

Duke's SDN Journey

International OpenFlow/SDN Test Beds

3/31/15

Charley.Kneifel@duke.edu

Part 1 - Planning

- Definitions
- Infrastructure Considerations
- Use Cases
- Service Delivery / Management Considerations
- User Engagement

Definitions

- SDN at Duke is the implementation of an OpenFlow software controller that manages network traffic flow on a set of network devices.
- It is focused on the edge of the network more than traffic within a data center.
- The primary goal is to improve the speed, reliability, and overall performance of the network used by researchers.

What Is The Current State at Duke?

- SDN Switches deployed in production
 - Hub and Spoke Model
- Production controller – Ryu based (forking our code – Vex?)
- Production rule manager – SwitchBoard (Mark to Demo)
- Funding through EAGER and CC-NIE
- perfSONAR nodes deployed across campus
 - In the middle of upgrading to new version (Puppet'izing)
- Efforts led to redesign of Duke core network
- Duke uses an MPLS core and can switch to a VRF easily – so routing is everywhere

Infrastructure Considerations

- Dedicated Science Network?
- Converged/Unified Network?
- Fiber Infrastructure?
- Needs at the Core?
- Needs at the Edge?

Planning For SDN – Lessons Learned At Duke

Test, Test, Test in controlled fashion

perfSONAR is your friend (more about that later)

Oversubscription – Accidental or Intentional

Span Ports, Layer 2/3 Domains

10G Cannons aimed at your network

Measurement of Real Bandwidth Availability

Firewalls – what are the real limits – per stream / overall

How do they fall over / fail?

IPS – where is traffic inspected, white listing, how often?

How do they fall over / fail open/closed?

IDS – Passive

Planning for SDN (Continued)

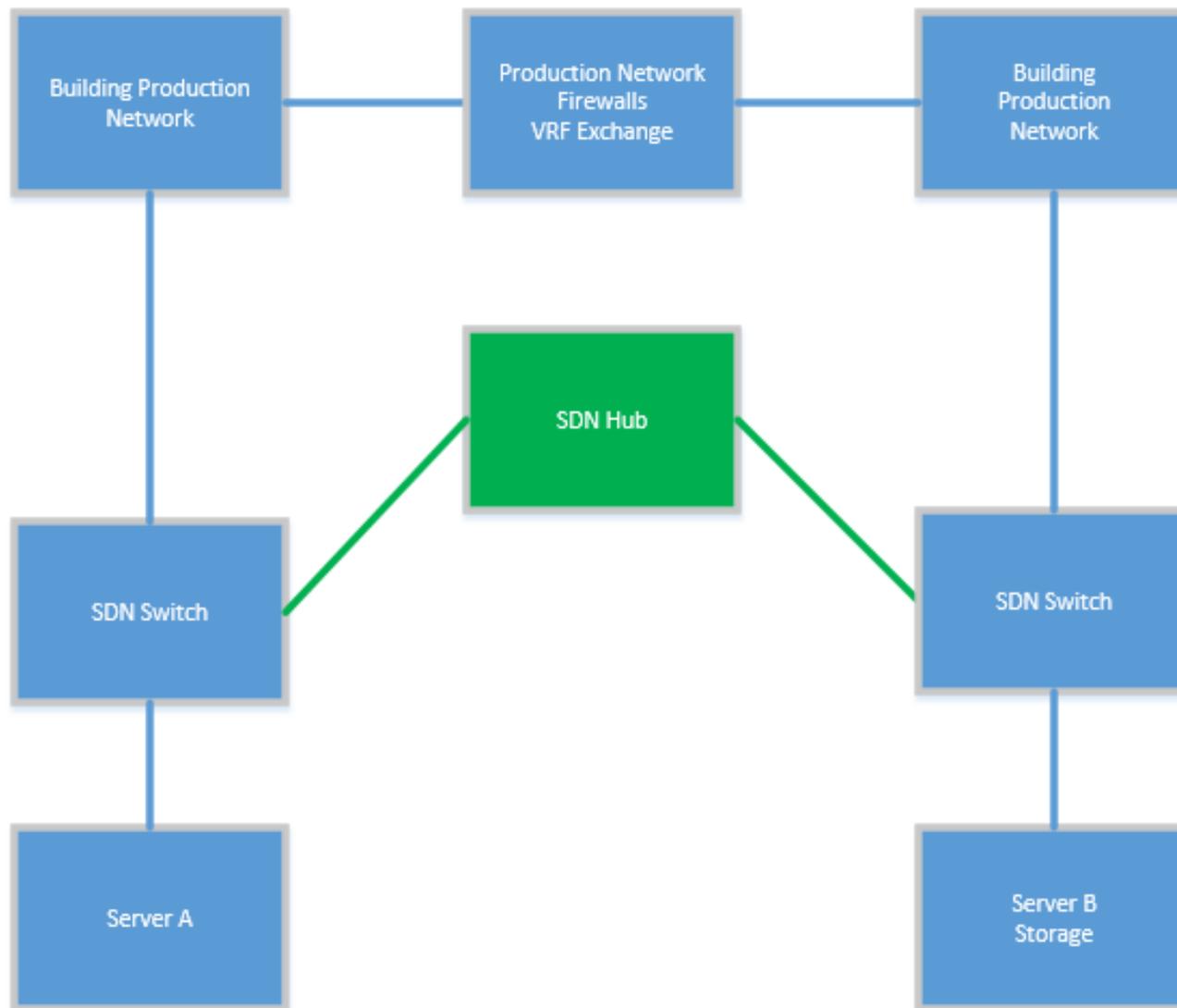
- It's another network upgrade
 - But it's not fully documented 😊
 - And it's changing
 - Keep the core a significant multiplier of the edge
- QOS Is Important if you have converged services
 - Voice, Video
- Use case documentation
 - Bypass Network / Large Predictable Data Flows
 - Science DMZ
 - Protected Data
 - Data Migration
 - Health System – University “Bridges”

General SDN Model At Duke

- Integrated – hosts connect to A network
- Hybrid – network fabric has multiple options for routing
- Did not want to build, deploy, and manage a separate infrastructure
- Leverage MPLS Core and VRFs to route traffic to production network

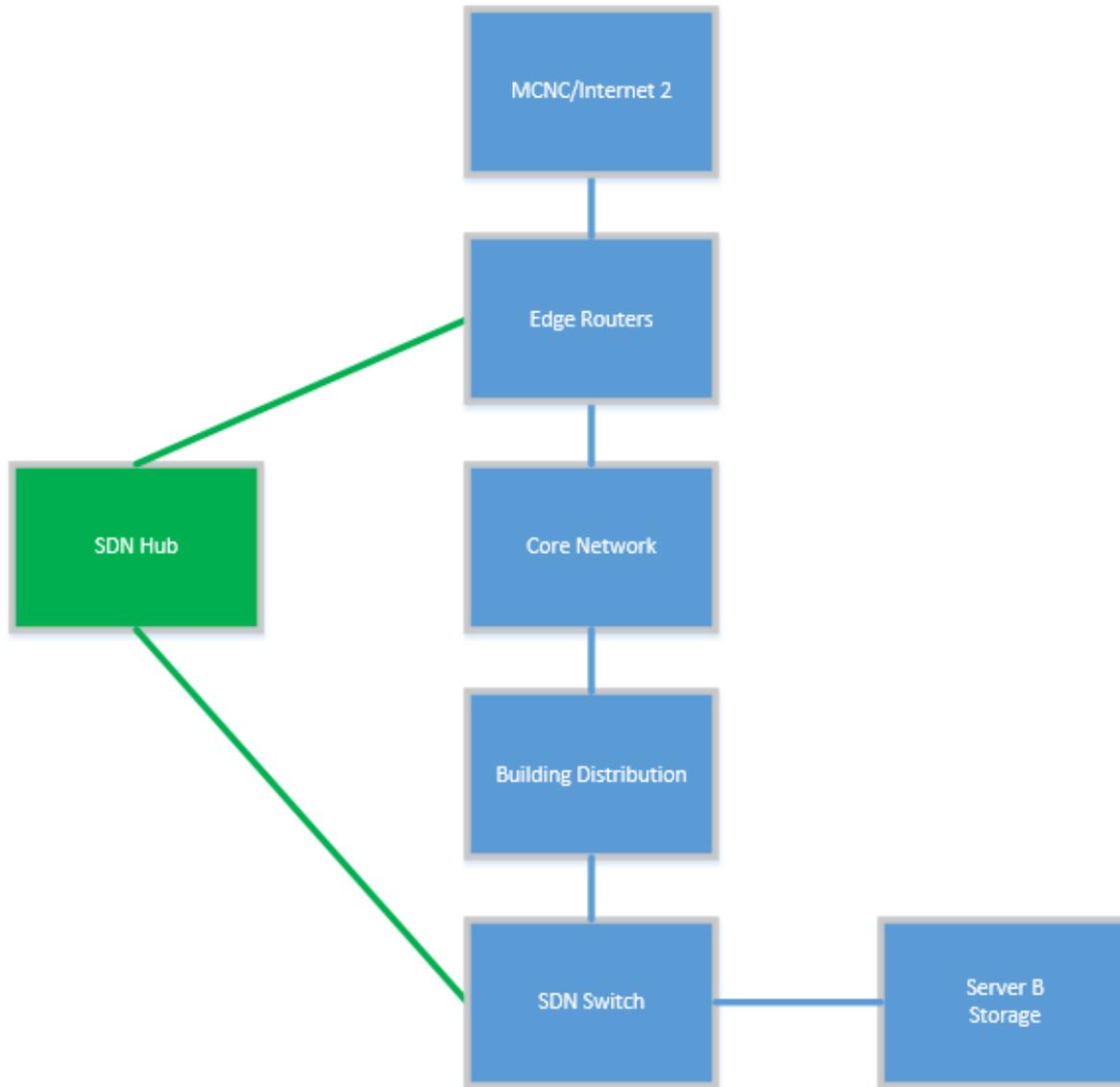
General SDN Model At Duke (Continued)

- Hosts connect to SDN Enabled switches at 10G
- Default path for traffic is to the production network
- Application applies rules to controller to route certain traffic over alternate path.
- Typically subnet to subnet or host to host
- VLAN tagging supported
- Want to add more functionality (port restrictions, VLAN flipping)



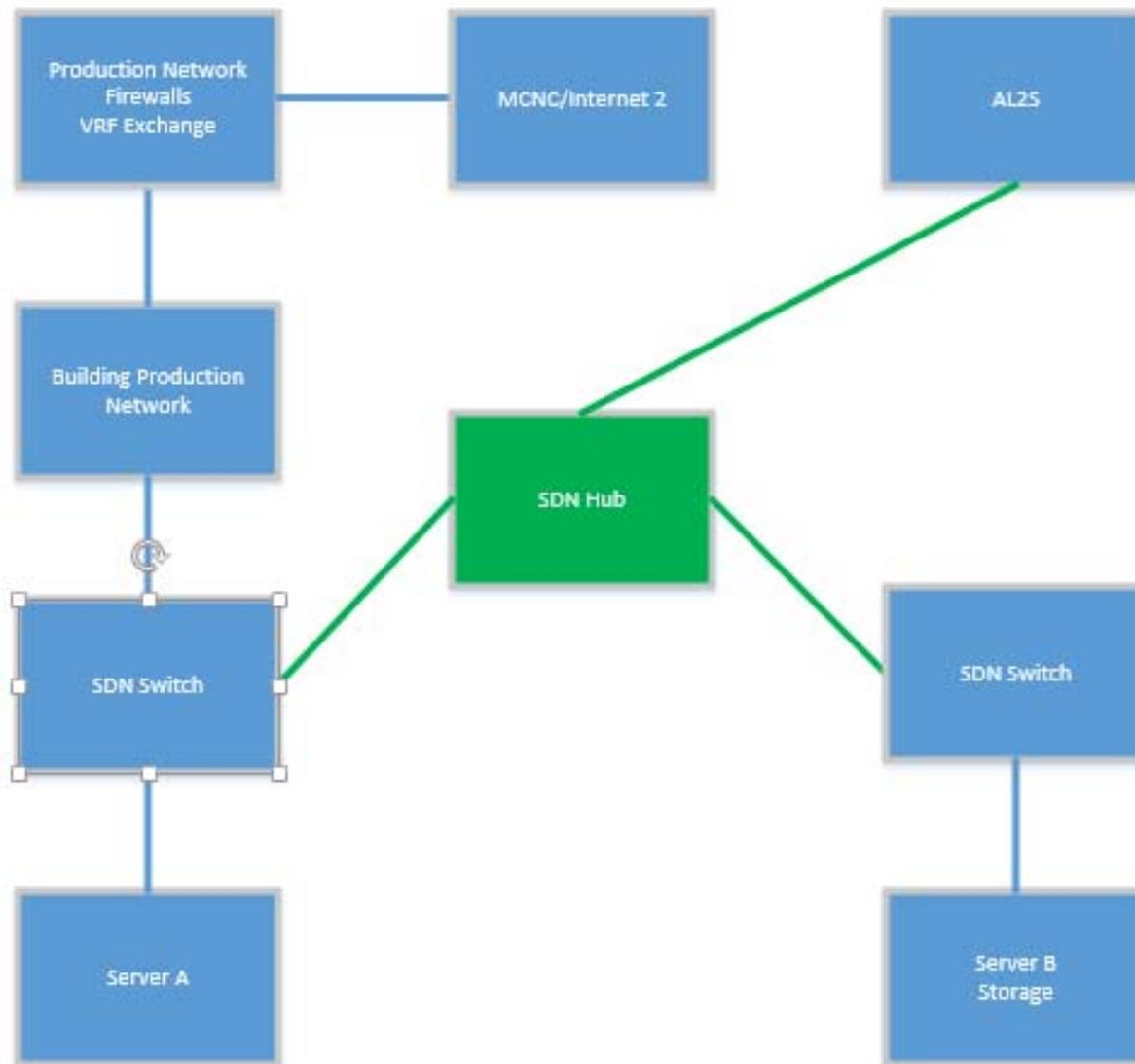
Use Case – Science DMZ

- Traffic coming from off campus to a specially routed IP address
- Route traffic to SDN Hub
- If rule setup – allow host to communicate w/o any additional inspection/overhead/firewalls



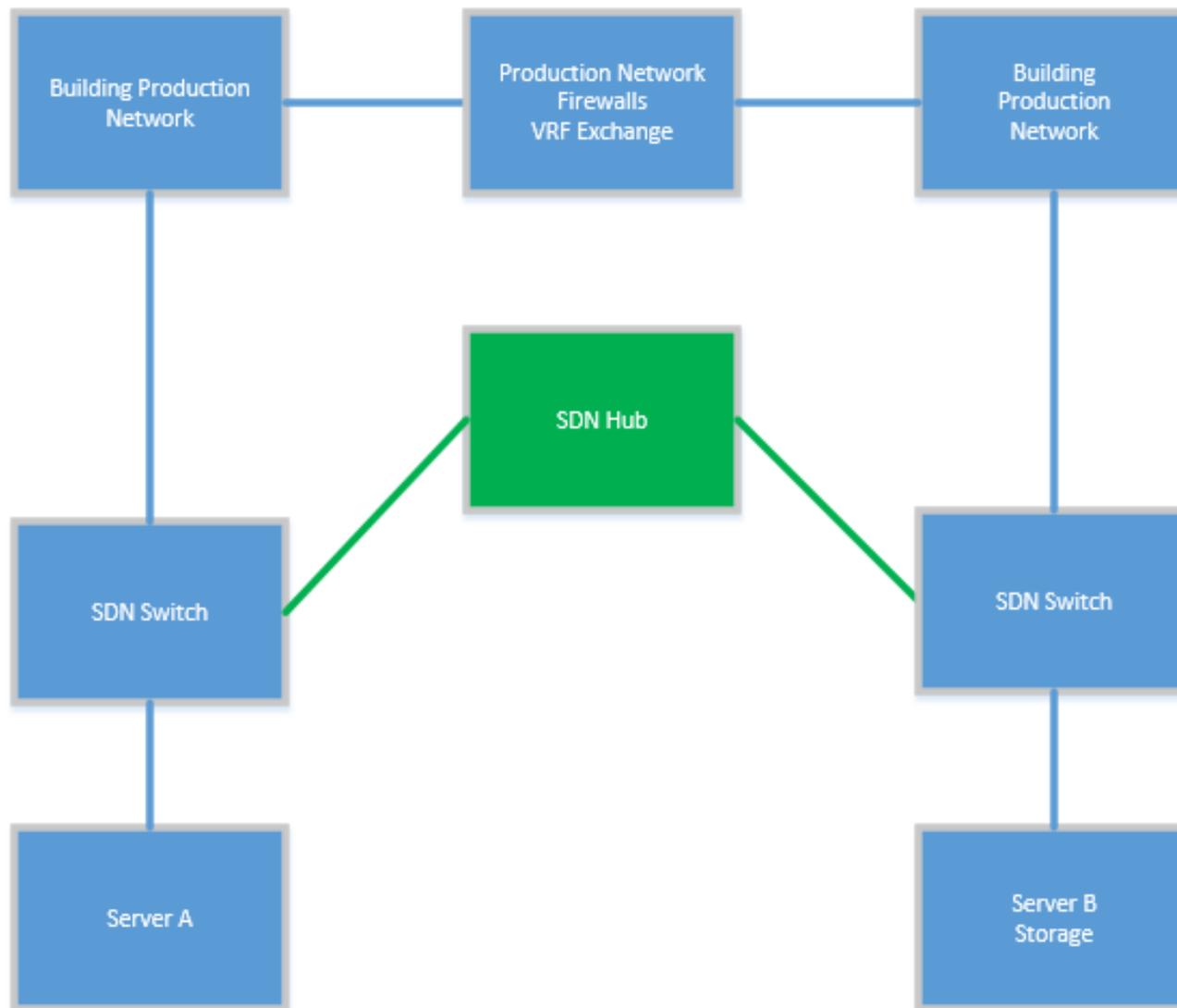
Use Case - AL2S Path

- Allow for dynamically built connections to resources at multiple universities on essentially a flat layer 2 network.
- AL2S Connection Terminates on SDN Hub Switch
- Traffic mapped to correct VLAN and rules enabled to route traffic



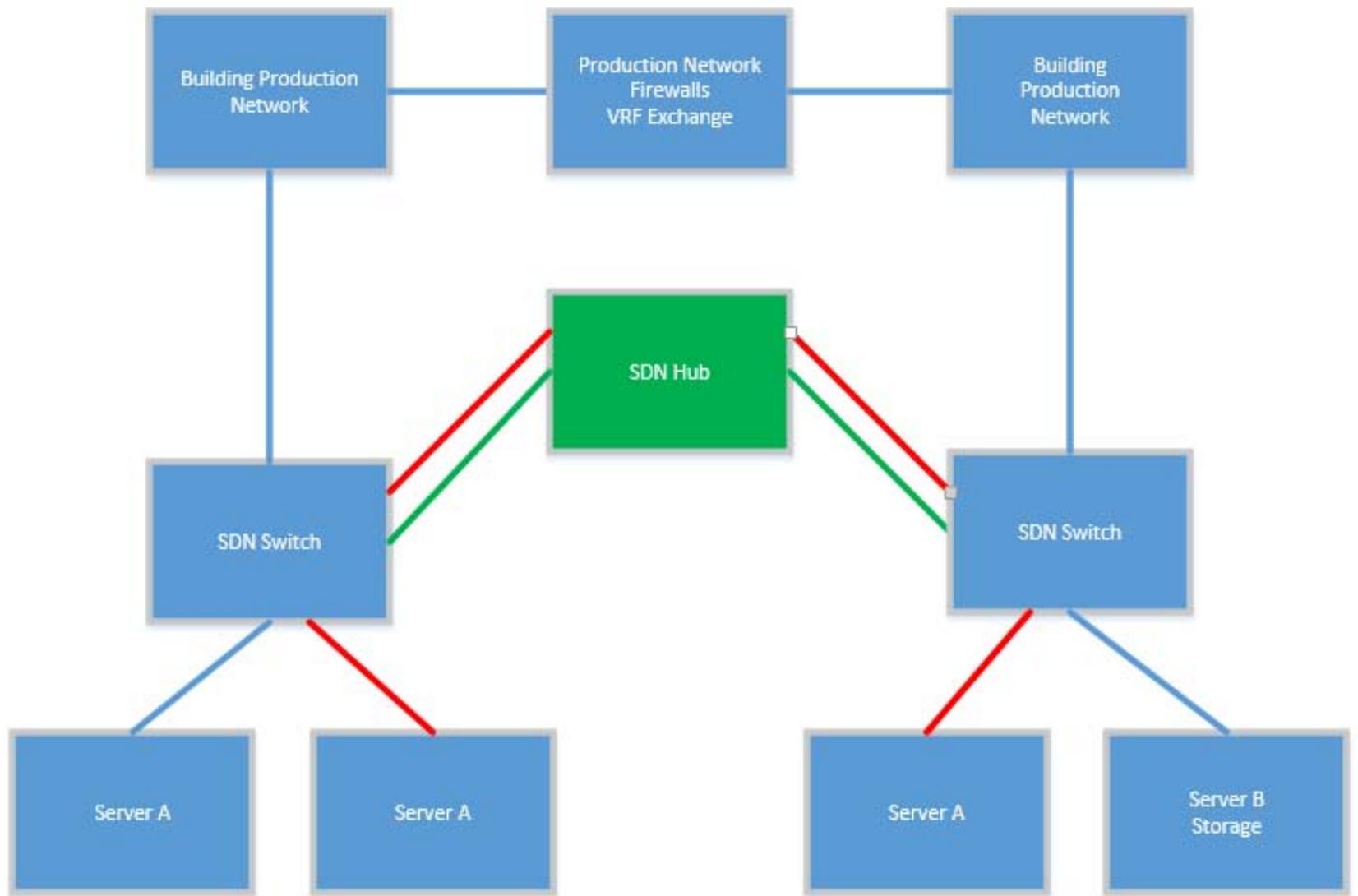
Use Case – Inspection Bypass

- Avoid latency of IPS and Firewalls
- Avoid saturation of IPS and Firewalls for “routine traffic”
- Map connections with source/destination networks
- Examples
 - Storage replication from instruments to central storage
 - Storage replication from central storage to dedicated HPC resources
 - Backups!
 - Latency sensitive apps in multiple buildings



Use Case - Bandwidth/Capacity Expansion

- Add capacity to a specific lab and route traffic over a second pair of fiber interconnects
- Provide alternate paths for different performance scenarios
 - Low Latency path
 - High Bandwidth path
 - Same switch

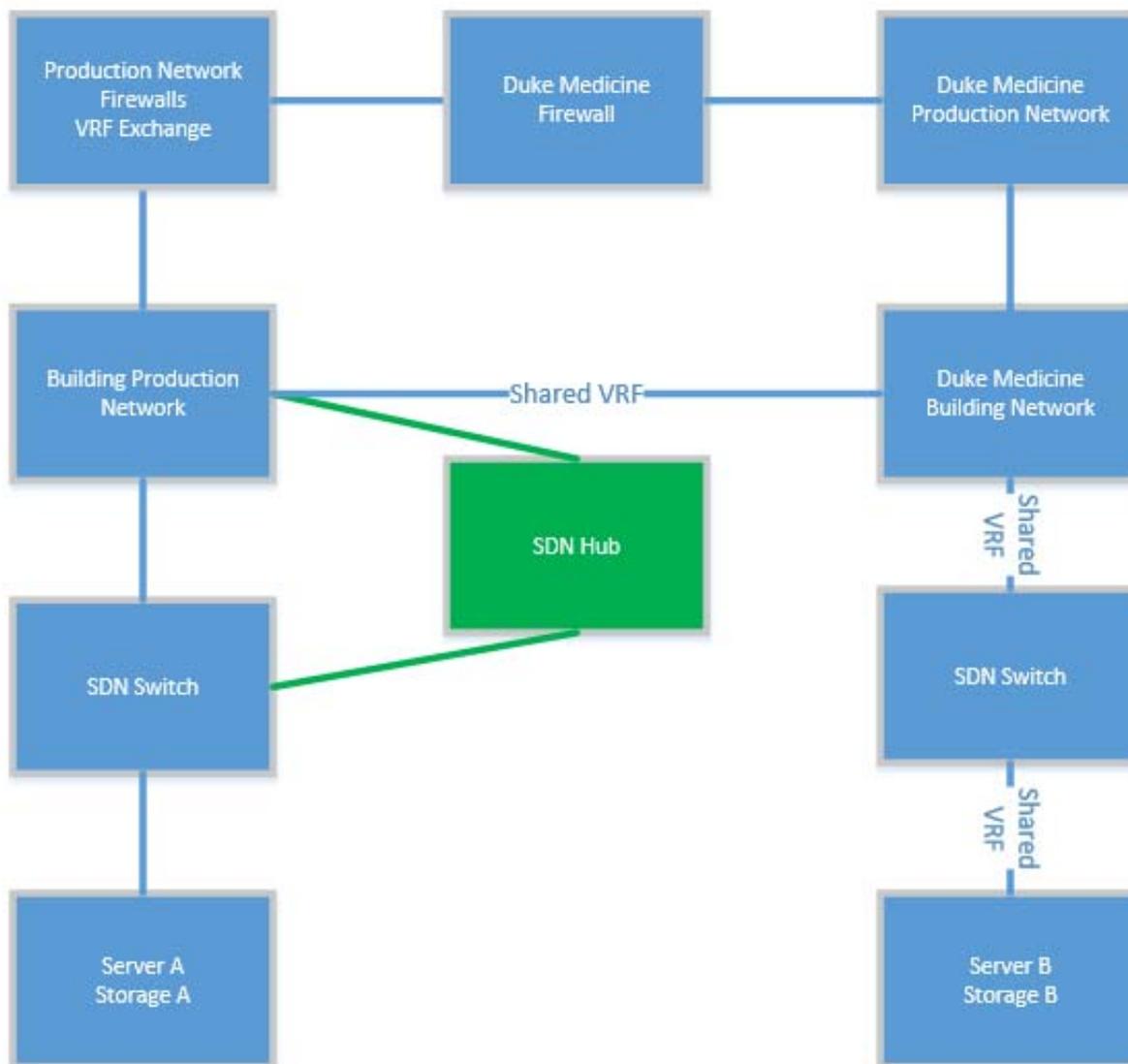


Use Case @ Duke – Data Migrations

- Duke has a protected research data network
- VMs provisioned
- Firewalled
- ACLs on routing between subnets
- VPN for general access
- Jump boxes/SSH/RDP servers
- Bottlenecks on getting data into the network
- Use an SDN path to migrate encrypted traffic into the network
 - Could apply to both external and internal connections
 - One time or potentially recurring use (external protected feeds)

Use Case - Research Support – Duke Medicine and University

- Separate IT support services
- Separate Networks to the server and desktop
- Many shared services – Health System include Sch of Med/Nursing
 - Common core financial system - SAP
 - Common student system – PeopleSoft
 - University provides/manages all email
 - University manages internet exchange
- Common technology (recent change) – MPLS
- Different Security Postures



Security Planning - Where Does Security Fit In

- Science DMZ = No Security?
 - Not really, limited set of connections, dynamic/open
- Use for Protected Data – encrypted blobs
- Separation of Data Plane and Control Plane
 - Control plane should only have controller and switch management
 - Duke uses a dedicated VRF for Control Plane
- Proxy/Firewall and Controller Separation
 - Multiple Controllers Can Be Used (FlowVisor) – But ...
- All the usual hardening (OS/WebApp/DB)

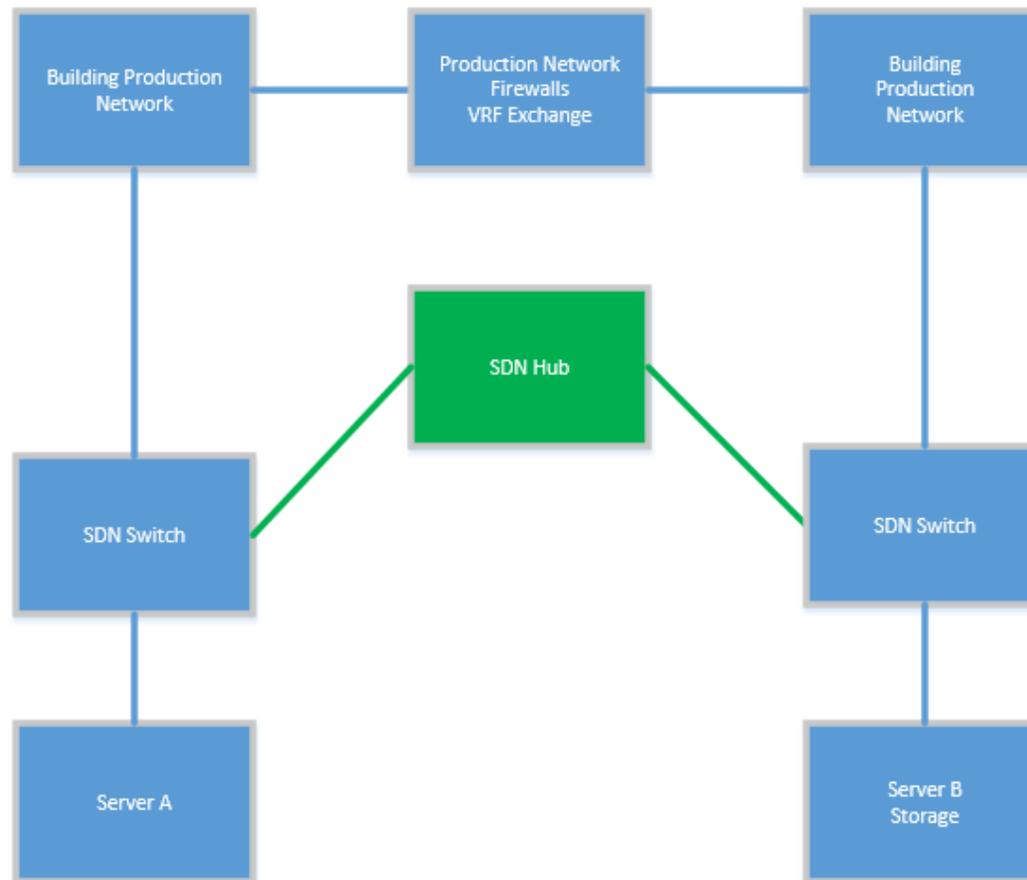
Security – Continued

- Who is allowed to communicate with Controller?
 - Mark will talk about Switchboard next – is that the only thing?
 - Other management servers (OSCARS or ...) for AL2S?
 - Rest – of course
- Accidental or intentional DOS attacks?
 - You can shoot off your own foot – easy to put traffic on the switch that has to query the controller and insert rule – can overload the switch

User Engagement

- Monthly planning luncheons with participation from Computer Science, OIT management, OIT network, OIT Systems
- Good broad discussions
- Need to get better at identifying additional research needs/users
- More about this later

Production Topology – Setting Up Mark



Part 2 – Rollout Strategies: Benefits & Gotchas

- Switch Evaluations
- Infrastructure Readiness
 - perfSONAR is your friend
- Operational Readiness
 - Reliability of the controller and management tools
 - Fiber evaluation – likely will need to be re-terminated
- Staff Readiness – Network, DevOps, Security
 - Culture – Programmers and Systems guys loose in the network!
- Services/System Security
- Controller – it can be a religious choice
- Rule Management
- End point networks/subnets/hosts – permissions model

Switch Evaluation

- What Did Duke do to evaluate switches?
- Test switches under load from:
 - Rules – 100/300/700 rules/second while transferring files between multiple servers
 - Saturate the network
 - Load up the CPU
 - Use Ryu Simple Switch or POX
 - Look for packet drops
 - Look for the switch to fall over
 - Confirm that the switches supported simple systems
 - Measure traffic flows/cpu load with SNMP polls

What did we evaluate?

- NEC
- Brocade – at the time - didn't support flood all (needed for ARP)
- Arista
- Cisco 4500X
 - Support hybrid mode – more on that later

A Test Lab Is A Requirement!

- We have had a test lab for evaluating switches since the beginning
- Load tests in the lab bled to campus network very unexpectedly
- Code black for the hospital
- Not a good day
- Consolidate test lab
- Isolated from the network
- Influenced the redesign of the network
- Testing done without impact

Test Lab Needs to Map to Typical Edge Usage

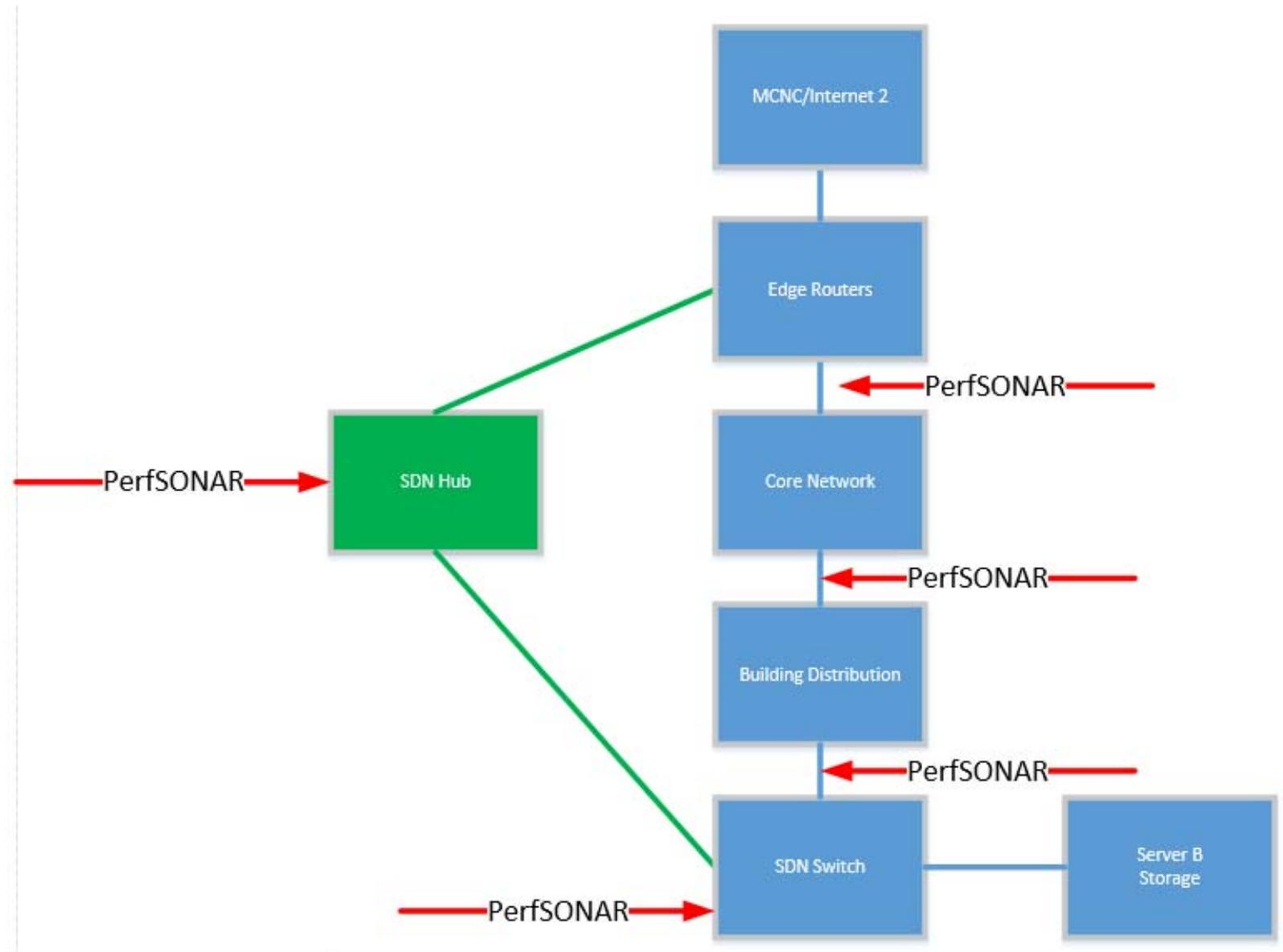
- Lab has switches from NEC, Arista, Cisco
- Lab will be connected to the core the same as research building
- Lab has pre-production controller and switchboard environment
- Lab has 12 dedicated Dell blades with 10G networking

Network Redesign

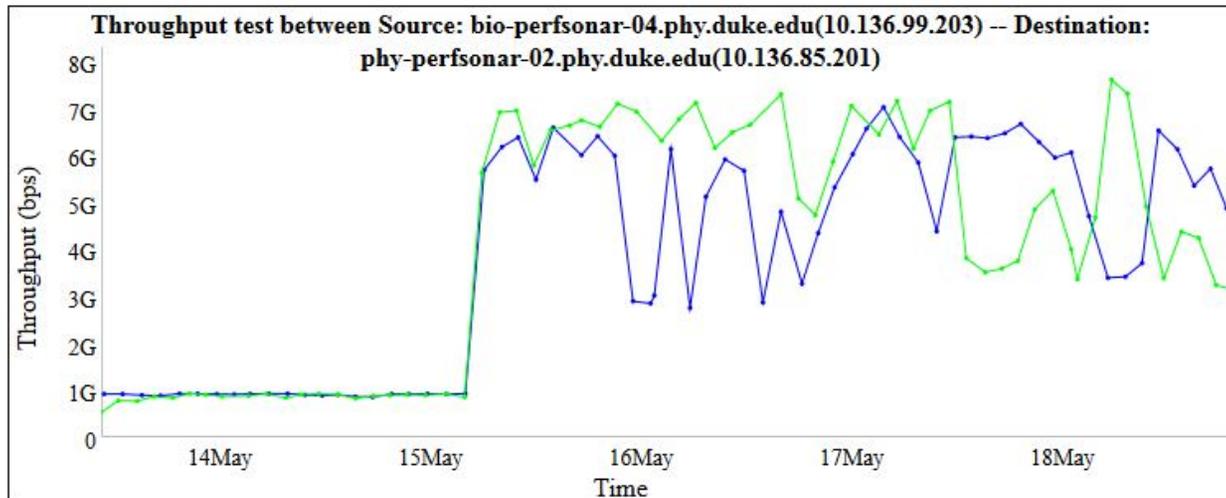
- Keep MPLS
- Separate Functions (VRF transitions, Routing, Aggregation, Edge)
- Move IPS to the edge
- But still keep the dorms and “foreign” networks behind the IPS
- Use IDS internally
- Add an edge routing layer
- Provide 10G or better connectivity to Science buildings

perfSONAR is your friend

- Designed for WAN connections typically
- Measures latency and bandwidth on a regular schedule
- Place nodes at multiple places on the network
- Older version had to split measurements of bandwidth and latency
- New version allow you to split across different interfaces
- Use it to prove that your network is capable of passing the traffic you expect – bandwidth measurements are very useful
- Help to prove that your network performs as expected
- Puppet or other management integration important (but challenging)



Bandwidth Measurements



Graph Key

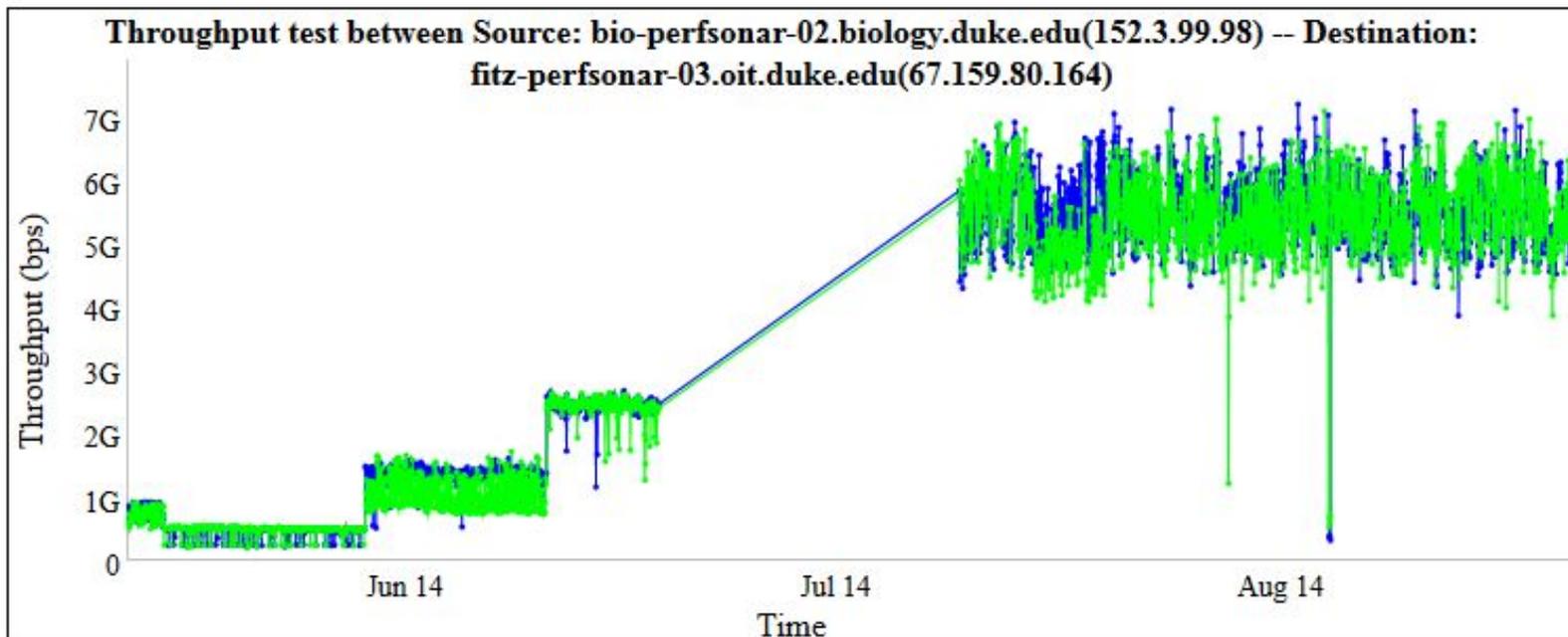
- Src-Dst throughput
- Dst-Src throughput

[< 1 month](#)

[1 month ->](#)

Timezone: Standard Time)

Bandwidth Examples



Graph Key

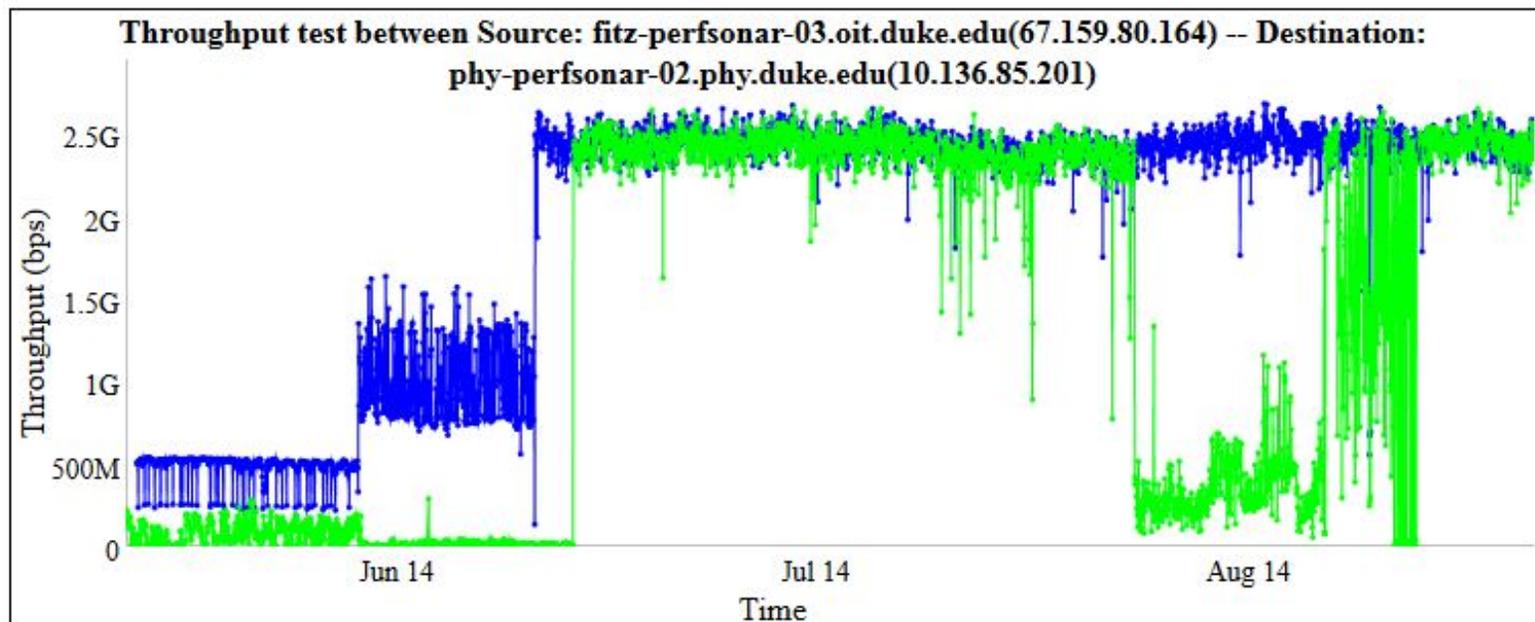
- Src-Dst throughput
- Dst-Src throughput

[< 1 month](#)

[1 month >](#)

Timezone: Standard Time)

Bandwidth Examples – Used for Operational Monitoring



Graph Key

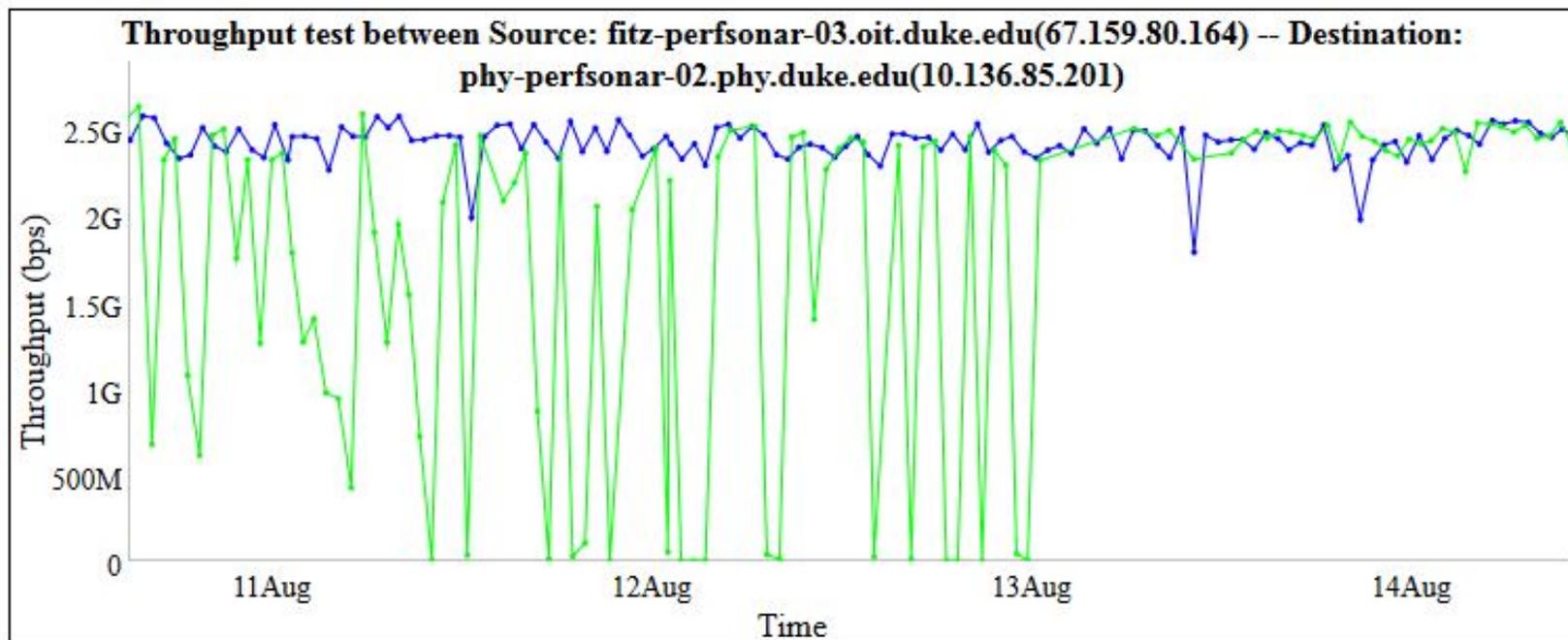
- Src-Dst throughput
- Dst-Src throughput

[< 1 month](#)

[1 month >](#)

Timezone: Standard Time)

Bandwidth Examples



Graph Key

- Src-Dst throughput
- Dst-Src throughput

[< 1 month](#)

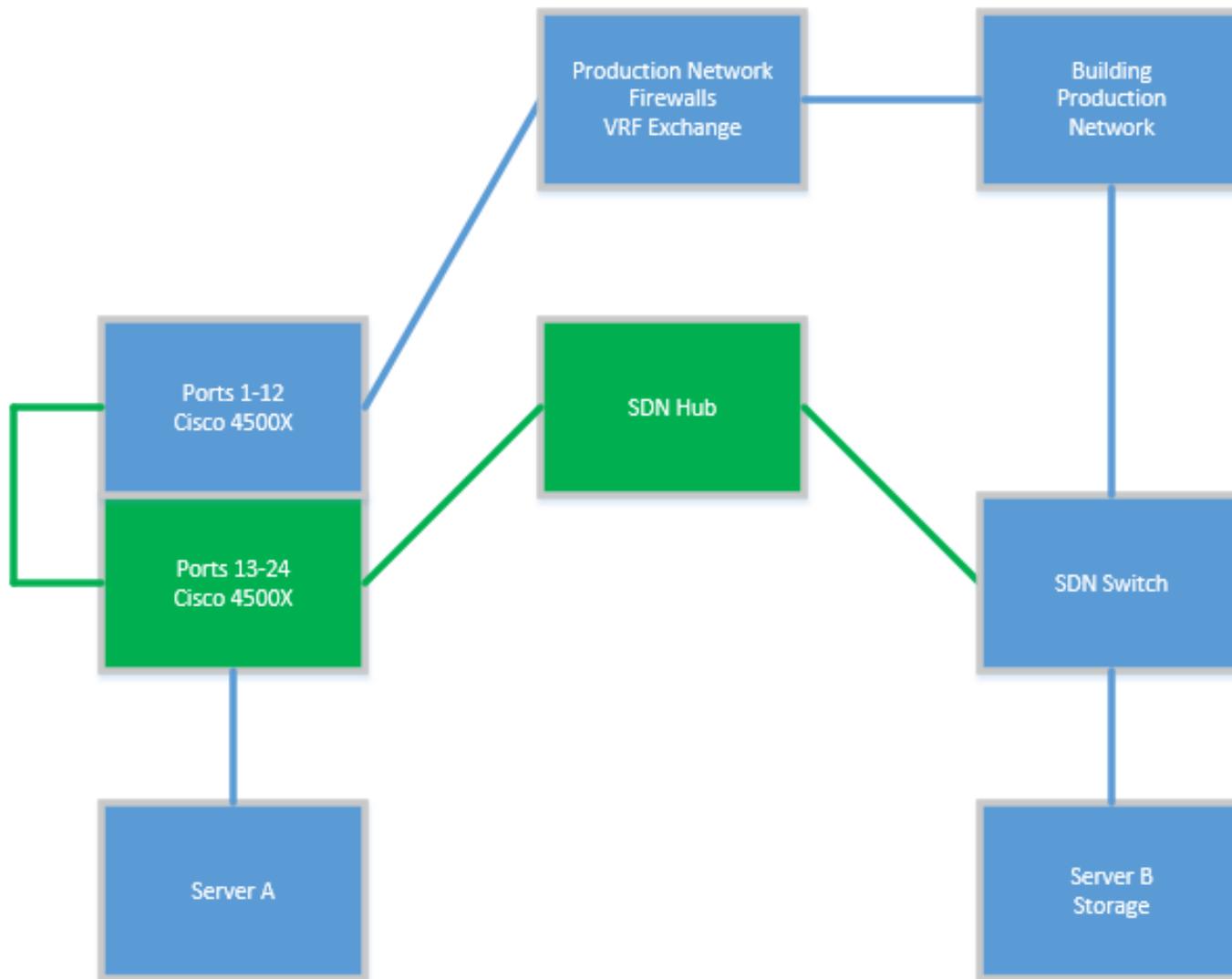
[1 month ->](#)

End User Engagement

- Sometimes it's all about having the fastest connection
- But that doesn't always help
- Complicated workflows that get around current or legacy issues – things are done to avoid problems that were fixed years ago
- If the system doesn't work as expected, science will go on
 - Usual stuff – disk drives from Best Buy, Sneakernet with Thumbdrives

Cisco 4500X – An Interesting Beast

- OpenFlow implemented as a virtual machine inside the switch
- A single switch can support both traditional Cisco IOS ports and OpenFlow ports.
- How would this work?



Cisco 4500X – Hybrid Mode

- Duke has deployed 4500X as the standard building aggregation switch
- Can enable SDN services to any building using a 4500X

SDN Still is a work in progress

- Mixed support for required and optional pieces of the OpenFlow standards
- Netflow data is not “built in”
- Accidental DOS attacks are possible – deliberate as well
- Need to program in services like DHCP
- As we scale up – need to be able to have effective time-out of rules that are not active – only so much capacity

Acknowledgements

- **Work Supported by NSF on the following grants:**

NSF OCI-1246042 - CC-NIE

NSF CNS 1243315 - EAGER