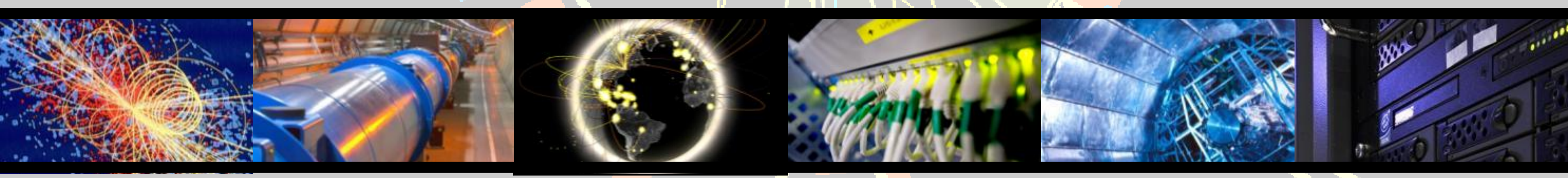
Networking in the WLCG Facilities

Michael Ernst
Brookhaven National Laboratory



WLCG Facilities (1/2)

- Today in WLCG our primary concern regarding resources is CPU and storage capacities
  - All LHC experiments are increasingly relying on networks to provide connectivity between distributed facility elements across all levels
    - E.g. ATLAS uses today ~50% of T2 storage for primary datasets
  - Wide Area Network infrastructure is complex
    - Besides the (mostly) predictable service of the OPN the vast majority of traffic (i.e. for ATLAS and CMS) has been using the GPN
    - With LHCONE there is a 3rd piece of network infrastructure
      - Meant to serve T1 ⇔ T2/T3, T2 ⇔ T2, T2 ⇔ T3 needs

WLCG Facilities (2/2)

- With evolving computing models traffic flows change significantly (e.g. due to vanishing T1/T2 differences)
  - With the changes in the ATLAS CM, what would not work given the current network infrastructure?
  - The LHC community should conduct a study to determine what kind of networks and how much networking is necessary – today and in the next ~5 years – Key questions include
    - Do we need either LHCOPN or LHCONE if there was a sufficiently provisioned commodity IP network?
      - » All commercial and R&E providers have 100 Gbps backbones deployed
      - » Issue is typically the "last mile"
    - Do we need LHCOPN if we have a globally funded LHCONE private IP network?
    - Do we need LHCONE given we have the LHCOPN and today's commodity IP network with increased capacities?
      - » Who will be paying for LHCONE circuits in the long-term?

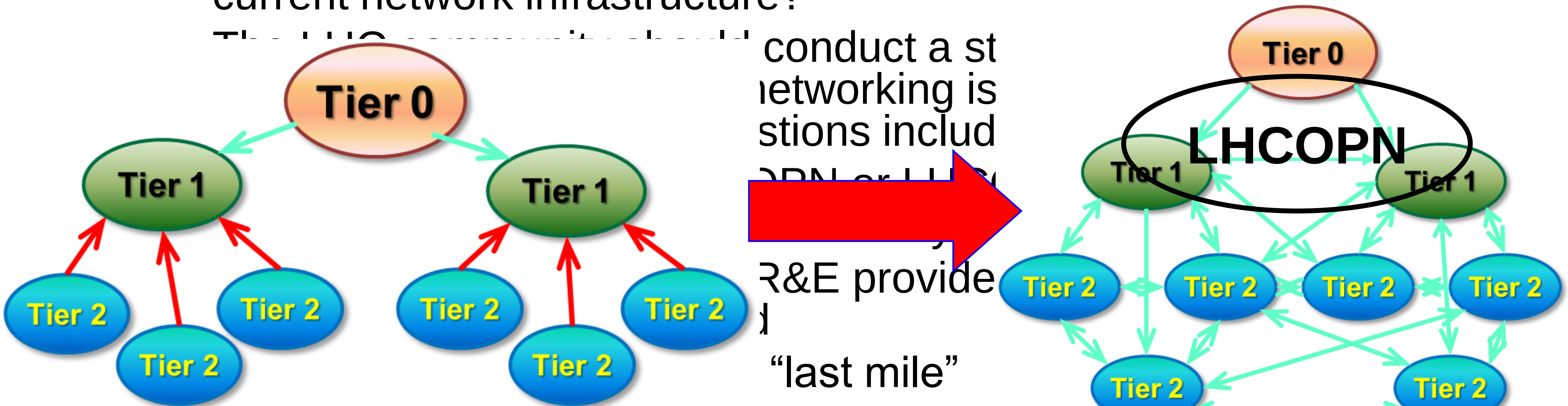
WLCG Facilities (2/2)

- With evolving computing models traffic flows change significantly (e.g. due to vanishing T1/T2 differences)
  - With the changes in the ATLAS CM, what would not work given the current network infrastructure?

conduct a st
ıetworking is
stions includ

R&E provide

“last mile”

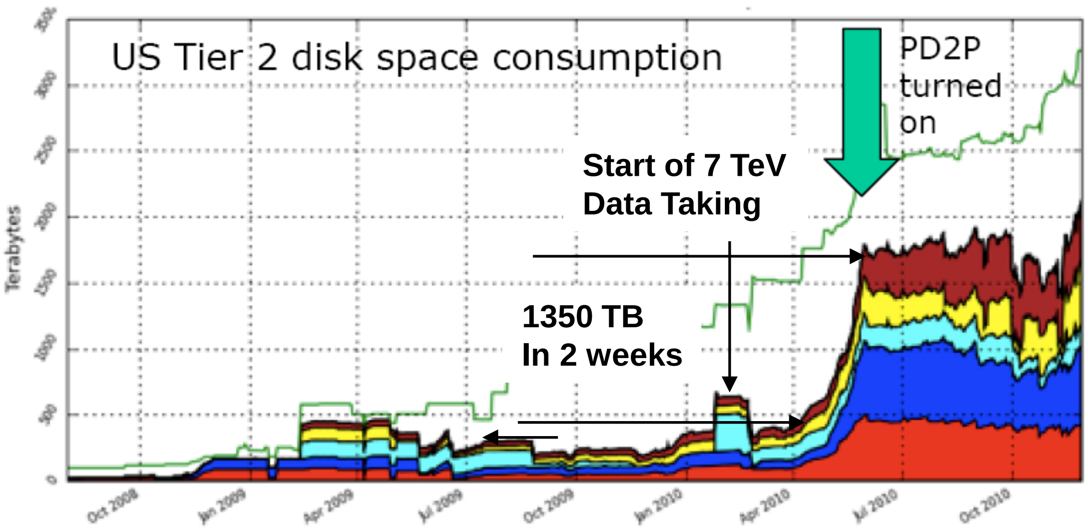


– Do we need LHCOPN if we have a g
private IP network?

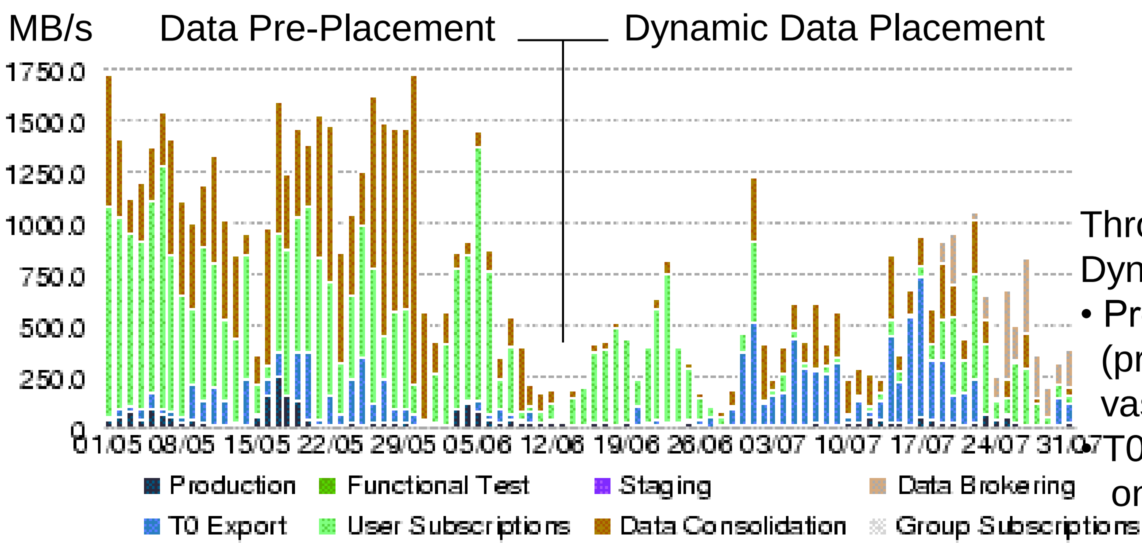
– Do we need LHCONE given we have the LHCOPN and today’s commodity IP network with increased capacities?

» Who will be paying for LHCONE circuits in the long-term?

CM Changes: Data Distribution - Moving from push to pull model



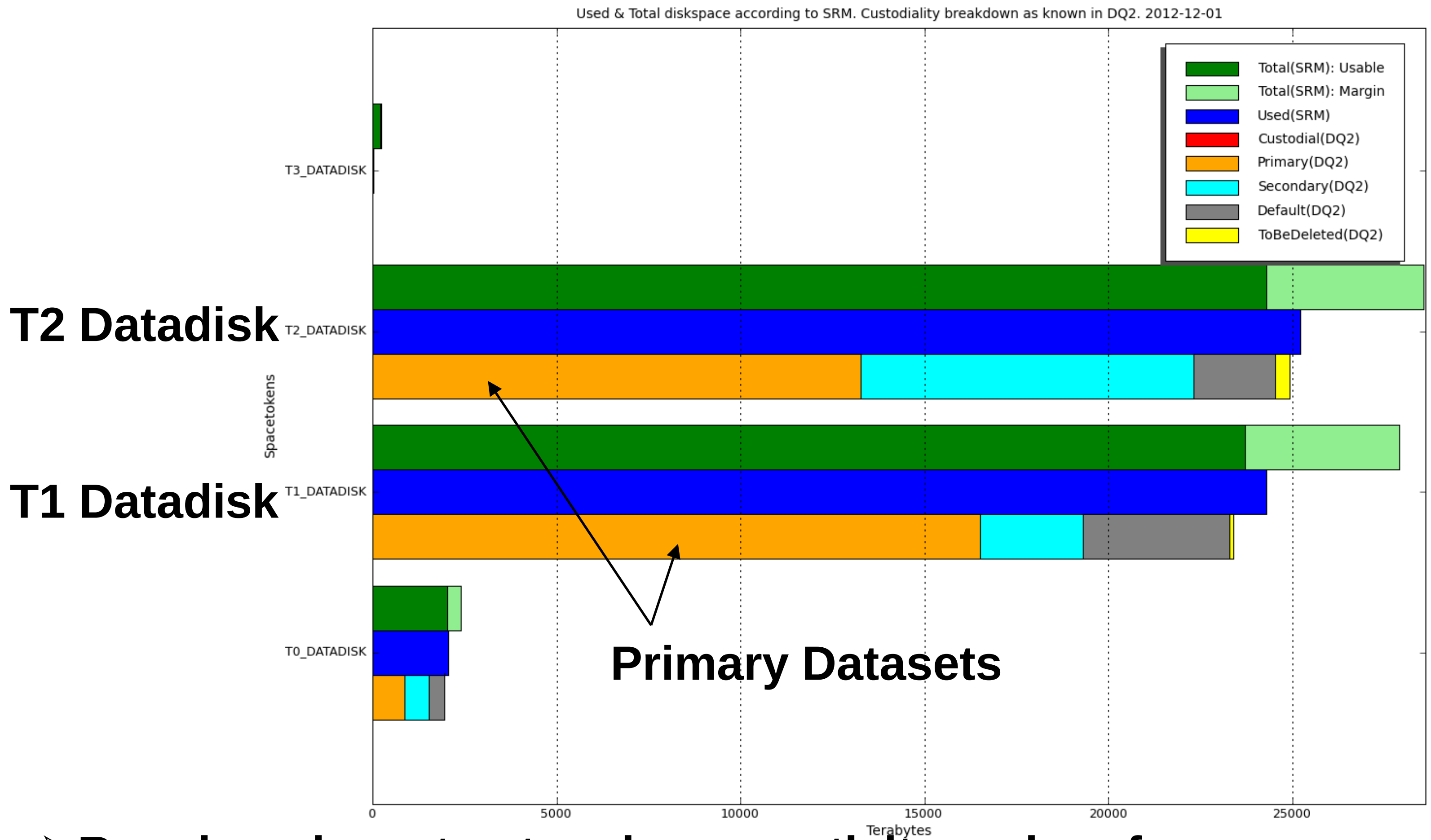
- We realized that, while disks filled up quickly, only a small fraction of the data that was programmatically subscribed to sites was actually used by analysis jobs
- Since July 2010: Workload management's dynamic data placement component (PD2P) sends data to Tier-2s for analysis on the basis of usage
  - ➢ Requires excellent networks



Throughput by Activity before and after Dynamic data placement (PD2P) turned on

- Predominant 'Data Consolidation' component (programmatic replication of ESD) vastly reduced after activation of PD2P
- T0 Export (blue) is sent from CERN to BNL T1 only

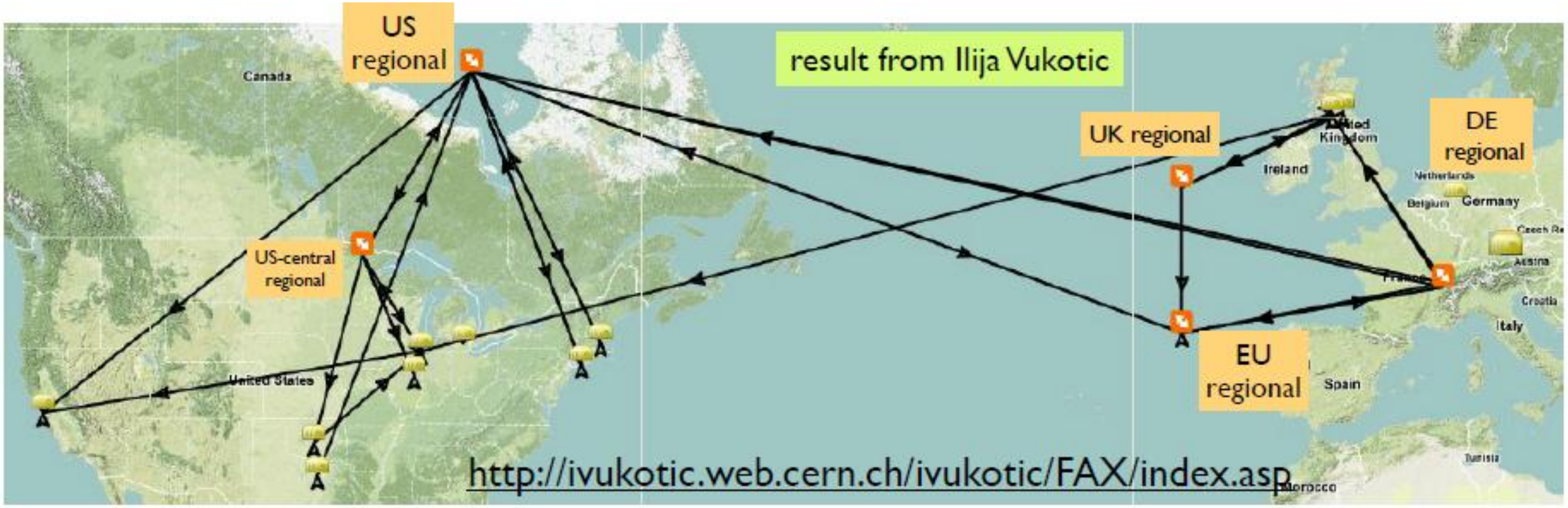
CM Changes: Data Placement – Using T2 disk to store primary datasets (not just as cache)



➢**Requires decent network connectivity and performance**

CM Changes: Federated Data Stores

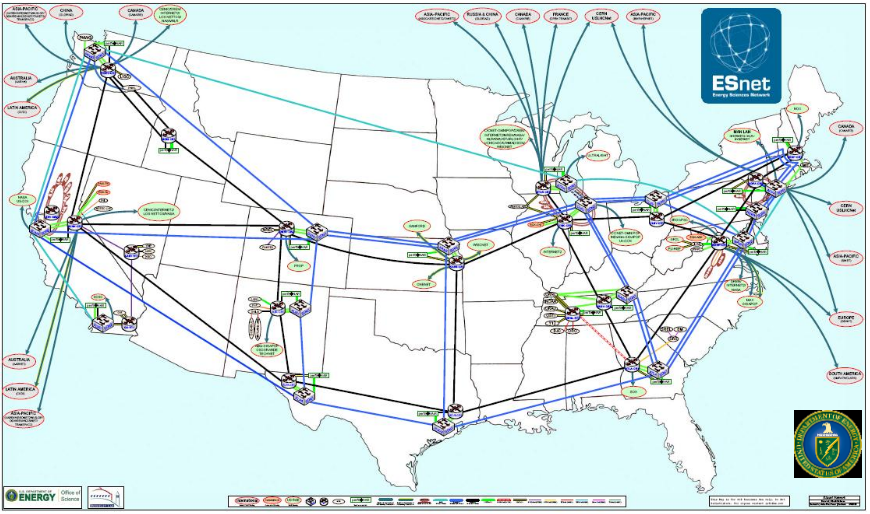
- **Common ATLAS namespace across all storage sites, accessible from anywhere**
- **Easy to use, homogeneous access to data**
- **Use as failover for existing systems**
- **Gain access to more CPUs using WAN direct read access**
- **Use as caching mechanism at sites to reduce local data management tasks**
- ➢ **Requires excellent networking**

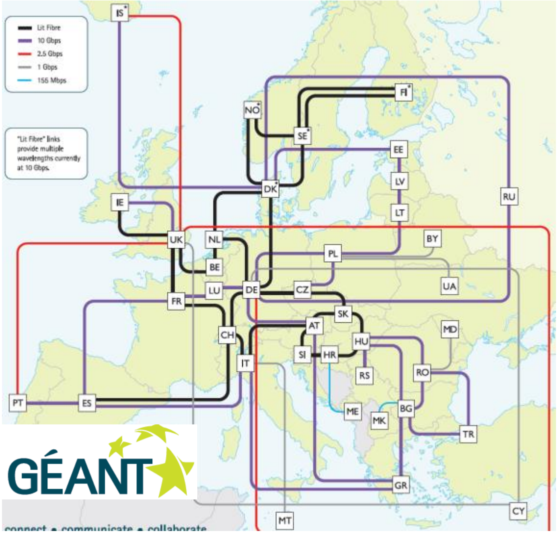


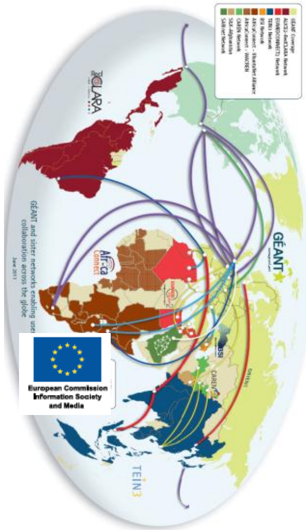
Much more in Rob Gardner's session

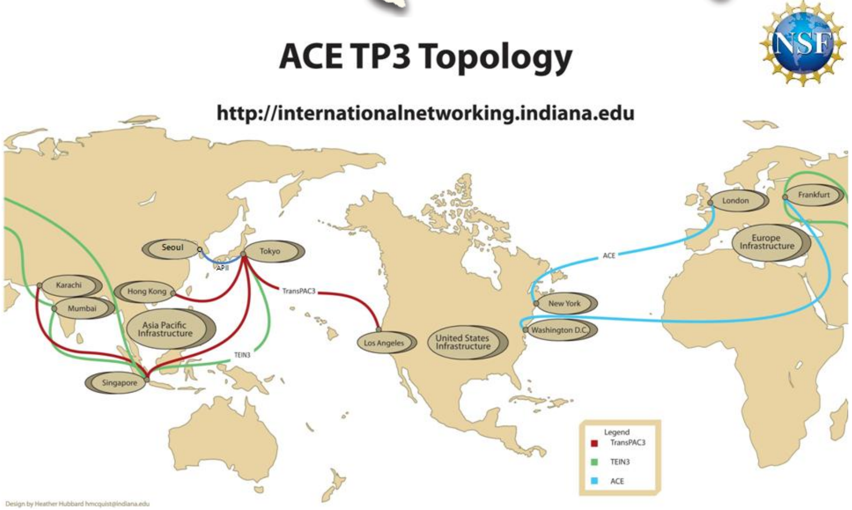
Regional and Global Connectivity

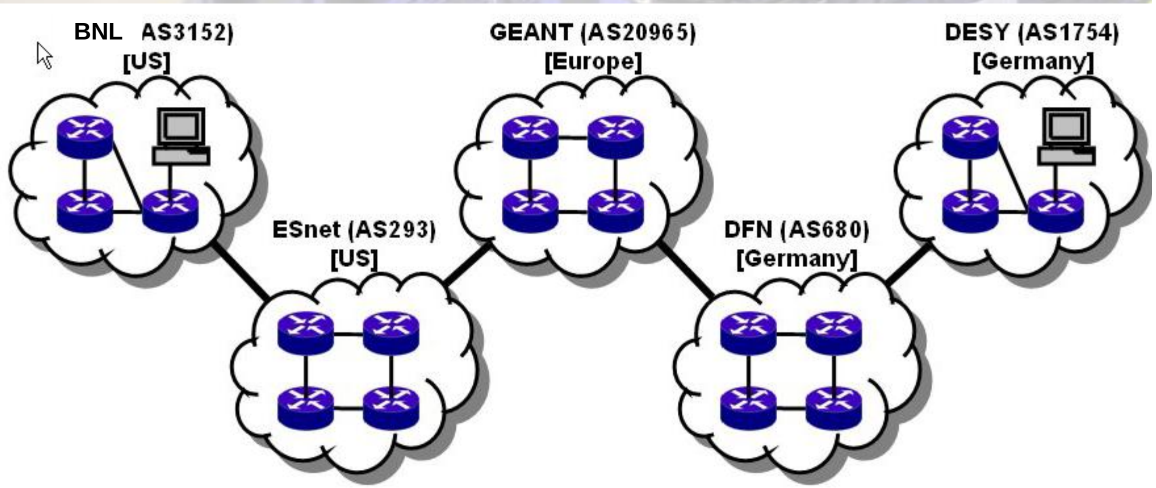
(only a subset is shown)







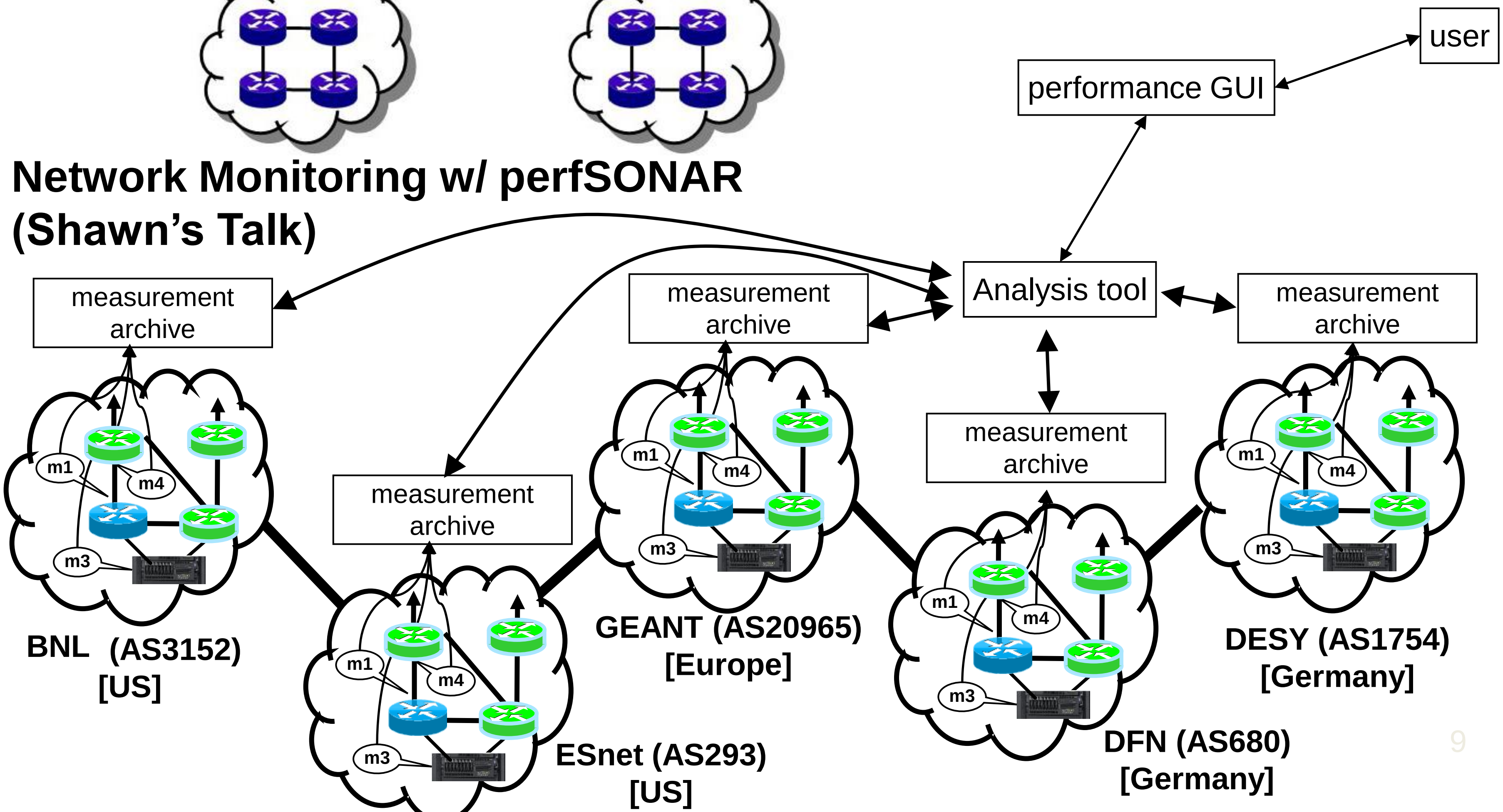




Collaborations in WLCG are inherently multi-domain, so for an end-to-end monitoring tool to work everyone must participate in the monitoring infrastructure

- Deployment being coordinated by WLCG Operations Coordination Team

Network Monitoring w/ perfSONAR (Shawn's Talk)



LHCONE – A new Network Infrastructure for HEP

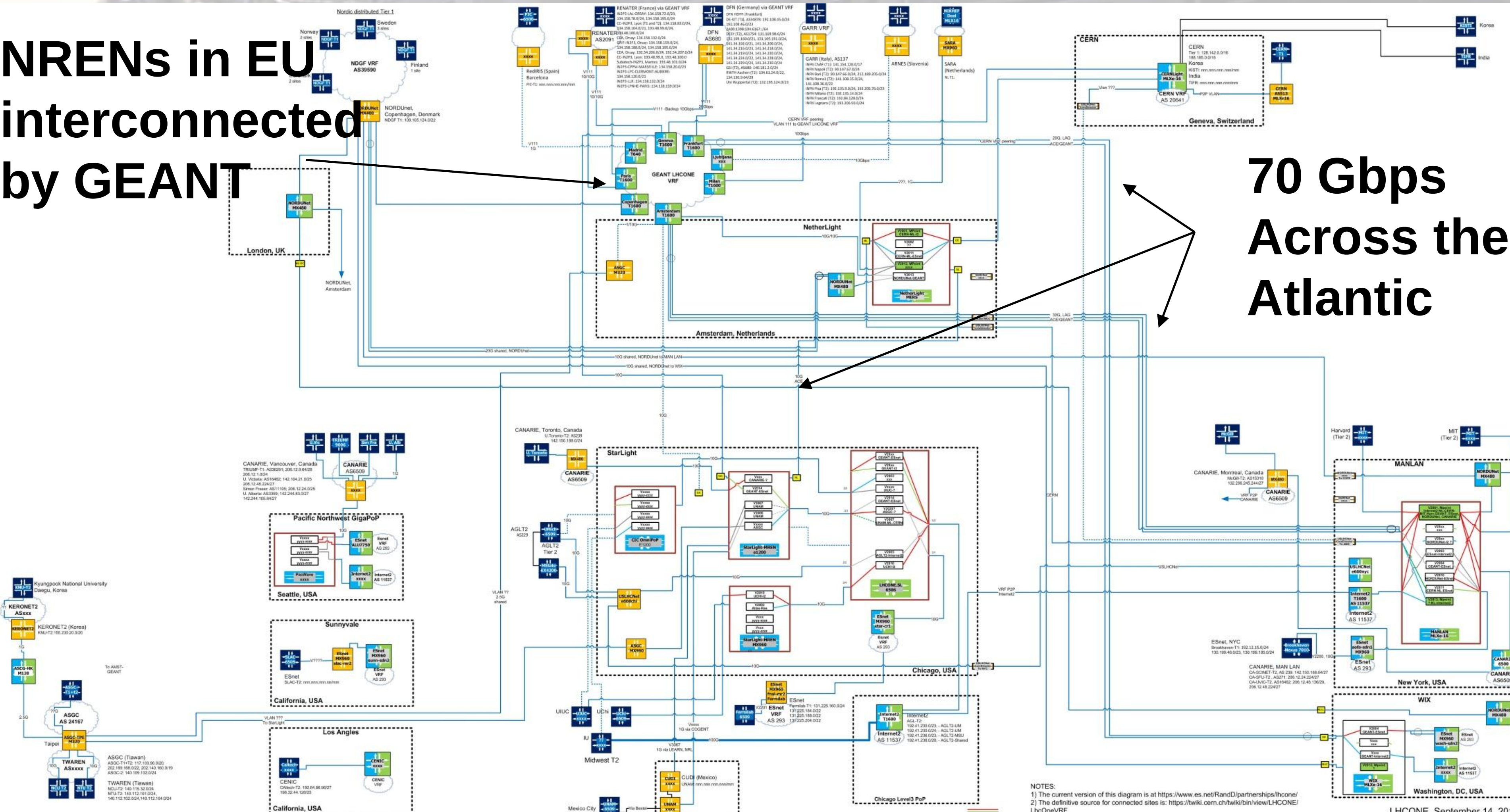
- **LHCONE was created to address two main issues:**
  - **To ensure that the services to the science community maintain their quality and reliability**
  - **To protect existing R&E infrastructures against very large data flows that look like 'denial of service' attacks**

- **LHCONE is expected to**
  - **Provide some guarantees of performance**
    - **Large data flows across managed bandwidth that would provide better determinism than shared IP networks**
    - **Segregation from competing traffic flows**
    - **Manage capacity as # sites x Max flow/site x # Flows increases**
  - **Provide ways for better utilization of resources**
    - **Use all available resources, especially transatlantic**
    - **Provide Traffic Engineering and flow management capability**
  - **Leverage investments being made in advanced networking**

LHCONE as of September 2012

NRENs in EU interconnected by GEANT

70 Gbps Across the Atlantic



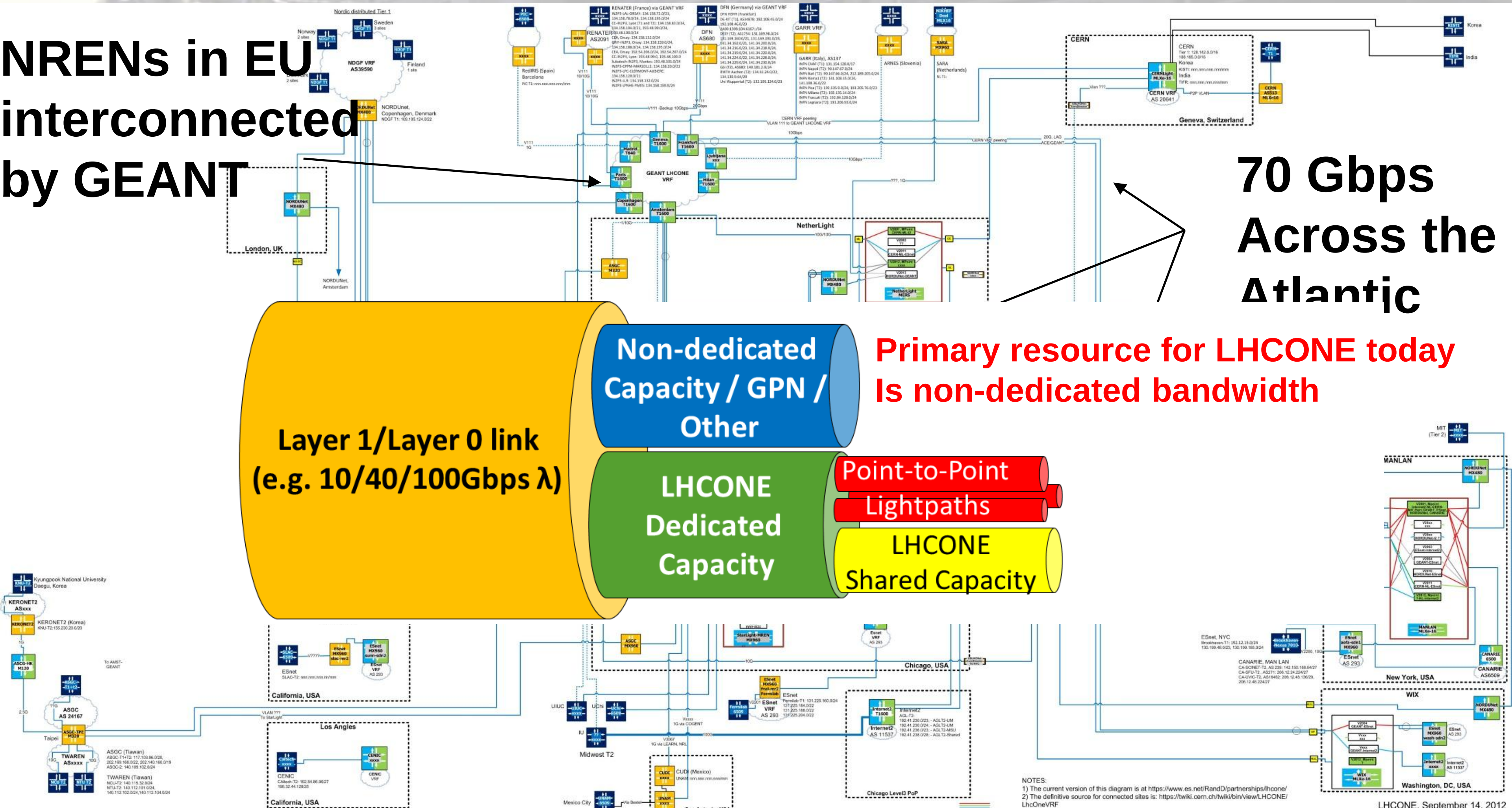
LHCONE is an 'Overlay' network infrastructure

B. Johnston/ESnet

LHCONE as of September 2012

NRENs in EU interconnected by GEANT

70 Gbps Across the Atlantic



Primary resource for LHCONE today
Is non-dedicated bandwidth

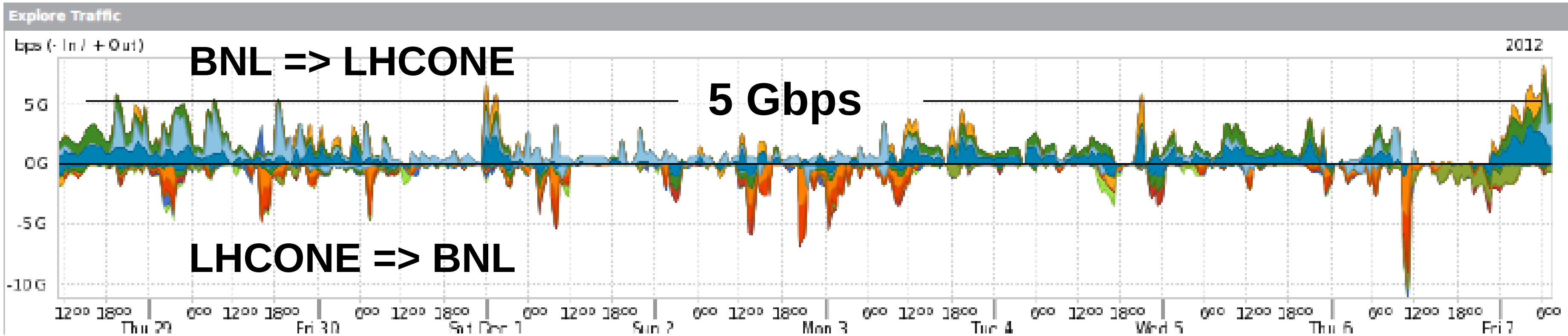
LHCONE is an 'Overlay' network infrastructure

B. Johnston/ESnet

mernst@bnl.gov

BROOKHAVEN NATIONAL LABORATORY

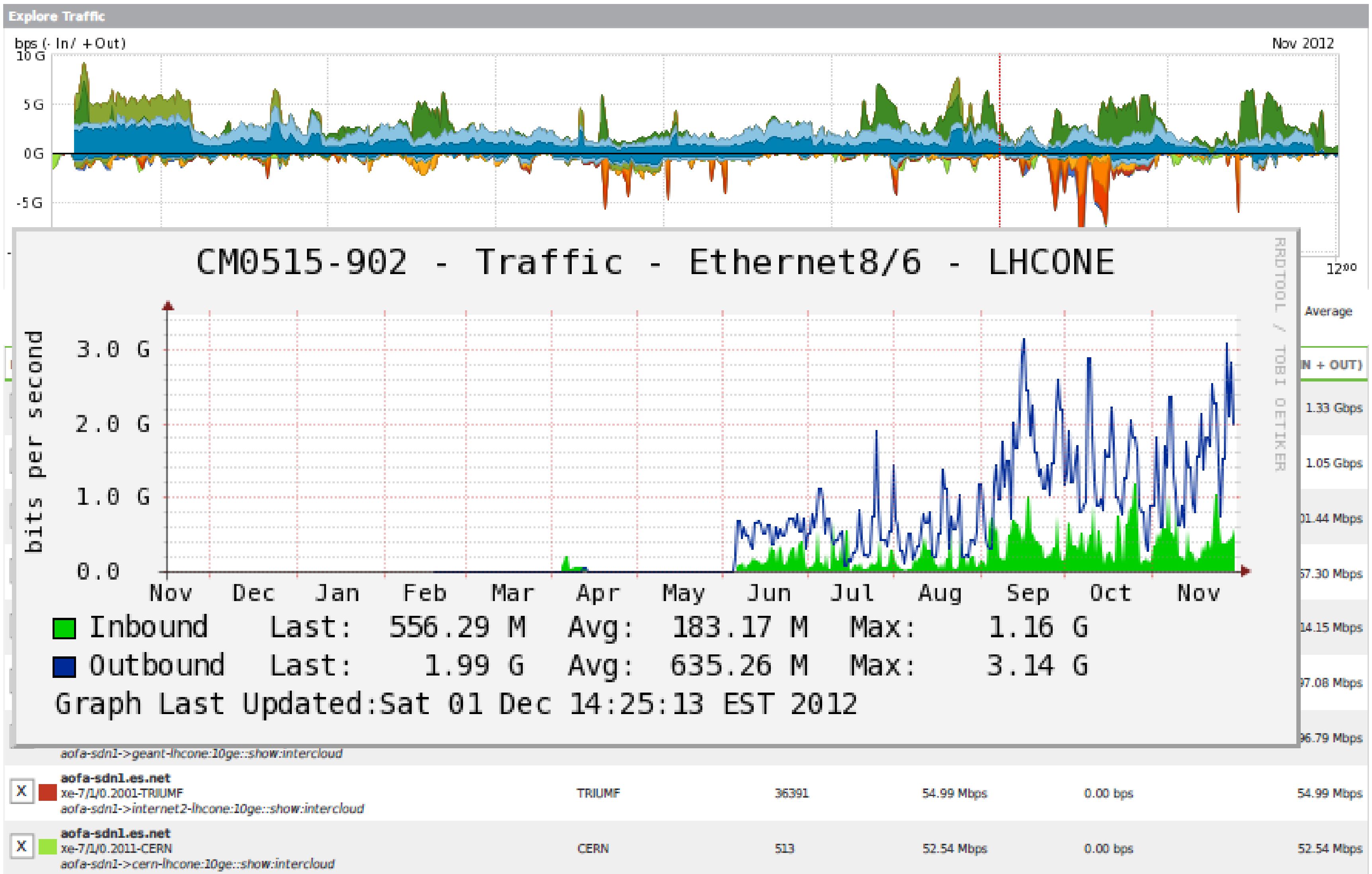
BNL LHCONE Traffic
(as reported by ESnet/Arbor)



Oct 29 **Nov 7**

INTERFACE	AS	ASN	IN	OUT	TOTAL (IN + OUT)
aofa-sdn1.es.net xe-7/1/0.2001-CANARIE-NTN *aofa-sdn1->internet2-lhcone:10ge::show:intercloud*	CANARIE-NTN	6509	126.47 Mbps	485.25 Mbps	611.73 Mbps
aofa-sdn1.es.net xe-7/1/0.2004-ASGARR *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	ASGARR	137	71.83 Mbps	447.09 Mbps	518.91 Mbps
aofa-sdn1.es.net xe-7/1/0.2001-BCNET *aofa-sdn1->internet2-lhcone:10ge::show:intercloud*	BCNET	271	105.52 Mbps	363.41 Mbps	468.93 Mbps
aofa-sdn1.es.net xe-7/1/0.2004-DFN-IP *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	DFN-IP	680	182.79 Mbps	0.00 bps	182.79 Mbps
aofa-sdn1.es.net xe-7/1/0.2004-DESY-HAMBURG *aofa-sdn1->geant-lhcone:10ge::show:intercloud*	DESY-HAMBURG	1754	161.01 Mbps	0.00 bps	161.01 Mbps
aofa-sdn1.es.net xe-7/1/0.2011-CERN *aofa-sdn1->cern-lhcone:10ge::show:intercloud*	CERN	513	144.03 Mbps	0.00 bps	144.03 Mbps

BNL LHCONE Traffic



LHCONE Operations Forum

- Meets bi-weekly on Mondays
  - Addressing higher level operational questions
  - Chaired by Dale Finkelson (Internet2) and Mike O'Connor (ESnet)
- Developing documentation (handbook)
  - What information is relevant?
  - Where the information is stored?
  - Who is responsible to keep the information up to date?
  - A reliable mechanism to broadcast information?

Evolving Application Requirements

- Application requirements extend beyond what's typically provided by the "network"
  - Maybe a controversial statement, but …
  - … we're now at the front edge of a paradigm shift driven by applications and enabled by emerging technologies allowing tighter coupling between big data, big computation and big networks
  - There is an opportunity that these requirements could be satisfied by a combination of intelligent network services, network embedded resources (processing & storage), in addition to our own processes and resources

Chin Guok's talk in this session

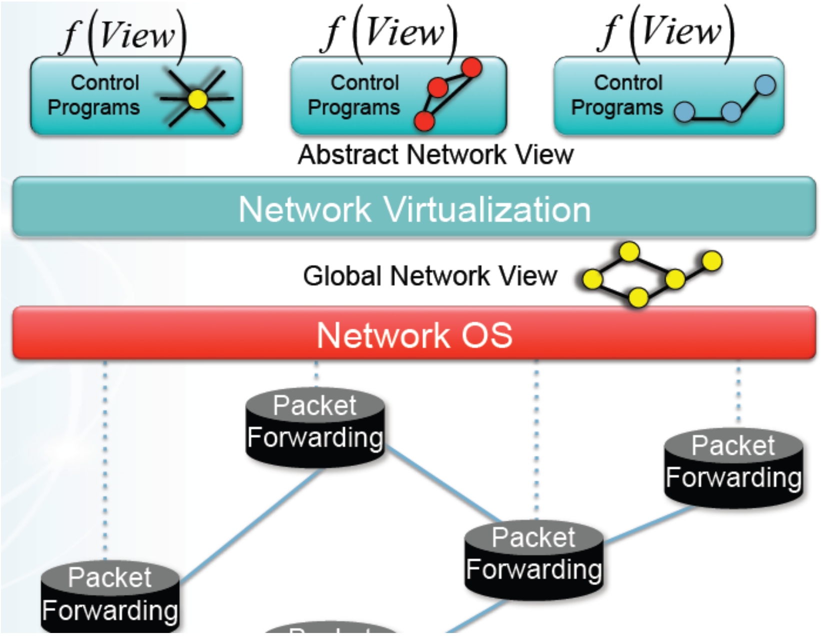
mernst@bnl.gov

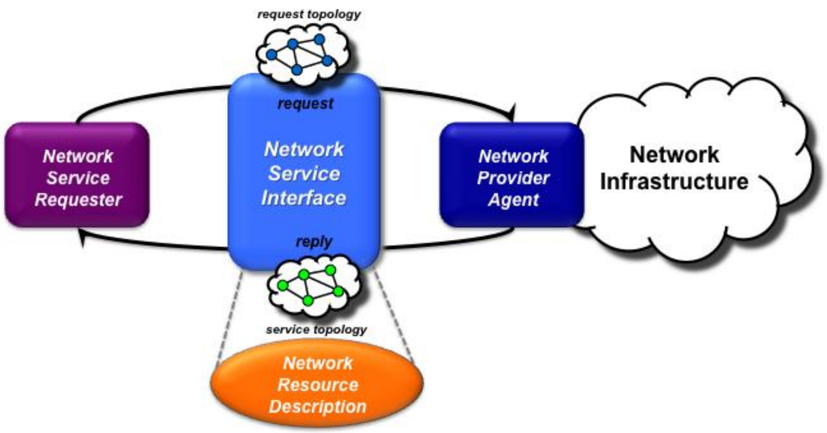
Point-to-Point Networking Activities

- **Build on national and regional projects for the basic Dynamic Circuit technology**
  - **OSCARS (ESnet, RNP), ION (Internet2), DRAC(SURFNet), AutoBAHN (GEANT and some EU NRENs)**
- **Extending into campus**
  - **DYNES (Switch and Control Server Equipment)**
- **Interfacing with LHC experiments/sites**
  - **Next Generation Workload Management and Analysis System for Big Data; based on PanDA (PI: A. Klimentov)**
  - **DYNES; primarily intended to improve T2/T3 connectivity**
  - **ANSE; new NSF funded project aiming at integration of Advanced Network Services with Experiments' data management/workflow SW (Implement Network Element in PanDA)**

Software Defined Networking (SDN) – This looks like a promising Technology

- SDN Paradigm - Network control by applications; provides an API to externally define network functionality
  - Enabler for applications to fully exploit available network resources
  - OpenFlow, a SDN implementation, widely adopted by Industry (Network Equipment Manufacturers, Google, Cloud Computing)





Network Services Interface (NSI) Framework

Inder Monga's talk at this week's GDB (14:30)

Jerry Sobieski's talk at PtP W/S on Thursday

A possible path forward

- We are close to a 2-year shutdown of the LHC machine
  - As our understanding of our computing needs evolves we have an opportunity to find out how we can benefit from the middleware that supports flexible & nimble operations, including the network as a fully integrated facility component
  - Work in close collaboration with the Network Providers
    - Define & implement Operational Interface between Users and Providers
      - E.g. the Working Group on LHCONE Operations has started activities
    - Besides providing production services most of the providers participate in pilot and research activities
      - with the motivation of providing new types of services to the user communities
      - A good opportunity for the LHC program with its applications to explore possibilities and influence directions developments are taking

Emerging Technology Discussion - Opportunities

- An opportunity to integrate capabilities of big data, big processing and big networks
- Core networking technologies and architecture concepts
  - SDN – network is re-programmable in response to changing application requirements or network traffic conditions through standardized API. OpenFlow is primary focus in this area
  - Dedicated Transfer Facility/Nodes – Data Center placed at location for most efficient data transfer for performance and economic benefit
  - Fast Networks – 100 Gbps (and terabit/s in the future) as enabler for new services
  - Next Generation Network Features
- Emerging Technologies in Application and User domain space
  - Cloud Computing – emergence of rapidly provisionable virtual machines with standard APIs for flexible access to computation power; include network as another virtualizable resource
  - Super Computer Center Evolution – interplay between cloud models and these infrastructures will allow increased access in flexible manner
  - Embedded Network Storage – storage tightly coupled with network
  - Advanced Security – "last mile" issues cloud/users drives network performance, virtualization with strong isolation and rapid reconfiguration