

Deploying and Operating a 100G Nationwide SDN WAN

Luke Fowler

<luke@iu.edu>



GlobalNOC
Global Research Network Operations Center



Background

Network Development and Deployment Initiative (NDDI)

 10g x 6 NEC PF8520 nationwide network

Advanced Layer 2 Service (AL2S)

 100G multi-vendor BTOP funded national network

OESS

 Dynamic circuit management with OpenFlow backend

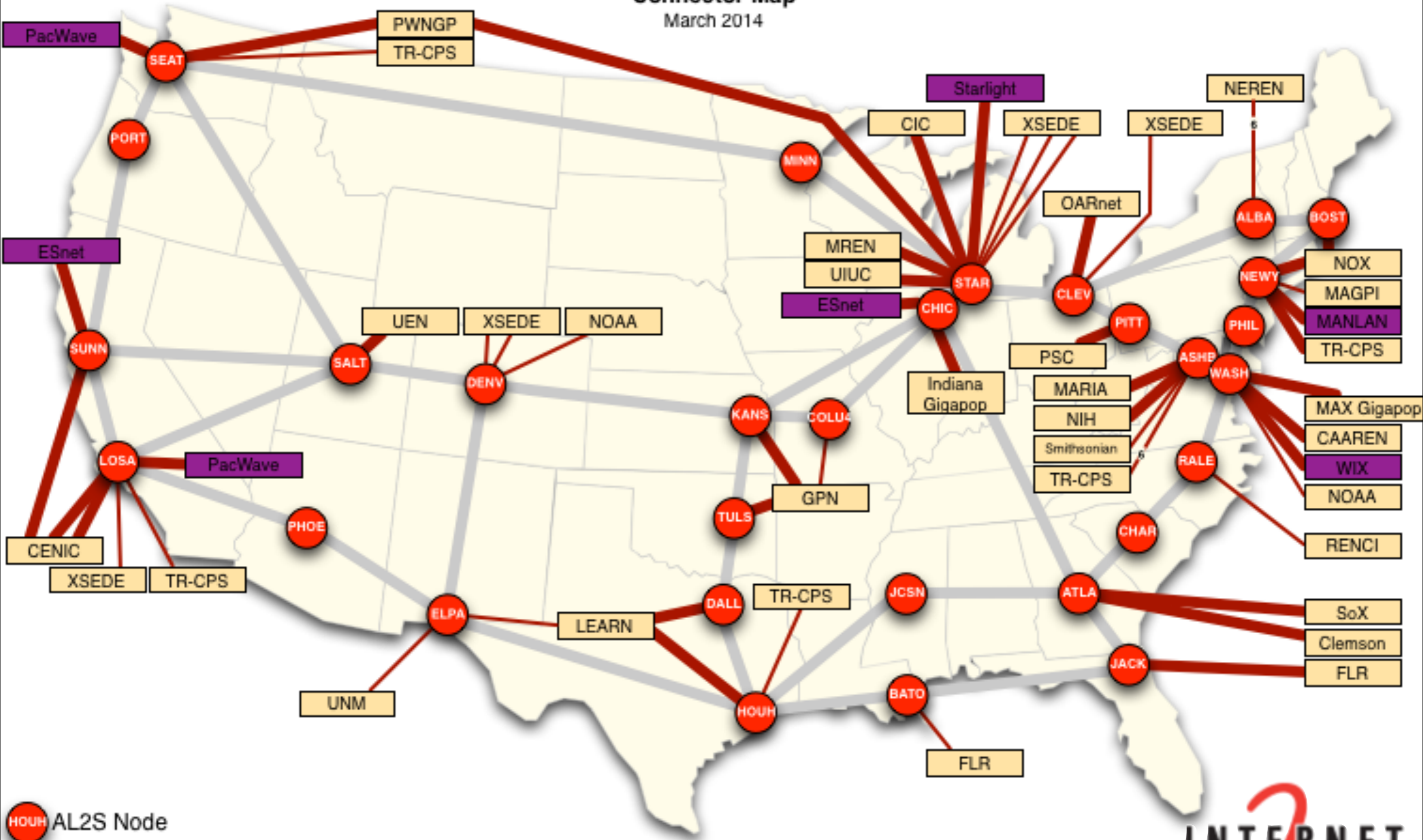
Virtualization: Want to support multiple OpenFlow apps (e.g. GENI apps)

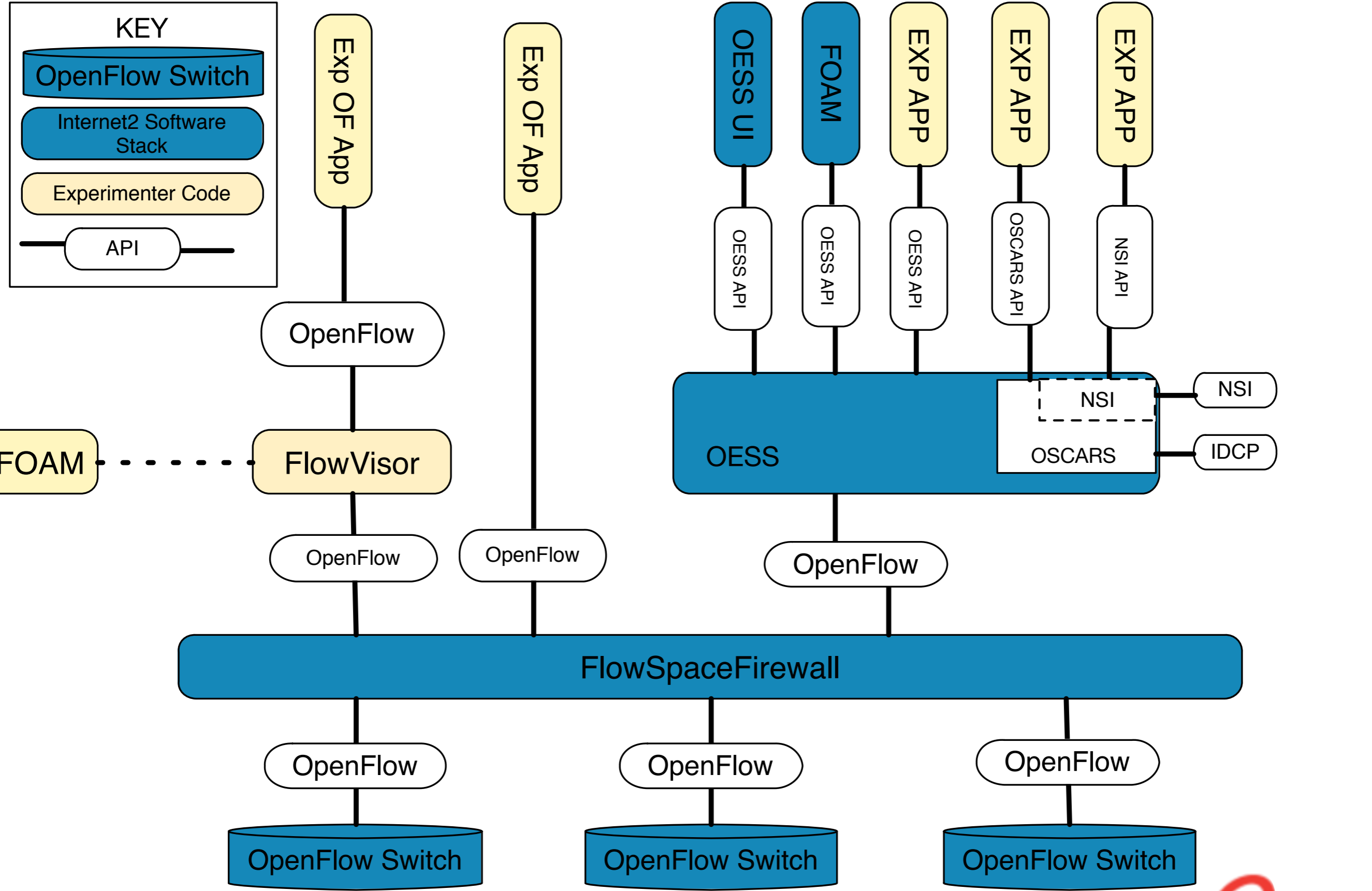


Internet2 Network

Advanced Layer2 Services Connector Map

March 2014





OE-SS Software

- 🔗 sub-second provisioning
- 🔗 auto failover to backup paths
- 🔗 100% OpenFlow
- 🔗 auto discovery of new devices and circuits
- 🔗 IDCP support for inter-domain
- 🔗 integrated measurement
- 🔗 Open Source: <http://globalnoc.iu.edu/sdn/oess.html>

The screenshot displays the OESS web interface. At the top, it says 'The Open Science, Scholarship & Services Exchange' and 'Workgroup: Indiana GigaPOP'. The main content area is divided into several sections:

- Active VLANs**: A navigation menu with options like Network Status, Available Resources, Users, Actions, and ACL.
- Network Map**: A map of the United States showing a network topology with nodes and connecting lines.
- Link Status Table**:

Link	Status
I2-CLEV-STAR-100GE-07736	up
I2-LOSA-SUNN-100GE-07755	up
I2-DENV-KANS-100GE-07746	up
I2-HOUH-TULS-100GE-07751	up
I2-KANS-TULS-100GE-07753	up
I2-CHIC-KANS-100GE-07745	up
I2-NEWY32AOA-WASH-100GE-07759	up
I2-DENV-SALT-100GE-07747	up
I2-SALT-SUNN-100GE-07760	up
- Switch Status Table**:

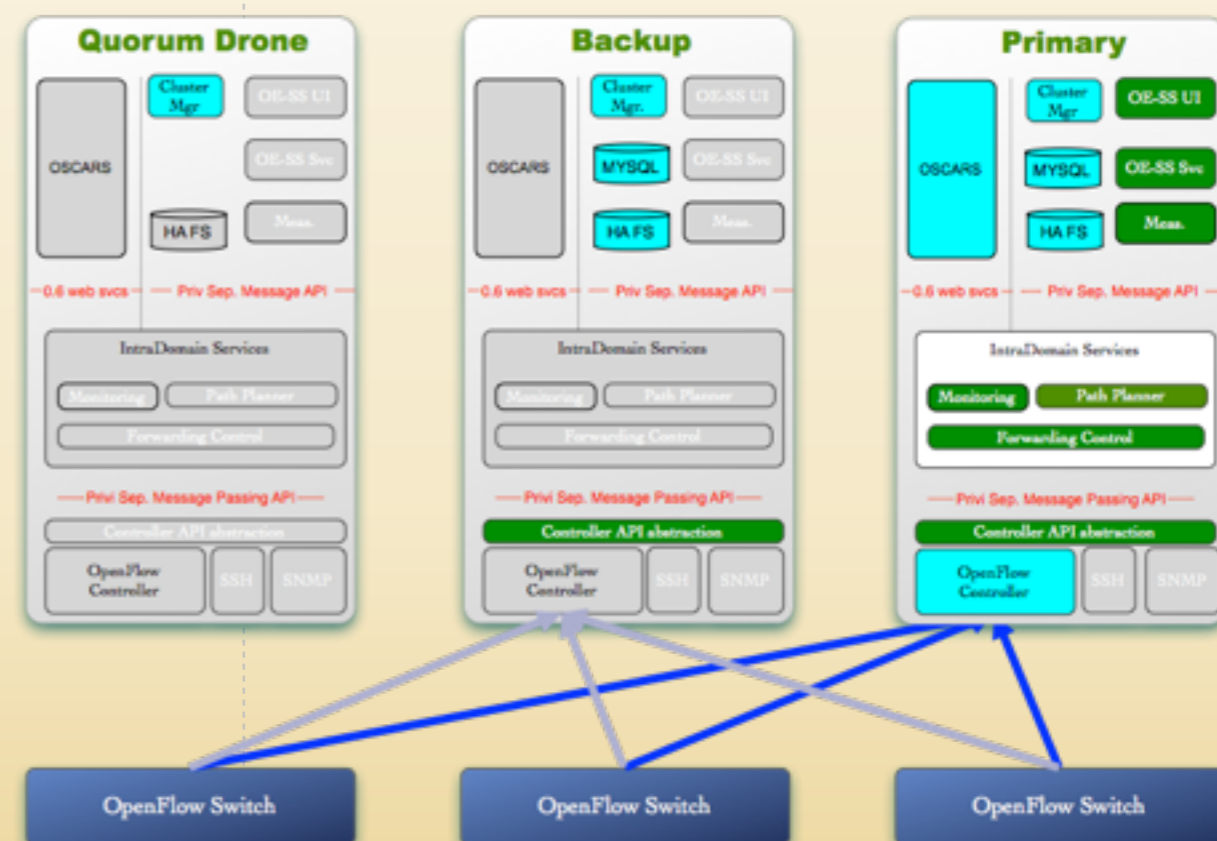
Switch	Status
sdn-sw.elpa.net.internet2.edu	up
sdn-sw.denv.net.internet2.edu	up
sdn-sw.star.net.internet2.edu	up
sdn-sw.salt.net.internet2.edu	up
sdn-sw.wash.net.internet2.edu	up
sdn-sw.losa.net.internet2.edu	up
sdn-sw.sunn.net.internet2.edu	up
sdn-sw.kans.net.internet2.edu	up
sdn-sw.tuls.net.internet2.edu	up
- Circuit Status Table**:

name	Status
[TR-CPS] Indiana Gigapop	primary
Indiana Gigapop R&E	primary
[CPS] Indiana Gigapop CPS-IPv6	primary
[LHCONE] Indiana Gigapop	primary
ING TEST	backup

At the bottom of the interface, there are logos for GlobalNOC and Internet2 Network, along with navigation links like 'Global Research NOC - Indiana University - Internet2 - OESS - About'.



OE-SS HA Architecture

- 🔸 controller is designed as a cluster
- 🔸 corosync & friends
- 🔸 multi-master mysql
- 🔸 drbd






Initial Challenges

Timeframe

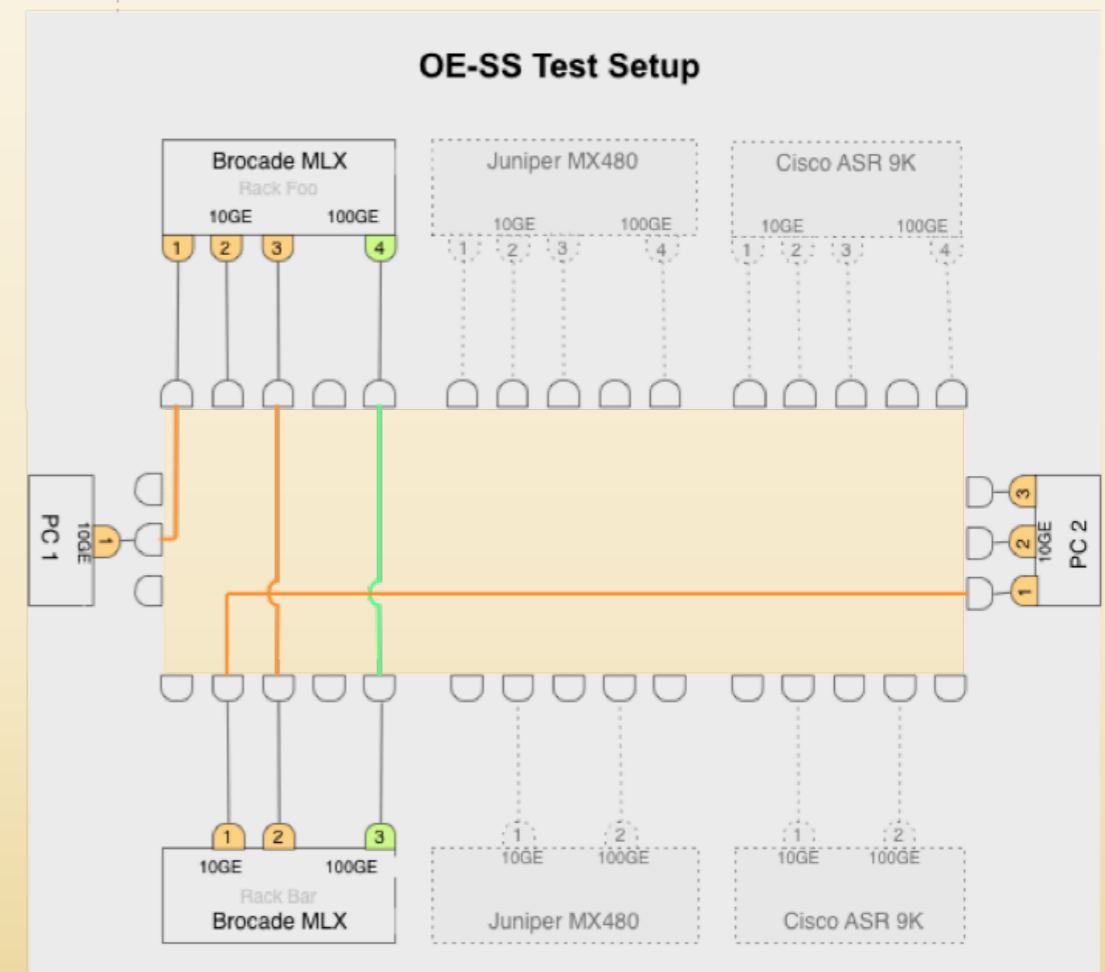
-  9 months for production code and deployed NDDI network
-  10 additional months to deploy AL2S network

Ecosystem

-  OpenFlow largely untested
-  brand new software stack
-  Multiple Vendors with differing SDN implementations/capabilities



System Testing

- ❏ Automated_(mostly) test suite using Jenkins
- ❏ Programmable topology with glimmerglass
- ❏ Several test points
- ❏ 2 devices per vendor (including GENI iDREAM funded systems)
- ❏ test every new vendor or tool chain software release






Types of testing

General

-  OFTest for general protocol adherence
-  Implementation behaviors not mandated in spec

OE-SS / Flowspace Firewall

-  base functionality
-  performance
-  burn in / stability

Things we have seen

- 🍯 **flow_mod processing speed limits (improving over time!)**
- 🍯 **incomplete OpenFlow spec support**
 - 🍯 layer2 or layer3 matching but not both
 - 🍯 no viable QoS mechanisms
 - 🍯 key actions not always supported
- 🍯 **inconsistencies in behavior**
- 🍯 **Implementation bugs in early versions of network device's OpenFlow support (and bugs in our controller implementation!)**
- 🍯 **Many of the problems have been “boring” / “traditional” — e.g. hardware failures, backhoes, etc.**

Controller Placement

🏠 Primary Controller cluster in Chicago

🏠 hot standby with synched state (corosync and DRBD)

🏠 failover in cluster causes controller to switch reset

🏠 Second redundant cluster in Bloomington

🏠 controller to switch latency

🏠 28ms RTT avg

🏠 64ms RTT worst case

Controller Placement

- ❏ **OE-SS mostly proactive, reacting to topology changes**
 - ❏ backup path preconfigured, failover involves 1 flow mod @ each ingress
 - ❏ controller takes ~70ms to receive a packet_down message and send a flow_mod in response. (some low hanging fruit here)
- ❏ **~100ms to respond to failover assuming avg latency** (ignoring time in switch)
 - ❏ Multiple seconds for IGP or Rapid Spanning Tree
 - ❏ MPLS path protect: ~100s ms
 - ❏ MPLS Fast Reroute: ~50 ms

Lessons

- ❏ **The architectural simplicity of a central controller is very attractive**
- ❏ **Central controller means management net is **critical**, if a fiber cut disrupts management and OpenFlow net then OESS failover blocks on management net failover.**
- ❏ **If management net is resilient then, central location seem reasonable choice**
- ❏ **Distributed controller will still be heavily dependent on management network to respond to non-local events.**

Lessons

- ❏ **Expect to be the system integrator, inter-op is on you**
- ❏ **a dedicated test infrastructure is essential**
- ❏ **automation a **very** good investment**
- ❏ **make sure you can run concurrent tests if you have multiple vendors**
- ❏ **Don't use exclusively black box tests**


Lessons

- ❖ **OpenFlow is a youthful protocol, vague in places creating complexity as vendors get creative.**
- ❖ **No vendor supports all of the spec**
- ❖ **multi-vendor == lowest common denominator feature set**
- ❖ **need to test component together as system**


What's Next?



Virtualization support on AL2S

 Planning underway to deploy FlowSpace Firewall on AL2S.

 Supported as a production service

 ISO: early adopters to work with to get their SDN control apps running on the network.

 Lab testing/verification with Internet2 NOC

 Deployment on AL2S