ConCERNing SDN

Or: How Duke and MCNC collaborated to speed LHC data to Research Triangle Physicists through Software-Defined Networks

Victor Orlikowski, Duke OIT

Background, Round 1

- Duke's "traditional" network is **heavily** internally segmented using VRFs.
    - Individual departments often have their own VRF.
    - Transitions across VRFs incur significant overhead, since they must pass through one (or more) firewall contexts.
- Traffic entering through the campus edge incurs a pass through the IPS.
- These are "best practices" – offering good security, as well as flexibility of management for departmental administrators.
- But – they're crippling to large scientific traffic flows.

Science DMZ: A Solution?

- The Science DMZ model proposes creating "friction-free" network paths for scientific flows.
  - Minimizing "unnecessary constraints" that hamper high-performance applications.
  - "Reducing or eliminating" causes of packet loss.
  - Requires careful consideration of path from campus edge to end host.
- At first glance, antithetical to traditional model.
  - IPS/firewall inspection provides security, but is a finite resource.
  - When overloaded, these mechanisms **cause** loss and delay.
  - Non-starter: removing all controls from science flows.
- Can these two models be meshed?

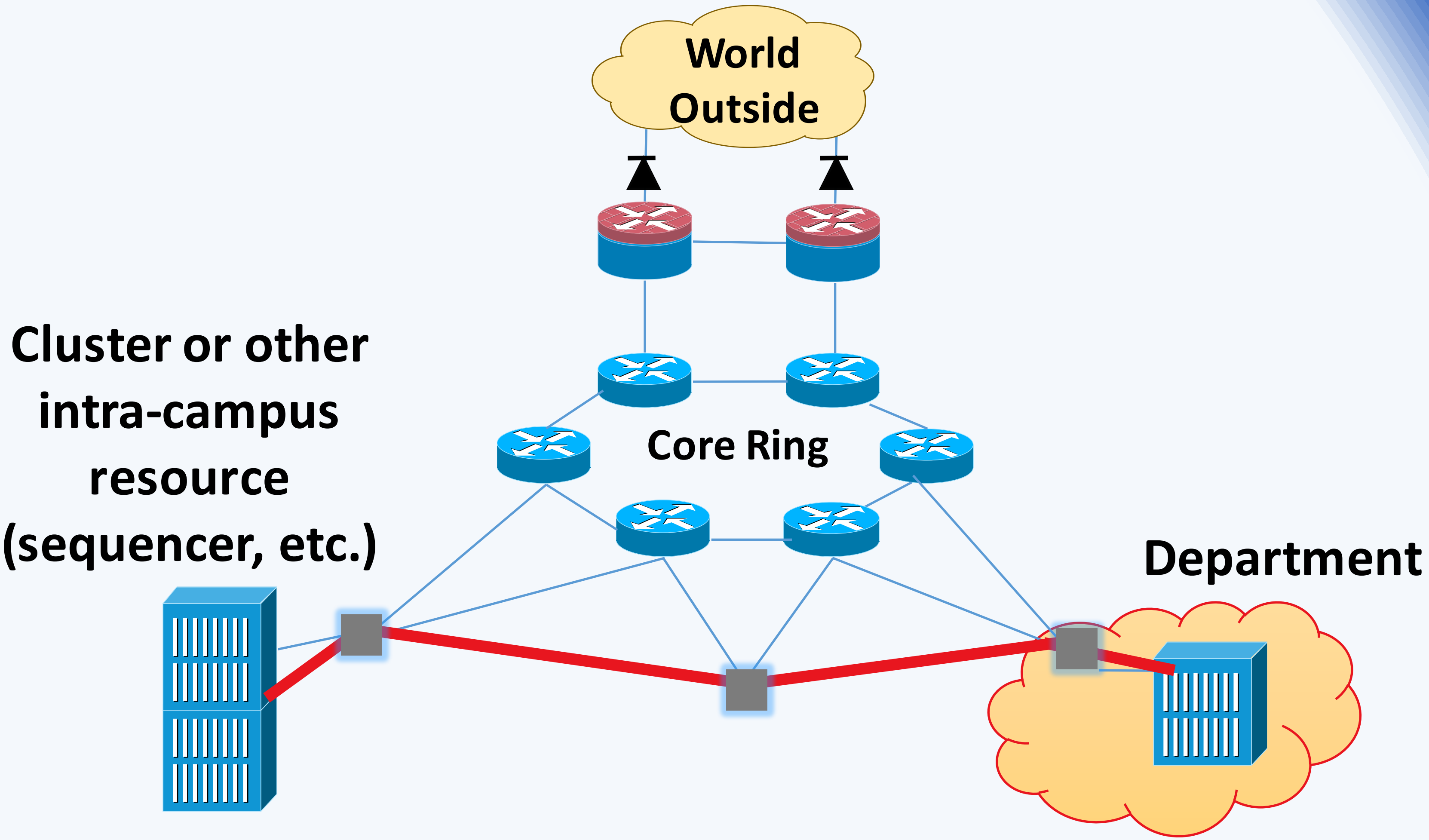
Building a Bypass, Part 1

- Duke married the “Science DMZ” and “traditional” models using a “bypass” network built on OpenFlow switches.
  - OpenFlow-enabled access switches direct traffic to the “traditional” production network by default.
  - At the instruction of Duke’s SDN controller, flows can be directed onto the “bypass” network – which obviates the firewalls and IPS.
  - Redirections are dynamic, require no re-configuration of the end host, and can only be requested by authorized personnel.

Building a Bypass, Part 2

- Both "production" and "test" bypass networks exist.
  - Production consists of "willing volunteer" departments (Computer Science, Physics, etc.)
  - Test is a "mirror" of production, consisting of "simulated departments" in our lab.
- "Hybrid" internal architecture - best of both worlds:
  - Majority of production traffic flows over traditional network; infrastructure investment preserved, benefits retained.
  - Science DMZ goals realized through bypasses that are friction-free, explicitly authorized, and audit logged.
  - Ability to dynamically create Science DMZ capable paths within Duke's network enabled.

A Bypass, Visualized



Traffic Control on the Bypass

- In order to control traffic on these OpenFlow networks, Duke wrote two applications:

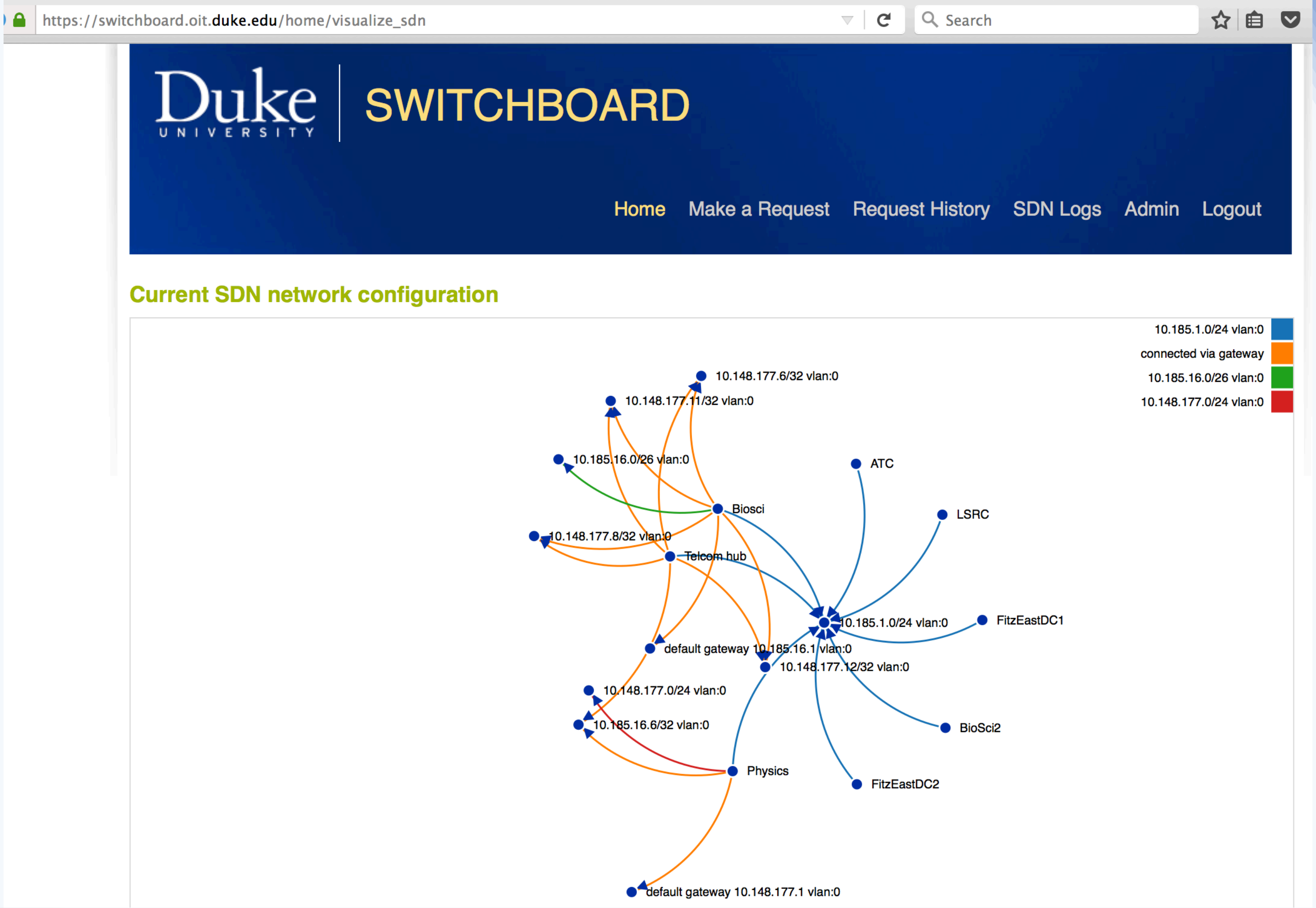
❖Switchboard

- Ruby on Rails web application that serves as central store of information and source of control.
- Written by Mark McCahill
- Available on GitHub at: https://github.com/mccahill/switchboard

❖Plexus

- An OpenFlow controller application (using Ryu as a base) that programs the OpenFlow switches comprising the networks at Switchboard's instruction.
- Written by your humble presenter
- Available on GitHub at: https://github.com/vjorlikowski/plexus

Switchboard’s View of the Production SDN

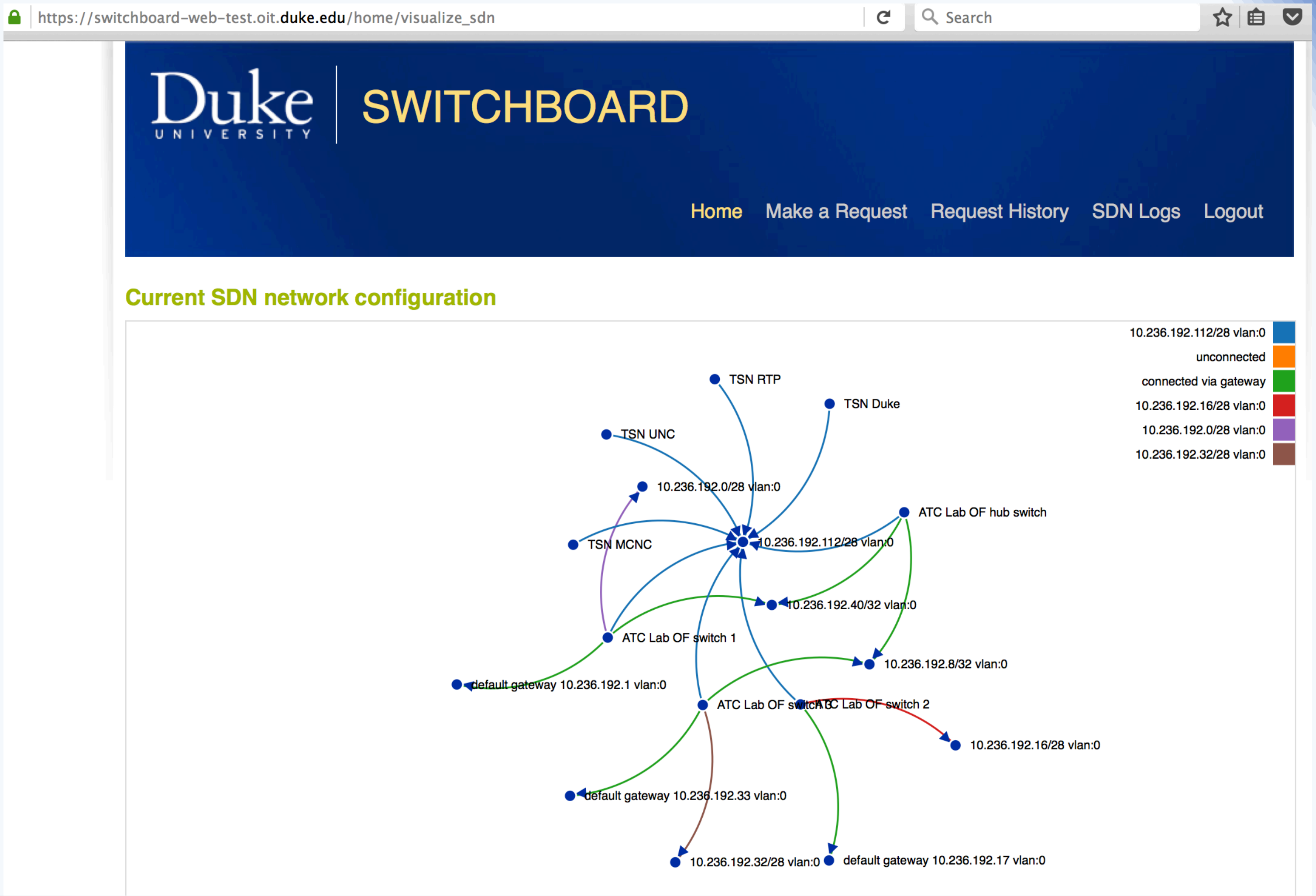


https://switchboard.oit.duke.edu/home/visualize_sdn

Duke UNIVERSITY | SWITCHBOARD

Current SDN network configuration

Switchboard’s View of the Test SDN



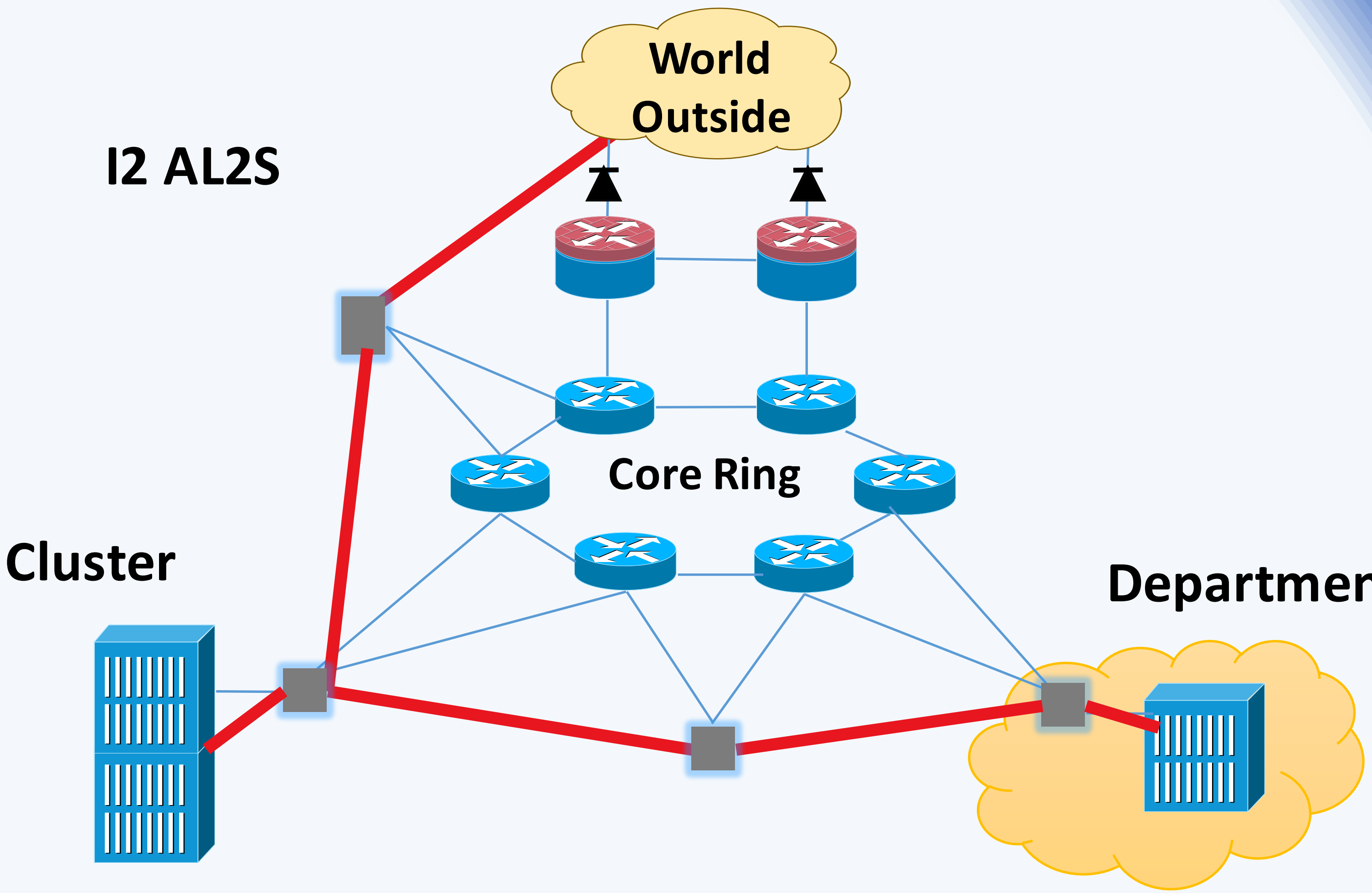
Current SDN network configuration

10.236.192.112/28 vlan:0	🟦
unconnected	🟧
connected via gateway	🟩
10.236.192.16/28 vlan:0	🟥
10.236.192.0/28 vlan:0	🟪
10.236.192.32/28 vlan:0	🟫

Duke’s Physicists Want More…

- Friction-free, fairly secure paths within Duke’s network: Great!
- …but could we have access to similar paths for off-campus resources? Say, data from the LHC?
- To do this, you’ll need to peer via LHCONE – the private layer 3 network associated with the LHC.
- First small challenge: allocating a layer 2 circuit over which to peer.
- Second small challenge: ensure that no unintended traffic “bleeds” into (or off of) the LHCONE private network.

The Physicists’ Ideal, Visualized

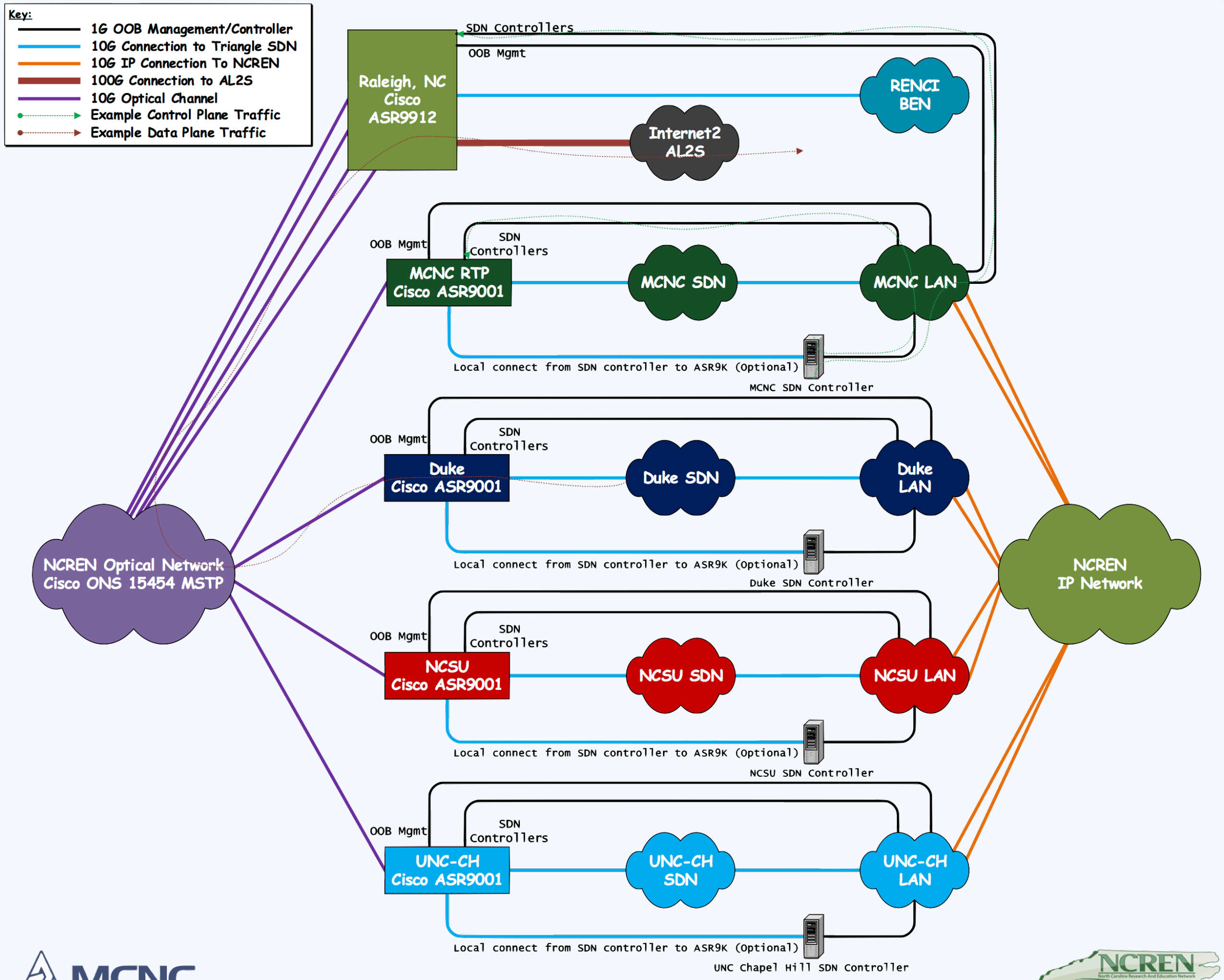


Background, Round 2

- MCNC (operator of NCREN) built an OpenFlow-based regional network, interconnecting the three Research Triangle universities and RENCI (the Renaissance Computing Institute).
- The intent? Allow the interconnected institutions to create dynamic, high-speed, peerings – under the control of the individual institutions.
- Each institution must run their own OpenFlow controller, with the control connection secured using TLS.
- AL2S circuits are also available through the regional SDN – allowing dynamic peerings beyond the Triangle.

A View of MCNC’s Regional SDN

MCNC Operated NCREN Software Defined Network Architecture


Key:
1G OOB Management/Controller
10G Connection to Triangle SDN
10G IP Connection To NCREN
100G Connection to AL2S
10G Optical Channel
Example Control Plane Traffic
Example Data Plane Traffic
SDN Controllers
OOB Mgmt
Raleigh, NC Cisco ASR9912
RENCI BEN
Internet2 AL2S
OOB Mgmt
SDN Controllers
MCNC RTP Cisco ASR9001
MCNC SDN
MCNC LAN
Local connect from SDN controller to ASR9K (Optional)
MCNC SDN Controller
OOB Mgmt
SDN Controllers
Duke Cisco ASR9001
Duke SDN
Duke LAN
Local connect from SDN controller to ASR9K (Optional)
Duke SDN Controller
NCREN Optical Network Cisco ONS 15454 MSTP
NCREN IP Network
OOB Mgmt
SDN Controllers
NCSU Cisco ASR9001
NCSU SDN
NCSU LAN
Local connect from SDN controller to ASR9K (Optional)
NCSU SDN Controller
OOB Mgmt
SDN Controllers
UNC-CH Cisco ASR9001
UNC-CH SDN
UNC-CH LAN
Local connect from SDN controller to ASR9K (Optional)
UNC Chapel Hill SDN Controller

Plumbing to CERN, Part 1

- In order to make the connection to LHCONE, Duke had to control its own portion of the regional SDN; how?
- Easy! Plumb the OpenFlow control connections to our controller, then inform Switchboard about the new switches.
- …Perhaps not so easy: our controller is on a private IP, in a secured portion of the network.
- Solution: use HAProxy as a TLS terminator, on a public IP.
  - HAProxy validates certificates presented by switch control connections.
  - After verification, control connections are proxied back through Duke’s network to the controller of choice.

Plumbing to CERN, Part 2

- Once the controller connection was completed (to the controller for the test network in the lab), a circuit needed to be plumbed over AL2S through the regional SDN.
- Michael O'Connor (of ESNet) requested the AL2S circuit, after co-ordinating with us on what VLAN was compatible within the Triangle regional SDN.
- Duke's test controller was then used to orchestrate the plumbing of the VLAN through to the internal test SDN.
- The final hurdle: LHCONE is Layer 3; we have to speak BGP.

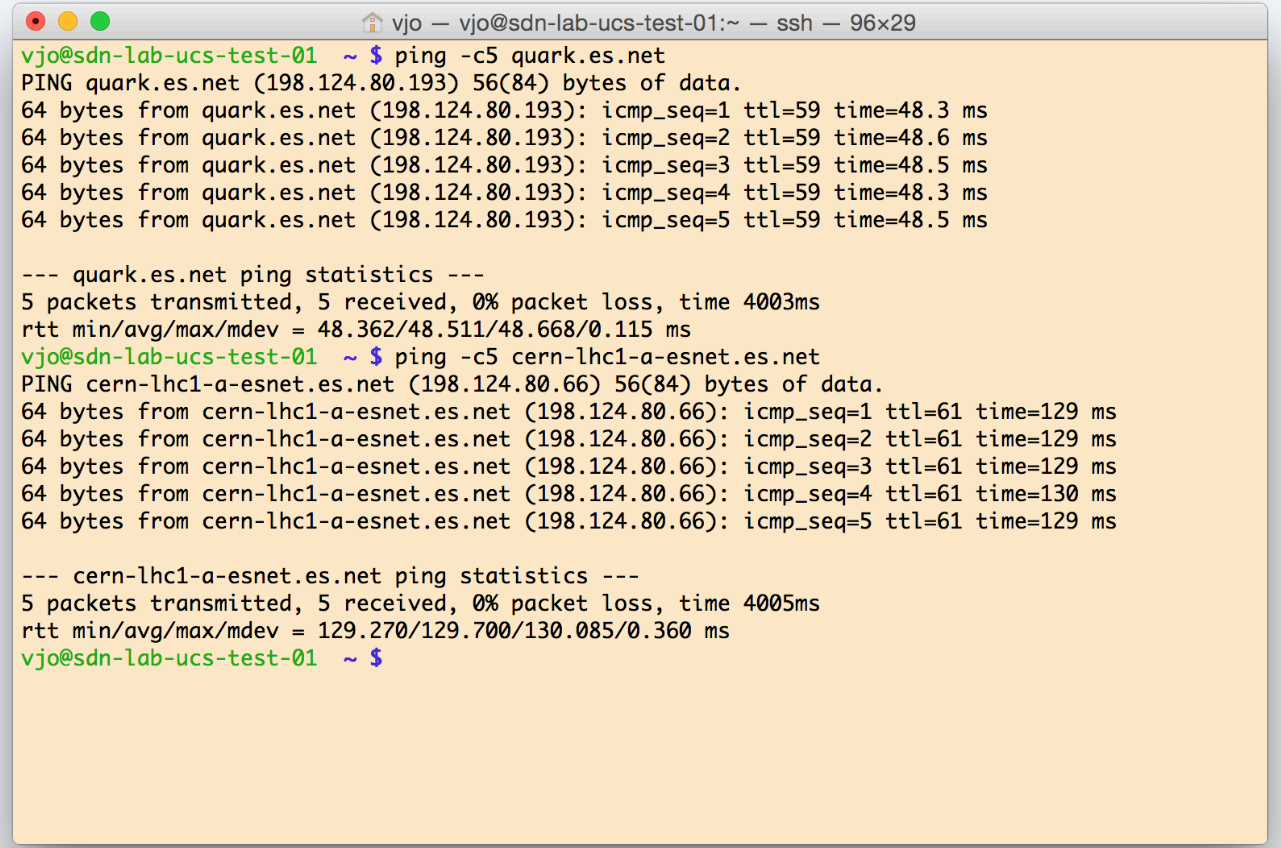
Plumbing to CERN, Part 3

- Our controller doesn’t speak BGP – yet.
- Short-term solution – use a loopback cable to bridge the OpenFlow and “traditional” sides of the “hub” OpenFlow switch in the lab SDN, and configure the BGP instance on that switch.
- BGP on “traditional” side of OpenFlow switch successfully peers over SDN path – but is intentionally prevented from communicating advertised routes to Duke’s production network.
- How do we communicate the advertised routes to hosts on the internal SDN?

Plumbing to CERN, Part 4

- Short-term solution: write a small daemon that pulls information from the switch BGP instance, and communicates it into Switchboard.
- Switchboard can then be used to authorize dynamic paths onto the advertised LHCONE routes, but only for the appropriate end hosts on Duke's network.
- Unauthorized traffic inbound from LHCONE will "fail secure" – unless Switchboard has specifically requested that the controller insert a path through the network, the default behavior is for traffic to be dropped.

Obligatory Successful Ping



```
vjo@sdn-lab-ucs-test-01  ~ $ ping -c5 quark.es.net
PING quark.es.net (198.124.80.193) 56(84) bytes of data.
64 bytes from quark.es.net (198.124.80.193): icmp_seq=1 ttl=59 time=48.3 ms
64 bytes from quark.es.net (198.124.80.193): icmp_seq=2 ttl=59 time=48.6 ms
64 bytes from quark.es.net (198.124.80.193): icmp_seq=3 ttl=59 time=48.5 ms
64 bytes from quark.es.net (198.124.80.193): icmp_seq=4 ttl=59 time=48.3 ms
64 bytes from quark.es.net (198.124.80.193): icmp_seq=5 ttl=59 time=48.5 ms

--- quark.es.net ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4003ms
rtt min/avg/max/mdev = 48.362/48.511/48.668/0.115 ms
vjo@sdn-lab-ucs-test-01  ~ $ ping -c5 cern-lhc1-a-esnet.es.net
PING cern-lhc1-a-esnet.es.net (198.124.80.66) 56(84) bytes of data.
64 bytes from cern-lhc1-a-esnet.es.net (198.124.80.66): icmp_seq=1 ttl=61 time=129 ms
64 bytes from cern-lhc1-a-esnet.es.net (198.124.80.66): icmp_seq=2 ttl=61 time=129 ms
64 bytes from cern-lhc1-a-esnet.es.net (198.124.80.66): icmp_seq=3 ttl=61 time=129 ms
64 bytes from cern-lhc1-a-esnet.es.net (198.124.80.66): icmp_seq=4 ttl=61 time=130 ms
64 bytes from cern-lhc1-a-esnet.es.net (198.124.80.66): icmp_seq=5 ttl=61 time=129 ms

--- cern-lhc1-a-esnet.es.net ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4005ms
rtt min/avg/max/mdev = 129.270/129.700/130.085/0.360 ms
vjo@sdn-lab-ucs-test-01  ~ $
```

Future Plans

- BGP support needs to be implemented in the controller; we'd like to remove the "hack" of using the BGP instance in the traditional side of the OpenFlow switch.
- Tag translation and cross-VLAN stitching; we'd like to be able to transit traffic from end hosts on one VLAN onto a different AL2S VLAN, within Duke's network.
- Out-of-band traffic inspection and reactive flow shutdown; we are interested in building on the work done by Indiana University with their SciPass application, so that we can increase security by inspecting science flows out-of-band, and shutting down those that violate the policies we define.

Grateful Thanks To…

- Funding:

NSF

Grants:
CNS-1243315
OCI-1246042
ACI-1440588

- Partners:

ESnet
Energy Sciences Network

MCNC
Connecting North Carolina's Future Today

NCREN
North Carolina Research And Education Network

- Infrastructure Providers:

INTERNET2®

CISCO™

ARISTA