NANOG 58

June 4th, 2013

Evolution of Services and Architecture at Internet2

Chris Spears

Sr. Network Planning Architect, Internet2

Internet2 Architecture

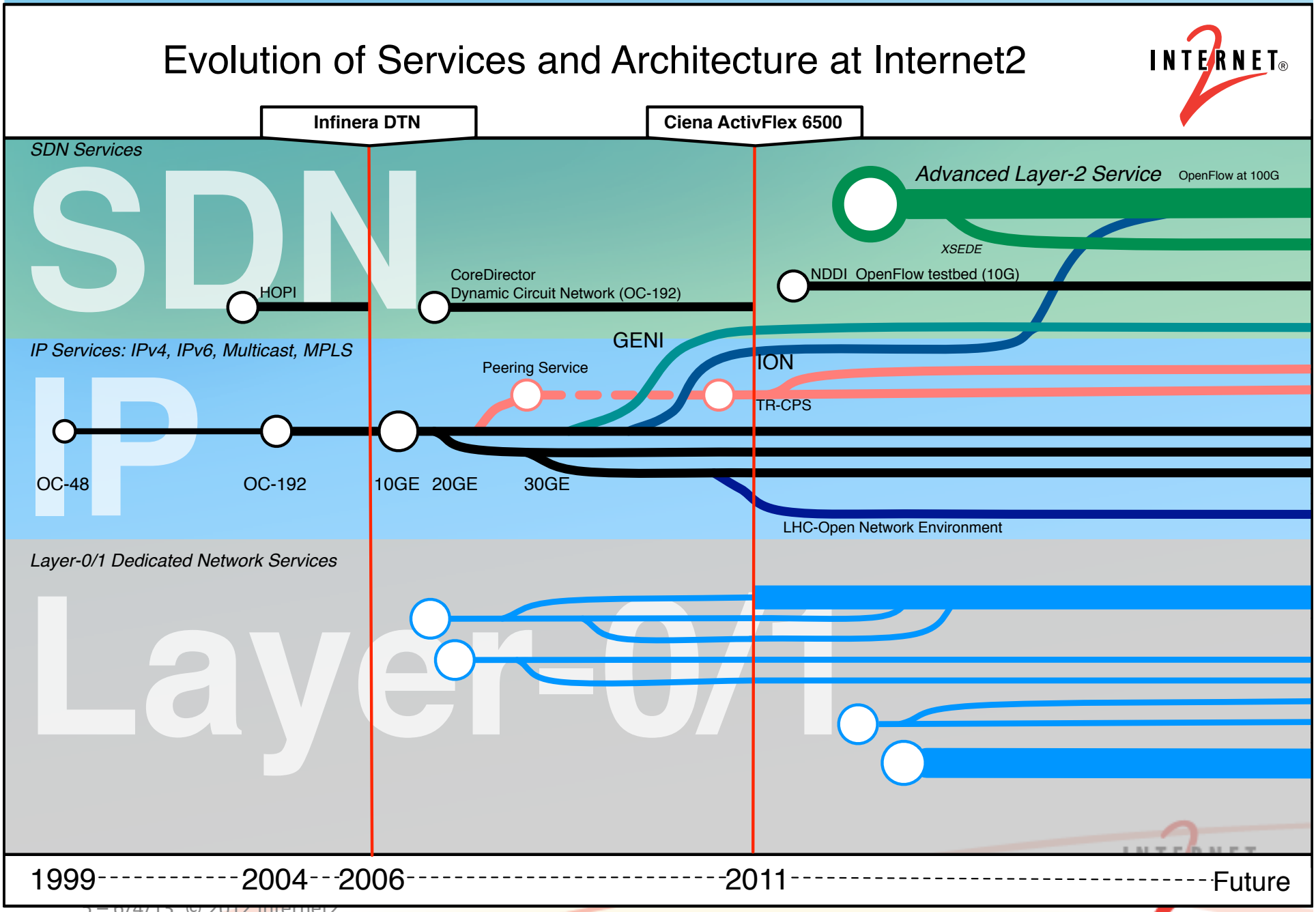
Internet2 Background

- National Research & Education Network (NREN)
- Formed in 1996 by 34 research universities
  - Need for a network focused on needs of researchers
- What is a Research Network?
  - http://www.nanog.org/meetings/nanog52/presentations/Monday/Oberman-NANOG-Research%20Networks-Final.pdf
- Different set of needs among researchers
- "Big Data" – and moreover "Big Science" – as driver
- More: internet2.edu/about

Internet2 Community Makeup

- 220 U.S. universities (over 4.5M enrollment)
- 60 leading corporations
- 70 government agencies
- 38 regional and state education networks (sponsored participants, K-12, etc)
- > 100 R&E partners, representing more than 50 countries

Evolution of Services and Architecture at Internet2



Internet2 Network Evolution: IP Services

- IP Services – multicast, IPv6, long ago…
- 1999 OC-48, partnership with Qwest
- 2004 OC-192c
- 2006 10G+, partnership with Level(3), Infinera DTN network
  - 100G (10x10) of capacity allowed growth outside of IP services
  - IP continued to grow, upto 30G inter-node capacity
  - New topologies to support research and experimentation
  - Peering service
- 2010 NTIA BTOP award
  - 2011 began building current optical infrastructure
  - Supports even greater scale of services, networks, and *applications*

Internet2 Optical Network Today

- 15,717 route miles of dark fiber (predominantly Level3, Zayo)
  - 51 optical add/drop sites (and growing)
  - 341 optical facilities across U.S.A.
- Ciena ActivFlex 6500 platform
  - 50GHz ITU grid spaced, 88-channels, DIA (directionless) in metros
- Partnered with ESnet at Layer-1
- 100G penetration:
  - 172 100G coherent DP-QPSK transponders deployed
  - Core L2/L3 interfaces: >70 (still adding nodes & links)
- 144 40G transponders – solely for OTU multiplexing of 10GE
- 100GE Firsts
  - Transcontinental (N.A.) - October 2011
  - Transatlantic – June 2013

Internet2 Network Evolution: Layer-0/1

- Custom Network Infrastructure supporting science & research
  - ESnet - http://es.net/
  - NOAA - http://noc.nwave.noaa.gov/
  - GENI - http://www.geni.net/
  - LHC-Open Network Environment - http://lhcone.net/
  - XSEDE - https://www.xsede.org/
- Shared Infrastructure partnerships
  - Dark fiber, spectrum, OTU multiplexers
- Dedicated Infrastructure for Internet2 networks

Internet2 Network Evolution: SDN

- History of experimenting with new technologies
- Dynamic network services, driven by software….. a.k.a. SDN
  - Hybrid Optical-Packet Infrastructure (HOPI)
  - GENI - Slice-able, experimental network substrate
  - Dynamic Circuit Network (DCN) –
    - 22 Node OC-192 network, Ciena CoreDirector
  - ION – ***I***nternet2-***ON*** demand circuits
    - Dynamic pseudowires, speaks OSCARS IDC protocol
  - NDDI OpenFlow testbed
  - Advanced Layer-2 Service (discussed later in this presentation)
    - 100GE backbone, 18 nodes; 25 by end of summer 2013
    - Built as an Open Exchange

Internet2 Network Infrastructure Topology

March 2013

IP router node (Layer 3 services)
Advanced Layer 2 Service (Layer 2 services)
Optical add/drop facility (Layer 1 services)

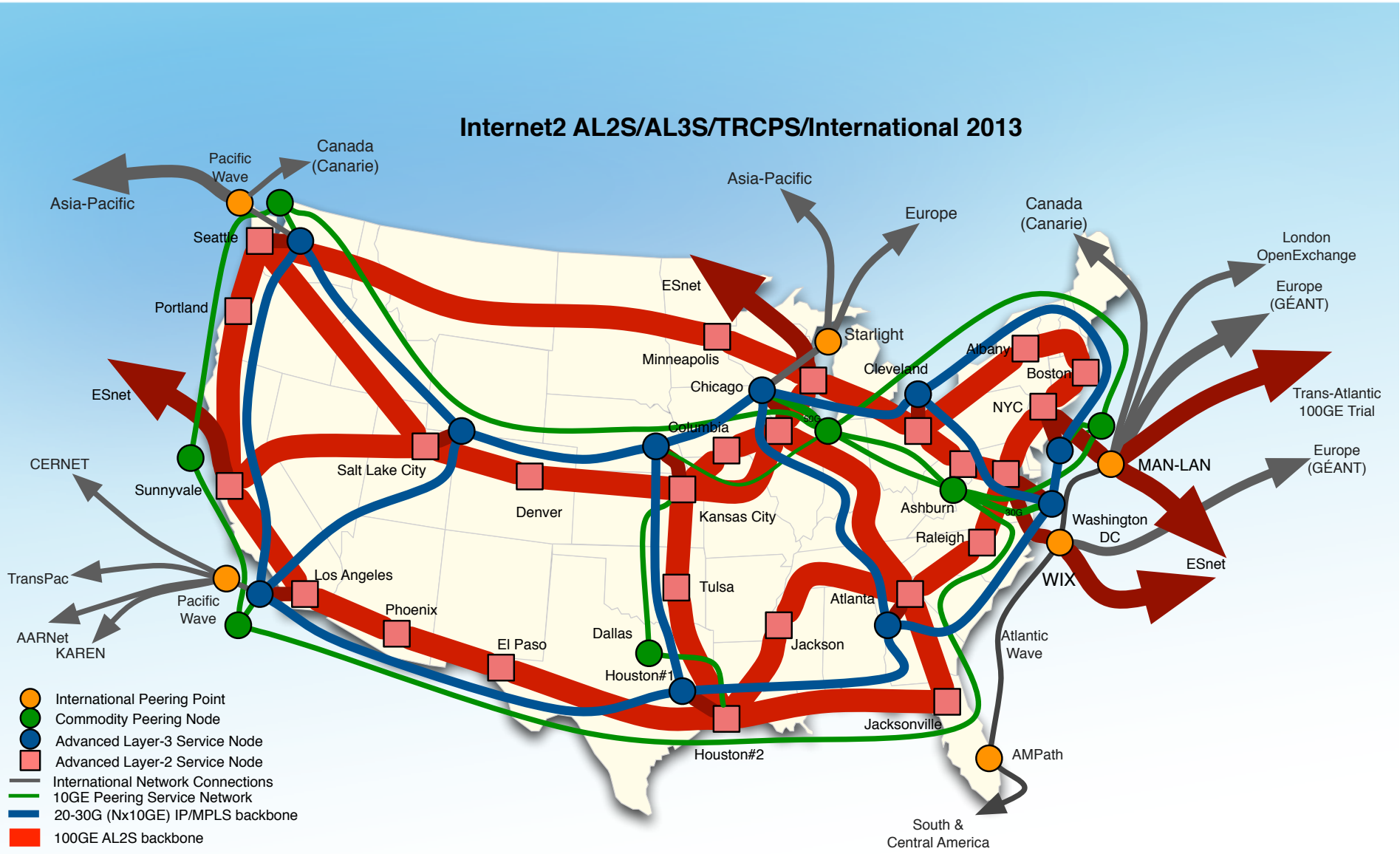
Internet2 Network *by the numbers*

17 Juniper MX960 routers in support of Advanced Layer 3 IP and Peering Network
21 Brocade and Juniper switches in support of Advanced Layer 2 Service Network
49 custom colocation facilities
250+ amplification racks
15,717 miles of newly acquired dark fiber
8.8 Tbps of optical capacity
100 Gbps of hybrid Layer 2 and Layer 3 capacity
300+ Ciena ActiveFlex 6500 network elements
2,400 miles partnered capacity with Zayo Communications in support of the Northern Tier region

IN SUPPORT OF U.S.UCAN

NETWORK PARTNERS

ciena
cisco
INDIANA UNIVERSITY
infinera
JUNIPER NETWORKS


Internet2 AL2S/AL3S/TRCPS/International 2013
Pacific Wave
Canada (Canarie)
Asia-Pacific
Seattle
Portland
ESnet
CERNET
Sunnyvale
TransPac
Pacific Wave
AARNet
KAREN
Los Angeles
Phoenix
El Paso
Salt Lake City
Denver
Dallas
Houston#1
Houston#2
Minneapolis
Chicago
Columbia
Kansas City
Tulsa
Jackson
Atlanta
Jacksonville
Asia-Pacific
Europe
Starlight
ESnet
Canada (Canarie)
London OpenExchange
Europe (GÉANT)
Trans-Atlantic 100GE Trial
Europe (GÉANT)
Albany
Cleveland
Boston
NYC
MAN-LAN
Ashburn
Washington DC
Raleigh
WIX
ESnet
Atlantic Wave
AMPath
South & Central America
International Peering Point
Commodity Peering Node
Advanced Layer-3 Service Node
Advanced Layer-2 Service Node
International Network Connections
10GE Peering Service Network
20-30G (Nx10GE) IP/MPLS backbone
100GE AL2S backbone

Edward Balas

Manager, Software Engineering, Indiana University GlobalNOC

SDN at Internet2

Origins of SDN at Internet2

- Historic projects have pushed for programmatic network control
  - HOPI
  - ION
  - NDDI
  - AL2S
- Motivated by desire to more quickly create new virtual networks
  - Give members ability to directly create
  - Remove unneeded provisioning delays
  - Concerned about quality and control
- The historic use case = National Exchange Fabric

Internet2 Innovation Platform

- Key Ingredients
  - Big Pipes(100g) with minimal aggregation
  - Open the network stack for non-vendor driven innovation
  - Domain expert involvement in developing new services
  - Means to separate experiment and production
- Goal
  - Create an improved experience for R&E Users
  - We want to find applications that better fill the pipes with science
  - Make it easier to move data so folks can focus on discovery
- Enabler
  - OpenFlow 1.0 today, 1.3 someday
  - Any cross platform SDN techniques we can find in future

Innovation Platform

- TestLab
  - Mixed vendor 8 switches and 6 test PCs
  - MEMS switch to control layer1 topology
  - Jenkins based test automation system
- NDDI
  - 5 NEC PF5820 switches
  - 10GE core
  - Ring Topology
- AL2S
  - 15 Brocade MXLe-16, 3 Juniper MX960
  - 100GE core
  - Partial mesh topology
  - OESS used to provide point and click provisioning

OS³E The Open Science, Scholarship & Services Exchange

BETA

Feedback | Admin

Workgroups > Home > **Circuit Details** Workgroup: GRNOC

Summary

Description
losa-salt test

Bandwidth
0 Mbps

Type
Local

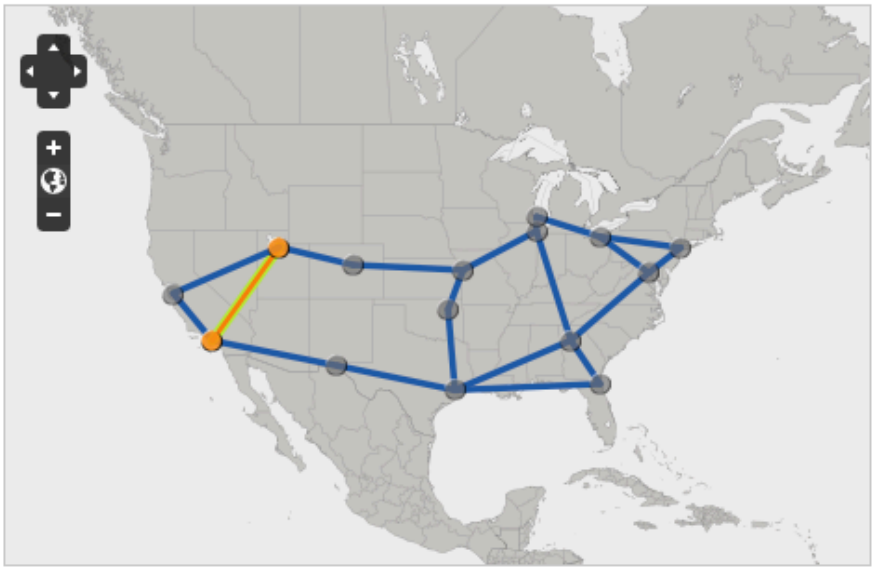
Status
active

Endpoints

Interface	VLAN
sdn-sw.losa.net.internet2.edu - e15/2	601
sdn-sw.salt.net.internet2.edu - e15/2	601

Edit Circuit

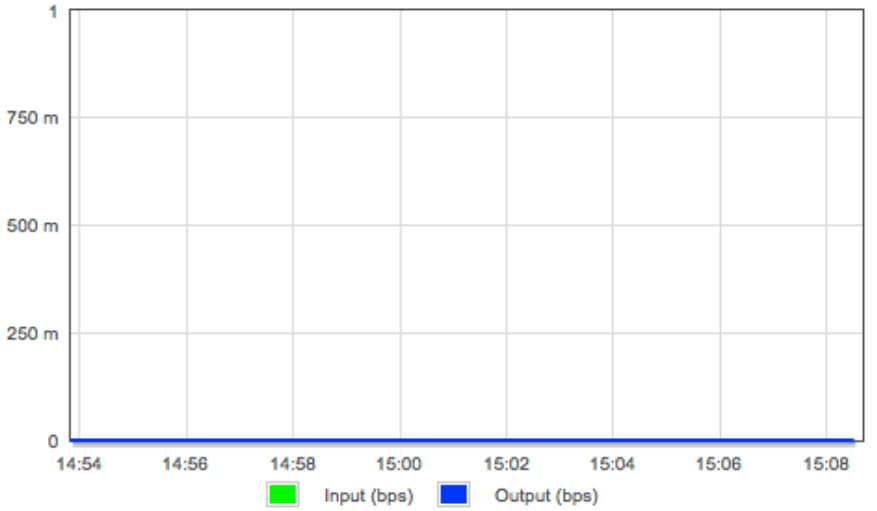
Remove Circuit



Circuit Endpoint *Unused Endpoint*
Primary Path *Secondary Path* *Available Link* *Down Link*

Traffic | User Events | Network Events

sdn-sw.salt.net.internet2.edu - e3/1



Past 10 Minutes

What the Innovation Platform is NOT

- Just a playground
  - We do encourage responsible experimentation
  - It is an involved process to get into the AL2S network
- Just a testbed
  - We do at-scale operation of OpenFlow Apps
    - Some are experimental
    - Others are considered production grade
  - There are risks that experiments will interfere with production traffic
  - We try to manage risk with technology and policy

Multi-Tenancy a key feature

- Running 2 separate production and research platforms too costly
- Goal
  - Run a production platform with a virtual SDN net built on top
  - Support multiple simultaneous applications / controllers
  - Minimize trust placed in applications
- Approach
  - Separate flow control by Switch / Port / Vlan Tag
  - Use FlowVisor etc to logically "**slice**" or partition the network
  - Each app gets a limited and non-overlapping "**flowspace**"
  - Customers define which apps can control their port's flow space
  - Traffic Engineering a concern in some cases
- Implementation
  - Evaluating FlowVisor
  - Exploring other options including use of overlay networks

Internet2 Innovative Application Award

- http://www.internet2.edu/network/innovative-application-awards.html
- Goal
  - Encourage development of SDN applications
  - Improve scientific data movement at 100G
  - Engage *.edu to developing network scale applications
- Sponsored by Juniper, Ciena and Brocade
- Modest (up to 10k) cash prize to support effort
- Apps must work on AL2S and be licensed modified Berkeley

OpenFlow Issues and Lessons Learned

Availability for last 6 months

- For last 6 months, **including** maintenance windows
  - 99.69% for circuits
  - 99.25% for nodes
- Single worst node event was 25 hr outage
  - Bug in controller related to corner case
  - Only alarm triggered was ISIS adjacency alarm
  - Prolonged by initial miss-diagnosis
- Circuit availability issues
  - Having 100G optic issues with some vendors
  - Non-trivial number of optical system upgrades during this period

Vendor Issues

- Partial support for specification
  - Match and act on both layer2 and layer3
  - Proper barrier support
  - Support for actions in hardware
- Stability problems
  - Various issues
- Performance issues
  - Port down event generation
    - > 1.5 sec for some!
  - Modify-State processing speed
    - ~100 / sec
  - Total number of supported rules
    - ~2,000

Protocol Issues

- OpenFlow 1.0 is not the best protocol
  - Too much left to vendor interpretation
- Inherent DoS risks, if you don't trust your north bound
  - No rate limits on packet in
  - No rate limits on packet out
  - Table space exhaustion
- Feature set lacking to replicate existing services
  - No viable QoS
  - No TTL decrement
  - No push / pop VLAN or MPLS tags
- Reacting to network events requires controller round trip
  - Fast Failover Port groups in 1.3 should be a win

Testing effort for last 6 months

- Vendor interaction still fairly intense
- Perform full system testing when we get a new code revs
  - Vendor code
    - 3 vendors , 6 total releases
    - 20 – 50 hours per test
  - Application Code
    - 1 vendor (us), 4 releases
    - 30 - 40 hours per test
  - Hypervisor/slicer code
    - 1 vendor, 2 releases
    - 20 – 50 hours per test
- More than 50% of lab time is spent helping vendors
- At least 50% of an FTE

Management Network

- Today use central controller cluster over dedicated management network
  - side band on the OSC channel
  - Limited bandwidth
- Management network disruptions impact OpenFlow operation
  - If shared fiber plant, OpenFlow restoration blocks on management network restoration
  - Traffic continues to flow, just black hole on failed link
  - Distributed controller architecture can help
    - Requires you mimic a routing protocol to avoid dependency
  - Port groups in 1.3 can also help

WAN OpenFlow Application Architecture

- Robust WAN capable apps are hard
- There is a reason for separating IGP from EGP
- Do WAN apps need to control the interior path?
  - If yes, do you trust to developer to perform TE
  - If no, how do you constrain bandwidth
- Considerations
  - Ability to function with partial management network disruption
  - Restoration performance
  - System complexity and cost of operation / testing

Future Challenges

- Working together to develop better testing regiments
- Migrating to 1.3 to get sought after features
- Developing better sw ecosystem
  - Truly distributed controllers
  - Standard north bound interfaces
- Refining our operations capability
  - Better monitoring and troubleshooting
    - What is the craft interface to an OpenFlow device or app?
  - Operations support team structure
  - With WAN multi-tenancy, where to Engineer Traffic?

Questions?