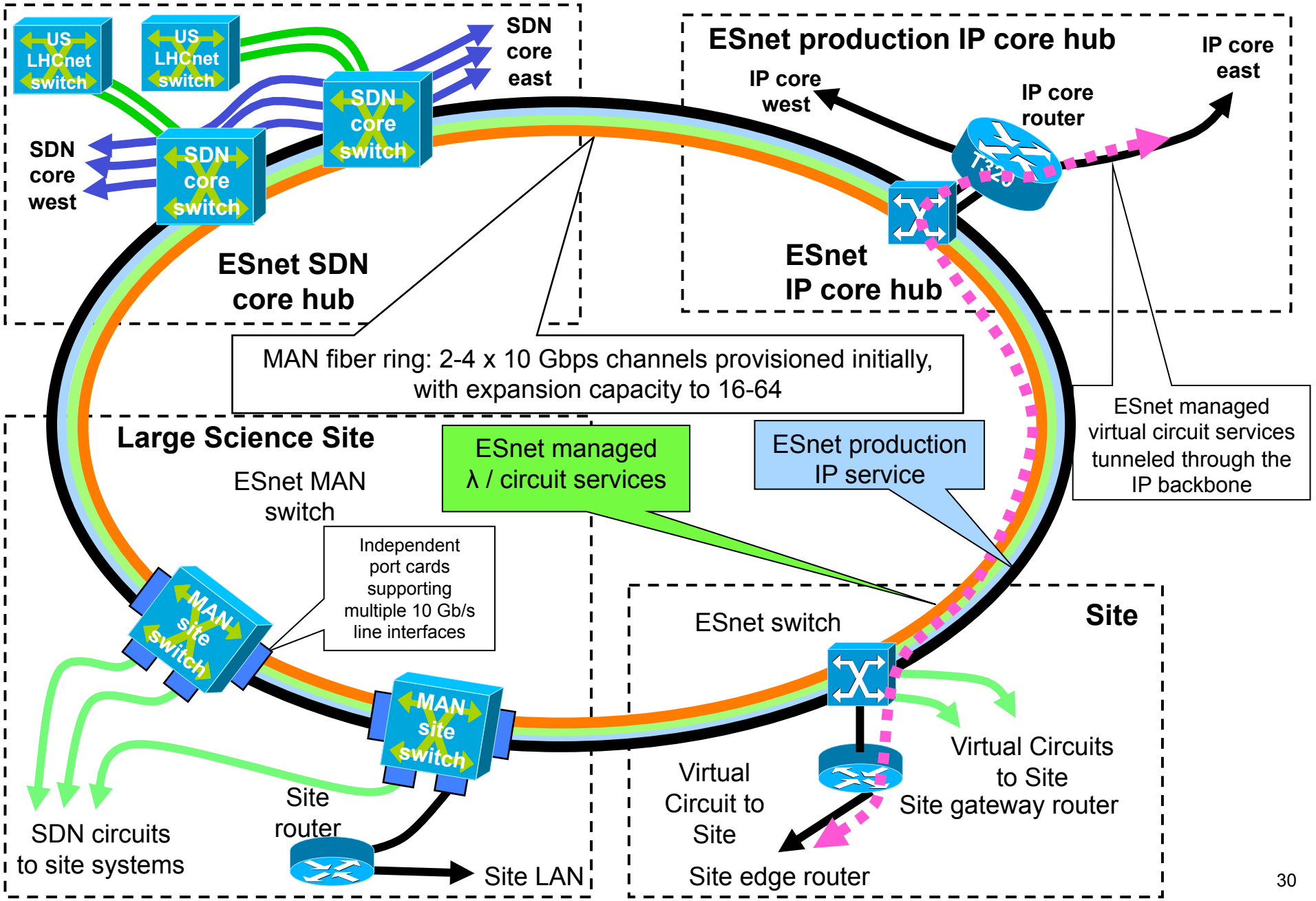
ESnet4

- ESnet will build its next generation IP network and its new circuit-oriented Science Data Network primarily on the Internet2 circuits (λs) that are dedicated to ESnet, together with a few National Lambda Rail and other circuits
  - ESnet will provision and operate its own routing and switching hardware that is installed in various commercial telecom hubs around the country, as it has done for the past 20 years
  - ESnet's peering relationships with the commercial Internet, various US research and education networks, and numerous international networks will continue and evolve as they have for the past 20 years

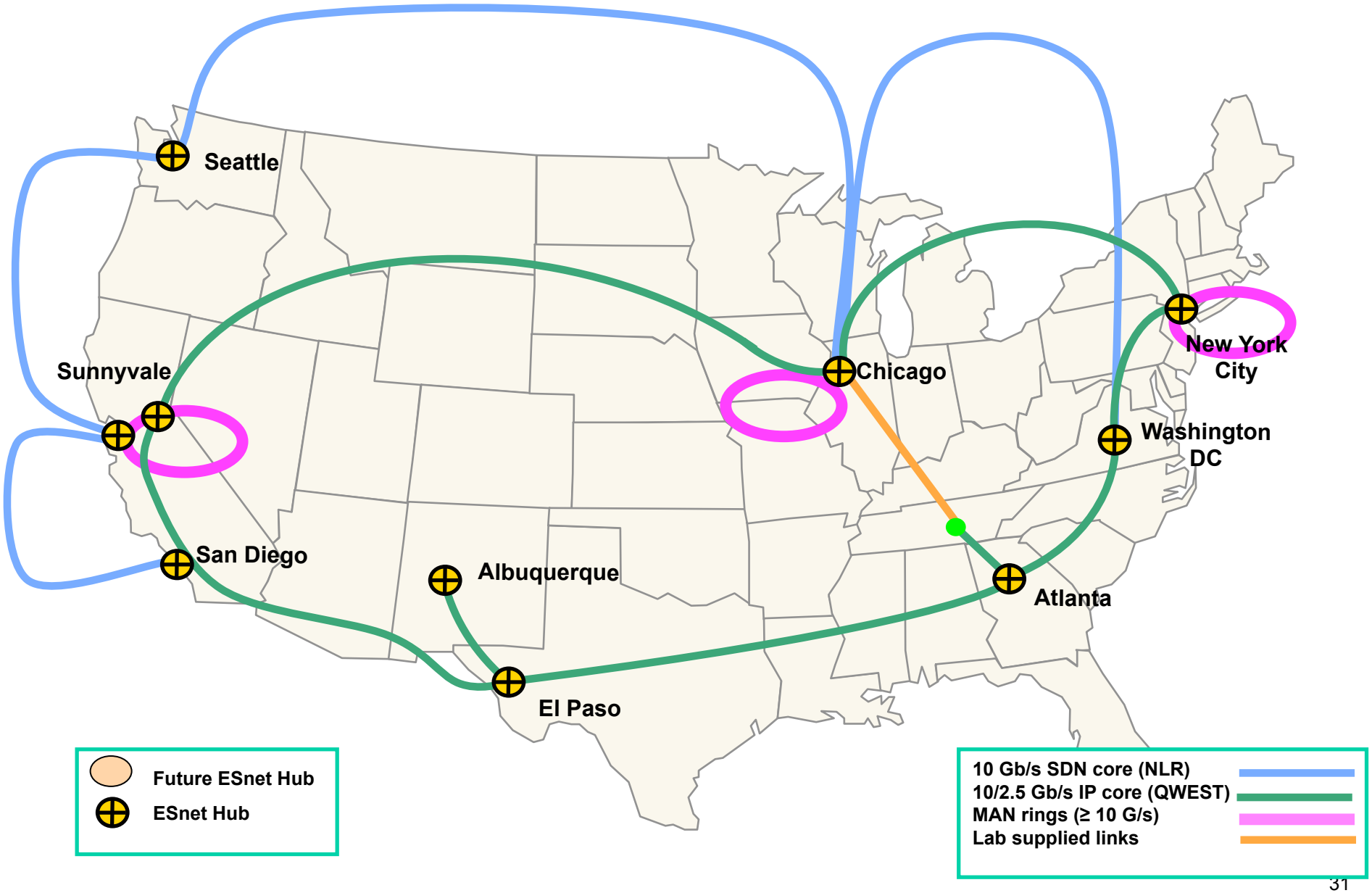
Internet2 and ESnet Optical Node



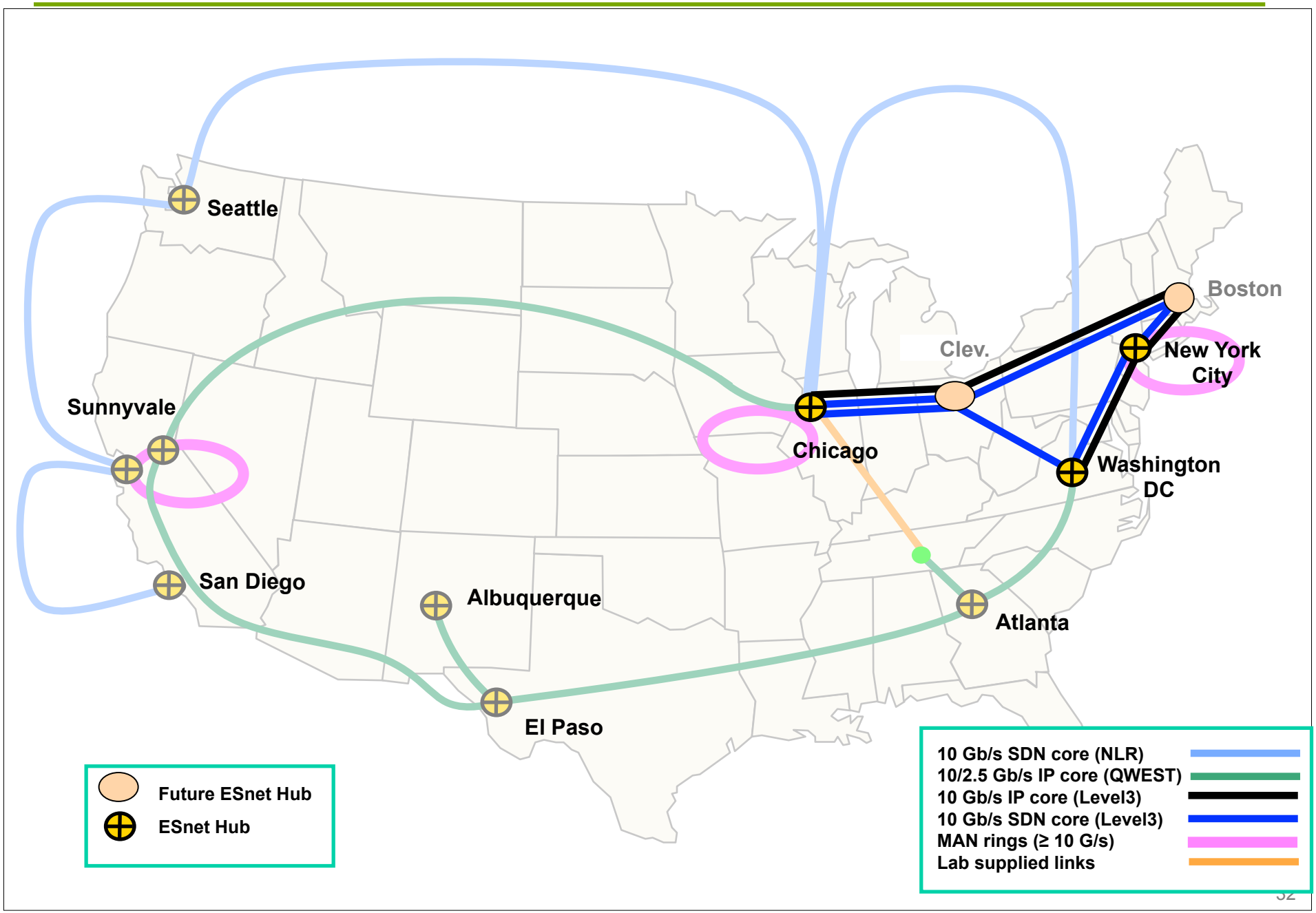
ESnet Metropolitan Area Network Ring Architecture for High Reliability Sites



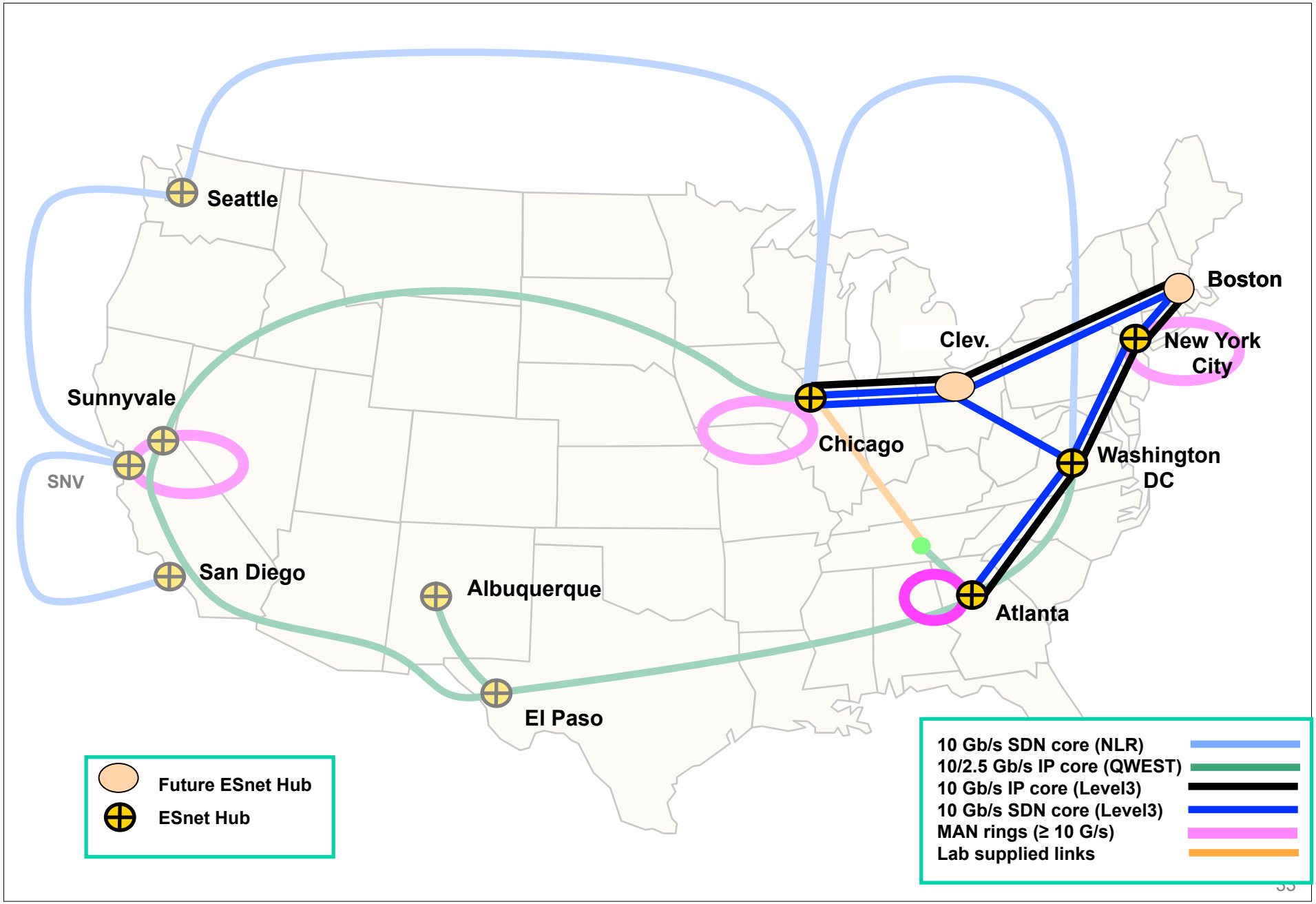
ESnet 3 Backbone as of January 1, 2007



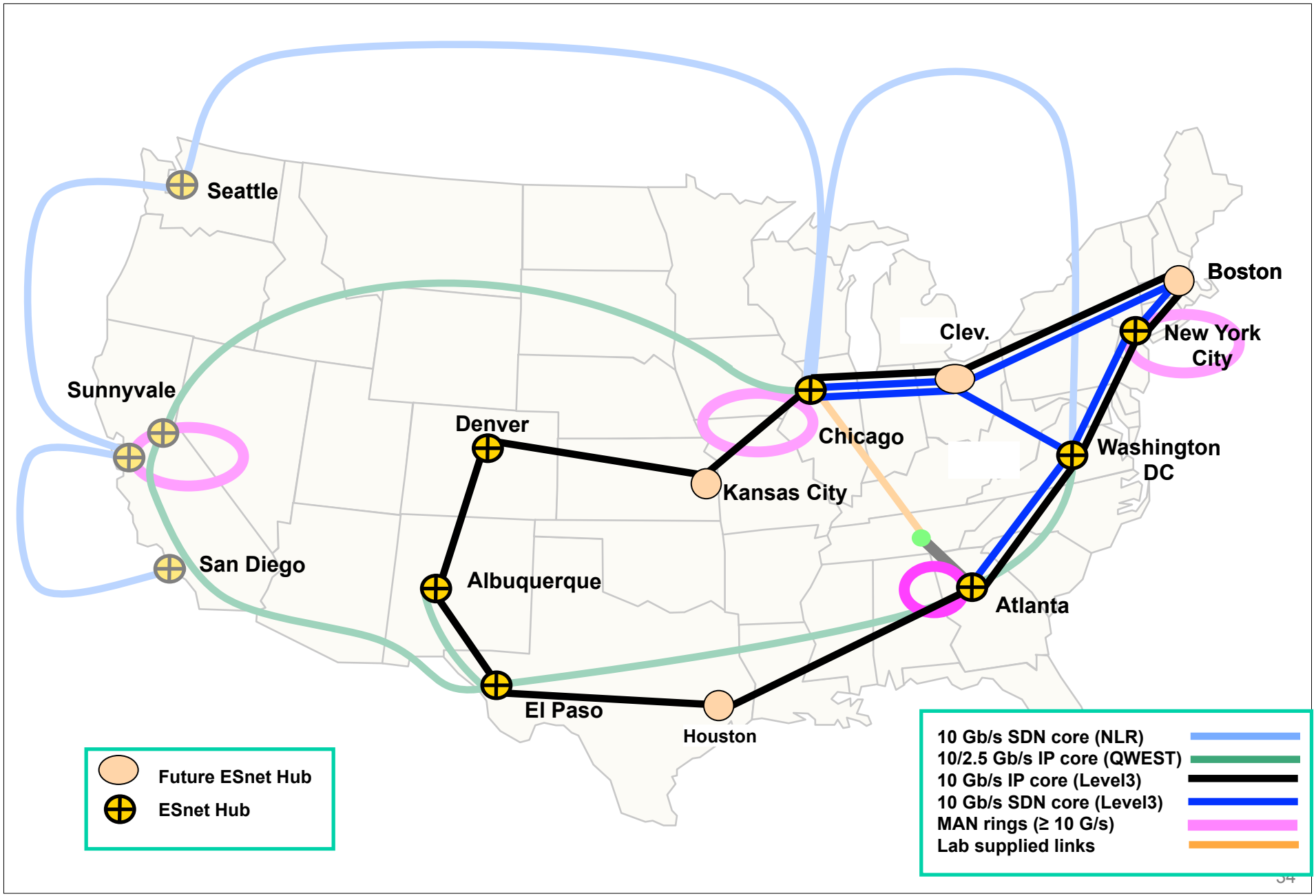
ESnet 4 Backbone as of April 15, 2007



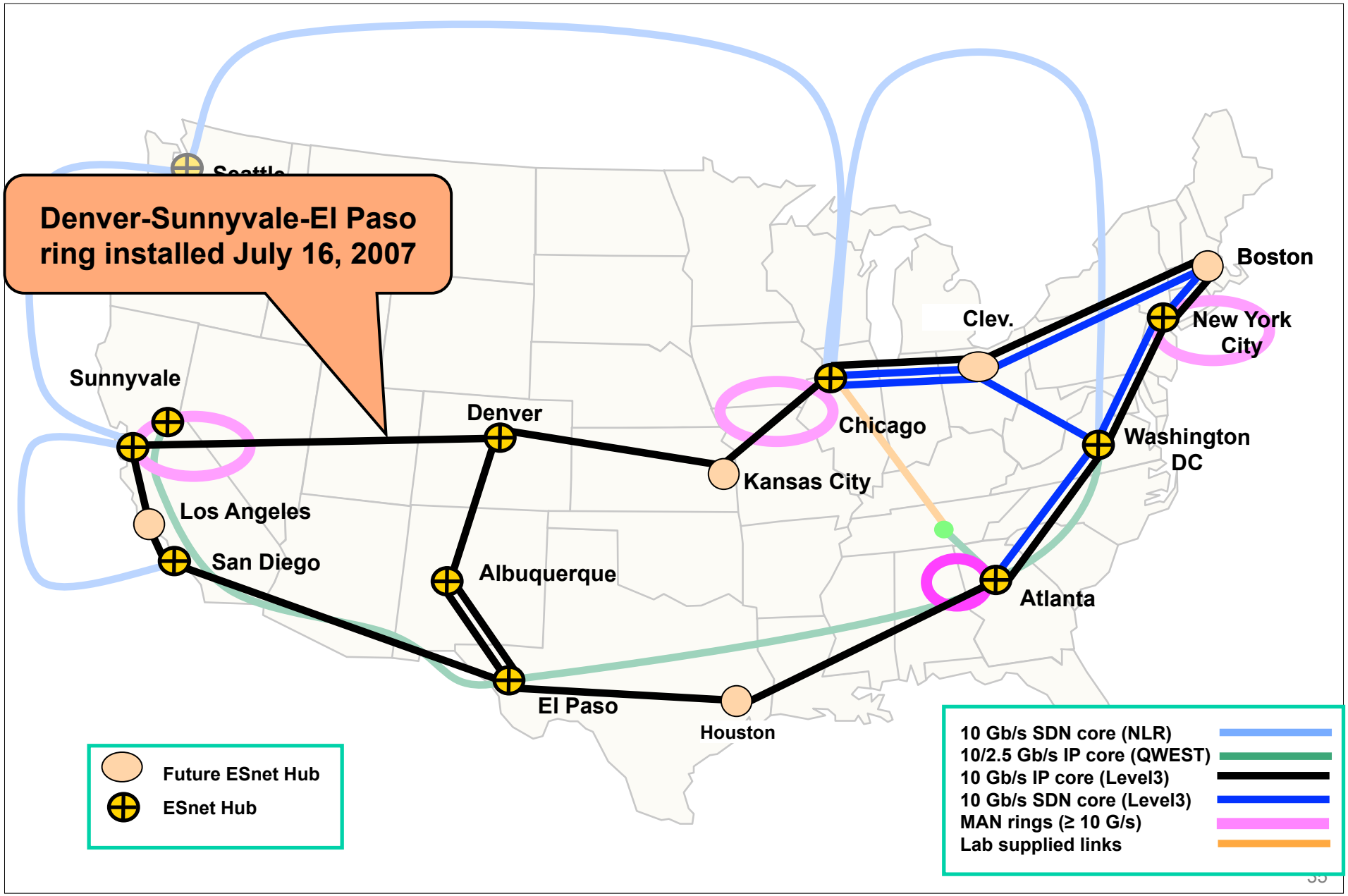
ESnet 4 Backbone as of May 15, 2007



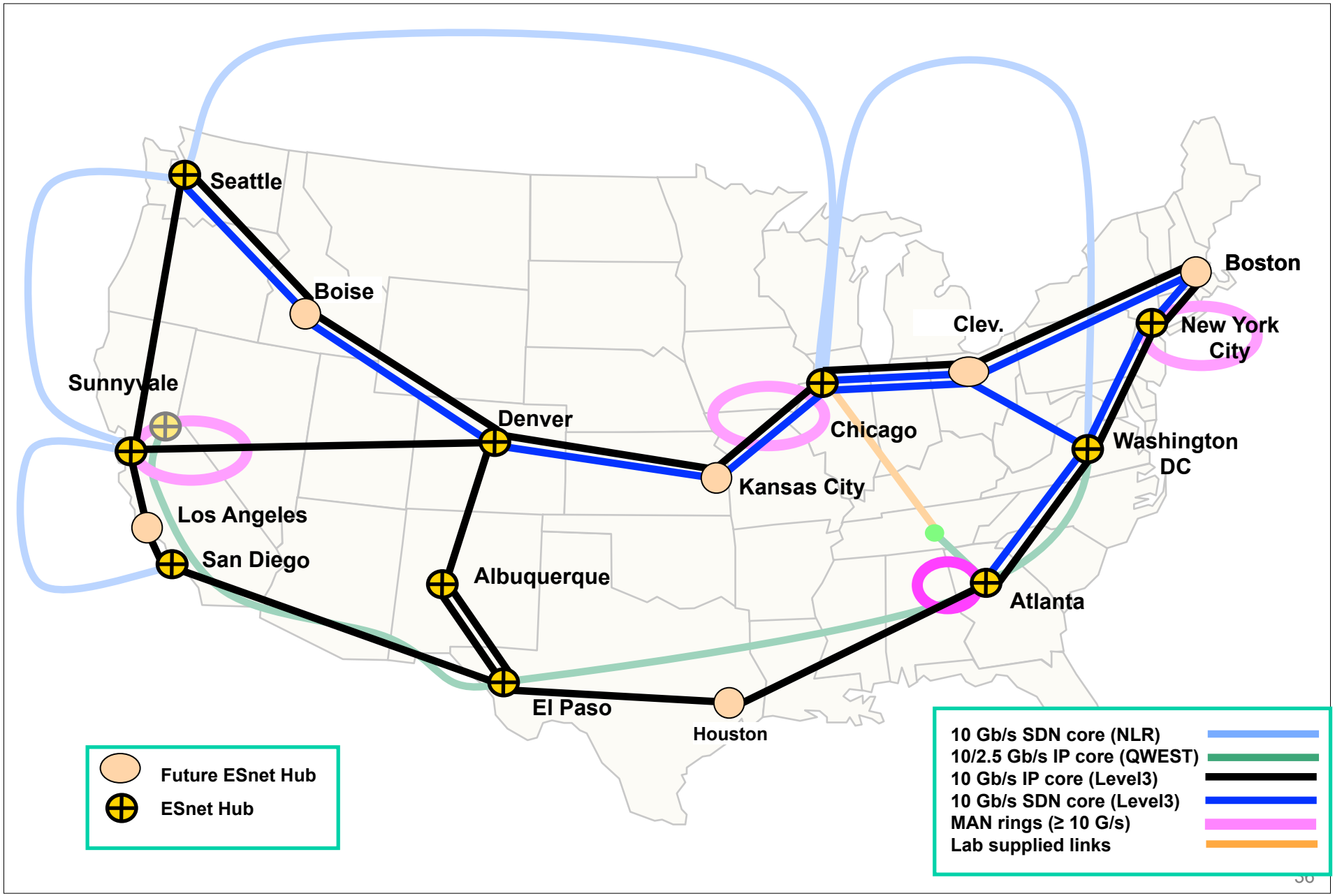
➢ ESnet 4 Backbone as of June 20, 2007



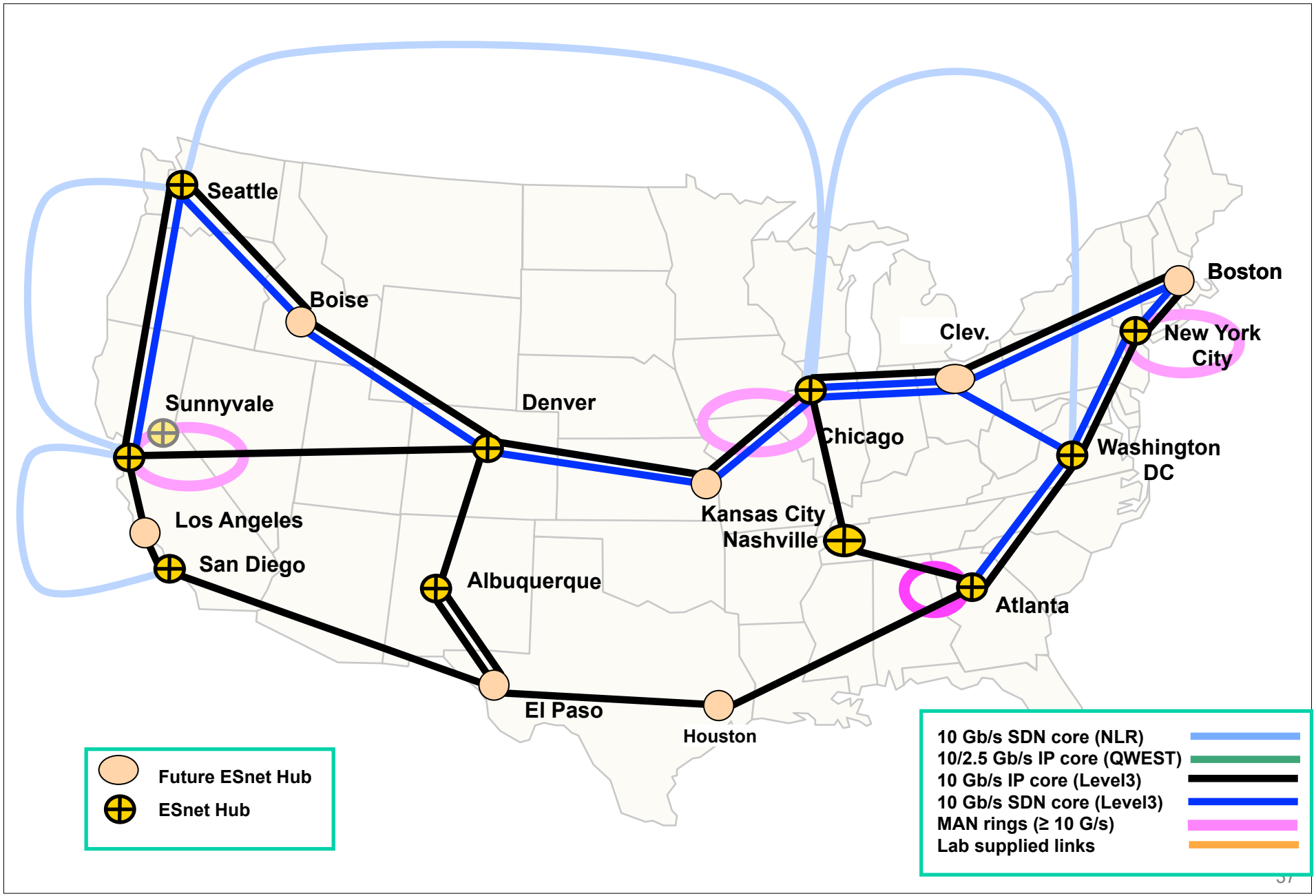
ESnet 4 Backbone Target August 1, 2007



ESnet 4 Backbone Target August 30, 2007

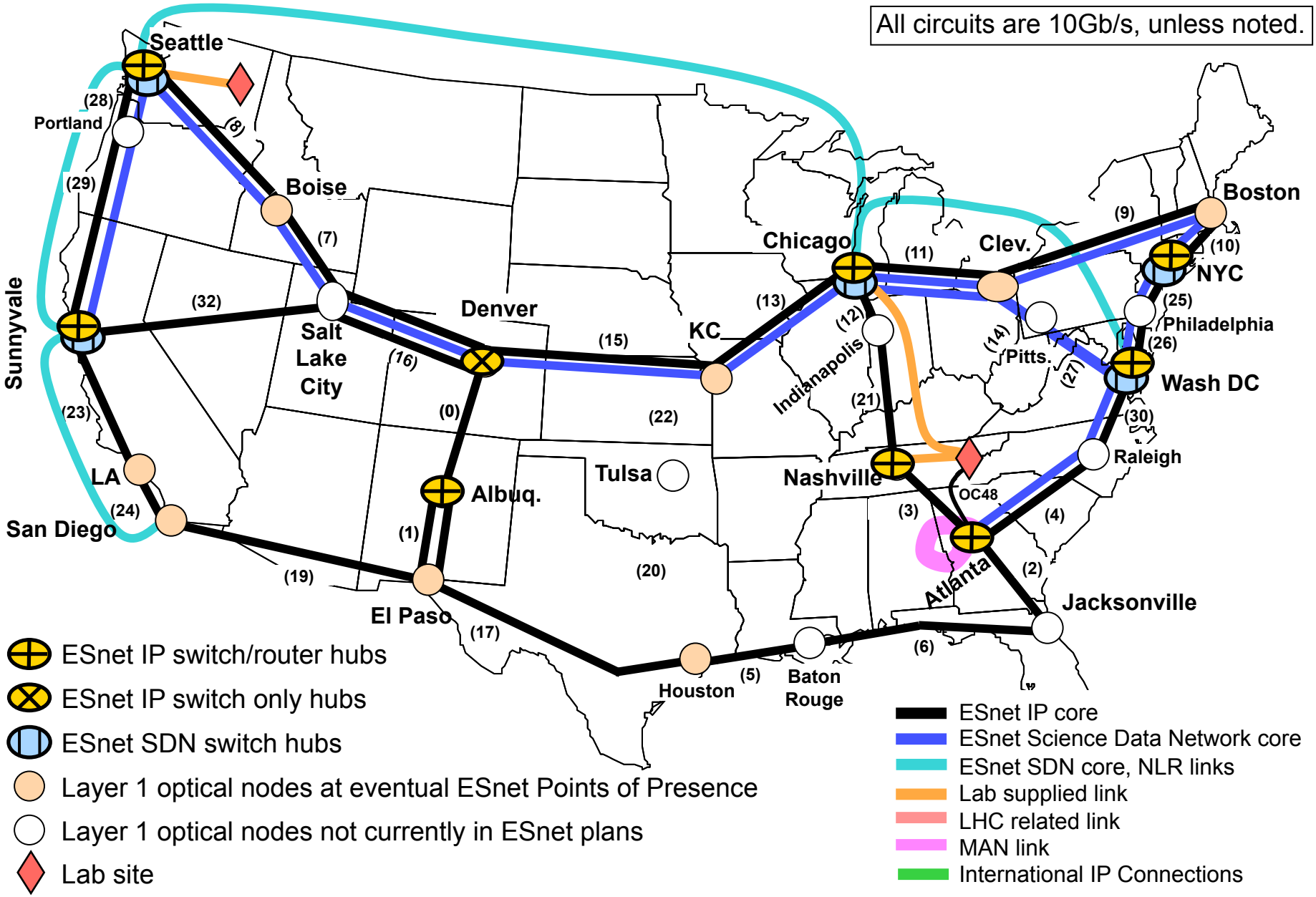


ESnet 4 Backbone Target September 30, 2007

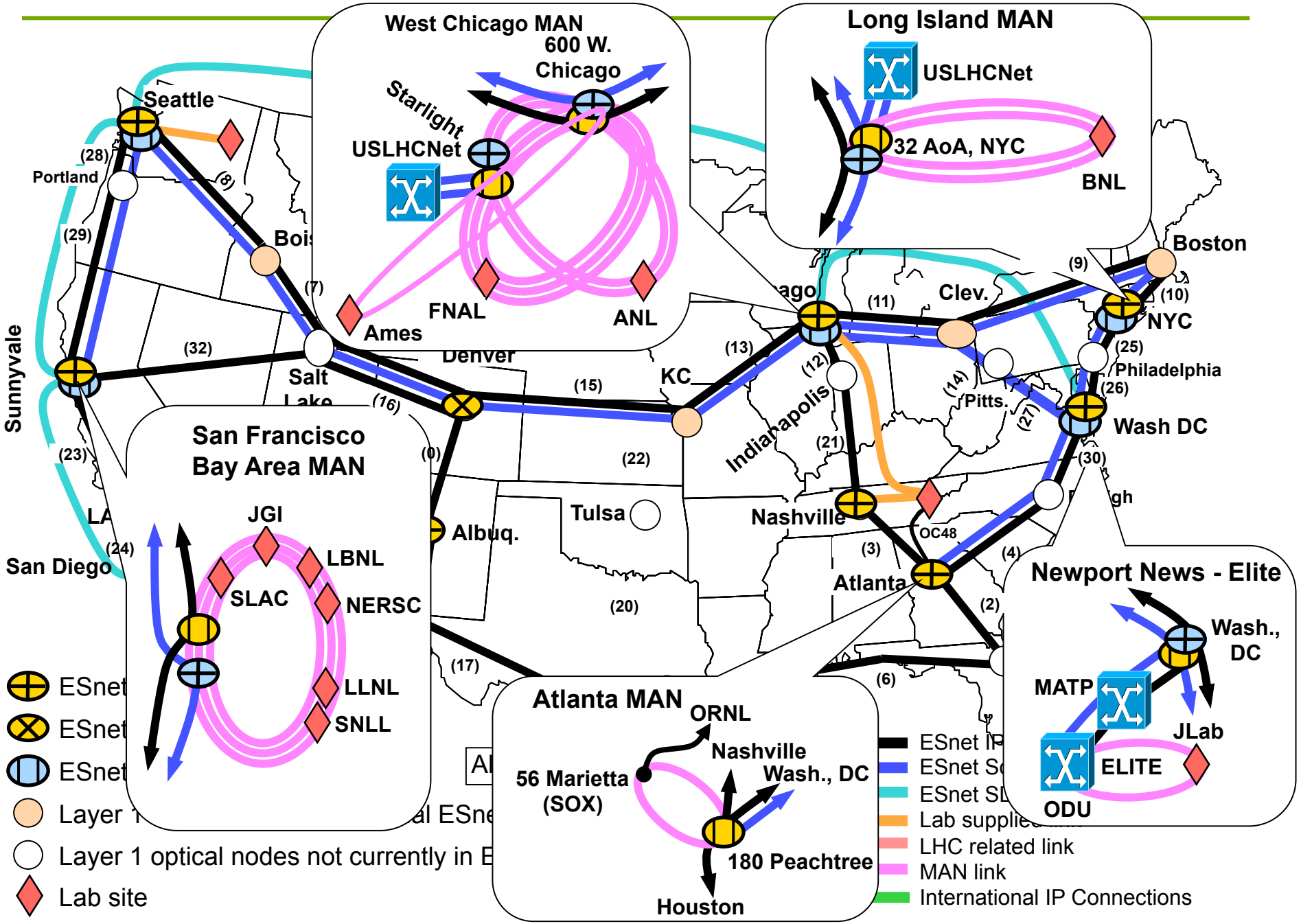


ESnet4 Roll Out

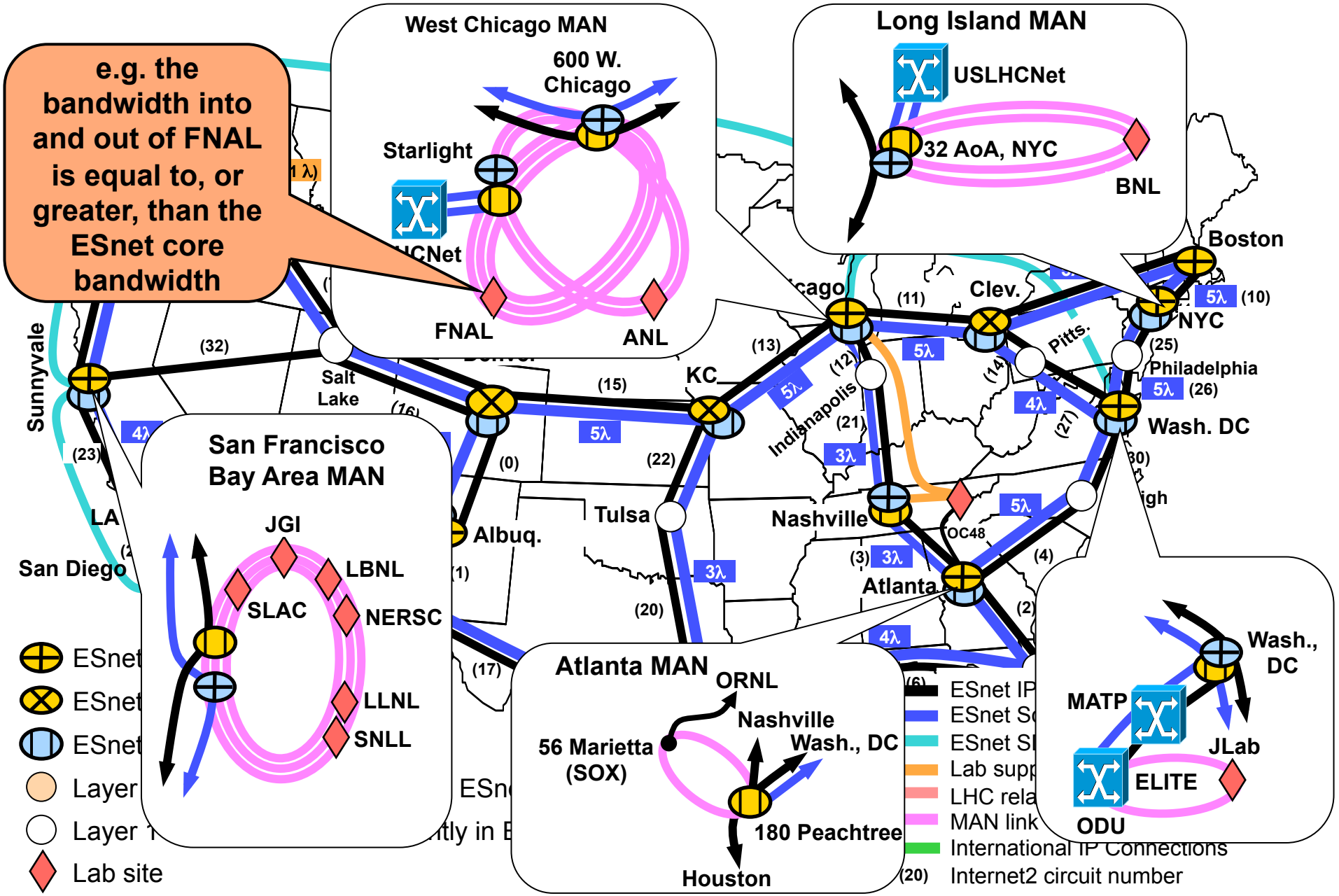
ESnet4 IP + SDN Configuration, mid-September, 2007



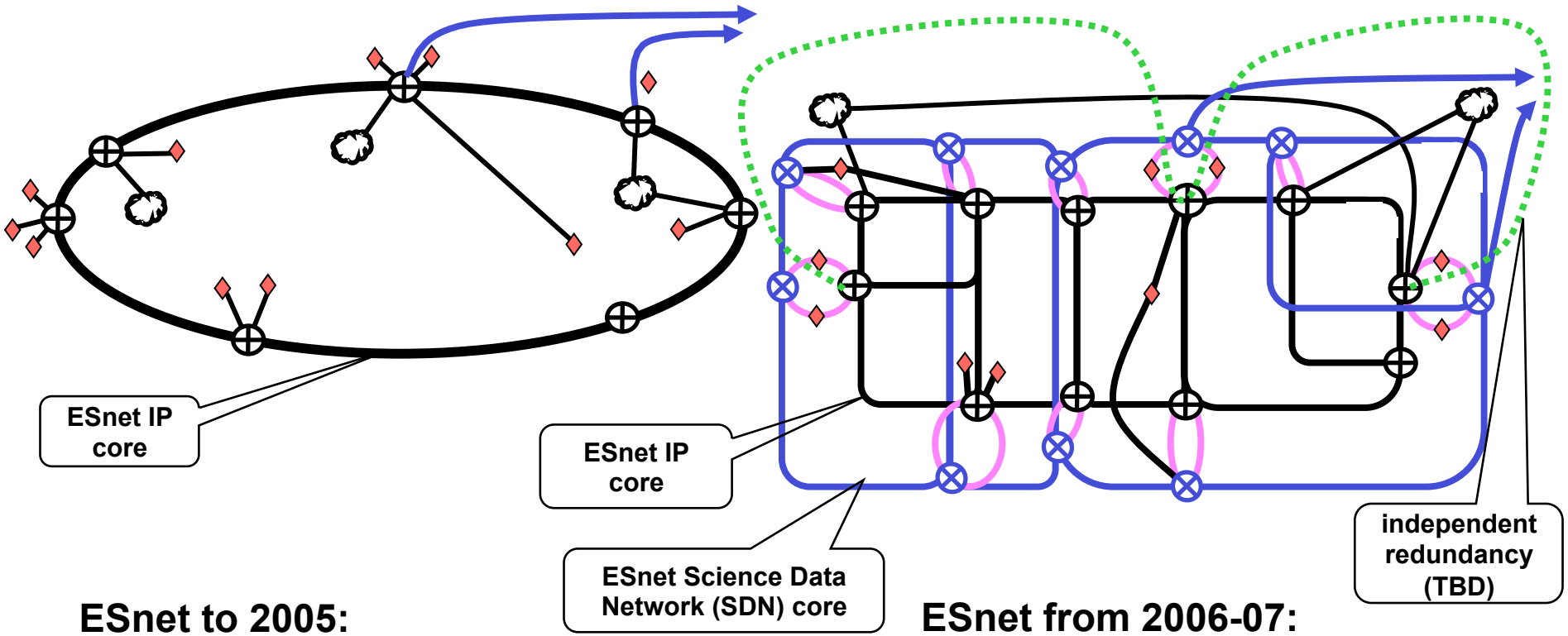
ESnet4 Metro Area Rings, 2007 Configurations



➢ Note that the major ESnet sites are now directly on the ESnet "core" network



The Evolution of ESnet Architecture



ESnet to 2005:

- A routed IP network with sites singly attached to a national core ring

ESnet sites

ESnet hubs / core network connection points

Metro area rings (MANs)

Other IP networks

Circuit connections to other science networks (e.g. USLHCNet)

ESnet from 2006-07:

- A routed IP network with sites dually connected on metro area rings or dually connected directly to core ring
- A switched network providing virtual circuit services for data-intensive science
- Rich topology offsets the lack of dual, independent national cores

ESnet 4 Factiods as of July 16, 2007

- Installation to date:
  - 10 new 10Gb/s circuits
  - ~10,000 Route Miles
  - 6 new hubs
  - 5 new routers 4 new switches
    - Total of 70 individual pieces of equipment shipped
      - Over two and a half tons of electronics
  - 15 round trip airline tickets for our install team
    - About 60,000 miles traveled so far….
    - 6 cities
      - 5 Brazilian Bar-B-Qs/Grills sampled

Typical ESnet 4 Hub

OWAMP Time Source

Power Controllers

10G Performance Tester

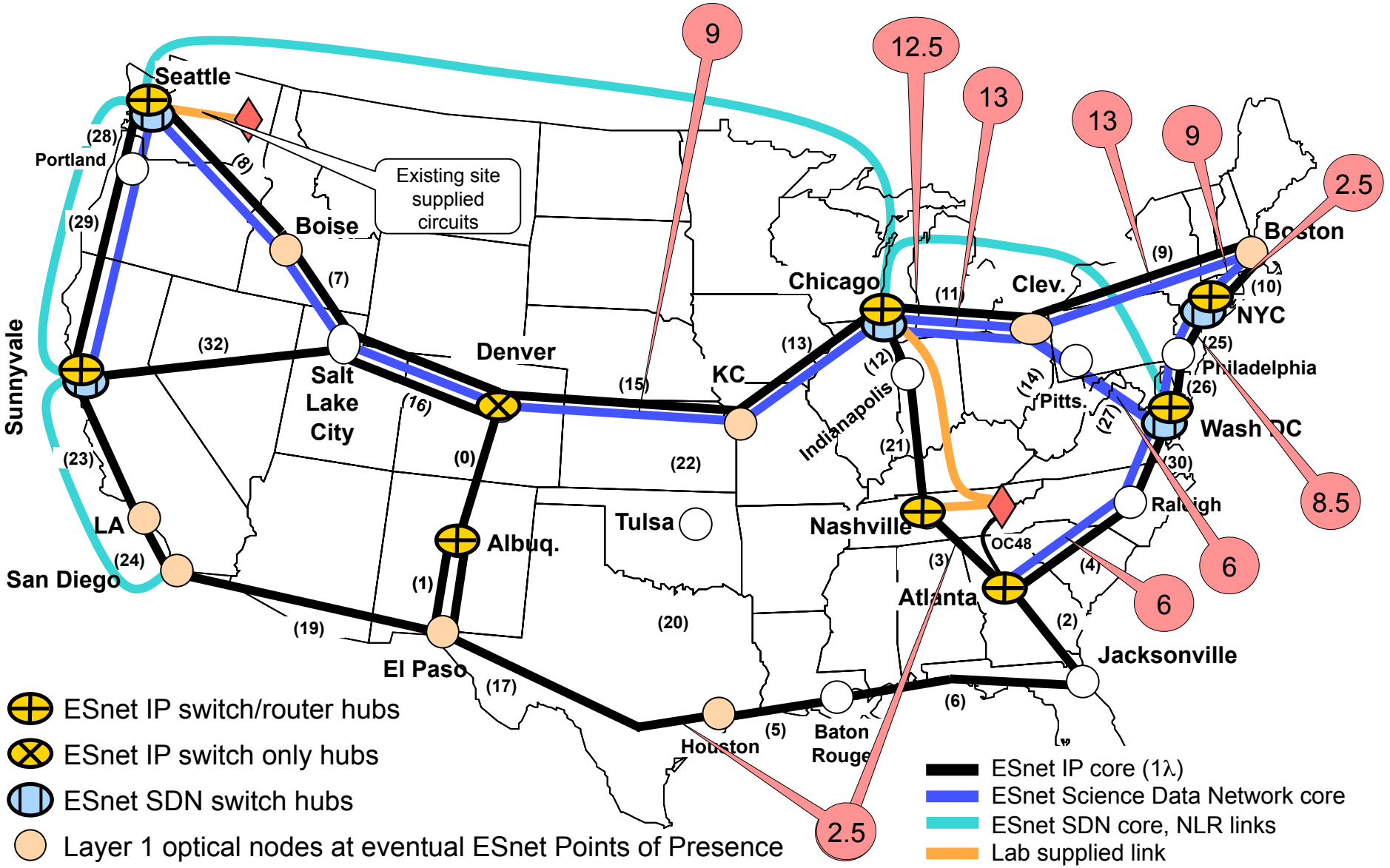
7609 Switch

Secure Term Server

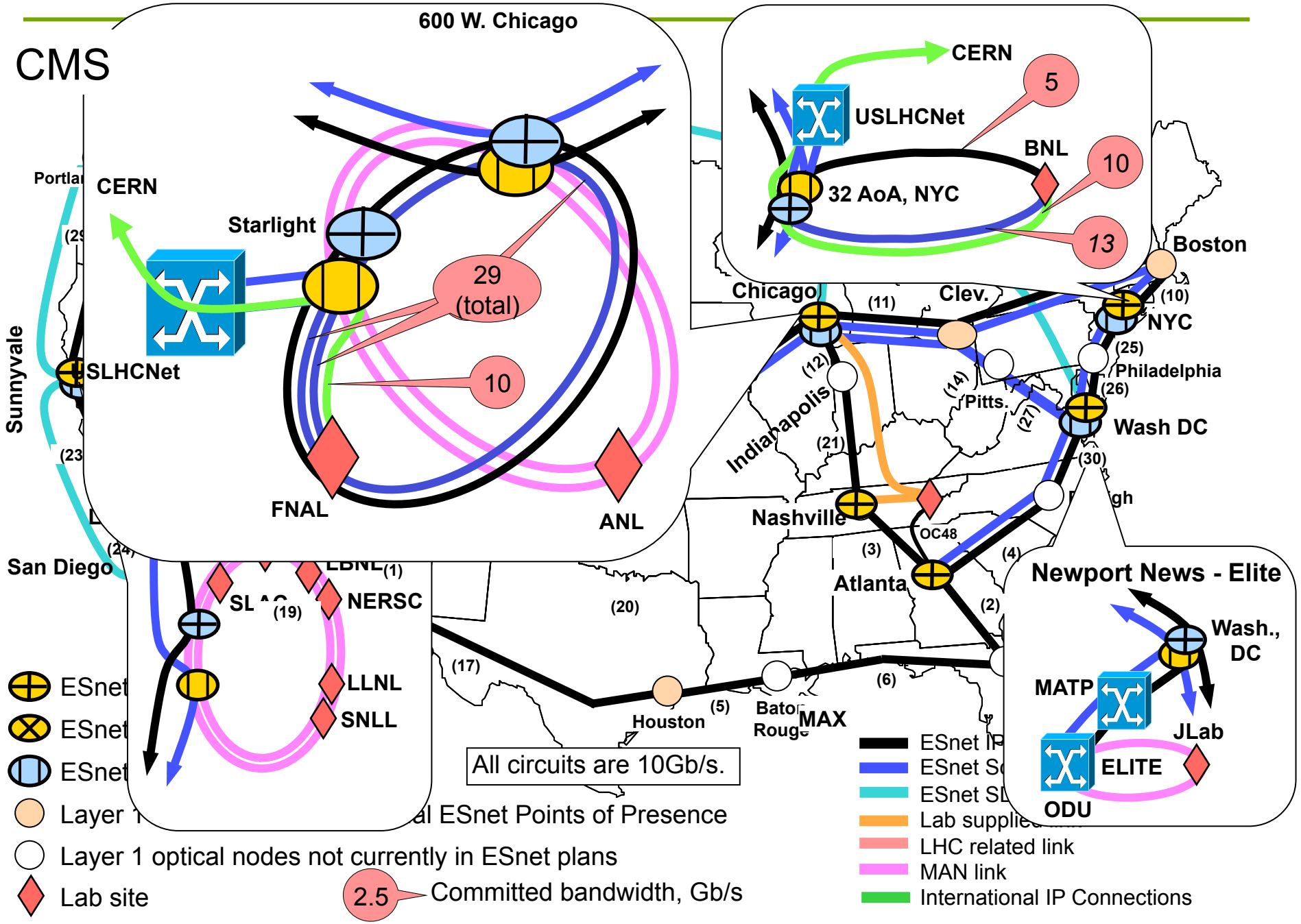
Peering Router

M320 Router

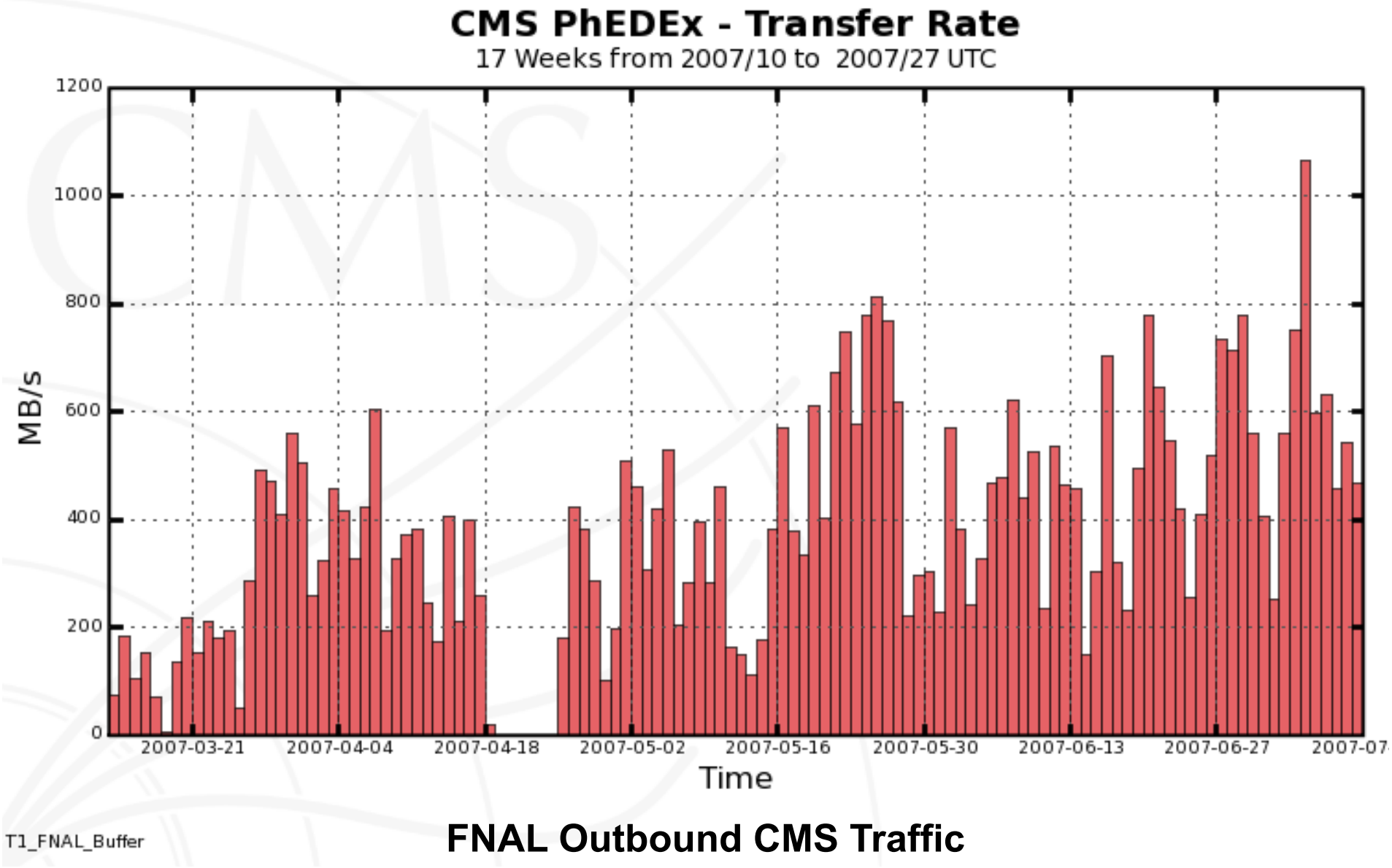
(2C) Aggregate Estimated Link Loadings, 2007-08



(2C) ESnet4 2007-8 Estimated Bandwidth Commitments



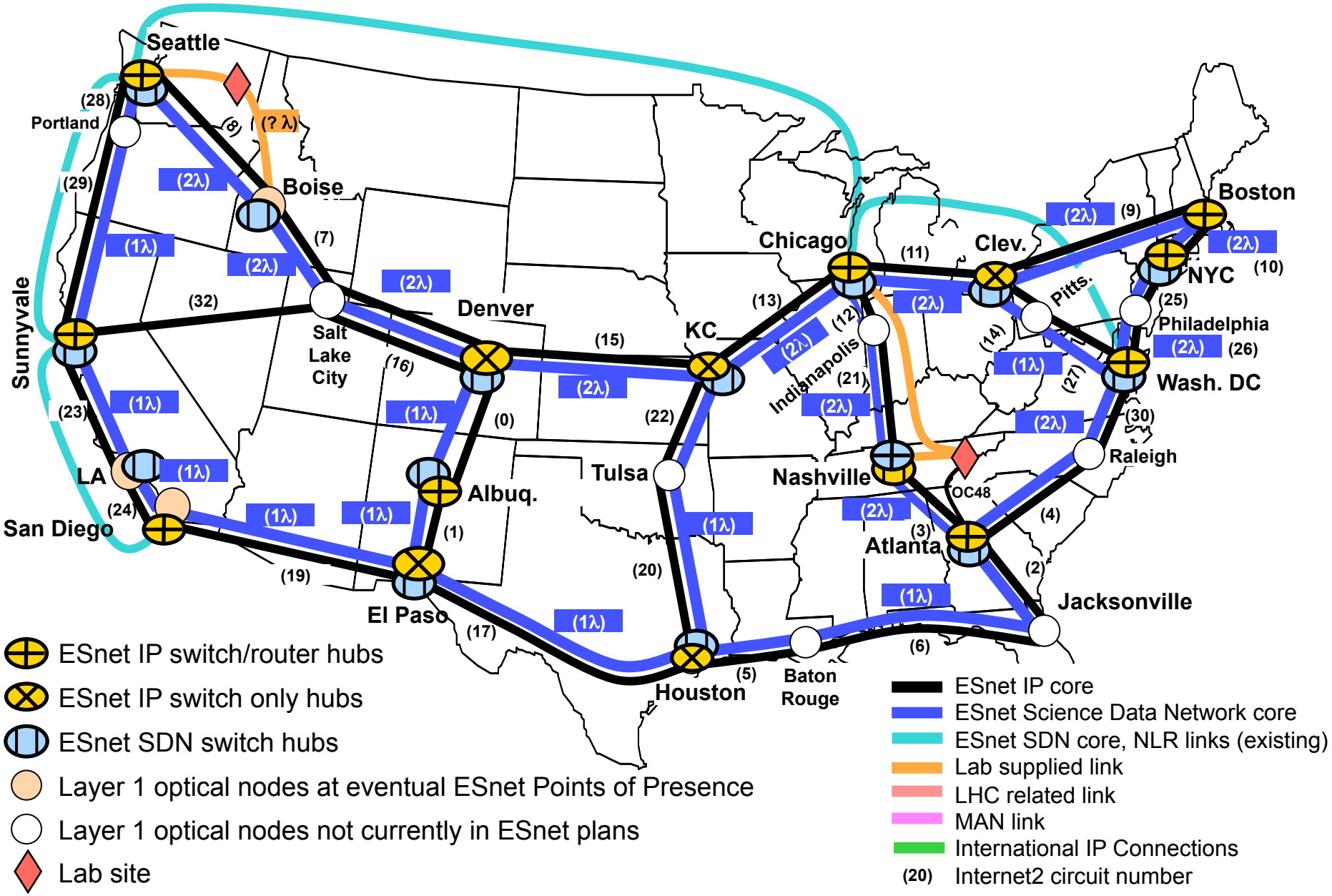
Are These Estimates Realistic? YES!



FNAL Outbound CMS Traffic

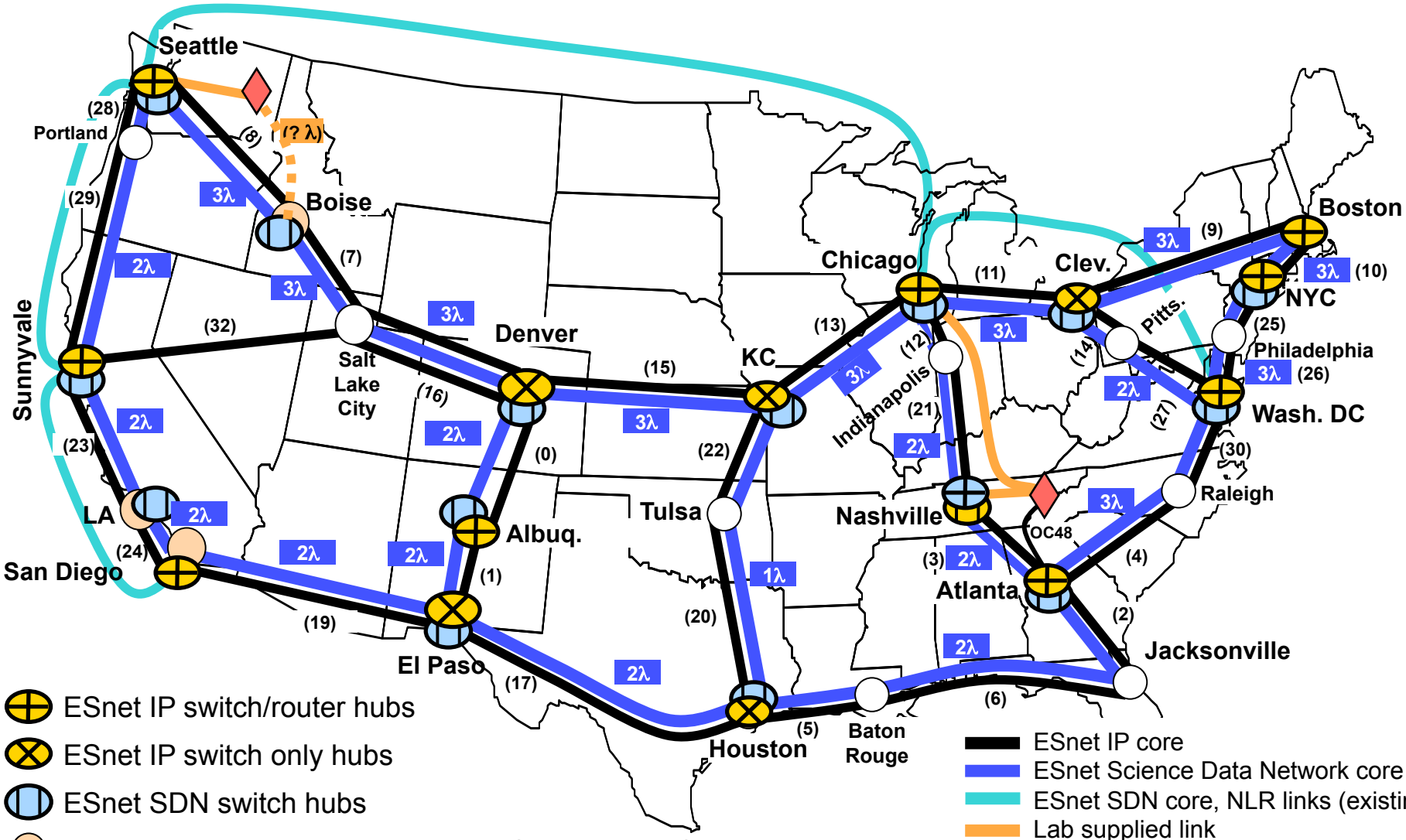
Max= 1064 MBy/s (8.5 Gb/s), Average = 394 MBy/s (3.2 Gb/s)

ESnet4 IP + SDN, 2008 Configuration



Estimated ESnet4 2009 Configuration

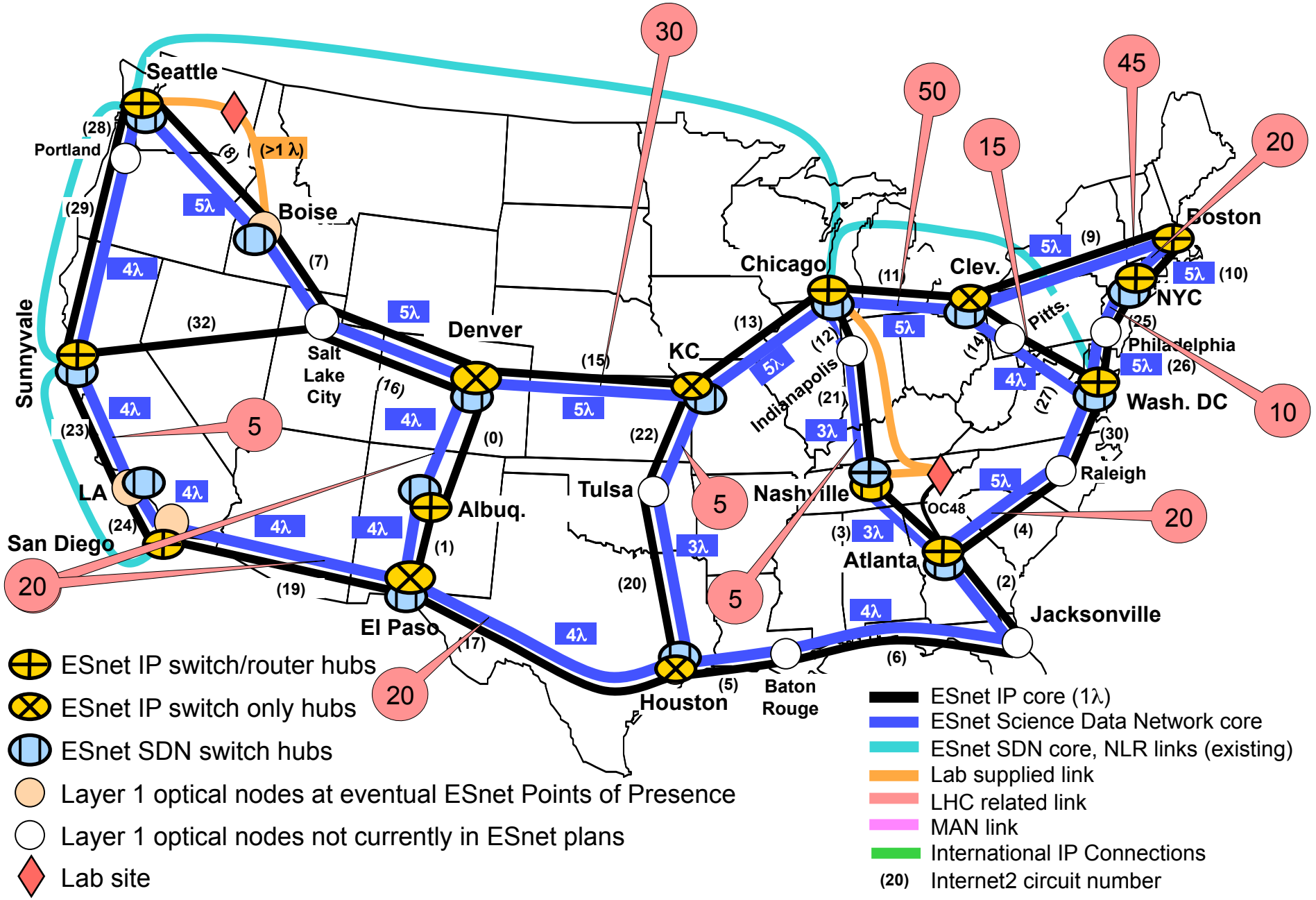
(Some of the circuits may be allocated dynamically from shared a pool.)



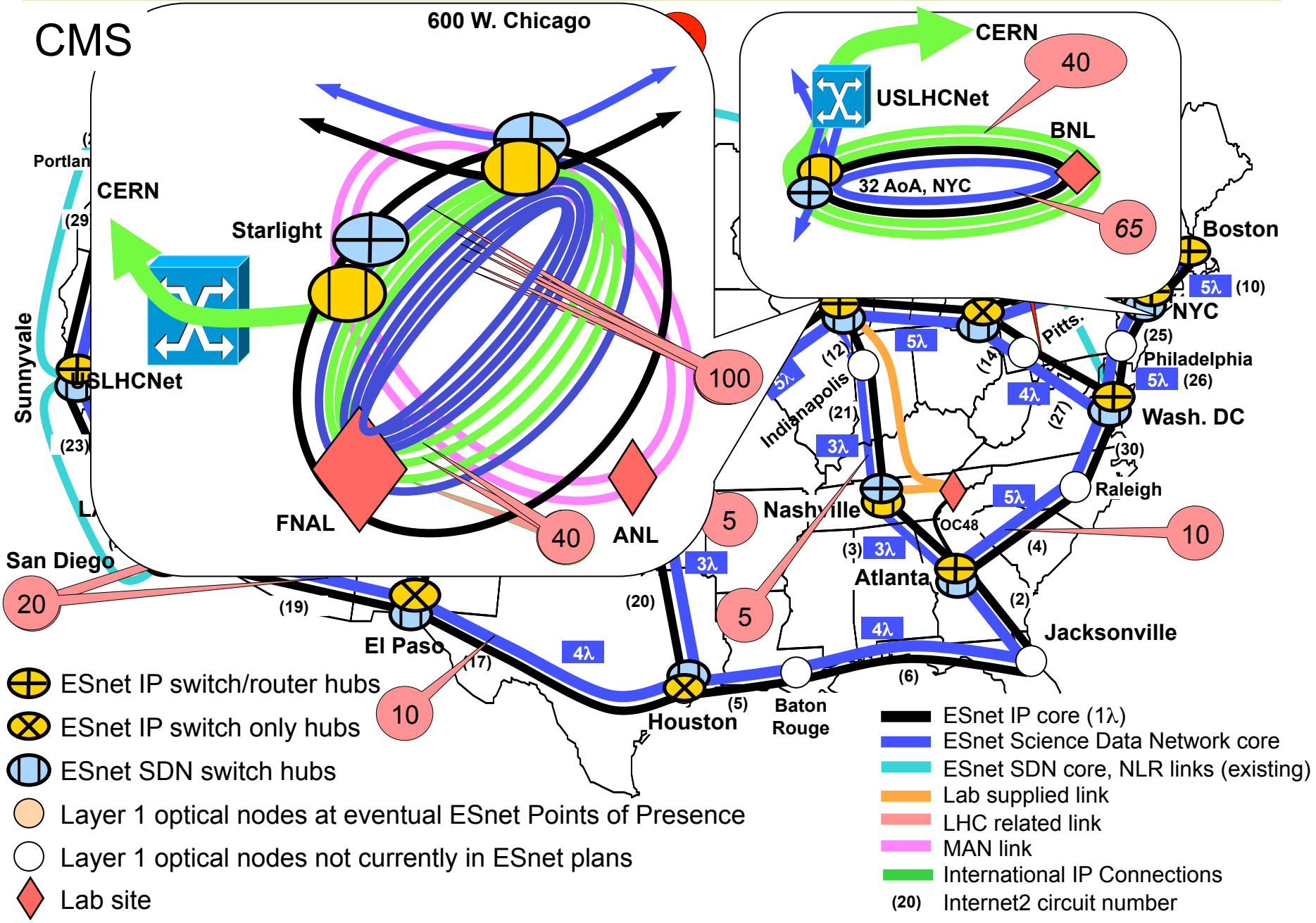
- ESnet IP switch/router hubs
- ESnet IP switch only hubs
- ESnet SDN switch hubs
- Layer 1 optical nodes at eventual ESnet Points of Presence
- Layer 1 optical nodes not currently in ESnet plans
- Lab site

- ESnet IP core
- ESnet Science Data Network core
- ESnet SDN core, NLR links (existing)
- Lab supplied link
- LHC related link
- MAN link
- International IP Connections
- (20) Internet2 circuit number

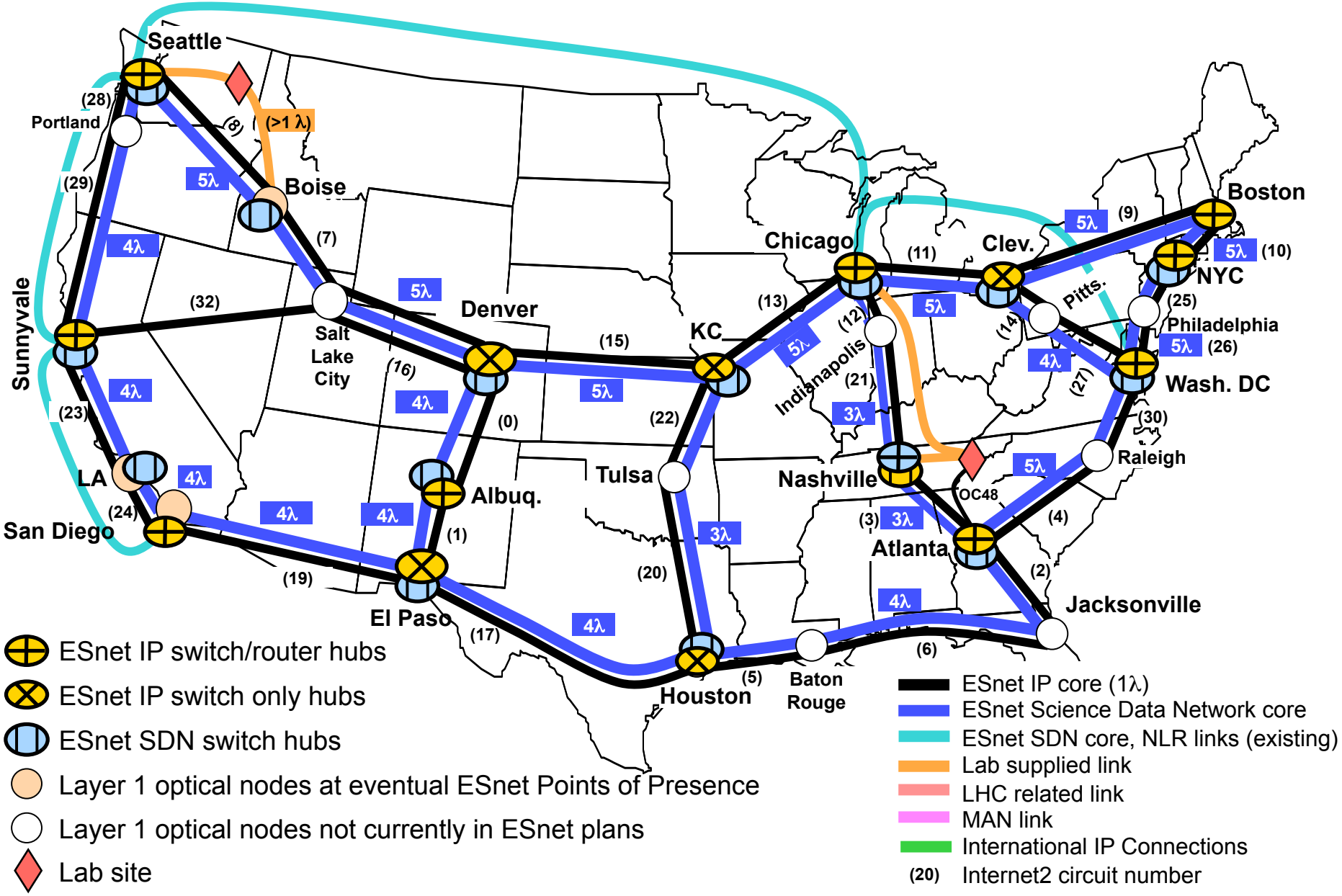
(2C) Aggregate Estimated Link Loadings, 2010-11



(2C) ESnet4 2010-11 Estimated Bandwidth Commitments

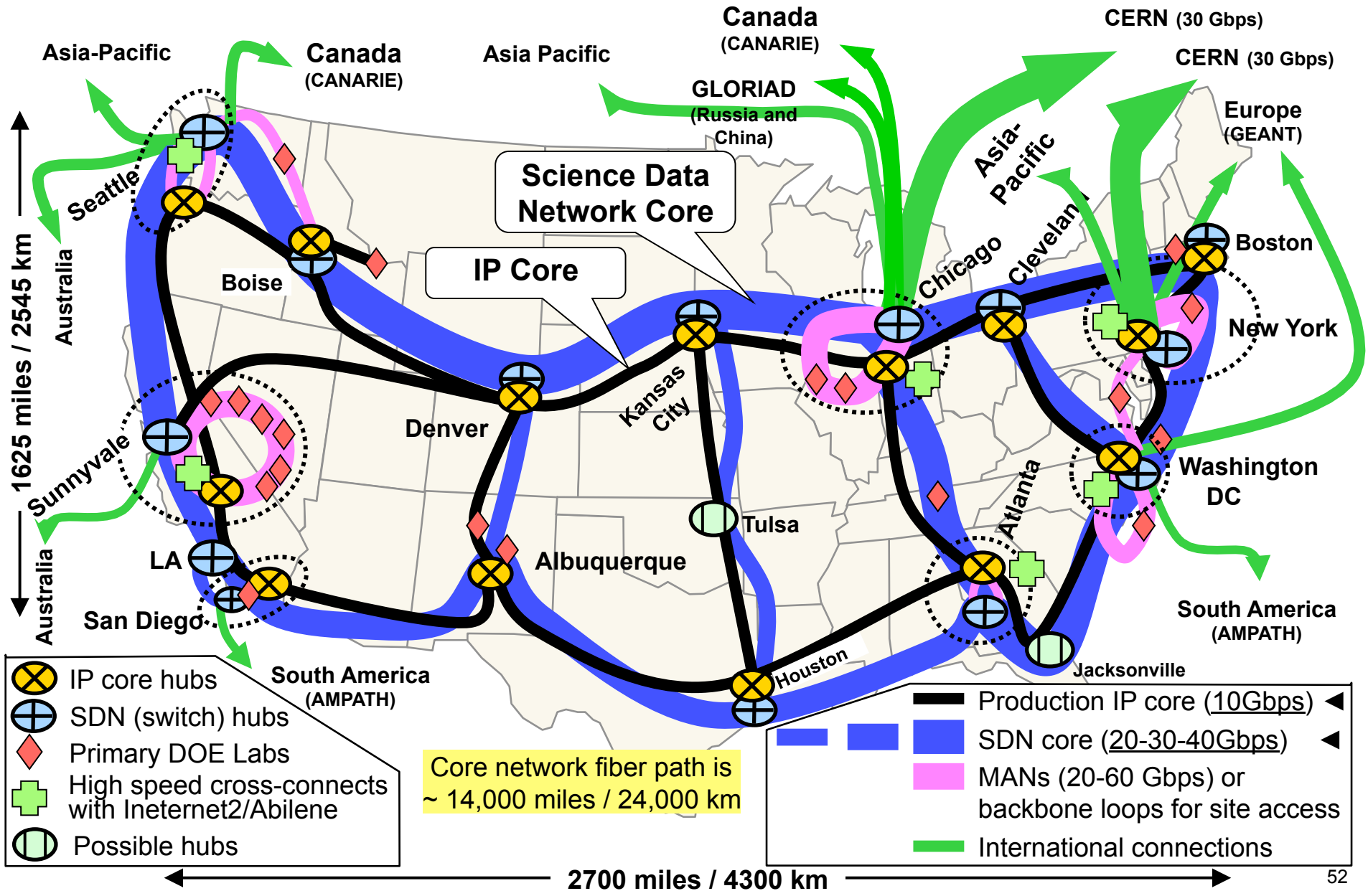


ESnet4 IP + SDN, 2011 Configuration



ESnet4 Planed Configuration

Core networks: 40-50 Gbps in 2009-2010, 160-400 Gbps in 2011-2012



➢ESnet Virtual Circuit Service

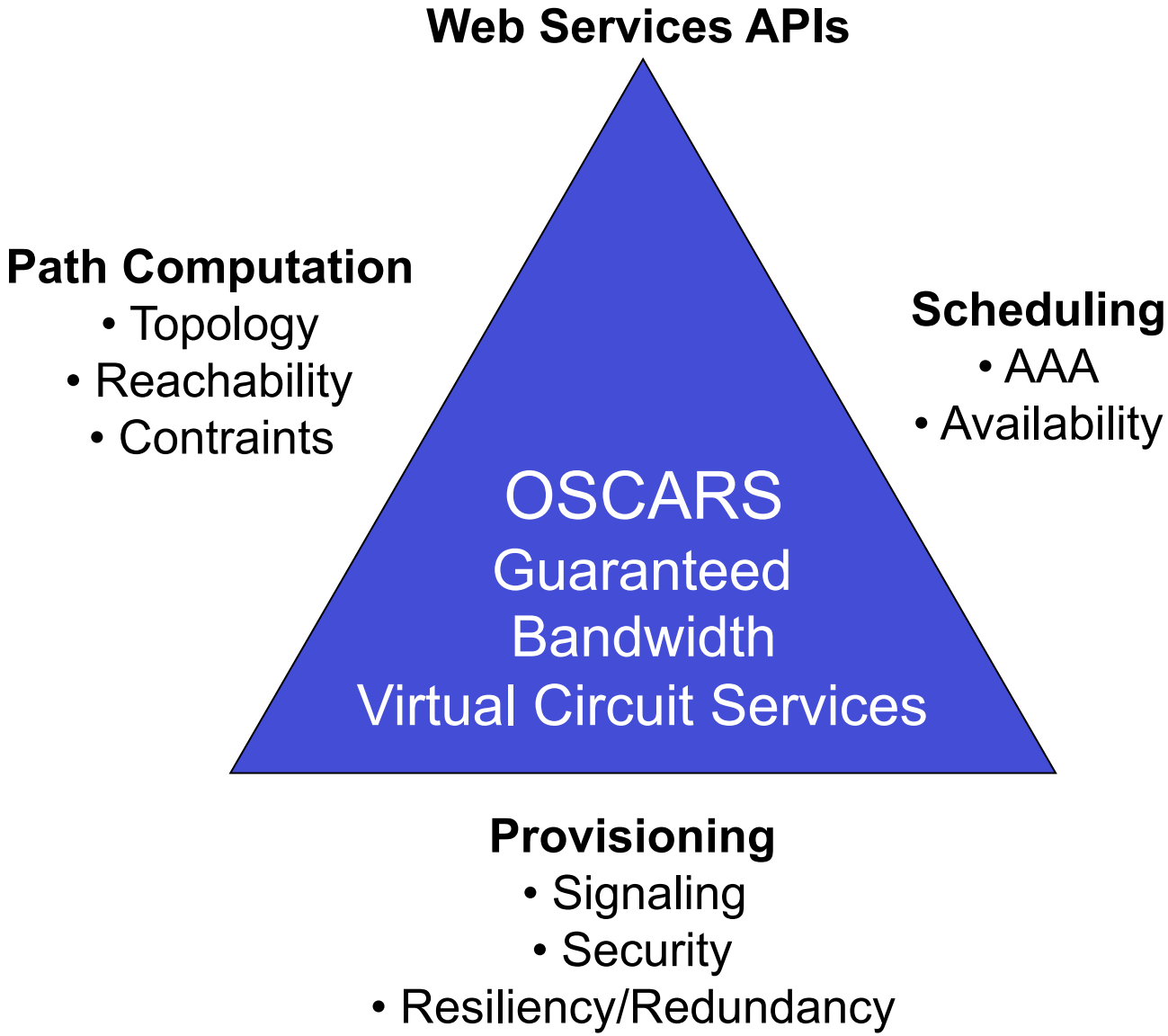
- Traffic isolation and traffic engineering
  - Provides for high-performance, non-standard transport mechanisms that cannot co-exist with commodity TCP-based transport
  - Enables the engineering of explicit paths to meet specific requirements
    - e.g. bypass congested links, using lower bandwidth, lower latency paths
- Guaranteed bandwidth (Quality of Service (QoS))
  - User specified bandwidth
  - Addresses deadline scheduling
    - Where fixed amounts of data have to reach sites on a fixed schedule, so that the processing does not fall far enough behind that it could never catch up – very important for experiment data analysis
- Secure
  - The circuits are “secure” to the edges of the network (the site boundary) because they are managed by the control plane of the network which is isolated from the general traffic
- Provides end-to-end connections between Labs and collaborator institutions

Virtual Circuit Service Functional Requirements

- Support user/application VC reservation requests
  - Source and destination of the VC
  - Bandwidth, start time, and duration of the VC
  - Traffic characteristics (e.g. flow specs) to identify traffic designated for the VC
- Manage allocations of scarce, shared resources
  - Authentication to prevent unauthorized access to this service
  - Authorization to enforce policy on reservation/provisioning
  - Gathering of usage data for accounting
- Provide circuit setup and teardown mechanisms and security
  - Widely adopted and standard protocols (such as MPLS and GMPLS) are well understood within a single domain
  - Cross domain interoperability is the subject of ongoing, collaborative development
  - secure and-to-end connection setup is provided by the network control plane
- Enable the claiming of reservations
  - Traffic destined for the VC must be differentiated from “regular” traffic
- Enforce usage limits
  - Per VC admission control polices usage, which in turn facilitates guaranteed bandwidth
  - Consistent per-hop QoS throughout the network for transport predictability

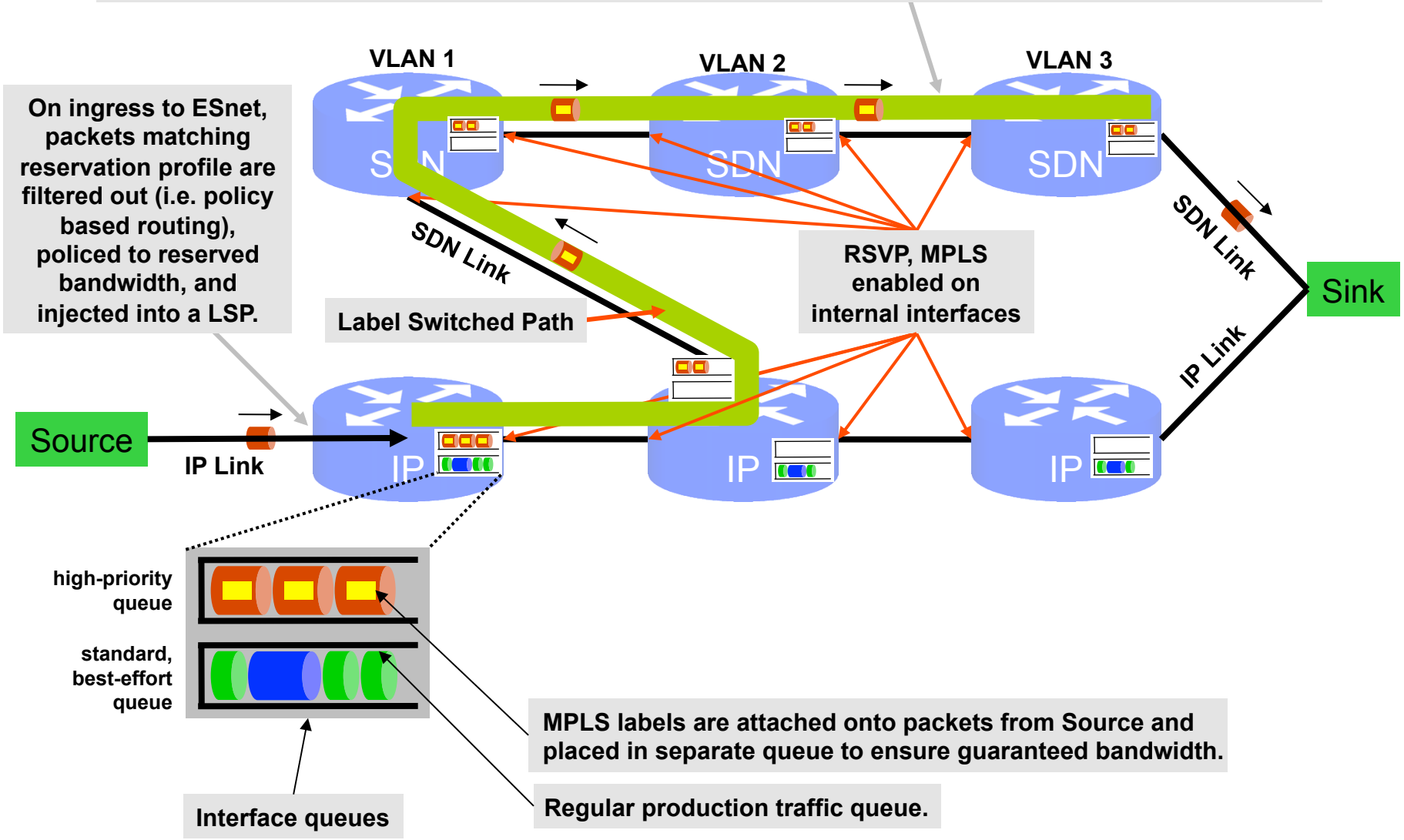
OSCARS Overview

On-demand Secure Circuits and Advance Reservation System



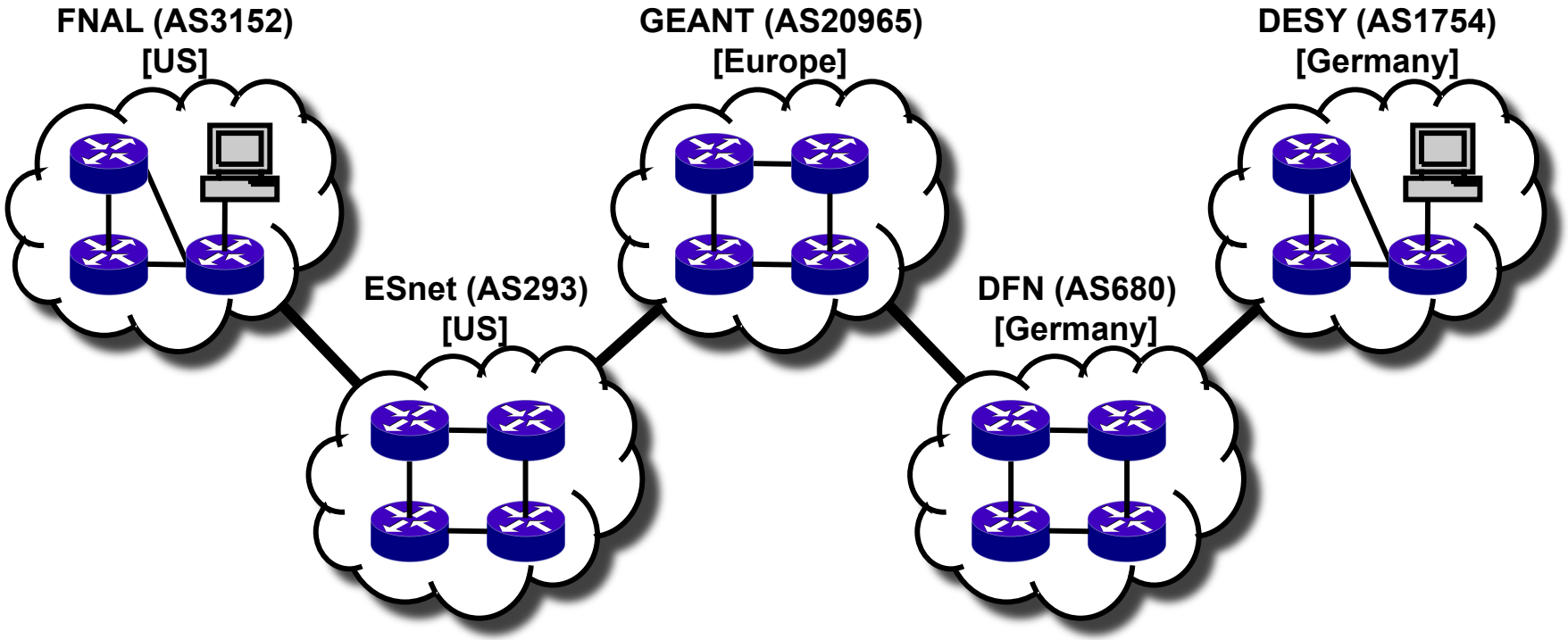
The Mechanisms Underlying OSCARS

Based on Source and Sink IP addresses, route of LSP between ESnet border routers is determined using topology information from OSPF-TE. Path of LSP can be explicitly directed to take SDN network. On the SDN Ethernet switches all traffic is MPLS switched (layer 2.5), which stitches together VLANs

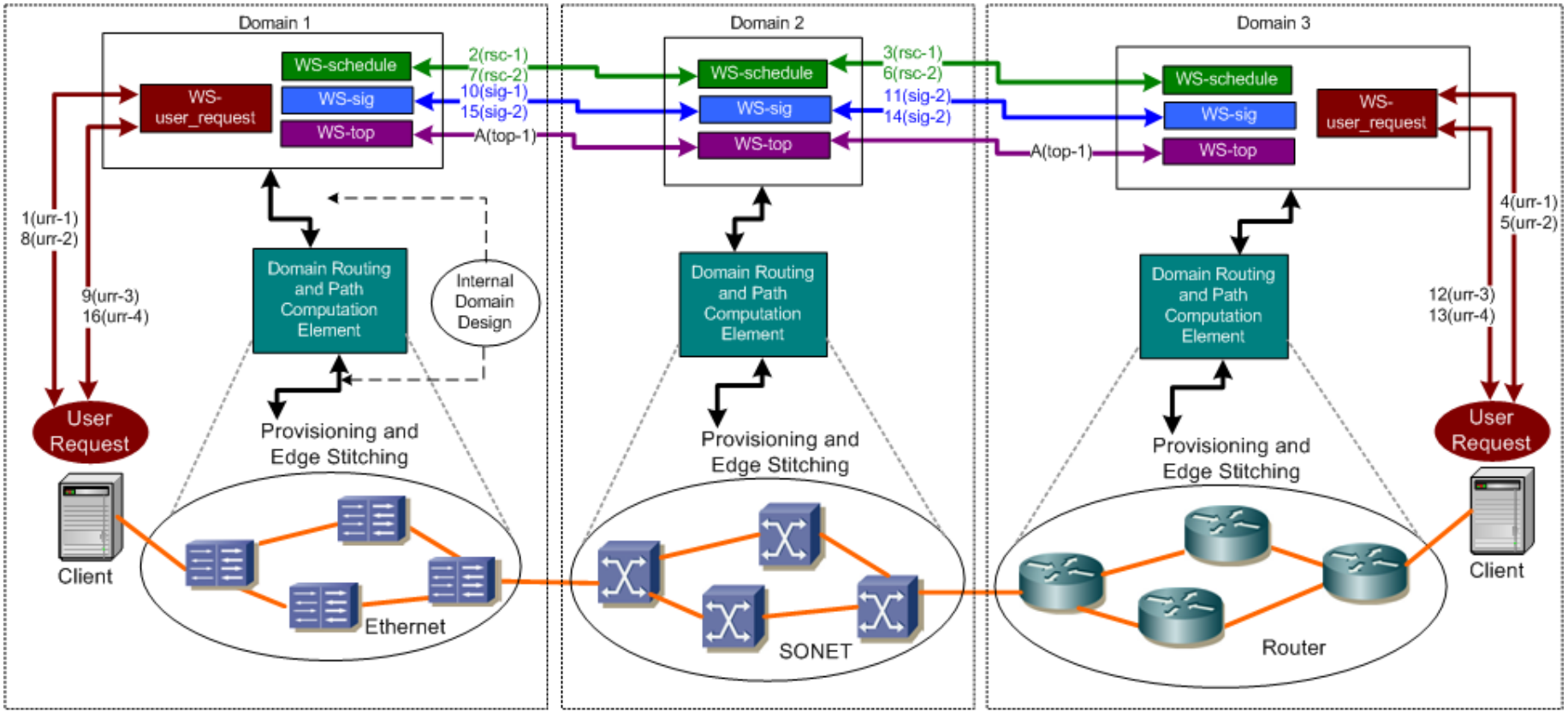


Environment of Science is Inherently Multi-Domain

- End points will be at independent institutions – campuses or research institutes - that are served by ESnet, Abilene, GÉANT, and their regional networks
  - Complex inter-domain issues – typical circuit will involve five or more domains - of necessity this involves collaboration with other networks
  - For example, a connection between FNAL and DESY involves five domains, traverses four countries, and crosses seven time zones



Interdomain Virtual Circuit Reservation Control Flow



Topology Exchange
top-1: getNetworkTopology

User Request/Response
urr-1: createReservation
urr-2: createReservationResponse
urr-3: createPath
urr-4: createPathResponse

Resource Scheduling
rsc-1: createReservation
rsc-2: createReservationResponse

Signaling
sig-1: createPath
sig-2: createPathReponse

Progress!

OSCARS Status Update

- ESnet Centric Deployment
  - Prototype layer 3 (IP) guaranteed bandwidth virtual circuit service deployed in ESnet (1Q05)
  - Layer 2 (Ethernet VLAN) virtual circuit service under development
- Inter-Domain Collaborative Efforts
  - Terapaths
    - Inter-domain interoperability for layer 3 virtual circuits demonstrated (3Q06)
    - Inter-domain interoperability for layer 2 virtual circuits under development
  - HOPI/DRAGON
    - Inter-domain exchange of control messages demonstrated (1Q07)
    - Initial integration of OSCARS and DRAGON has been successful (1Q07)
  - DICE
    - First draft of topology exchange schema has been formalized (in collaboration with NMWG) (2Q07), interoperability test scheduled for 3Q07
    - Drafts on reservation and signaling messages under discussion
  - UVA
    - Integration of Token based authorization in OSCARS under discussion
- Measurements
  - Hybrid dataplane testing with ESnet, HOPI/DRAGON, USN, and Tennessee Tech (1Q07)
- Administrative
  - Vangelis Haniotakis (GRNET) has taken a one-year sabbatical position with ESnet to work on interdomain topology exchange, resource scheduling, and signalling

➢ Monitoring Applications Move Networks Toward Service-Oriented Communications Services

- perfSONAR is a global collaboration to design, implement and deploy a network measurement framework.
  - Web Services based Framework
    - Measurement Archives (MA)
    - Measurement Points (MP)
    - Lookup Service (LS)
    - Topology Service (TS)
    - Authentication Service (AS)
  - Some of the currently Deployed Services
    - Utilization MA
    - Circuit Status MA & MP
    - Latency MA & MP
    - Bandwidth MA & MP
    - Looking Glass MP
    - Topology MA
  - This is an Active Collaboration
    - The basic framework is complete
    - Protocols are being documented
    - New Services are being developed and deployed.

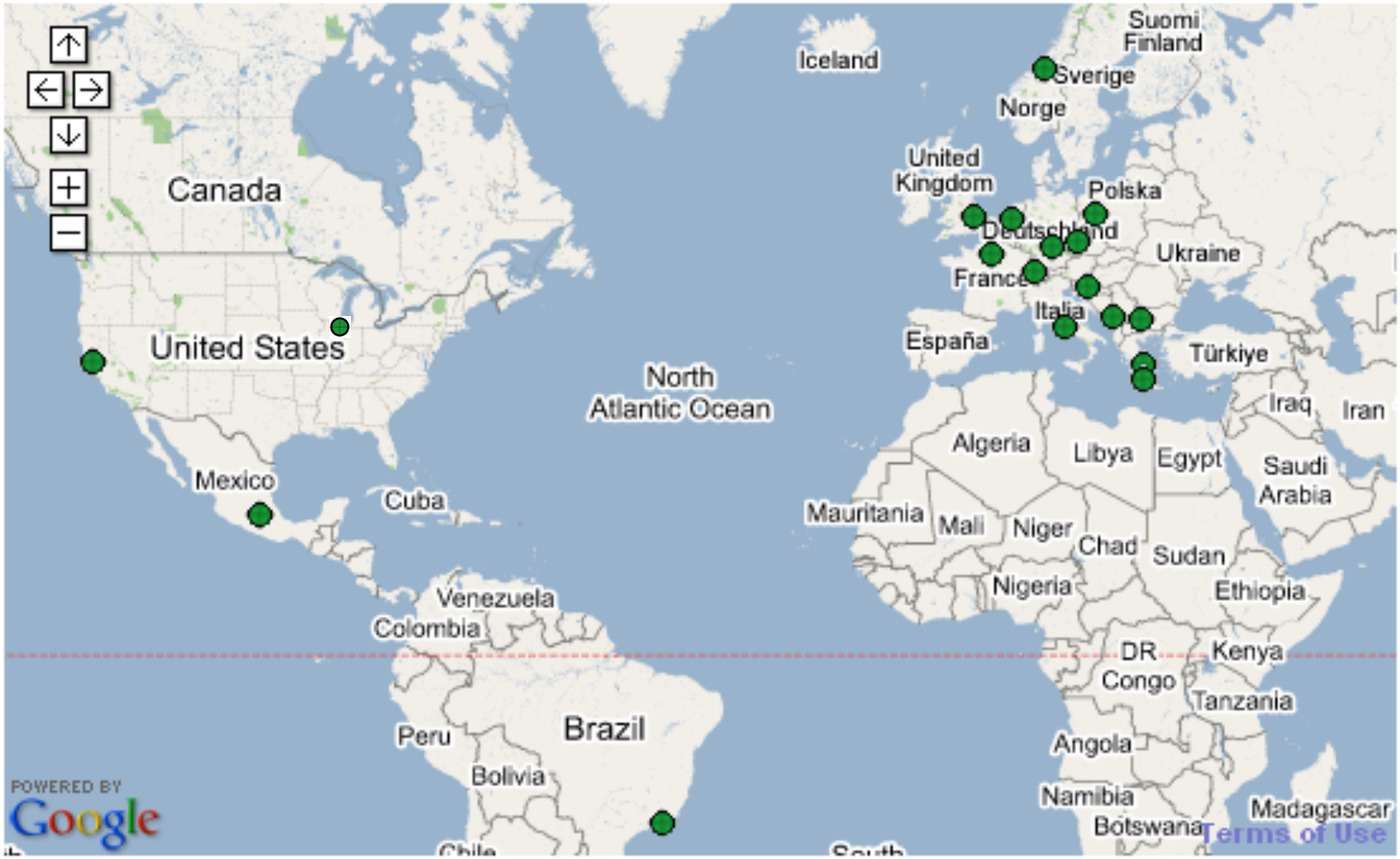
perfSONAR Collaborators

- ARNES
- Belnet
- CARnet
- CESnet
- Dante
- University of Delaware
- DFN
- ESnet
- FCCN
- FNAL
- GARR
- GEANT2
- Georga Tech
- GRNET
- Internet2
- IST
- POZNAN Supercomputing Center
- Red IRIS
- Renater
- RNP
- SLAC
- SURFnet
- SWITCH
- Uninett

* Plus others who are contributing, but haven't added their names to the list on the WIKI.

perfSONAR Deployments

16+ different networks have deployed at least 1 perfSONAR service (Jan 2007)



❖ E2Emon - a perfSONAR application

- E2Emon provides end-to-end path status in a service-oriented, easily interpreted way
  - a perfSONAR application used to monitor the LHC paths end-to-end across many domains
  - uses perfSONAR protocols to retrieve current circuit status every minute or so from MAs and MPs in all the different domains supporting the circuits
  - is itself a service that produces Web based, real-time displays of the overall state of the network, and it generates alarms when one of the MP or MA's reports link problems.

E2Emon: Status of E2E link CERN-LHCOPN-FNAL-001

Oper. State: Up
Admin. State: Normal Oper.

Domain	CERN		USLHCNET				
Link Structure	EP	◄·····	·····►	DP	◄——►	DP	◄·····
Type	EndPoint	ID Part.Info	ID Part.Info	Demarc	Domain Link	Demarc	ID Part.Info
Local Name	**CERN-T0**	S513-C-BE1	CERN-FERMI-LHCOPN-001-GVA-CERN	**USLHCNET-GEN**	CERN-FERMI-LHCOPN-001-GVA-CHI	**USLHCNET-CHI**	CERN-FERMI-LHCOPN-001-CHI-ESNET
State Oper.	-	Up	Up	-	Up	-	Up
State Admin.	-	Normal Oper.	Normal Oper.	-	Normal Oper.	-	Normal Oper.
Timestamp	-	2007-04-08 T05:04:08+02:00	2007-04-08 T05:04:11+02:00	-	2007-04-08 T05:04:53+02:00	-	2007-04-08 T05:03:59+02:00

Page generat

ESNET					FERMI			
·····►	DP	◄——►	DP	◄·····	·····►	DP	◄——►	EP
ID Part.Info	Demarc	Domain Link	Demarc	ID Part.Info	ID Part.Info	Demarc	Domain Link	EndPoint
CERN-FERMI-LHCOPN-001-STARLIGHT-Tail	**ESNET-STARLIGHT**	CERN-FERMI-LHCOPN-001-FERMI-STARLIGHT	**ESNET-FERMI**	CERN-FERMI-LHCOPN-001-Site-Tail	md8	**FERMI-ESNET**	md2	**FERMI-T1**
Up	-	Up	-	Up	Up	-	Up	-
Normal Oper.	-	Normal Oper.	-	Normal Oper.	Normal Oper.	-	Normal Oper.	-
2007-04-08 T01:40:37.0	-	2007-04-08T01:40:37.0	-	2007-04-08 T01:40:37.0	2007-04-08 T01:40:01.0-6:00	-	2007-04-08 T01:40:01.0-6:00	-

E2Emon generated view of the data for one OPN link [E2EMON]

❖ Path Performance Monitoring

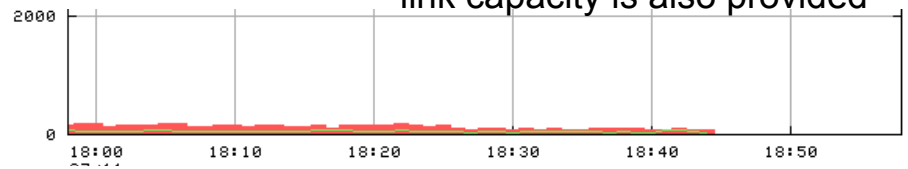
- Path performance monitoring needs to provide users/applications with the end-to-end, multi-domain traffic and bandwidth availability
  - should also provide real-time performance such as path utilization and/or packet drop
- Multiple path performance monitoring tools are in development
  - One example – Traceroute Visualizer [TrViz] ,developed by Joe Metzger, ESnet – has been deployed at about 10 R&E networks in the US and Europe that have at least some of the required perfSONAR MA services to support the tool

Traceroute Visualizer

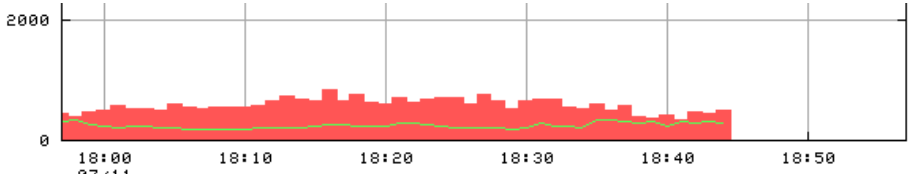
- Forward direction bandwidth utilization on application path from LBNL to INFN-Frascati (Italy)
  - traffic shown as bars on those network device interfaces that have an associated MP services (the first 4 graphs are normalized to 2000 Mb/s, the last to 500 Mb/s)

1 ir1000gw (131.243.2.1)
2 er1kgw
3 lbl2-ge-lbnl.es.net

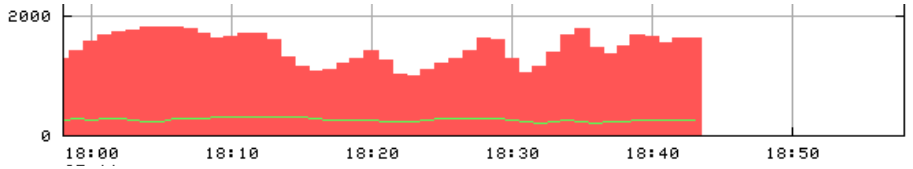
link capacity is also provided



4 slacmr1-sdn-lblmr1.es.net (GRAPH OMITTED)
5 snv2mr1-slacmr1.es.net (GRAPH OMITTED)
6 snv2sdn1-snv2mr1.es.net

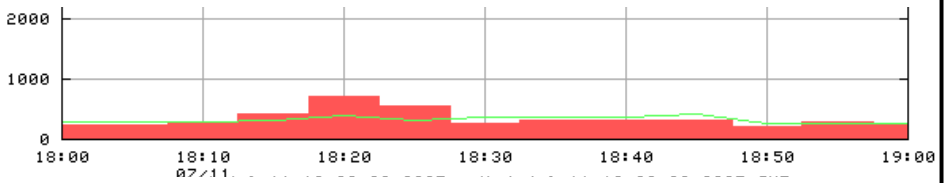


7 chislsdn1-oc192-snv2sdn1.es.net (GRAPH OMITTED)
8 chiccr1-chislsdn1.es.net

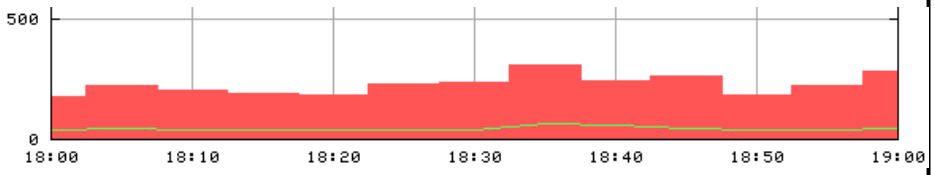


9 aofacr1-chicsdn1.es.net (GRAPH OMITTED)

10 esnet.rt1.nyc.us.geant2.net (NO DATA)
11 so-7-0-0.rt1.ams.nl.geant2.net (NO DATA)
12 so-6-2-0.rt1.fra.de.geant2.net (NO DATA)
13 so-6-2-0.rt1.gen.ch.geant2.net (NO DATA)
14 so-2-0-0.rt1.mil.it.geant2.net (NO DATA)
15 garr-gw.rt1.mil.it.geant2.net (NO DATA)
16 rt1-mi1-rt-mi2.mi2.garr.net



17 rt-mi2-rt-rm2.rm2.garr.net (GRAPH OMITTED)
18 rt-rm2-rc-fra.fra.garr.net (GRAPH OMITTED)
19 rc-fra-ru-lnf.fra.garr.net (GRAPH OMITTED)

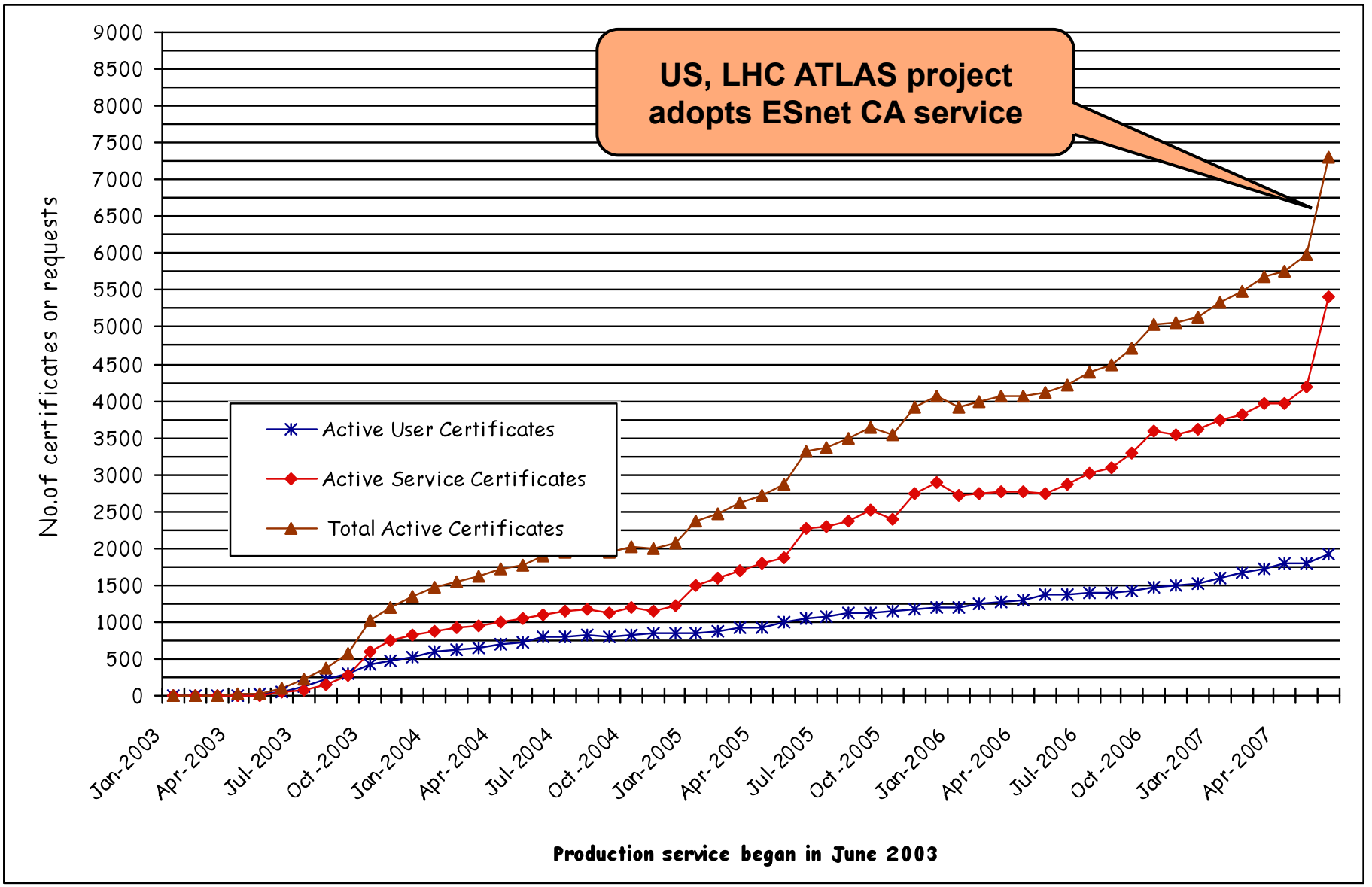


20
21 www6.lnf.infn.it (193.206.84.223) 189.908 ms 189.596 ms 189.684 ms

III. Federated Trust Services – Support for Large-Scale Collaboration

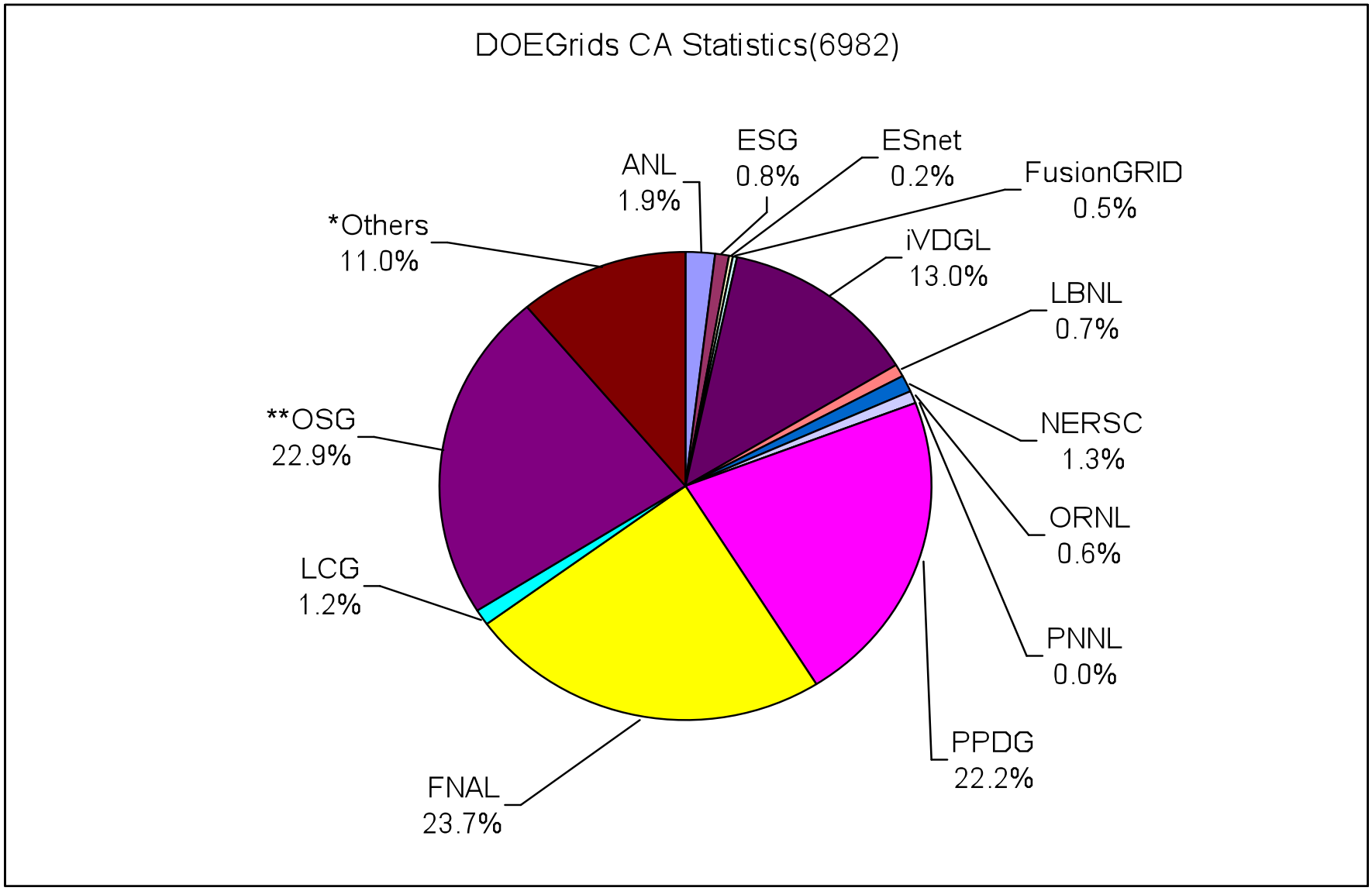
- Remote, multi-institutional, identity authentication is critical for distributed, collaborative science in order to permit sharing widely distributed computing and data resources, and other Grid services
- Public Key Infrastructure (PKI) is used to formalize the existing web of trust within science collaborations and to extend that trust into cyber space
  - The function, form, and policy of the ESnet trust services are driven entirely by the requirements of the science community and by direct input from the science community
- International scope trust agreements that encompass many organizations are crucial for large-scale collaborations
  - ESnet has lead in negotiating and managing the cross-site, cross-organization, and international trust relationships to provide policies that are tailored for collaborative science
    - ➢ This service, together with the associated ESnet PKI service, is the basis of the routine sharing of HEP Grid-based computing resources between US and Europe

DOEGrids CA (Active Certificates) Usage Statistics



* Report as of July 5, 2007

DOEGrids CA Usage - Virtual Organization Breakdown



* DOE-NSF collab. & Auto renewals

** OSG Includes (BNL, CDF, CMS, CompBioGrid,DES, DOSAR, DZero, Engage, Fermilab, fMRI, GADU, geant4, GLOW, GPN, GRASE, GridEx, GROW, GUGrid, i2u2, iVDGL, JLAB, LIGO, mariachi, MIS, nanoHUB, NWICG, OSG, OSGEDU, SBGrid,SDSS, SLAC, STAR & USATLAS)

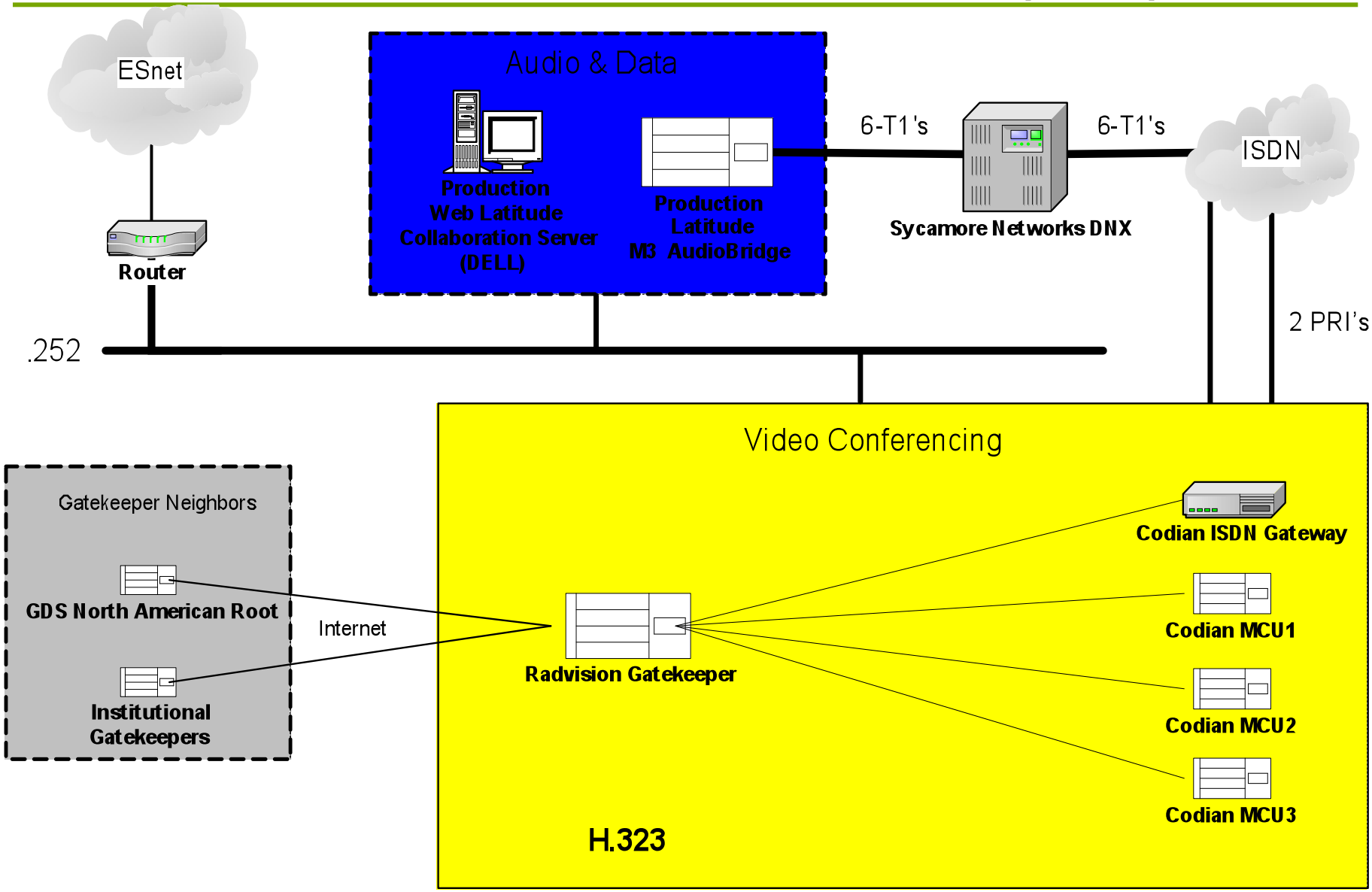
DOEGrids CA Adopts Red Hat CS 7.1

- Motivation: SunOne CMS (Certificate Management System) went End-of-Life 2 years ago
  - RH CS is a continuation of the original CA product and development team, and is fully supported by Red Hat
- Transition was over a year in negotiation, development, and testing
- 05 July 2007: Transition from SunONE CMS 4.7 to Red Hat
- Major transition - Minimal outage of 7 hours
- Preserved important assets
  - ***Existing DOEGrids signing key***
  - Entire history: over 35000 data objects transferred
  - UI (for subscriber-users and operators)
- New CA software will allow ESnet to develop more useful applications and interfaces for the user communities

IV. ESnet Conferencing Service (ECS)

- An ESnet Science Service that provides audio, video, and data teleconferencing service to support human collaboration of DOE science
  - Seamless voice, video, and data teleconferencing is important for geographically dispersed scientific collaborators
  - Provides the central scheduling essential for global collaborations
  - ESnet serves more than a thousand DOE researchers and collaborators worldwide
    - H.323 (IP) videoconferences (4000 port hours per month and rising)
    - audio conferencing (2500 port hours per month) (constant)
    - data conferencing (150 port hours per month)
    - Web-based, automated registration and scheduling for all of these services
  - Very cost effective (saves the Labs a lot of money)

ESnet Collaboration Services (ECS)



ECS Video Collaboration Service

- High Quality videoconferencing over IP and ISDN
- Reliable, appliance based architecture
- Ad-Hoc H.323 and H.320 multipoint meeting creation
- Web Streaming options on 3 Codian MCU's using Quicktime or Real
- 3 Codian MCUs with Web Conferencing Options
- 120 total ports of video conferencing on each MCU (40 ports per MCU)
- 384k access for video conferencing systems using ISDN protocol
- Access to audio portion of video conferences through the Codian ISDN Gateway

ECS Voice and Data Collaboration

- 144 usable ports
  - Actual conference ports readily available on the system.
- 144 overbook ports
  - Number of ports reserved to allow for scheduling beyond the number of conference ports readily available on the system.
- 108 Floater Ports
  - Designated for unexpected port needs.
  - Floater ports can float between meetings, taking up the slack when an extra person attends a meeting that is already full and when ports that can be scheduled in advance are not available.
- Audio Conferencing and Data Collaboration using Cisco MeetingPlace
- Data Collaboration = WebEx style desktop sharing and remote viewing of content
- Web-based user registration
- Web-based scheduling of audio / data conferences
- Email notifications of conferences and conference changes
- 650+ users registered to schedule meetings (not including guests)

ECS Service Level

- ESnet Operations Center is open for service 24x7x365.
- A trouble ticket is opened within15 to 30 minutes and assigned to the appropriate group for investigation.
- Trouble ticket is closed when the problem is resolved.
- ***ECS support*** is provided Monday to Friday, 8AM to 5 PM Pacific Time excluding LBNL holidays
  - Reported problems are addressed within 1 hour from receiving a trouble ticket during ECS support period
  - ESnet does NOT provide a real time (during-conference) support service

Typical Problems Reported to ECS Support

Video Conferencing

- User E.164 look up
- Gatekeeper registration problems – forgotten IP address or user network problems
- Gateway Capacity for ISDN service expanded to 2 full PRI's = 46 x 64kbps chs
- For the most part, problems are with user-side network and systems configuration.

Voice and Data Collaboration

- Scheduling Conflicts and Scheduling Capacity has been addressed by expanding overbooking capacity to 100% of actual capacity
  - Future equipment plans will allow for optimal configuration of scheduling parameters.
- Browser compatibility with Java based data sharing client – users are advised to test before meetings
- Lost UserID and/or passwords

Words of Wisdom

We advise users that at least two actions must be taken in advance of conferences to reduce the likelihood of problems:

A) testing of the configuration to be used for the audio, video and data conference.

B) appropriate setup time must be allocated BEFORE the conference to ensure punctuality and correct local configuration. (at least 15 min recommended)

Real Time ECS Support

- A number of user groups have requested “real-time” conference support (monitoring of conferences while in-session)
- Limited Human and Financial resources currently prohibit ESnet from:

  A) Making real time information available to the public on the systems status (network, ECS, etc) This information is available only on some systems to our support personnel

  B) 24x7x365 real-time support

  C) Addressing simultaneous trouble calls as in a real time support environment.

  - This would require several people addressing multiple problems simultaneously

Real Time ECS Support

- Proposed solution
  - A fee-for-service arrangement for real-time conference support
  - Such an arrangement could be made by ***contracting directly with TKO Video Communications, ESnet's ECS service provider***
  - Service offering would provide:
    - Testing and configuration assistance prior to your conference
    - Creation and scheduling of your conferences on ECS Hardware
    - Preferred port reservations on ECS video and voice systems
    - Connection assistance and coordination with participants
    - Endpoint troubleshooting
    - Live phone support during conferences
    - Seasoned staff and years of experience in the video conferencing industry
    - ESnet community pricing at $xxx per hour (Commercial Price: $yyy/hr)

➢Summary

- **ESnet is currently satisfying its mission by enabling SC science that is dependant on networking and distributed, large-scale collaboration:**

  "The performance of ESnet over the past year has been excellent, with only minimal unscheduled down time. The reliability of the core infrastructure is excellent. Availability for users is also excellent" - DOE 2005 annual review of LBL

- **ESnet has put considerable effort into gathering requirements from the DOE science community, and has a forward-looking plan and expertise to meet the five-year SC requirements**
  - **A Lehman review of ESnet (Feb, 2006) has strongly endorsed the plan presented here**

References

1. High Performance Network Planning Workshop, August 2002
   - http://www.doecollaboratory.org/meetings/hpnpw
2. Science Case Studies Update, 2006 (contact eli@es.net)
3. DOE Science Networking Roadmap Meeting, June 2003
   - http://www.es.net/hypertext/welcome/pr/Roadmap/index.html
4. Science Case for Large Scale Simulation, June 2003
   - http://www.pnl.gov/scales/
5. Planning Workshops-Office of Science Data-Management Strategy, March & May 2004
   - http://www-conf.slac.stanford.edu/dmw2004
6. For more information contact Chin Guok (chin@es.net). Also see
   - http://www.es.net/oscars

[LHC/CMS] http://cmsdoc.cern.ch/cms/aprom/phedex/prod/Activity::RatePlots?view=global

[ICFA SCIC] "Networking for High Energy Physics." International Committee for Future Accelerators (ICFA), Standing Committee on Inter-Regional Connectivity (SCIC), Professor Harvey Newman, Caltech, Chairperson.
- http://monalisa.caltech.edu:8080/Slides/ICFASCIC2007/

[E2EMON] Geant2 E2E Monitoring System –developed and operated by JRA4/WI3, with implementation done at DFN
http://cnmdev.lrz-muenchen.de/e2e/html/G2_E2E_index.html
http://cnmdev.lrz-muenchen.de/e2e/lhc/G2_E2E_index.html

[TrViz] ESnet PerfSONAR Traceroute Visualizer
https://performance.es.net/cgi-bin/level0/perfsonar-trace.cgi

And Ending on a Light Note….

NON SEQUITUR - BY WILEY