



# Introduction to Bandwidth on Demand Concepts

Inder Monga  
Chief Technologist, ESnet  
Co-chair, NSI working group

LHCONE P2P workshop  
Geneva, December 2012

# Myth Busters



- Bandwidth:
  - Capacity of a given network

10  
Gbps

100  
Gbps

*Network property*

- Throughput:
  - How many bits/second can your carried between any two points of the network

From my.es.net



*Application or end-to-end property*

# On-Demand



- Bandwidth, not throughput
- “On-demand” is defined by **time-scale**
  - Not O(days or months)
  - But O(seconds to minutes)
- Additional property
  - Software-controlled
  - GUI or Application initiated

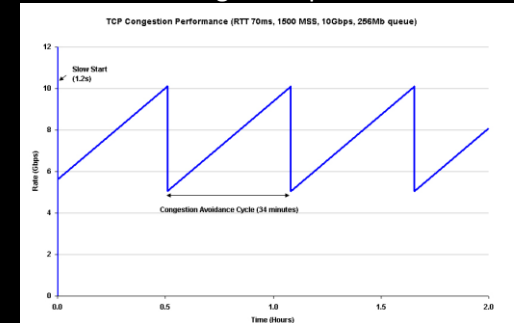
A screenshot of a web application titled "On-demand Secure Circuits and Advance Reservation System". The page includes a navigation menu with tabs like "Reservations", "Reservation Details", "Create Reservation", "User Profile", "User List", "Add User", "Attributes", "Institutions", "Authorizations", "Authorization Details", and "Login/Logout". The main content area is a "Reservation creation form" with various input fields for source, destination, path, bandwidth, description, start/end dates, and ports. There are also checkboxes for "Create Reservation" and "Production circuit", and a "Reset form fields" button. The page footer contains links for "Documentation", "ESnet", "Berkeley Lab", "Notice to Users", and "Contacts: Chris Guik, David Robertson".

# Some more truths



- Application can try engineer their end-to-end throughput
- Application currently can request point-to-point bandwidth
  - As supported in most installations today
  - Point-to-multi-point investigated
- Throughput ? Bandwidth
  - $\neq$
  - $\leq$

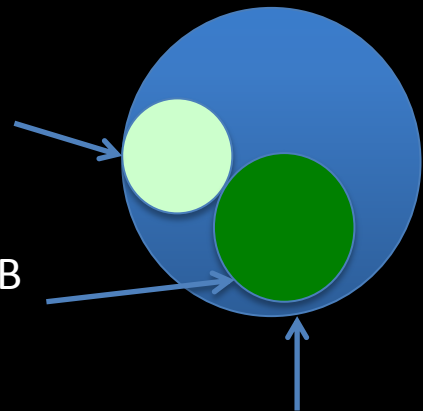
TCP congestion performance



Application A  
12 Gbps

Application B  
28 Gbps

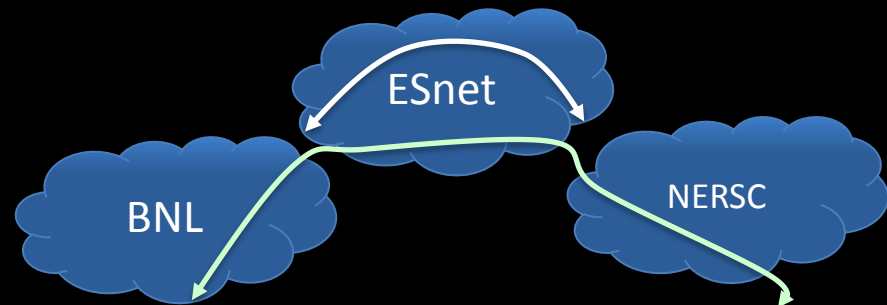
100 Gbps



# “Guaranteed” Bandwidth



- Bandwidth reserved for a single flow or set of application flows
  - Edge-to-edge
  - End-to-end
- Related concepts
  - Best Effort
    - Public Internet
  - Oversubscription
    - Aggregation

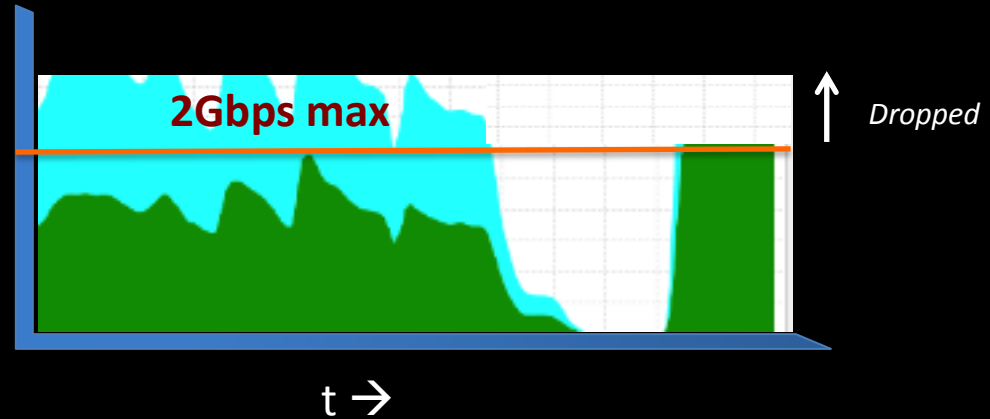


# Styles of Guarantee



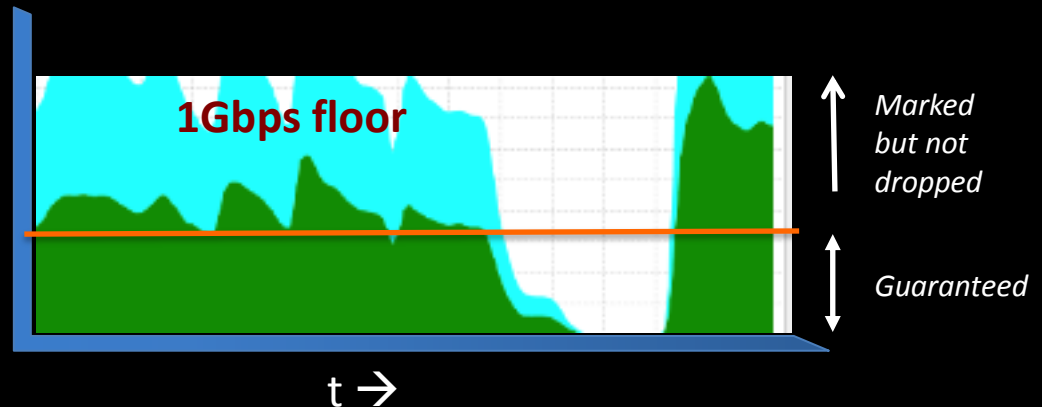
- **Strict or exact**

- Ceiling
- Peak



- **Floor**

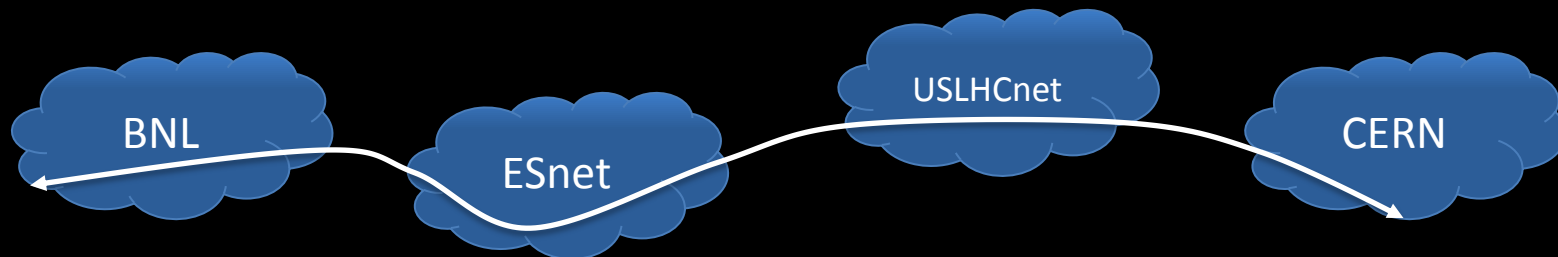
- Traffic above floor marked for discard if congestion



# Domains (networks)



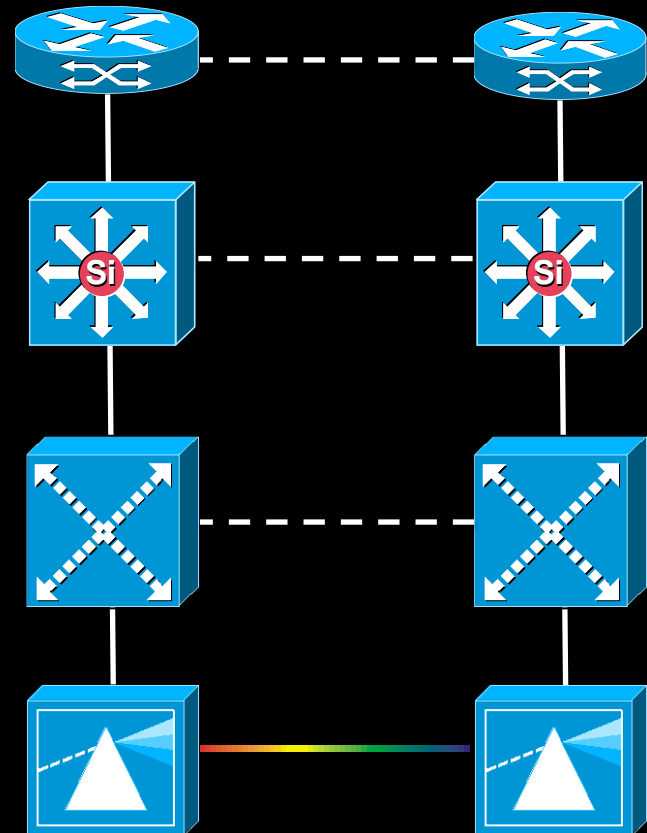
- Administrative boundaries are broad and can be arbitrary
- Single Domain
  - Single administrative management entity
- Multi-domain
  - Multiple administrative management entities



# BoD Spans Network Layers



- Layer 3/2.5: MPLS
- Layer 2: Carrier Eth.
- Layer 1: Lightpath/OTN
- Layer 0: Wavelength





# Scheduling



- On-Demand
  - **Time**: Now
  - **Duration**: Till when I please
- Scheduled
  - **Time**: Start-time, specific time in the future
  - **Duration**: End-time OR time duration

# Reservation



- Guaranteed Bandwidth reserved for a certain application at a certain “schedule”
- **On-demand** is equivalent to instant bandwidth reservation with no specified end-time

# Path-Finding



- Two ends of the network
  - Start = **A** point
  - End = **Z** point
- Path
  - Creating a physical or virtual circuit between points **A** and **Z** passing through multiple network devices
- Can be single or multi-domain

# Topology



- Map of interconnected devices
  - Intra-domain
  - Inter-domain
- Network and technology characteristics
- Can represent the physical layer limitations

# Path Computation Engine (PCE)



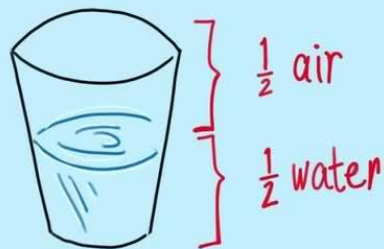
- Choose a path based on Topology, current state and certain criteria
- Current state
  - Available capacity and resource commitments
- Criteria
  - Bandwidth
  - Layer
  - Latency
  - Green energy
  - Your favorite metric

# Modify



- Ability to change certain parameters of the reservation
  - May or may not cause service outage
  - Focus on a *very limited* set of parameters
- Modify can be powerful for applications facing a varying workload
  - Can only be implemented at Layer 2 or above in the network
  - Duration is typically the most modified parameter where the connection remains the same
  - Bandwidth is the next most common

# Limitations



**technically,  
the glass is always  
full.**

# Physical Infrastructure

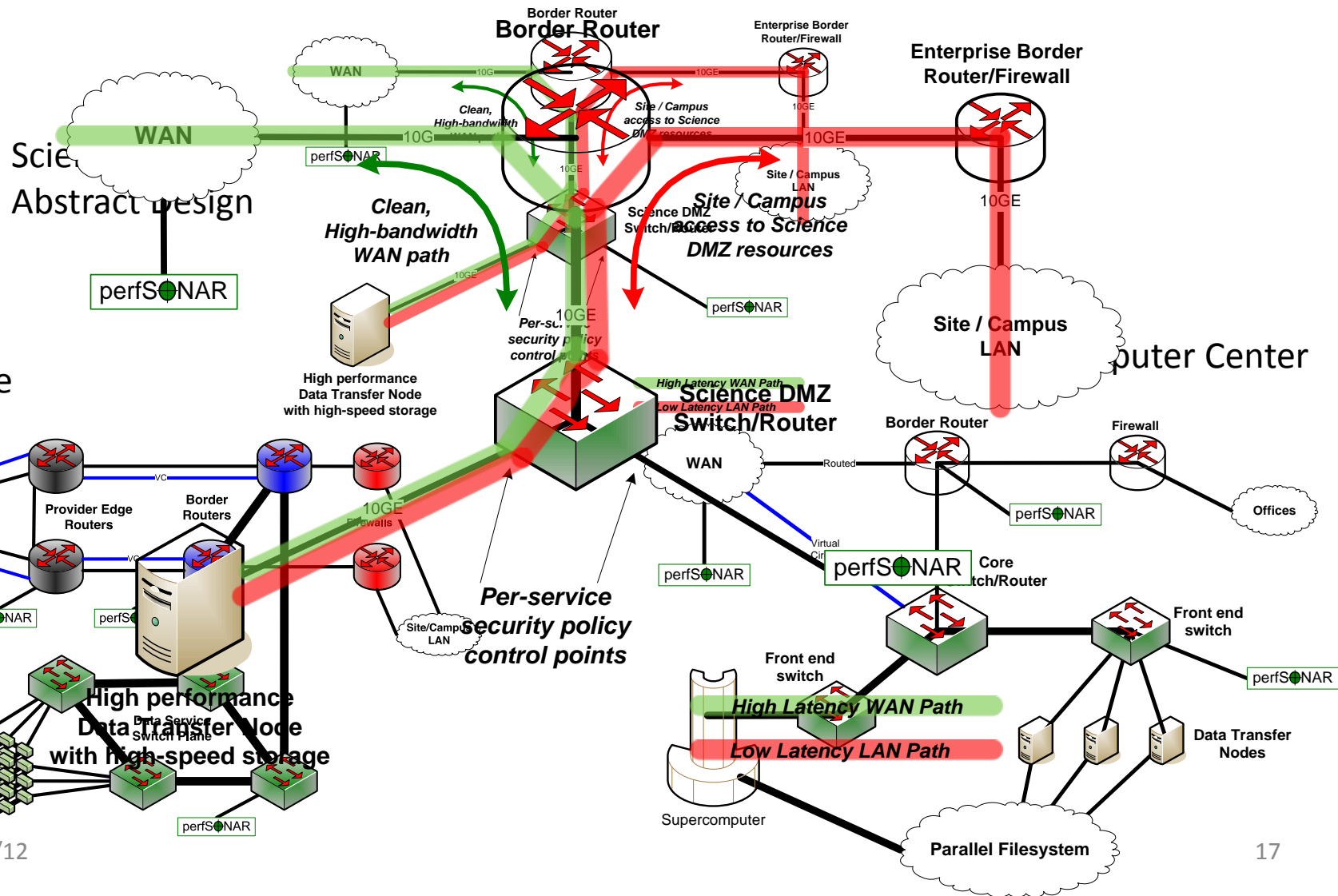


- Most applications not familiar with the topology might assume BoD capabilities that is not physically possible
  - For example:
    - 40G NIC on application host, 10G WAN connectivity of DC
    - Can't do 40G flows over nx10G connectivity



# Site Network Architecture

## Science DMZ design patterns



Data Site

Science Abstract design

Computer Center

# Blocking



- ‘Guaranteed Reservations’ of other applications might consume all resources
  - Even though the traffic profile indicates a lot of headroom
- Some of these reservations might be scheduled
  - Similar to hotel reservation
- An intermediate network domain might have resource constraints
  - Path finding needs to be intelligent
  - Path computation may take a lot of cycles if network is ‘reservation congested’

# Putting it all together



- Multi-layer, multi-domain, multi-vendor, On-Demand, Bandwidth Scheduling and Reservation **Service**
- **Service** is the most important aspect, **protocols** only are building blocks

# What else is needed for a service?



- Authorization and Authentication
  - Global federated system that works well with applications
- Service Level Agreements
  - What is the lowest common denominator across the multi-domain network?
- Service Definition
  - Consistent view of the end-to-end service
  - Homogenous service over heterogeneous technologies
- Monitoring and measurement
  - End-to-end as well
- Multi-domain debugging
  - How do you find errors, report them so they can be debugged and fixed?

# Top to Bottom view

## **Service Plane:**

End-to-end services, application-oriented



## **Management Plane:**

Provisioning and management of the network devices and as a system

## **Control Plane:**

distributed management of forwarding

## **Data Plane:**

carries real bits, electrical and optical

# Does this really work?

## An ESnet perspective



- OSCARS was introduced into ESnet as a proto-production service in early 2007, by mid-late 2008 was a supported production service
- The DICE collaboration (DANTE, Internet2, ESnet) had a prototype of the inter-domain control protocol (IDCP) working by late 2008 – early 2009
- The service has managed all LHCOPN Tier 0 – Tier 1 traffic since mid-2008 in ESnet, most T2-T1 traffic since 2009
- A lot of what OSCARS is about from ESnet's point of view is capacity management
- A lot of what OSCARS is about from the user's point of view is capacity guarantees

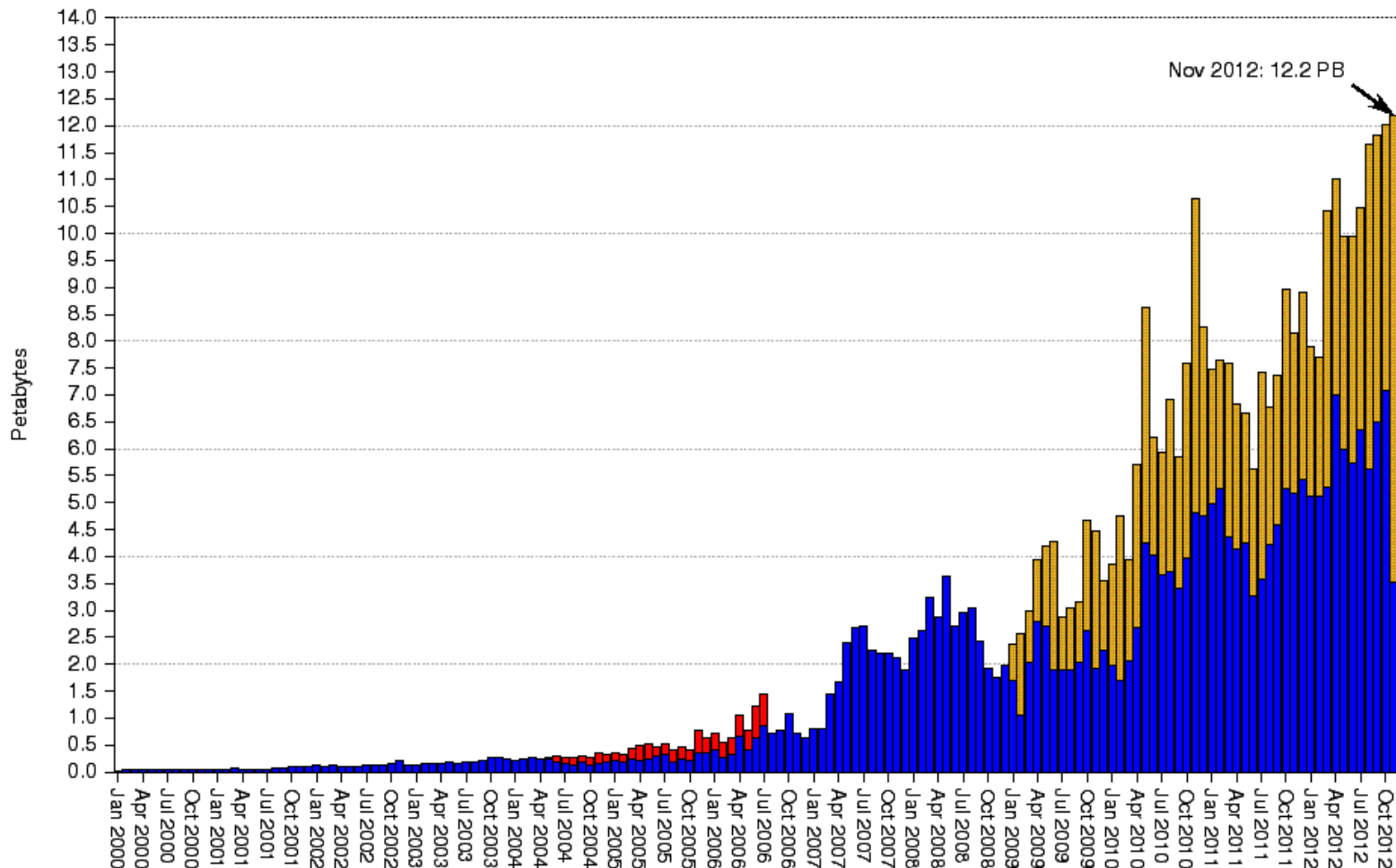
# OSCARS service carries 50% of big-data flows



## ESnet Accepted Traffic: Jan 2000 - Nov 2012

Petabytes/Month, Maximum Volume: 12.2 PB

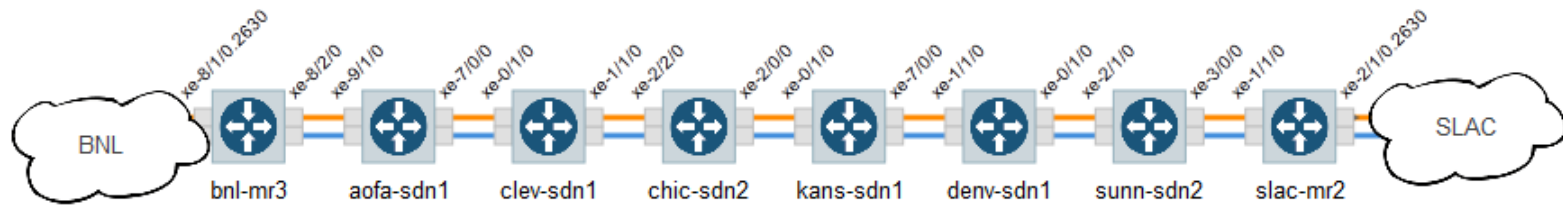
- Traffic Accepted
- OSCARS Accepted
- Top 1000 Host-Host Accepted



# Monitoring is Done Automatically



es.net-2708 | BNL ATLAS SLAC | 10-13-2011 To 10-14-2020



Circuit Traffic [Refresh](#)

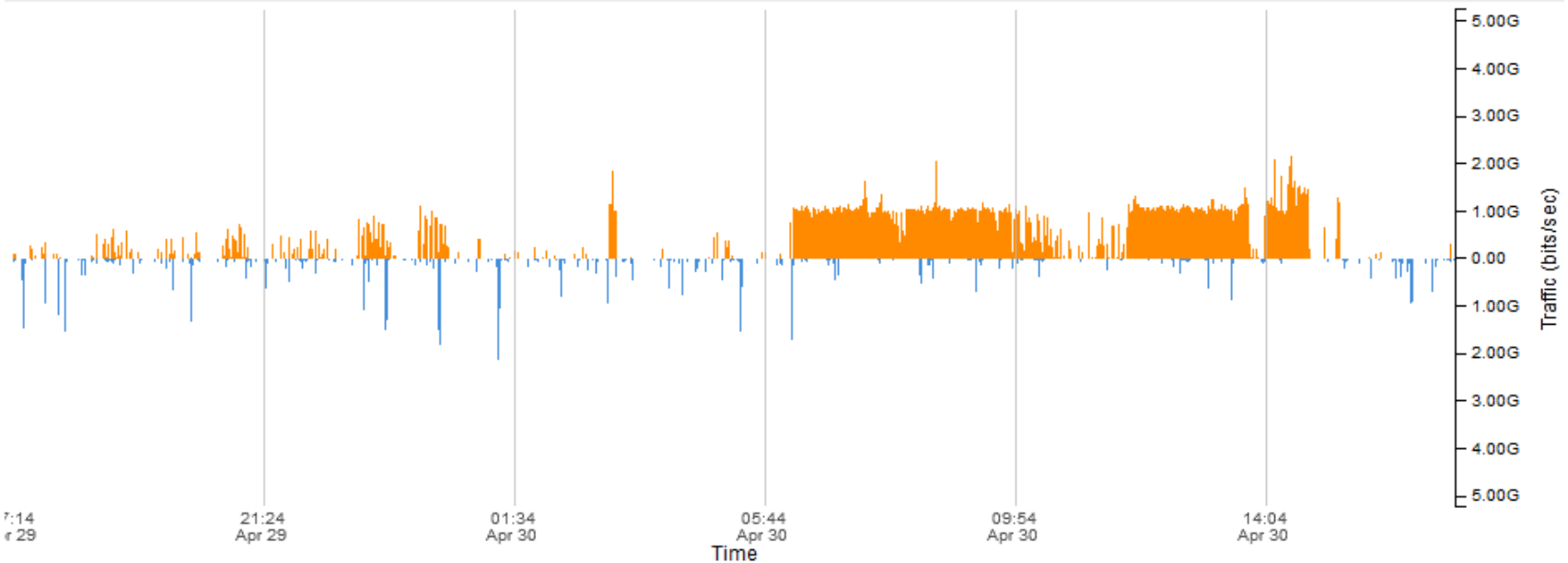
**1day** [Last hour](#)

[A to Z Delivered](#)

[Z to A Delivered](#)

Circuit Capacity 1.00G

[Reset zoom](#)



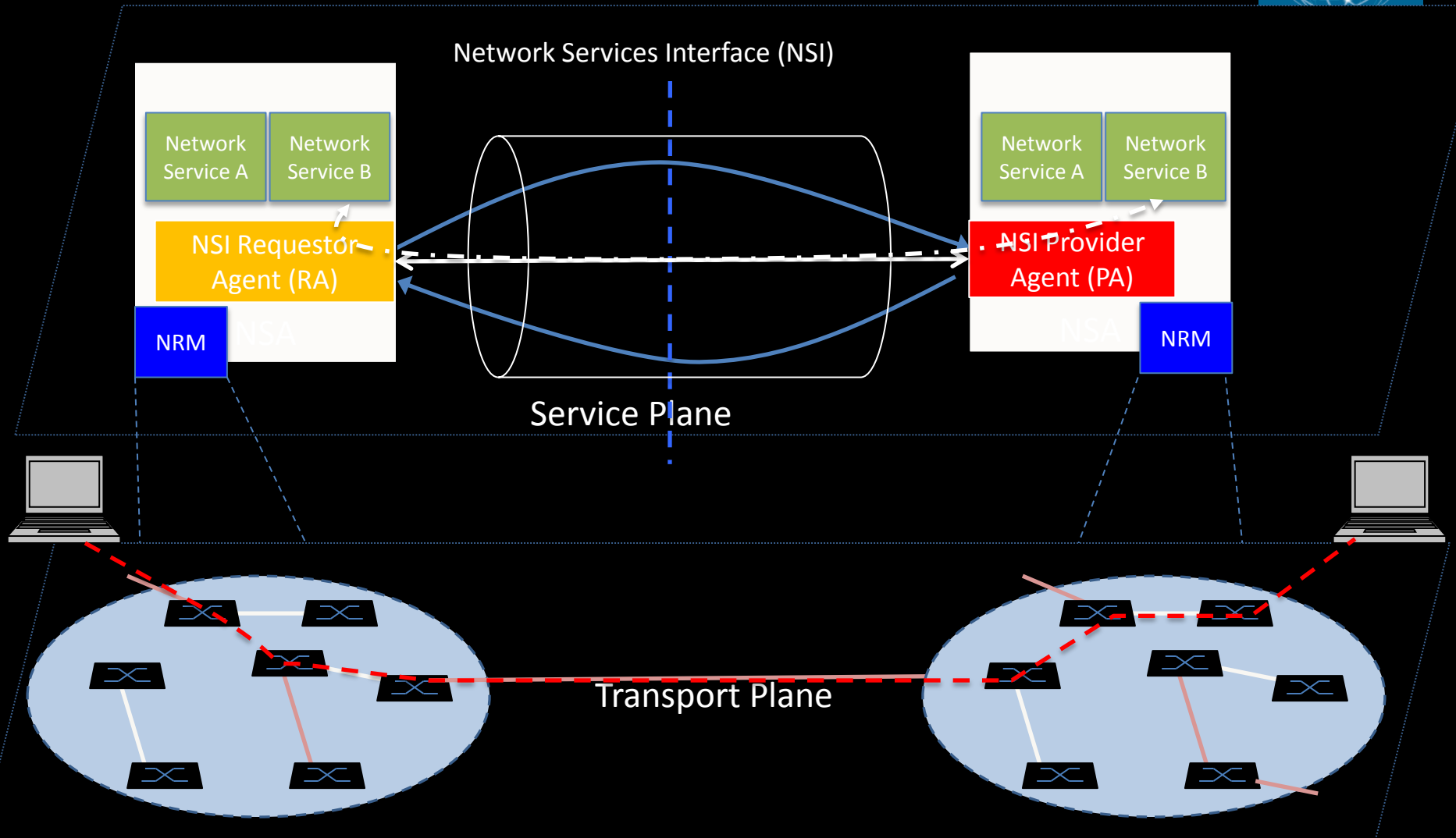


# Network Services Interface



- Standards process in OGF
- Strong participation from NREN community

# Network Service Framework concepts



\* Slides contain animation, does not show in pdf

NSA = Network Services Agent  
NRM = Network Resource Manager



Imonga at es.net

**QUESTIONS?**