



ESnet4: Networking for the Future of Science

ASCAC
August 9, 2006

William E. Johnston

ESnet Department Head and Senior Scientist
Lawrence Berkeley National Laboratory



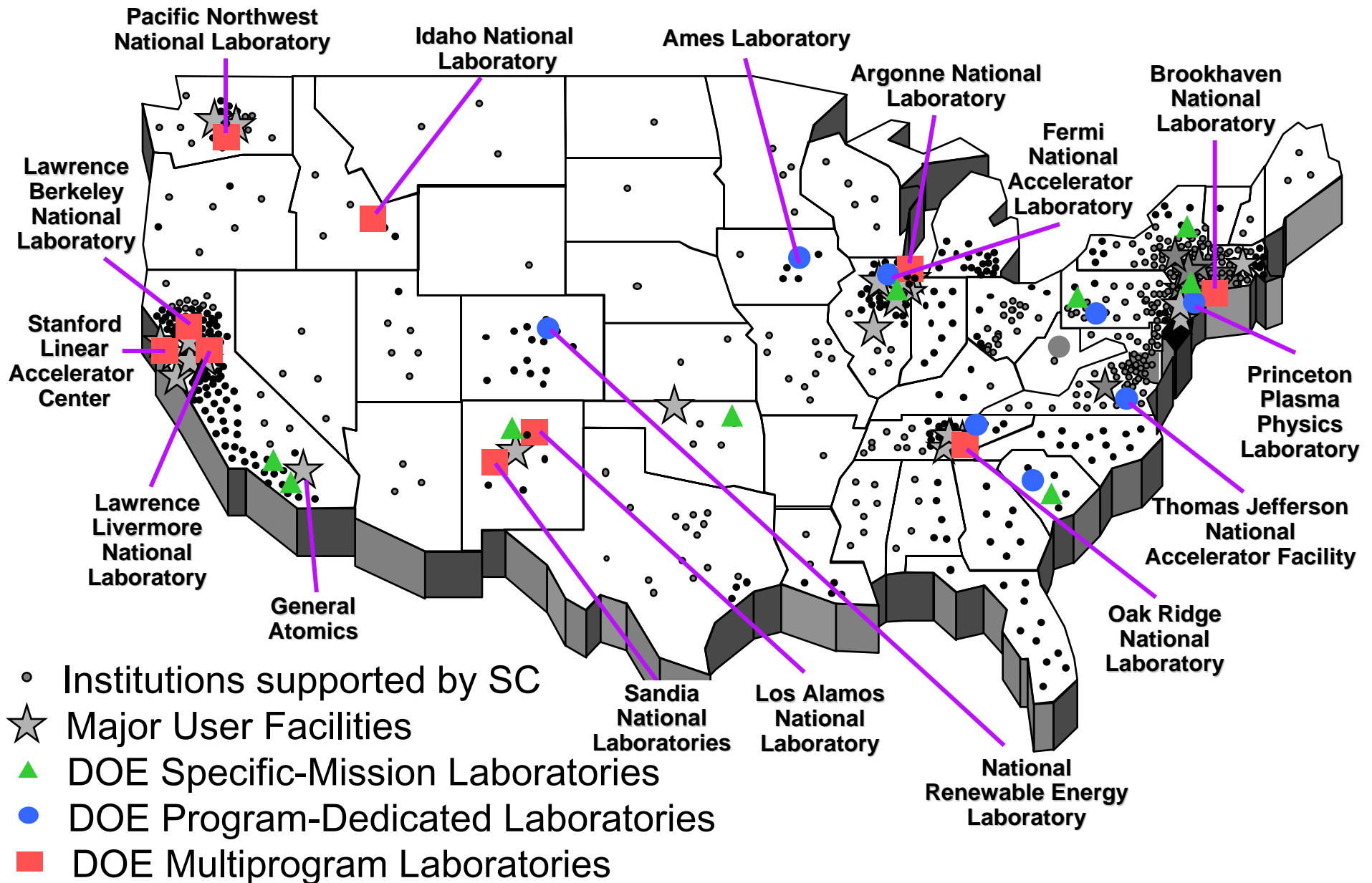
www.es.net



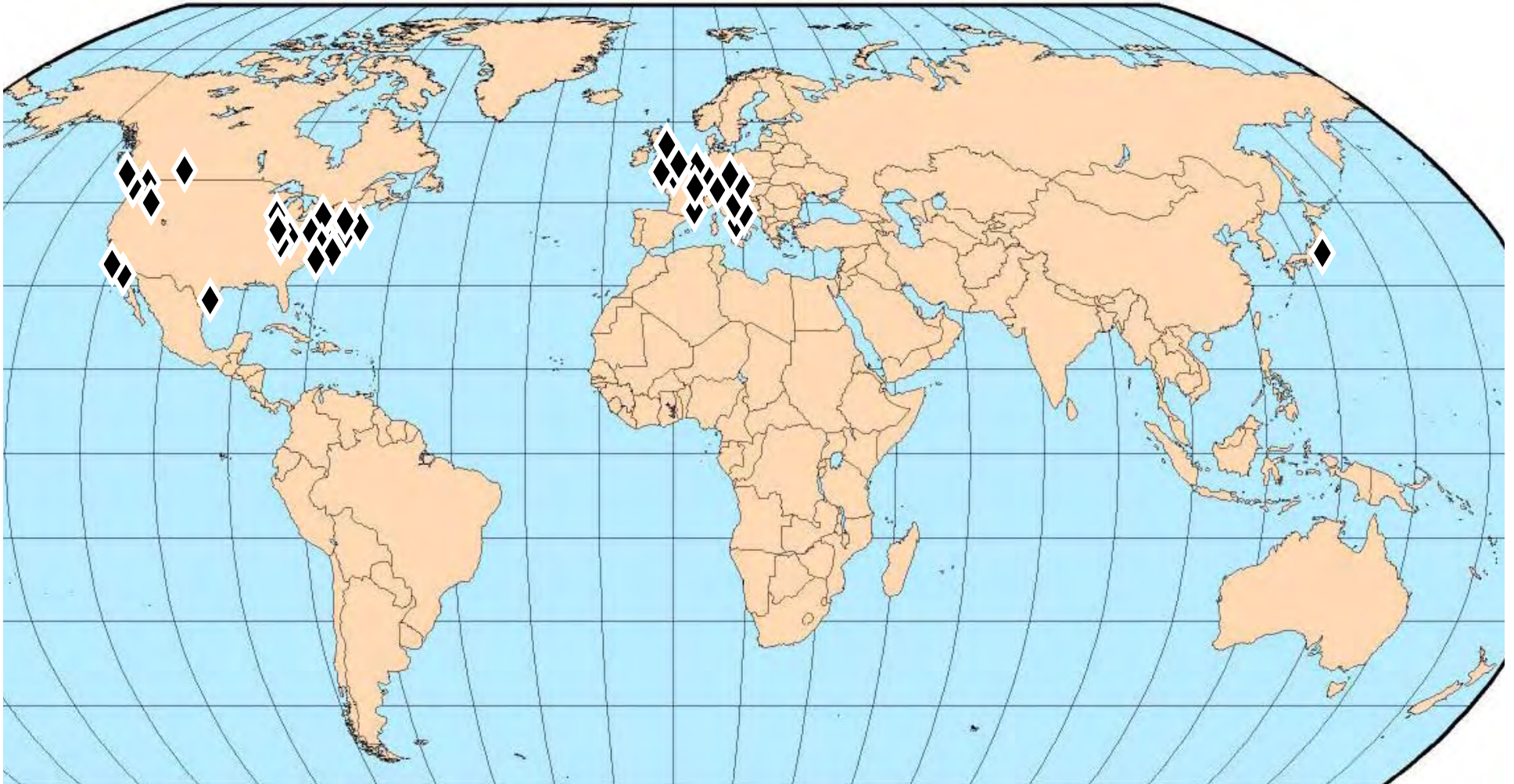
➤ DOE Office of Science and ESnet – the ESnet Mission

- **ESnet's primary mission is to enable the large-scale science that is the mission of the Office of Science (SC):**
 - Sharing of massive amounts of data
 - Supporting thousands of collaborators world-wide
 - Distributed data processing
 - Distributed data management
 - Distributed simulation, visualization, and computational steering
- ESnet also provides network and collaboration services to DOE laboratories and other DOE programs in cases where this is cost effective.

Office of Science US Community Drives ESnet Design for Domestic Connectivity



Footprint of Largest SC Data Sharing Collaborators Drives the International Footprint that ESnet Must Support

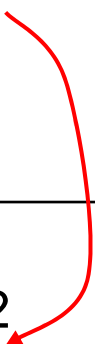


- Top 100 data flows generate 50% of all ESnet traffic (ESnet handles about 3×10^9 flows/mo.)
- 91 of the top 100 flows are from the Labs to other institutions (shown) (CY2005 data)

ESnet History

ESnet0/MFENet mid-1970s-1986	ESnet0/MFENet	56 Kbps microwave and satellite links
ESnet1 1986-1995	ESnet formed to serve the Office of Science	56 Kbps, X.25 to 45 Mbps T3
ESnet2 1995-2000	Partnered with Sprint to build the first national footprint ATM network	IP over 155 Mbps ATM net
ESnet3 2000-2007	Partnered with Qwest to build a national Packet over SONET network and optical channel Metropolitan Area Networks	IP over 10Gbps SONET
ESnet4 2007-2012	Partner with US Research & Education community to build a dedicated national optical network	IP and virtual circuits on a configurable optical infrastructure with at least 5-6 optical channels of 10-100 Gbps each

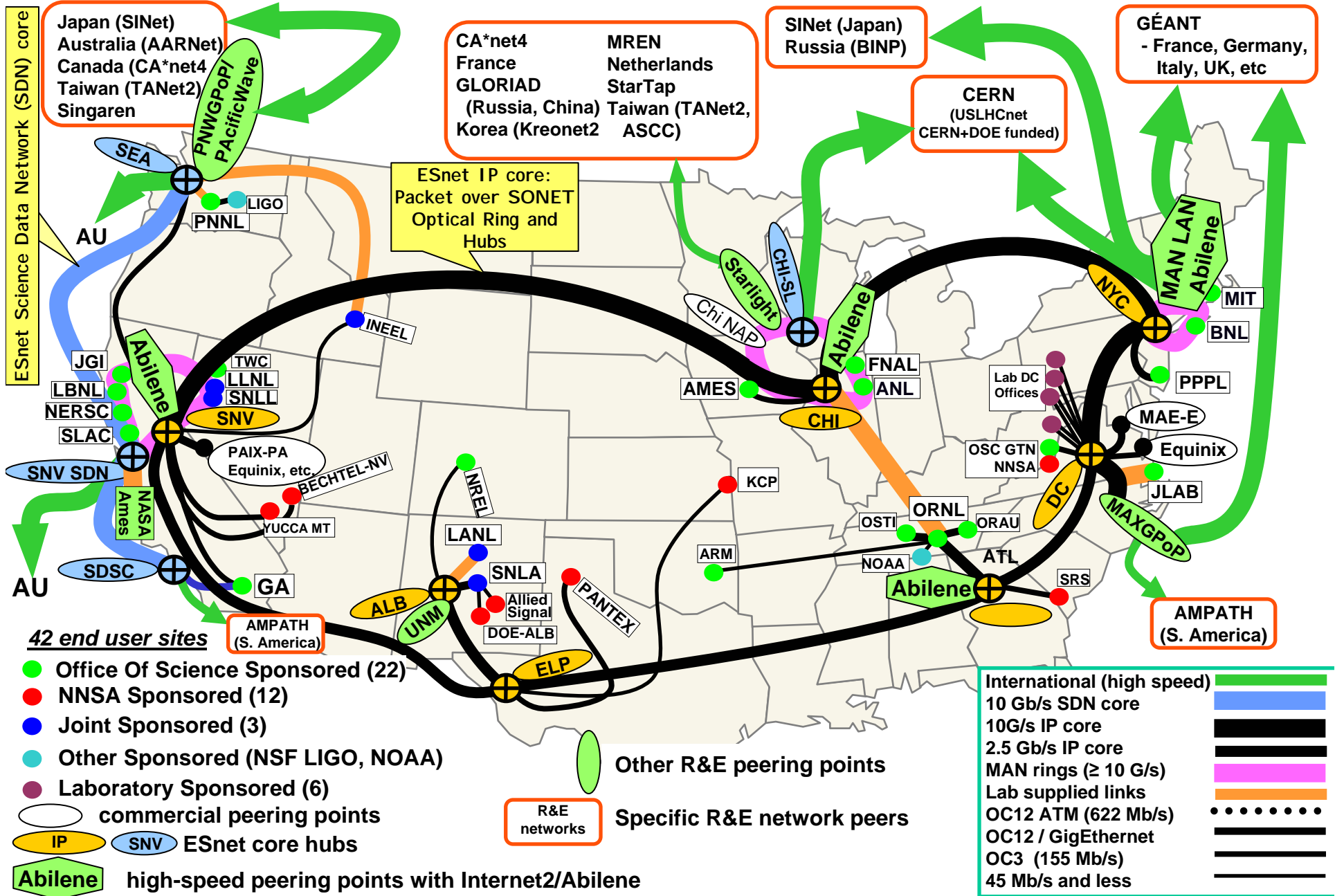
transition
in progress



ESnet's Place in U. S. and International Science

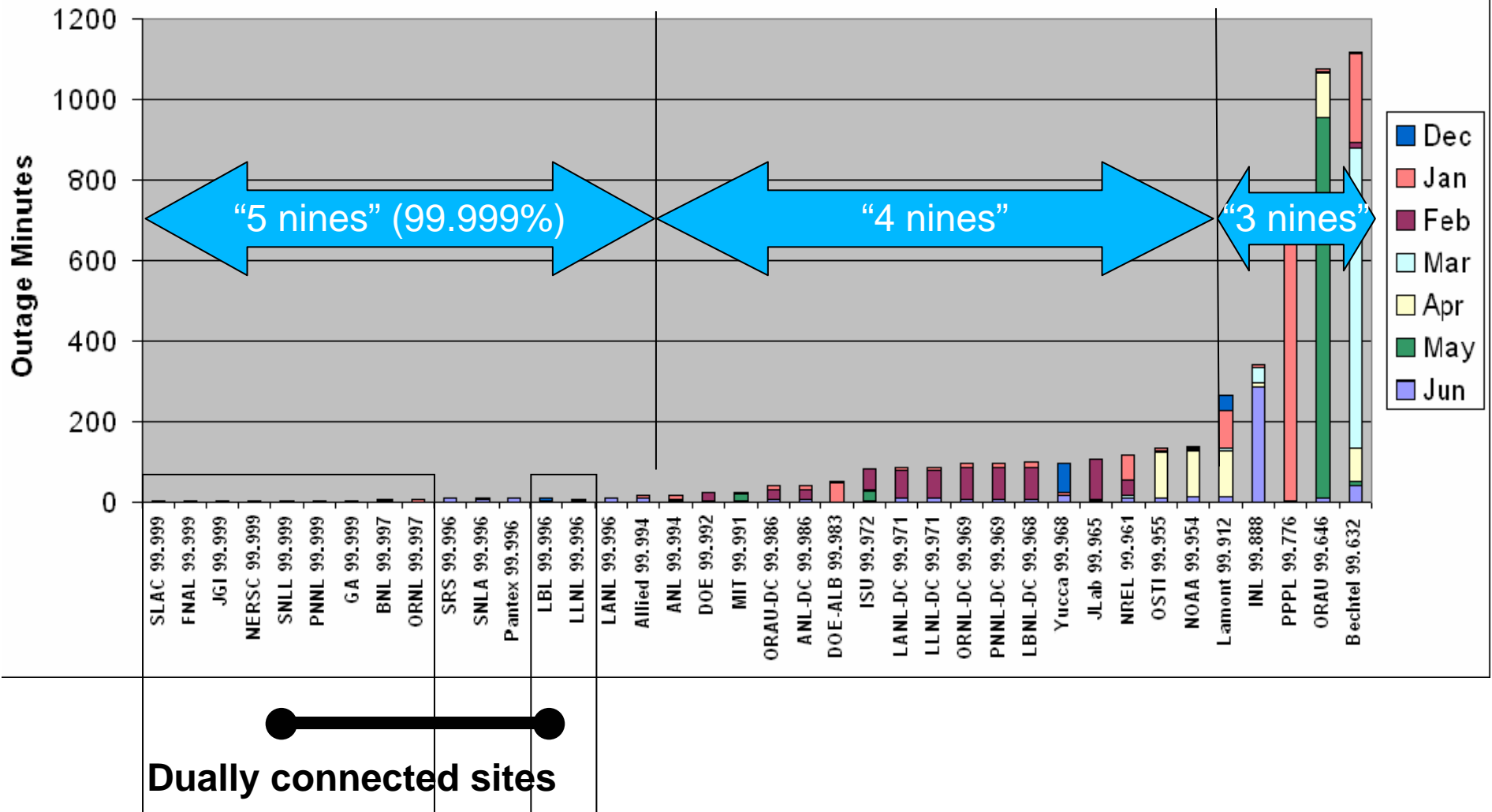
- ESnet, Internet2/Abilene, and National Lambda Rail (NLR) provide most of the nation's transit networking for basic science
 - Abilene provides national transit networking for most of the US universities by interconnecting the regional networks (mostly via the GigaPoPs)
 - ESnet provides national transit networking and ISP service for the DOE Labs
 - NLR provides various science-specific and network R&D circuits
- GÉANT plays a role in Europe similar to Abilene and ESnet in the US – it interconnects the European National Research and Education Networks (NRENs), to which the European R&E sites connect
 - GÉANT currently carries all non-LHC ESnet traffic to Europe, and this is a significant fraction of all ESnet traffic

ESnet3 Today Provides Global High-Speed Internet Connectivity for DOE Facilities and Collaborators (summer, 2006)



ESnet is a Highly Reliable Infrastructure

ESnet Availability 12/2005 through 6/2006



➤ **A Changing Science Environment is the Key Driver of the Next Generation ESnet**

- Large-scale collaborative science – big facilities, massive data, thousands of collaborators – is now a significant aspect of the Office of Science (“SC”) program
- SC science community is almost equally split between Labs and universities
 - SC facilities have users worldwide
- Very large international (non-US) facilities (e.g. LHC and ITER) and international collaborators are now a key element of SC science
- Distributed systems for data analysis, simulations, instrument operation, etc., are essential and are now common (in fact dominate this data analysis that now generates 50% of all ESnet traffic)

Planning the Future Network - ESnet4

There are many stakeholders for ESnet

1. SC programs

- Advanced Scientific Computing Research
- Basic Energy Sciences
- Biological and Environmental Research
- Fusion Energy Sciences
- High Energy Physics
- Nuclear Physics
- Office of Nuclear Energy

2. Major scientific facilities

- At DOE sites: large experiments, supercomputer centers, etc.
- Not at DOE sites: LHC, ITER

3. SC supported scientists not at the Labs (mostly at US R&E institutions)

4. Other collaborating institutions (mostly US, European, and AP R&E)

5. Other R&E networking organizations that support major collaborators

- Mostly US, European, and Asia Pacific networks

6. Lab operations and general population

7. Lab networking organizations

These account
for 85% of all
ESnet traffic

Planning the Future Network - ESnet4

- **Requirements of the ESnet stakeholders are primarily determined by**
 - 1) Data characteristics of instruments and facilities that will be connected to ESnet**
 - What data will be generated by instruments coming on-line over the next 5-10 years?
 - How and where will it be analyzed and used?
 - 2) Examining the future process of science**
 - How will the processing of doing science change over 5-10 years?
 - How do these changes drive demand for new network services?
 - 3) Studying the evolution of ESnet traffic patterns**
 - What are the trends based on the use of the network in the past 2-5 years?
 - How must the network change to accommodate the future traffic patterns implied by the trends?

(1) Requirements from Instruments and Facilities

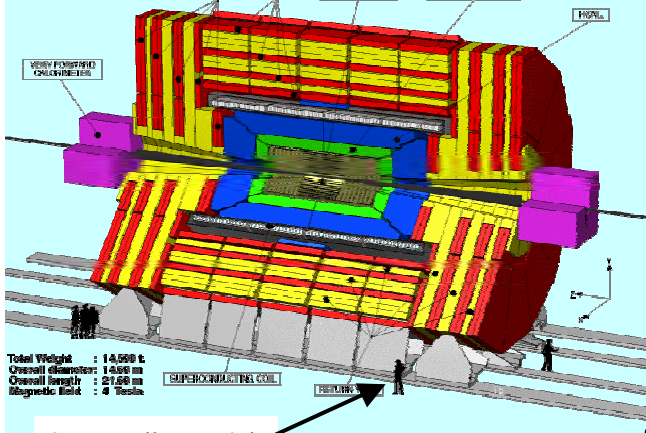
DOE SC Facilities that are, or will be, the top network users

- Advanced Scientific Computing Research
 - National Energy Research Scientific Computing Center (NERSC) (LBNL)*
 - National Leadership Computing Facility (NLCF) (ORNL)*
 - Argonne Leadership Class Facility (ALCF) (ANL)*
- Basic Energy Sciences
 - National Synchrotron Light Source (NSLS) (BNL)
 - Stanford Synchrotron Radiation Laboratory (SSRL) (SLAC)
 - Advanced Light Source (ALS) (LBNL)*
 - Advanced Photon Source (APS) (ANL)
 - Spallation Neutron Source (ORNL)*
 - National Center for Electron Microscopy (NCEM) (LBNL)*
 - Combustion Research Facility (CRF) (SNLL)*
- Biological and Environmental Research
 - William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) (PNNL)*
 - Joint Genome Institute (JGI)
 - Structural Biology Center (SBC) (ANL)
- Fusion Energy Sciences
 - DIII-D Tokamak Facility (GA)*
 - Alcator C-Mod (MIT)*
 - National Spherical Torus Experiment (NSTX) (PPPL)*
 - ITER
- High Energy Physics
 - Tevatron Collider (FNAL)
 - B-Factory (SLAC)
 - Large Hadron Collider (LHC, ATLAS, CMS) (BNL, FNAL)*
- Nuclear Physics
 - Relativistic Heavy Ion Collider (RHIC) (BNL)*
 - Continuous Electron Beam Accelerator Facility (CEBAF) (JLab)*

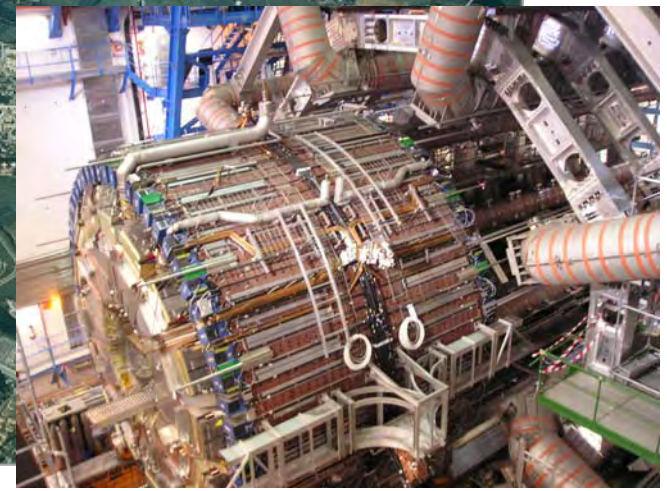
*characterized by current case studies

The Largest Facility: Large Hadron Collider at CERN

LHC CMS detector
15m X 15m X 22m, 12,500 tons, \$700M



human (for scale)



(2) Requirements from Examining the Future Process of Science

- In a major workshop [1], and in subsequent updates [2], requirements were generated by asking the science community how their process of doing science will / must change over the next 5 and next 10 years in order to accomplish their scientific goals
- Computer science and networking experts then assisted the science community in
 - analyzing the future environments
 - deriving middleware and networking requirements needed to enable these environments
- These were compiled as case studies that provide specific 5 & 10 year network requirements for bandwidth, footprint, and new services

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)	<ul style="list-style-type: none"> • DOE sites • US Universities • Industry 	200+ Mbps	1 Gbps	<ul style="list-style-type: none"> • Bulk data • Remote control 	<ul style="list-style-type: none"> • Guaranteed bandwidth • Guaranteed QoS • Deadline scheduling
NERSC and ACLF	-	<ul style="list-style-type: none"> • DOE sites • US Universities • International • Other ASCR supercomputers 	10 Gbps	20 to 40 Gbps	<ul style="list-style-type: none"> • Bulk data • Remote control • Remote file system sharing 	<ul style="list-style-type: none"> • Guaranteed bandwidth • Guaranteed QoS • Deadline Scheduling • PKI / Grid
NLCF	-	<ul style="list-style-type: none"> • DOE sites • US Universities • Industry • International 	Backbone Band width parity	Backbone band width parity	<ul style="list-style-type: none"> • Bulk data • Remote file system sharing 	
Nuclear Physics (RHIC)	-	<ul style="list-style-type: none"> • DOE sites • US Universities • International 	12 Gbps	70 Gbps	<ul style="list-style-type: none"> • Bulk data 	<ul style="list-style-type: none"> • Guaranteed bandwidth • PKI / Grid
Spallation Neutron Source	High (24x7 operation)	<ul style="list-style-type: none"> • DOE sites 	640 Mbps	2 Gbps	<ul style="list-style-type: none"> • 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	-	<ul style="list-style-type: none"> • DOE sites • US Universities • Industry 	1 TB/day 300 Mbps	5 TB/day 1.5 Gbps	<ul style="list-style-type: none"> • Bulk data • Remote control 	<ul style="list-style-type: none"> • Guaranteed bandwidth • PKI / Grid
Bioinformatics	-	<ul style="list-style-type: none"> • DOE sites • US Universities 	625 Mbps 12.5 Gbps in two years	250 Gbps	<ul style="list-style-type: none"> • Bulk data • Remote control • Point-to-multipoint 	<ul style="list-style-type: none"> • Guaranteed bandwidth • High-speed multicast
Chemistry / Combustion	-	<ul style="list-style-type: none"> • DOE sites • US Universities • Industry 	-	10s of Gigabits per second	<ul style="list-style-type: none"> • Bulk data 	<ul style="list-style-type: none"> • Guaranteed bandwidth • PKI / Grid
Climate Science	-	<ul style="list-style-type: none"> • DOE sites • US Universities • International 	-	5 PB per year 5 Gbps	<ul style="list-style-type: none"> • Bulk data • Remote control 	<ul style="list-style-type: none"> • Guaranteed bandwidth • PKI / Grid
Immediate Requirements and Drivers						
High Energy Physics (LHC)	99.95+% (Less than 4 hrs/year)	<ul style="list-style-type: none"> • US Tier1 (FNAL, BNL) • US Tier2 (Universities) • International (Europe, Canada) 	10 Gbps	60 to 80 Gbps (30-40 Gbps per US Tier1)	<ul style="list-style-type: none"> • Bulk data • Coupled data analysis processes 	<ul style="list-style-type: none"> • Guaranteed bandwidth • Traffic isolation • PKI / Grid

Changing Science Environment ⇒ New Demands on Network

- **Increased capacity**

- Needed to accommodate a large and steadily increasing amount of data that must traverse the network

- **High network reliability**

- Essential when interconnecting components of distributed large-scale science

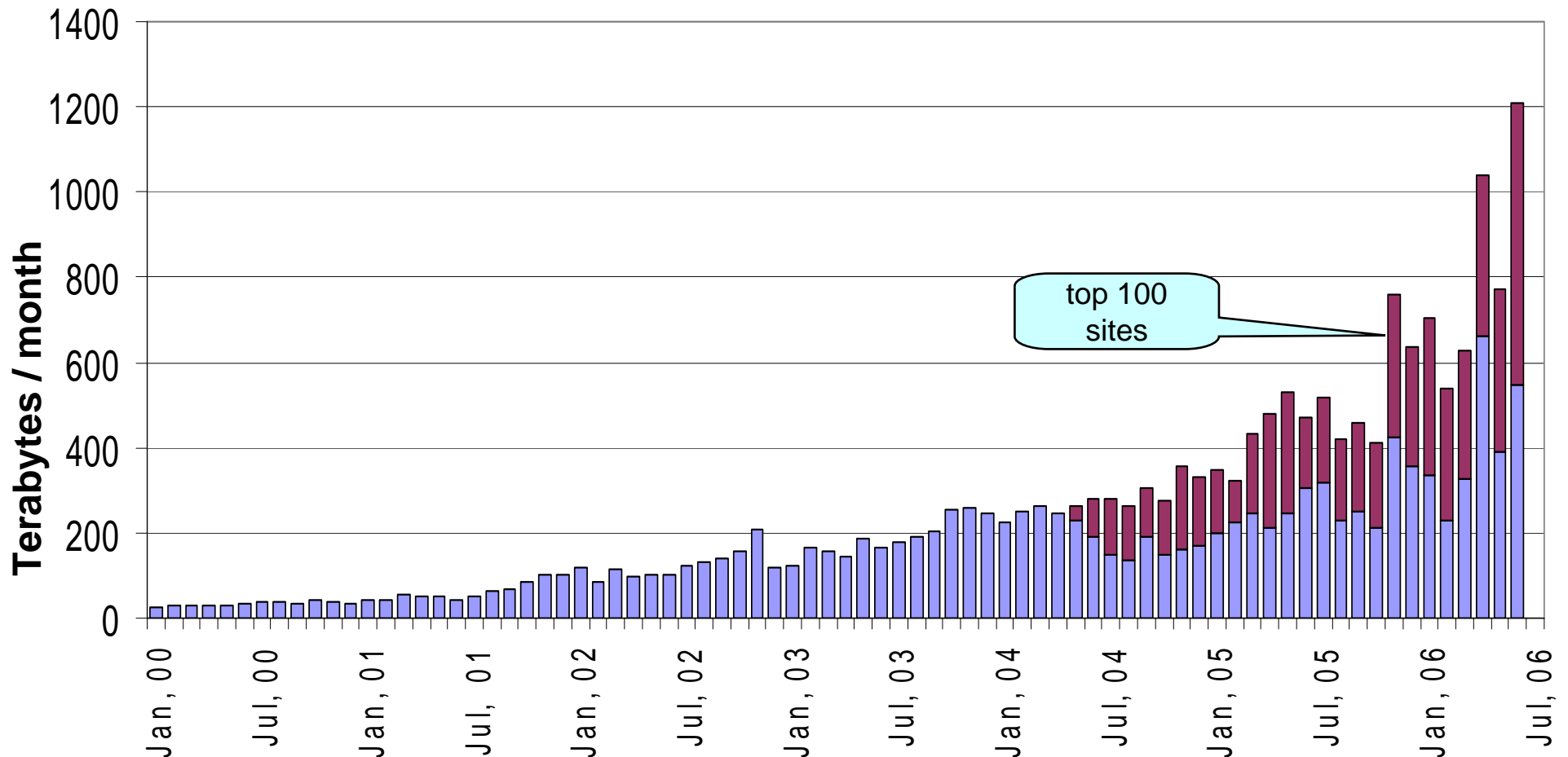
- **High-speed, highly reliable connectivity between Labs and US and international R&E institutions**

- To support the inherently collaborative, global nature of large-scale science

- **New network services to provide bandwidth guarantees**

- Provide for data transfer deadlines for
 - remote data analysis, real-time interaction with instruments, coupled computational simulations, etc.

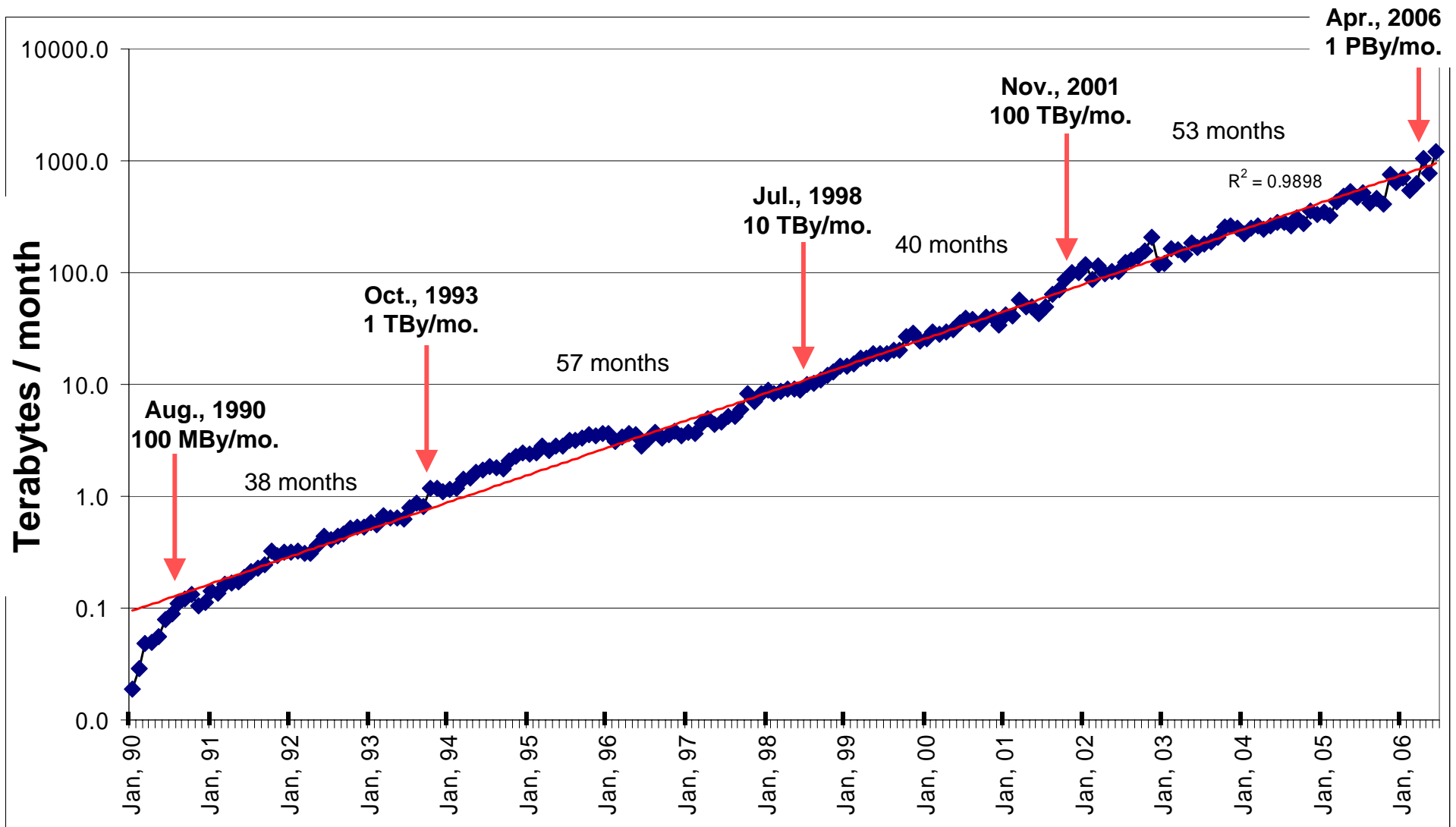
3) These Trends are Seen in Observed Evolution of Historical ESnet Traffic Patterns



ESnet Monthly Accepted Traffic, January, 2000 – June, 2006

- ESnet is currently transporting more than 1 petabyte (1000 terabytes) per month
- More than 50% of the traffic is now generated by the top 100 sites

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



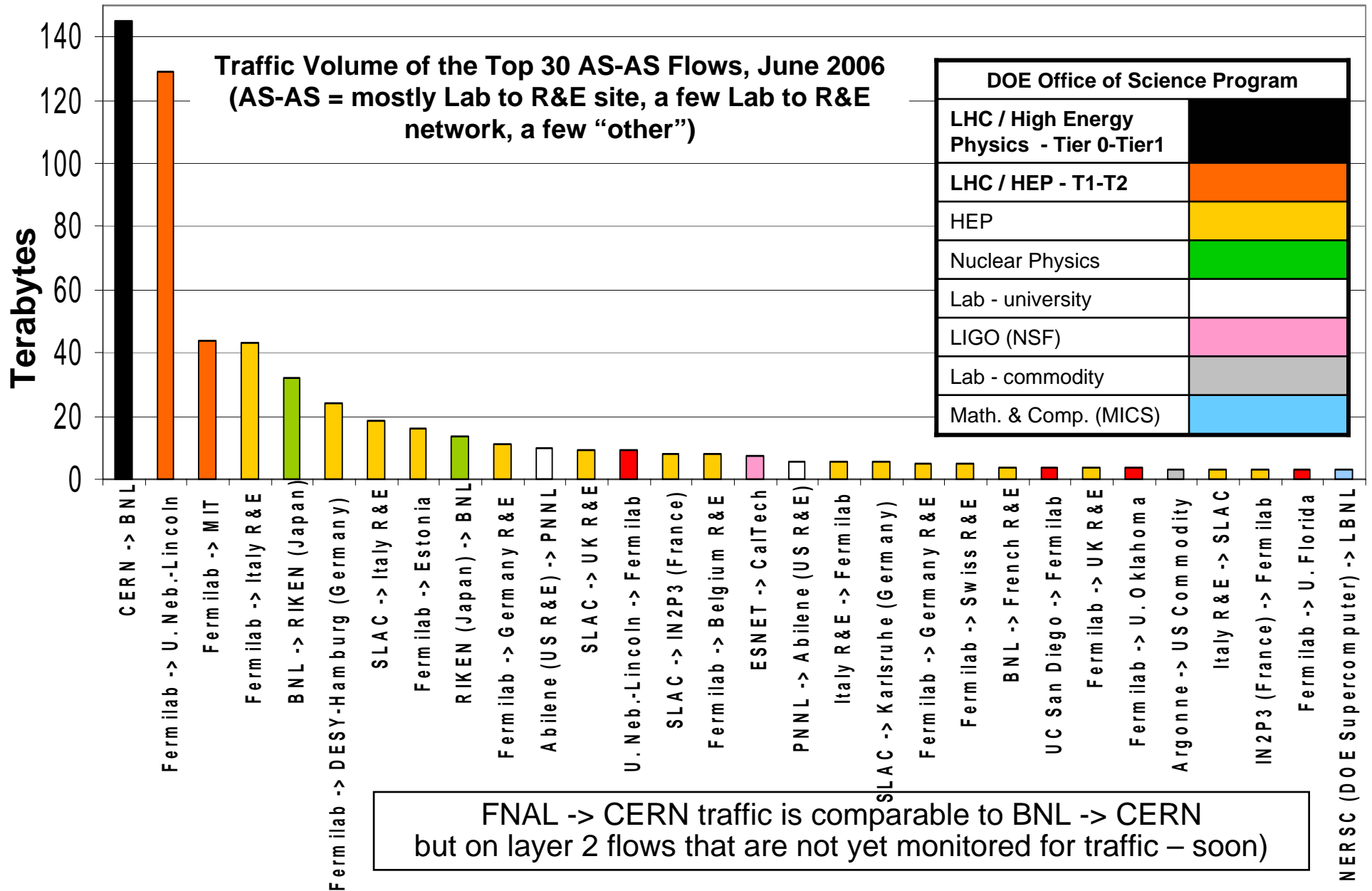
Log Plot of ESnet Monthly Accepted Traffic, January, 1990 – June, 2006

Requirements from Network Utilization Observation

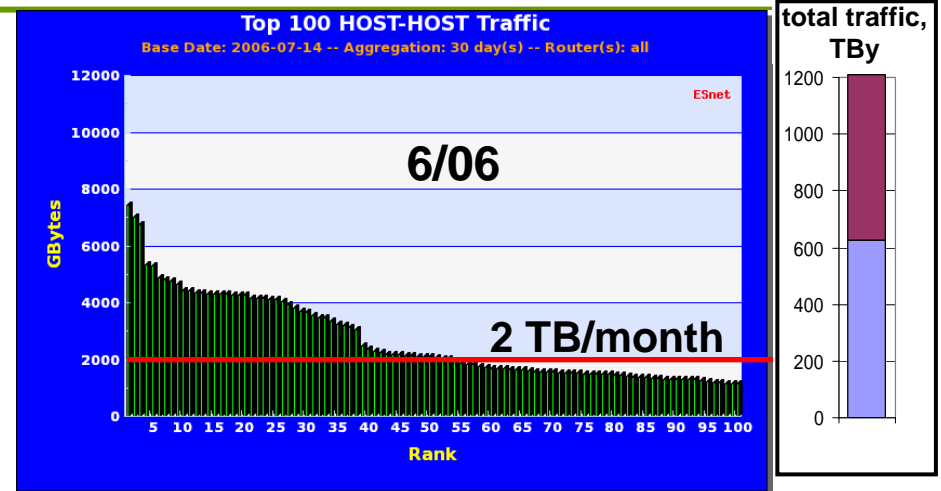
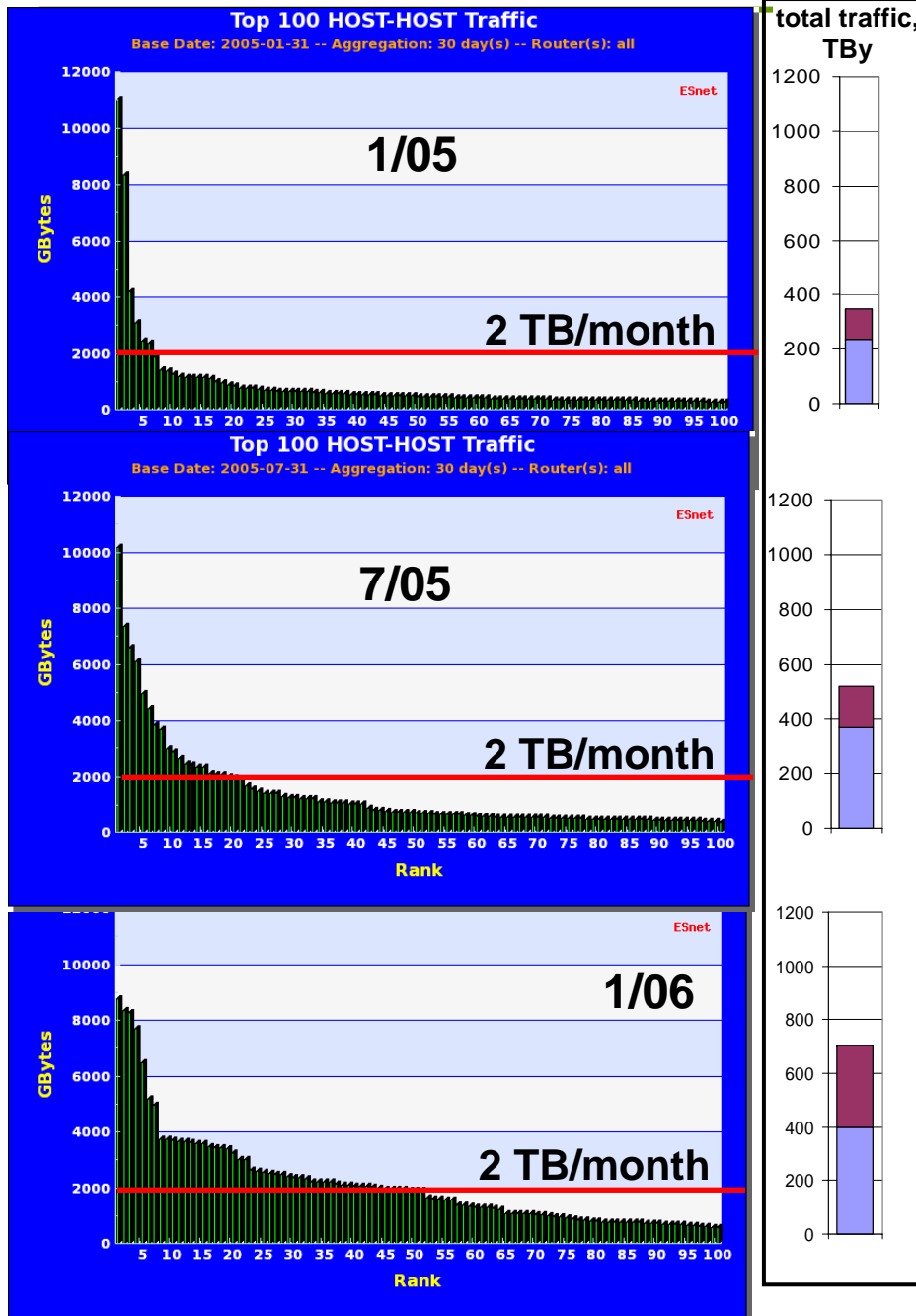
- In 4 years, we can expect a 10x increase in traffic over current levels *without the addition of production LHC traffic*
 - Nominal average load on busiest backbone links is ~1.5 Gbps today
 - In 4 years that figure will be ~15 Gbps based on current trends
- Measurements of this type are science-agnostic
 - 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
- Bandwidth trends drive requirement for a new network architecture
 - New architecture/approach must be scalable in a cost-effective way

Large-Scale Flow Trends, June 2006

Subtitle: "Onslaught of the LHC"



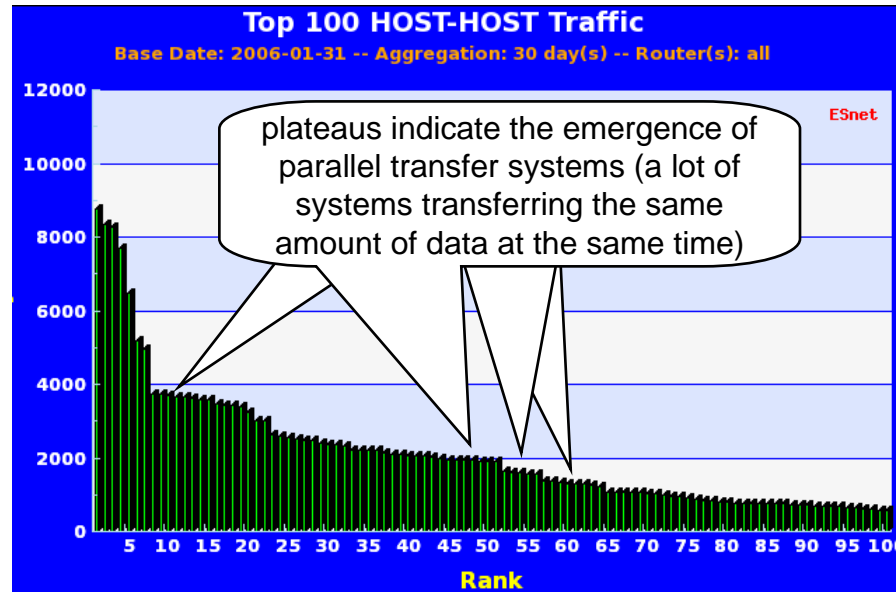
➤ Traffic Patterns are Changing Dramatically



- While the total traffic is increasing exponentially
 - Peak flow – that is system-to-system – bandwidth is decreasing
 - The number of large flows is increasing

The Onslaught of Grids

Question: Why is peak flow bandwidth decreasing while total traffic is increasing?



Answer: Most large data transfers are now done by parallel / Grid data movers

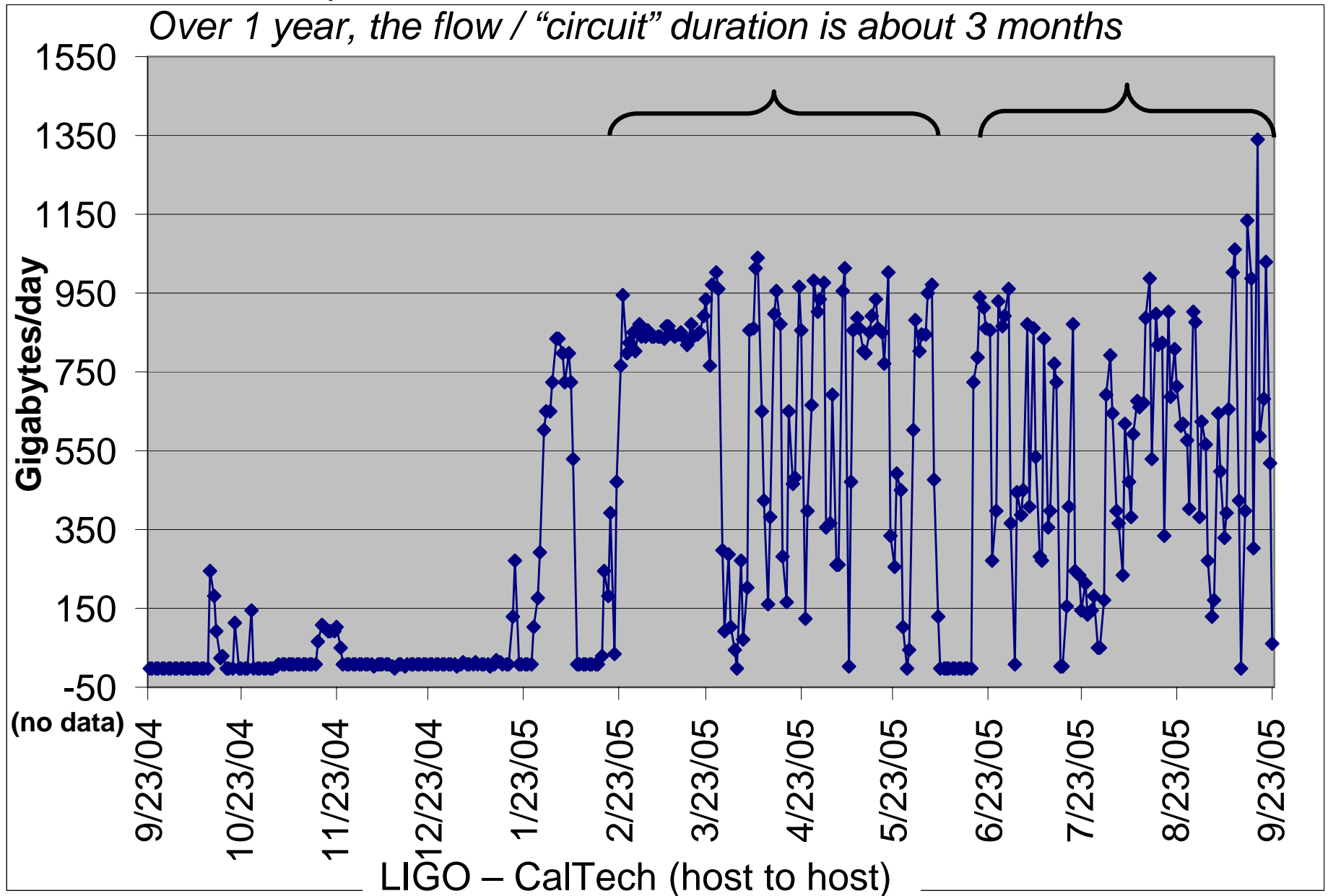
- In June, 2006 **72%** of the hosts generating the 1000 flows were involved in parallel data movers (Grid applications)
- ***This is the most significant traffic pattern change in the history of ESnet***
- This has implications for the network architecture that favor path multiplicity and route diversity

Requirements from Traffic Flow Observations

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

Network Observation – Circuit-like Behavior

Look at Top 20 Traffic Generator's Historical Flow Patterns
Over 1 year, the flow / "circuit" duration is about 3 months



➤ The ESnet Response to the Requirements

I) A new network architecture and implementation strategy

- 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 end-to-end interoperability

Next Generation ESnet: I) Architecture and Configuration

- **Main architectural elements and the rationale for each element**

1) A **High-reliability IP core** (e.g. the current ESnet core) to address

- General science requirements
- Lab operational requirements
- Backup for the SDN core
- Vehicle for science services
- Full service IP routers

2) **Metropolitan Area Network (MAN)** rings to provide

- Dual site connectivity for reliability
- Much higher site-to-core bandwidth
- Support for both production IP and circuit-based traffic
- Multiply connecting the SDN and IP cores

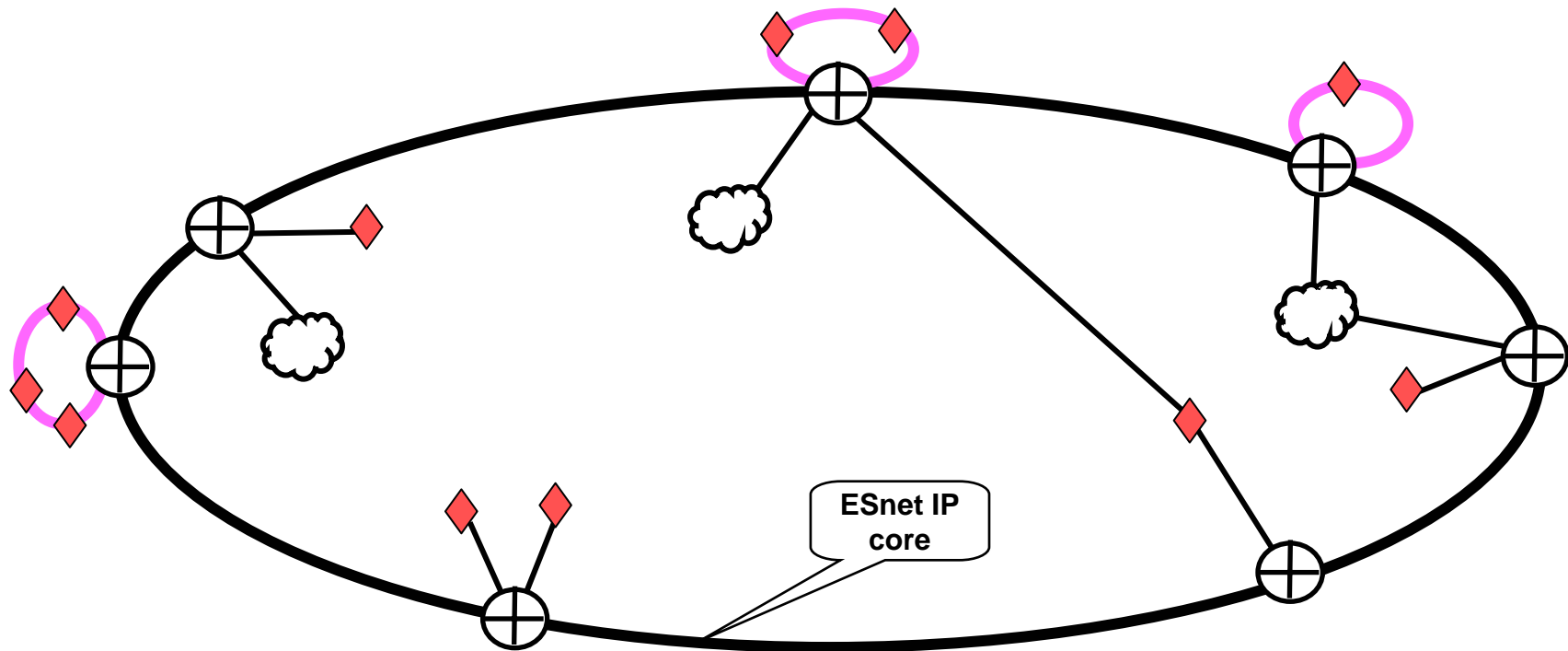
2a) **Loops off of the backbone** rings to provide





- For dual site connections where MANs are not practical

3) **A Science Data Network (SDN) core for**

- Provisioned, guaranteed bandwidth circuits to support large, high-speed science data flows
- Very high total bandwidth
- Multiply connecting MAN rings for protection against hub failure
- Alternate path for production IP traffic
- Less expensive router/switches
- Initial configuration targeted at LHC, which is also the first step to the general configuration that will address all SC requirements
- Can meet other unknown bandwidth requirements by adding lambdas

ESnet3 Core Architecture and Implementation



-  ESnet sites
-  ESnet hubs / core network connection points
-  Metro area rings (MANs)
-  Other networks

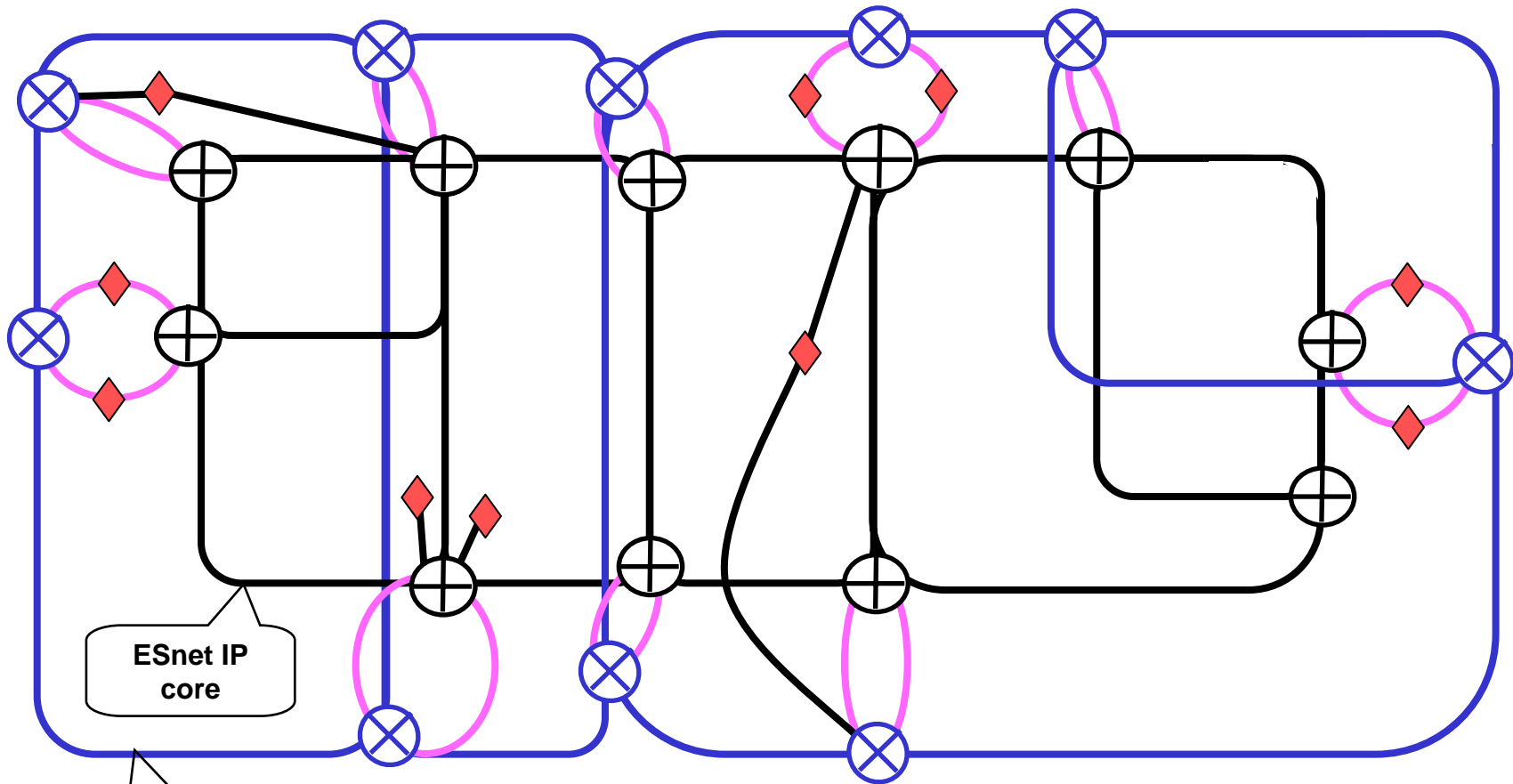
Strengths

- single cuts in any ring will not isolate any site
- most SC Labs are dually connected to the core ring so no single cut will isolate

Weaknesses

- core ring, MAN, and sites connect in a single hub - a single point of failure
- two cuts in the core ring on opposite sides of a hub partitions the network
- no route diversity

ESnet4 Core Architecture and Implementation



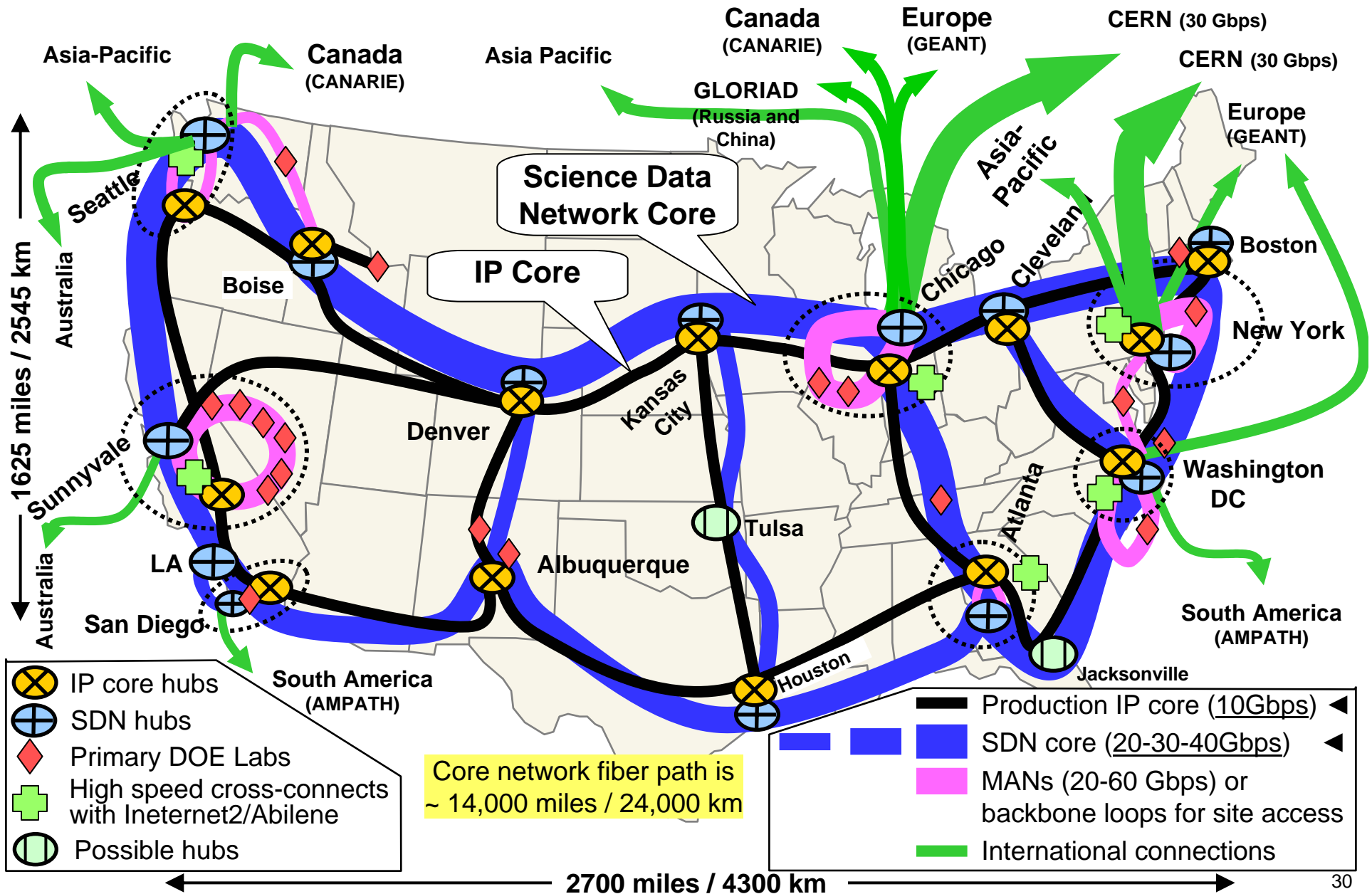
ESnet Science Data Network (SDN) core

- ◆ ESnet sites
- ⊕ ESnet IP routers
- ⊗ ESnet SDN circuit switches
- Metro area rings (MANs)

<p><u>Strengths</u></p> <ul style="list-style-type: none"> • No single circuit or equipment failure can isolate SC Labs • Route diversity permits traffic optimization (especially for circuits) for better network utilization of capacity and failover • Diverse footprint allows better access to US regional nets and LHC tier 2 sites (universities) 	<p><u>Weaknesses</u></p> <ul style="list-style-type: none"> • Both IP and SDN IP circuits (optical channels) are on a single fiber so simultaneous cuts on the opposite sides of an outer ring will still partition the network
--	--

ESnet4 Configuration

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



Next Generation ESnet: II) Virtual Circuits

- 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
- 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
- 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

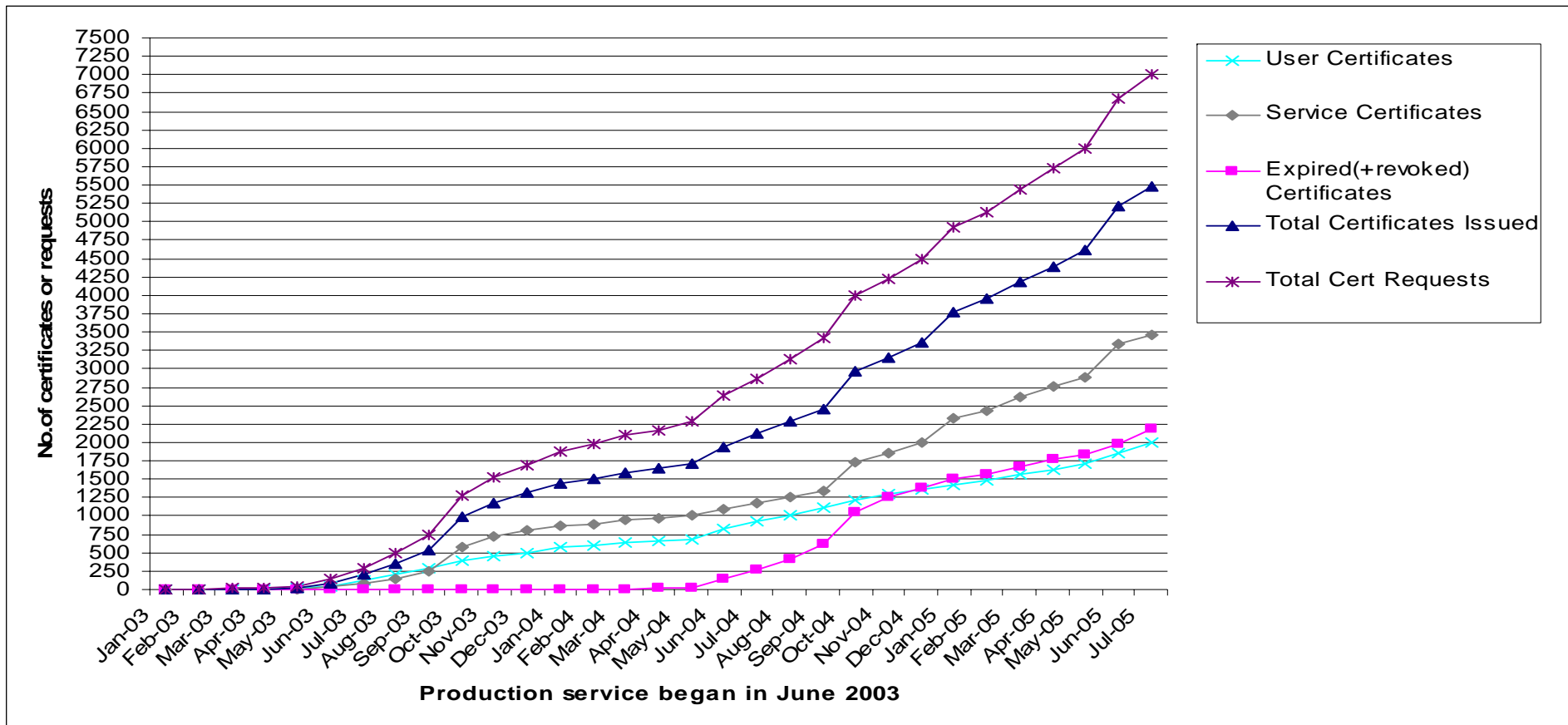
OSCARS: Guaranteed Bandwidth VC Service For SC Science

- ESnet Virtual Circuit project: On-demand Secured Circuits and Advanced Reservation System (OSCARS)
- To ensure compatibility, the design and implementation is done in collaboration with the other major science R&E networks and end sites
 - Internet2: Bandwidth Reservation for User Work (BRUW)
 - Development of common code base
 - GEANT: Bandwidth on Demand (GN2-JRA3), Performance and Allocated Capacity for End-users (SA3-PACE) and Advance Multi-domain Provisioning System (AMPS) extends to NRENs
 - BNL: TeraPaths - A QoS Enabled Collaborative Data Sharing Infrastructure for Peta-scale Computing Research
 - GA: Network Quality of Service for Magnetic Fusion Research
 - SLAC: Internet End-to-end Performance Monitoring (IEPM)
 - USN: Experimental Ultra-Scale Network Testbed for Large-Scale Science
- In its current phase this effort is being funded as a research project by the Office of Science, Mathematical, Information, and Computational Sciences (MICS) Network R&D Program
- A prototype service has been deployed as a proof of concept
 - To date more than 20 accounts have been created for beta users, collaborators, and developers
 - More than 100 reservation requests have been processed

➤ 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

Usage Statistics for the DOEGrids Certification Authority that Issues Identity Certificates for Grid Authentication



User Certificates	1999	Total No. of Certificates	5479
Host & Service Certificates	3461	Total No. of Requests	7006
ESnet SSL Server CA Certificates			38
DOEGrids CA 2 CA Certificates (NERSC)			15
FusionGRID CA certificates			76

* Report as of Jun 15, 2005

➤ Summary

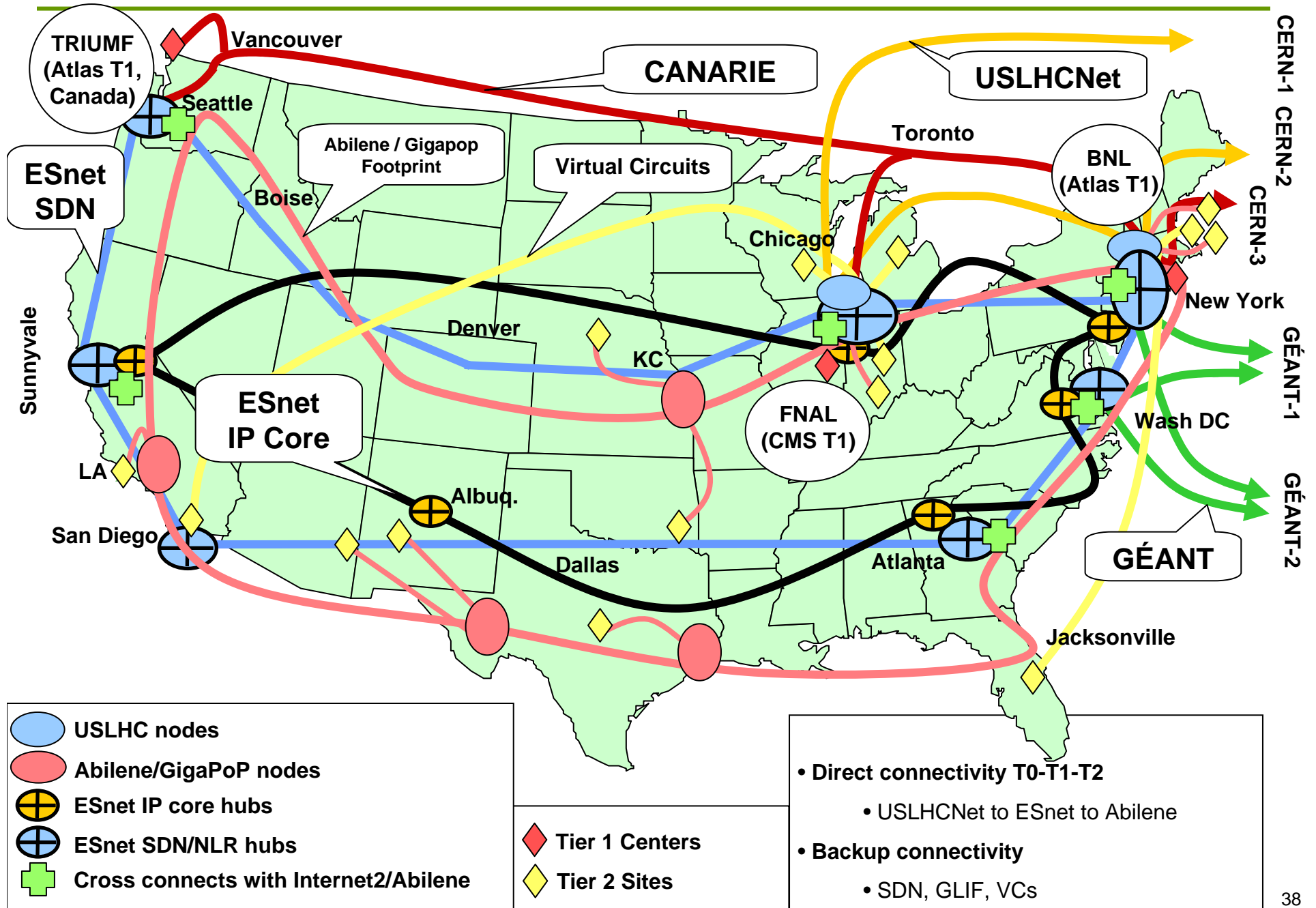
- **ESnet is currently satisfying its mission by enabling SC science that is dependant on networking and distributed, large-scale collaboration:**
 - “The performance of ESnet over the past year has been excellent, with only minimal unscheduled down time. The reliability of the core infrastructure is 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. DOE Workshop on Ultra High-Speed Transport Protocols and Network Provisioning for Large-Scale Science Applications, April 2003
 - <http://www.csm.ornl.gov/ghpn/wk2003>
5. Science Case for Large Scale Simulation, June 2003
 - <http://www.pnl.gov/scales/>
6. Workshop on the Road Map for the Revitalization of High End Computing, June 2003
 - <http://www.cra.org/Activities/workshops/nitrd>
 - http://www.sc.doe.gov/ascr/20040510_hecrtf.pdf (public report)
7. ASCR Strategic Planning Workshop, July 2003
 - <http://www.fp-mcs.anl.gov/ascr-july03spw>
8. Planning Workshops-Office of Science Data-Management Strategy, March & May 2004
 - <http://www-conf.slac.stanford.edu/dmw2004>

Additional Information

LHC Tier 0, 1, and 2 Connectivity Requirements Summary



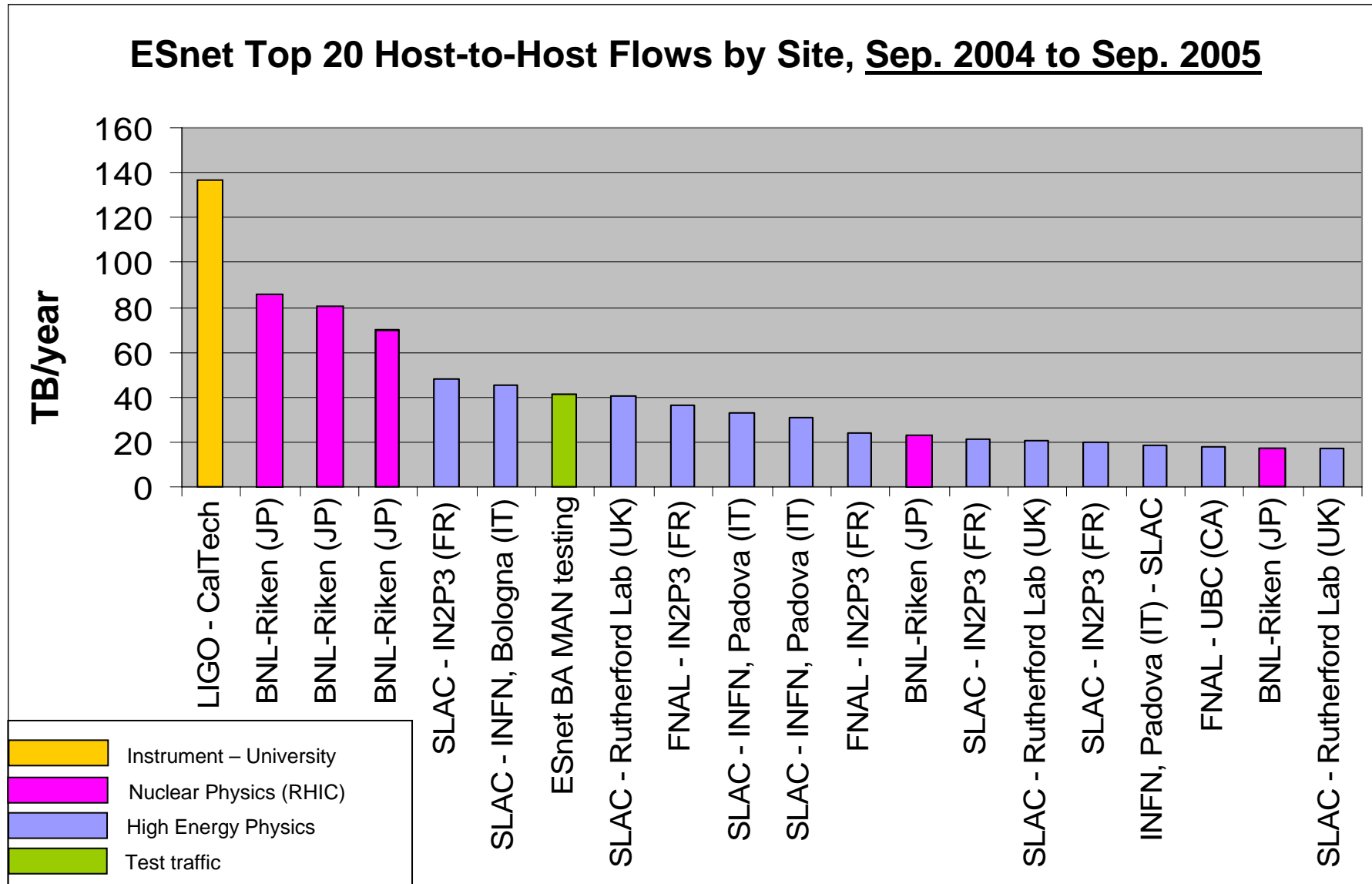
Example Case Study Summary Matrix: Fusion

- Considers instrument and facility requirements, the process of science drivers and resulting network requirements cross cut with timelines

Feature	Anticipated Requirements			
Time Frame	Science Instruments and Facilities	Process of Science	Network	Network Services and Middleware
Near-term	<ul style="list-style-type: none"> • Each experiment only gets a few days per year - high productivity is critical • Experiment episodes (“shots”) generate 2-3 Gbytes every 20 minutes, which has to be delivered to the remote analysis sites in two minutes in order to analyze before next shot • Highly collaborative experiment and analysis environment 	<ul style="list-style-type: none"> • Real-time data access and analysis for experiment steering (the more that you can analyze between shots the more effective you can make the next shot) • Shared visualization capabilities 		<ul style="list-style-type: none"> • PKI certificate authorities that enable strong authentication of the community members and the use of Grid security tools and services. • Directory services that can be used to provide the naming root and high-level (community-wide) indexing of shared, persistent data that transforms into community information and knowledge • Efficient means to sift through large data repositories to extract meaningful information from unstructured data.
5 years	<ul style="list-style-type: none"> • 10 Gbytes generated by experiment every 20 minutes (time between shots) to be delivered in two minutes • Gbyte subsets of much larger simulation datasets to be delivered in two minutes for comparison with experiment • Simulation data scattered across United States • Transparent security • Global directory and naming services needed to anchor all of the distributed metadata • Support for “smooth” collaboration in a high-stress environment 	<ul style="list-style-type: none"> • Real-time data analysis for experiment steering combined with simulation interaction = big productivity increase • Real-time visualization and interaction among collaborators across United States • Integrated simulation of the several distinct regions of the reactor will produce a much more realistic model of the fusion process 	<ul style="list-style-type: none"> • Network bandwidth and data analysis computing capacity guarantees (quality of service) for inter-shot data analysis • Gbits/sec for 20 seconds out of 20 minutes, guaranteed • 5 to 10 remote sites involved for data analysis and visualization 	<ul style="list-style-type: none"> • Parallel network I/O between simulations, data archives, experiments, and visualization • High quality, 7x24 PKI identity authentication infrastructure • End-to-end quality of service and quality of service management • Secure/authenticated transport to ease access through firewalls • Reliable data transfer • Transient and transparent data replication for real-time reliability • Support for human collaboration tools
5+ years	<ul style="list-style-type: none"> • Simulations generate 100s of Tbytes • ITER – Tbyte per shot, PB per year 	<ul style="list-style-type: none"> • Real-time remote operation of the experiment • Comprehensive integrated simulation 	<ul style="list-style-type: none"> • Quality of service for network latency and reliability, and for co-scheduling computing resources 	<ul style="list-style-type: none"> • Management functions for network quality of service that provides the request and access mechanisms for the experiment run time, periodic traffic noted above.

Large-Scale Flow Trends – 2004/2005

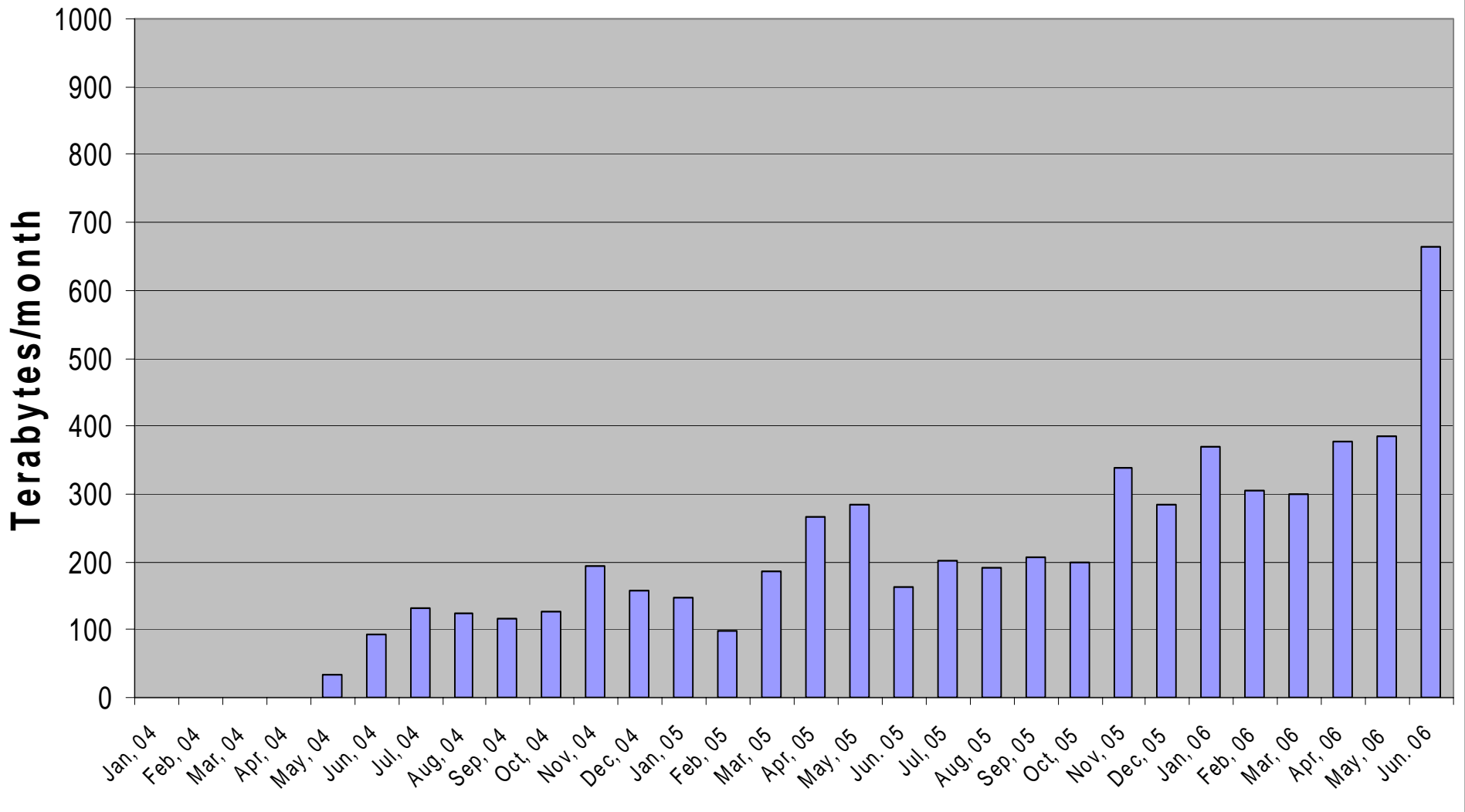
(Among other things these observations help establish the network footprint requirements)



The Increasing Dominance of Science Traffic

Traffic Volume of the Top 100 AS-AS Flows by Month

(Mostly Lab to R&E site, a few Lab to R&E network – all science)



Parallel Data Movers now Predominate

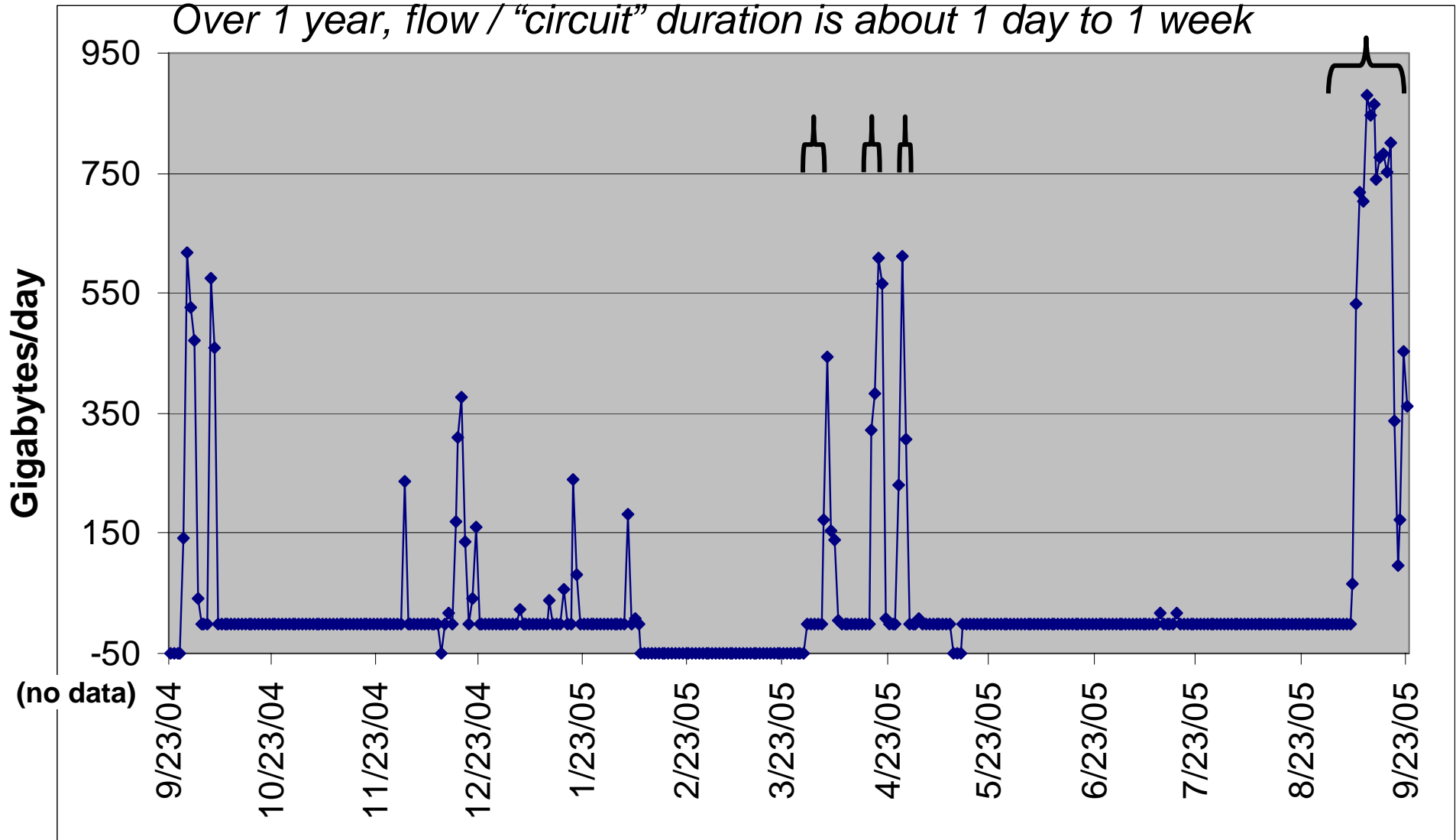
Look at the hosts involved in 2006-01-31—the plateaus in the host-host top 100 flows are all parallel transfers (thx. to Eli Dart for this observation)

A132023.N1.Vanderbilt.Edu	Istore1.fnal.gov	5.847	bbr-xfer07.slac.stanford.edu	babar2.fzk.de	2.113
A132021.N1.Vanderbilt.Edu	Istore1.fnal.gov	5.884	bbr-xfer05.slac.stanford.edu	babar.fzk.de	2.254
A132018.N1.Vanderbilt.Edu	Istore1.fnal.gov	6.048	bbr-xfer04.slac.stanford.edu	babar.fzk.de	2.294
A132022.N1.Vanderbilt.Edu	Istore1.fnal.gov	6.39	bbr-xfer07.slac.stanford.edu	babar.fzk.de	2.337
A132021.N1.Vanderbilt.Edu	Istore2.fnal.gov	6.771	bbr-xfer04.slac.stanford.edu	babar2.fzk.de	2.339
A132023.N1.Vanderbilt.Edu	Istore2.fnal.gov	6.825	bbr-xfer05.slac.stanford.edu	babar2.fzk.de	2.357
A132022.N1.Vanderbilt.Edu	Istore2.fnal.gov	6.86	bbr-xfer08.slac.stanford.edu	babar2.fzk.de	2.471
A132018.N1.Vanderbilt.Edu	Istore2.fnal.gov	7.286	bbr-xfer08.slac.stanford.edu	babar.fzk.de	2.627
A132017.N1.Vanderbilt.Edu	Istore1.fnal.gov	7.62	bbr-xfer04.slac.stanford.edu	babar3.fzk.de	3.234
A132017.N1.Vanderbilt.Edu	Istore2.fnal.gov	9.299	bbr-xfer05.slac.stanford.edu	babar3.fzk.de	3.271
A132023.N1.Vanderbilt.Edu	Istore4.fnal.gov	10.522	bbr-xfer08.slac.stanford.edu	babar3.fzk.de	3.276
A132021.N1.Vanderbilt.Edu	Istore4.fnal.gov	10.54	bbr-xfer07.slac.stanford.edu	babar3.fzk.de	3.298
A132018.N1.Vanderbilt.Edu	Istore4.fnal.gov	10.597	bbr-xfer05.slac.stanford.edu	bbr-datamove10.cr.cnaf.infn.it	2.366
A132018.N1.Vanderbilt.Edu	Istore3.fnal.gov	10.746	bbr-xfer07.slac.stanford.edu	bbr-datamove10.cr.cnaf.infn.it	2.519
A132022.N1.Vanderbilt.Edu	Istore4.fnal.gov	11.097	bbr-xfer04.slac.stanford.edu	bbr-datamove10.cr.cnaf.infn.it	2.548
A132022.N1.Vanderbilt.Edu	Istore3.fnal.gov	11.097	bbr-xfer08.slac.stanford.edu	bbr-datamove10.cr.cnaf.infn.it	2.656
A132021.N1.Vanderbilt.Edu	Istore3.fnal.gov	11.213	bbr-xfer08.slac.stanford.edu	bbr-datamove09.cr.cnaf.infn.it	3.927
A132023.N1.Vanderbilt.Edu	Istore3.fnal.gov	11.331	bbr-xfer05.slac.stanford.edu	bbr-datamove09.cr.cnaf.infn.it	3.94
A132017.N1.Vanderbilt.Edu	Istore4.fnal.gov	11.425	bbr-xfer04.slac.stanford.edu	bbr-datamove09.cr.cnaf.infn.it	4.011
A132017.N1.Vanderbilt.Edu	Istore3.fnal.gov	11.489	bbr-xfer07.slac.stanford.edu	bbr-datamove09.cr.cnaf.infn.it	4.177
babar.fzk.de	bbr-xfer03.slac.stanford.edu	2.772	bbr-xfer04.slac.stanford.edu	csfmove01.rl.ac.uk	5.952
babar.fzk.de	bbr-xfer02.slac.stanford.edu	2.901	bbr-xfer04.slac.stanford.edu	move03.gridpp.rl.ac.uk	5.959
babar2.fzk.de	bbr-xfer06.slac.stanford.edu	3.018	bbr-xfer05.slac.stanford.edu	csfmove01.rl.ac.uk	5.976
babar.fzk.de	bbr-xfer04.slac.stanford.edu	3.222	bbr-xfer05.slac.stanford.edu	move03.gridpp.rl.ac.uk	6.12
bbr-export01.pd.infn.it	bbr-xfer03.slac.stanford.edu	11.289	bbr-xfer07.slac.stanford.edu	csfmove01.rl.ac.uk	6.242
bbr-export02.pd.infn.it	bbr-xfer03.slac.stanford.edu	19.973	bbr-xfer08.slac.stanford.edu	move03.gridpp.rl.ac.uk	6.357
			bbr-xfer08.slac.stanford.edu	csfmove01.rl.ac.uk	6.48
			bbr-xfer07.slac.stanford.edu	move03.gridpp.rl.ac.uk	6.604

Network Observation – Circuit-like Behavior (2)

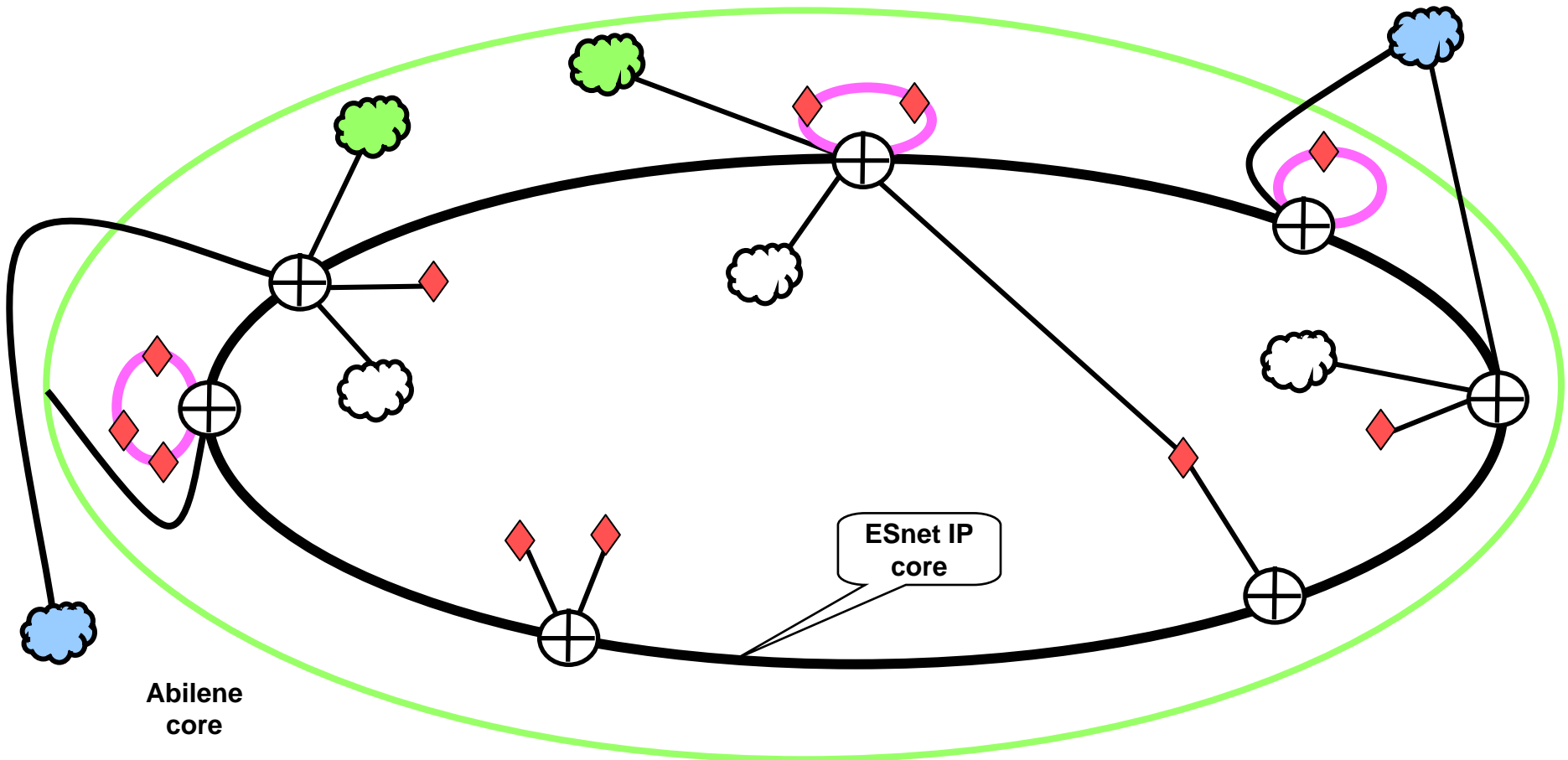
Look at Top 20 Traffic Generator's Historical Flow Patterns

Over 1 year, flow / "circuit" duration is about 1 day to 1 week



SLAC - IN2P3, France (host to host)

ESnet3 Overall Architecture



- ◆ ESnet sites
- ⊕ ESnet hubs / core network connection points
- Metro area rings (MANs)
- ☁ US R&E regional nets
- ☁ International R&E nets
- ☁ US commercial ISPs

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 end-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