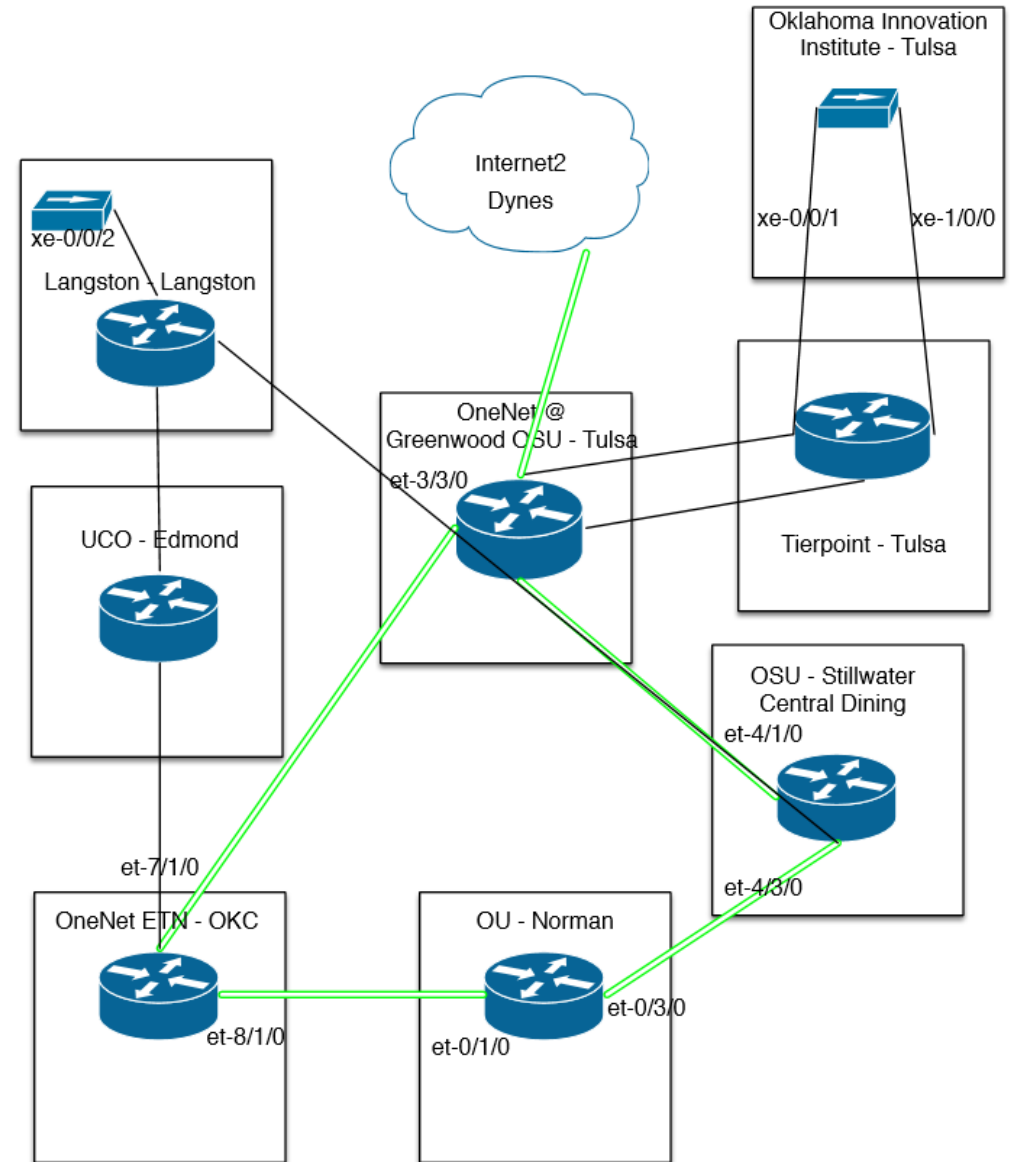


Abstract

This session presents a stage model that describes the ongoing deployment of a Science DMZ known as the OneOklahoma Friction Free Network (OFFN) in order to guide multi-campus collaborations on their journey to a robust cyberinfrastructure model. This session defines implementation stages and describes the deliverables, benefits, challenges, best practices, and metrics for each stage. The results of the initial implementations are expected to show that metrics including speed, stack size, context switching, signal deliveries, and socket message traffic is improved and latency and errors including page faults are reduced. It is hoped that this model will serve as a sustainable model for EPSCoR states hoping to maximize research efficiency and minimize risk in their engagement of statewide research interests.

Introduction

- Shared amongst the University of Oklahoma, Oklahoma State University, Langston University, University of Central Oklahoma, and the Tandy Supercomputing Center (TSC), as the Oklahoma Innovation Institute (OII), the ring and spur implementation of a NSF Cyber Connectivity - Networking Integration and Engineering (CC-NIE) grant, known as the OneOklahoma Friction Free Network (OFFN), is discussed in terms of its deployment as a multi-institutional Science DMZ for friction free and Software Defined Networking (SDN)[1].
- University of Central Oklahoma and University of Tulsa are in the approval process for membership in OFFN.



Drawing by David Brockus

— 10Gb —
— 100Gb —

What is a Science DMZ?

Science DMZs enable high speed file transfers to/from off-campus locations.

In 2011, the Campus Bridging task force of the National Science Foundation (NSF) Advisory Committee on Cyberinfrastructure (ACCI) recommended[2]:

“The NSF should create a new program funding high-speed (currently 10 Gbps) connections from campuses to the nearest landing point for a national network backbone. The design of these connections must include support for dynamic network provisioning services and must be engineered to support rapid movement of large scientific data sets”.

Moving a 1 Terabyte file[3]:

10 Mbps network	10 Mbps network	USB 2.0 portable disk	1 Gbps network	10 Gbps network
300 hours	30 hours	20-30 hours	3 hours	20 minutes



Photo courtesy of the Oklahoma Innovation Institute

What is the Science DMZ architecture?

As specified by Esnet[4a], three key components:

1. “Friction Free” network path -Allows scientists to perform research and send and receive data without constraints, such as firewalls, associated traditional enterprise networks.

- Highly capable network devices
- Virtual circuit connectivity option
- Security policy and enforcement specific to science workflows
- Located at or near site perimeter

2. Dedicated, high-performance Data Transfer Nodes (DTNs)

- Hardware, operating system, libraries optimized for transfer
- Optimized data transfer tools

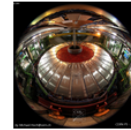
3. Performance measurement/test node

- perfSONAR (next slide)



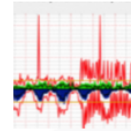
Science DMZ at the University of Florida

Introducing the Science DMZ ESnet's network design pattern, called the Science DMZ, solves local area network problems so that science researchers can more effectively use fast research networks to transmit large amounts of data. The Science DMZ allows data transfers to flow through a dedicated portion of a local area network thereby accelerating the data transfer rate. By enabling this... [READ MORE »](#)



Science DMZ Implemented at CU Boulder

The University of Colorado, Boulder campus was an early adopter of Science DMZ technologies. Their core network features an immediate split into a protected campus infrastructure (beyond a firewall), as well as a research network (RCNet) that delivers unprotected functionality directly to campus consumers. Figure 1 shows the basic breakdown of this network, along with the placement of measurement... [READ MORE »](#)



Science DMZ for Pennsylvania State University & Virginia Tech Transportation Institute

The Pennsylvania State University's College of Engineering (CoE) collaborates with many partners on jointly funded activities. The Virginia Tech Transportation Institute (VTTI), housed at Virginia Polytechnic Institute and State University, is one such partner. VTTI chooses to collocate computing and storage resources at Penn State, whose network security and management is implemented by local... [READ MORE »](#)



Science DMZ National Oceanic and Atmospheric Administration

The National Oceanic and Atmospheric Administration (NOAA) in Boulder houses the Earth System Research Lab, which supports a "reforecasting" project. The initiative involves running several decades of historical weather forecasts with the same current version of NOAA's Global Ensemble Forecast System (GEFS). Among the advantages associated with a long reforecast dataset is that model forecast... [READ MORE »](#)



Science DMZ Implemented at NERSC

In 2009, both NERSC and OLCF installed data transfer nodes (DTNs) to enable researchers who use their computing resources to move large data sets between each facility's mass storage systems. As a result, wide area network (WAN) transfers between NERSC and OLCF increased by at least a factor of 20 for many collaborations. As an example, a computational scientist in the OLCF Scientific Computing... [READ MORE »](#)



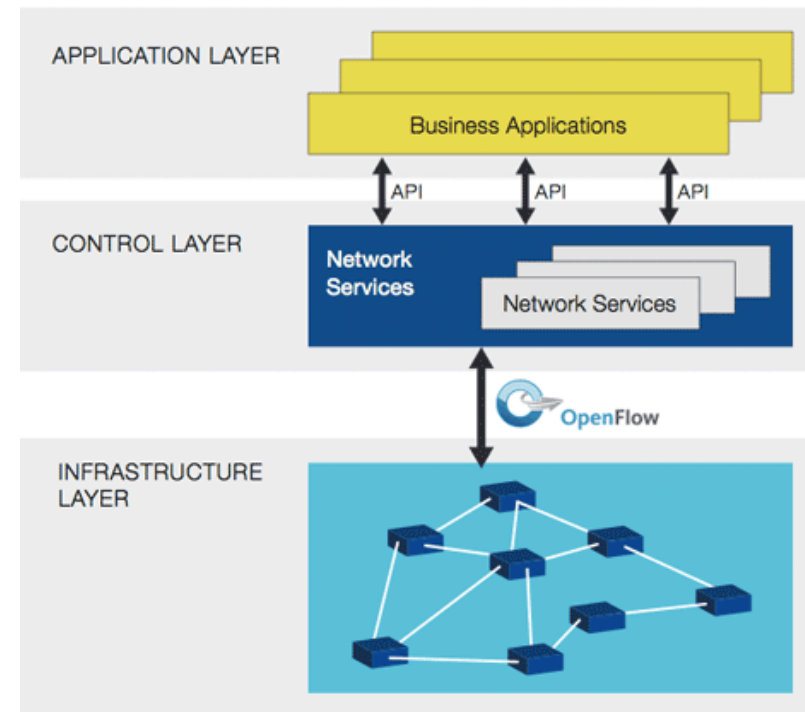
Science DMZ for ALS

Many beamline scientists at Berkeley Lab's Advanced Light Source (ALS) are or will be experiencing slower network speeds because of instrument upgrades. These new instruments, or more specifically detectors, are yielding immense data output rates on the scale of hundreds of megabytes per second, which are causing bottlenecks in many of the beamline workflows. For example, at one of the X-ray... [READ MORE »](#)

What is Software-Defined Networking (SDN)?

The OpenNetworking Foundation defines SDN as “The physical separation of the network control plane from the forwarding plane, and where a control plane controls several devices” [4].

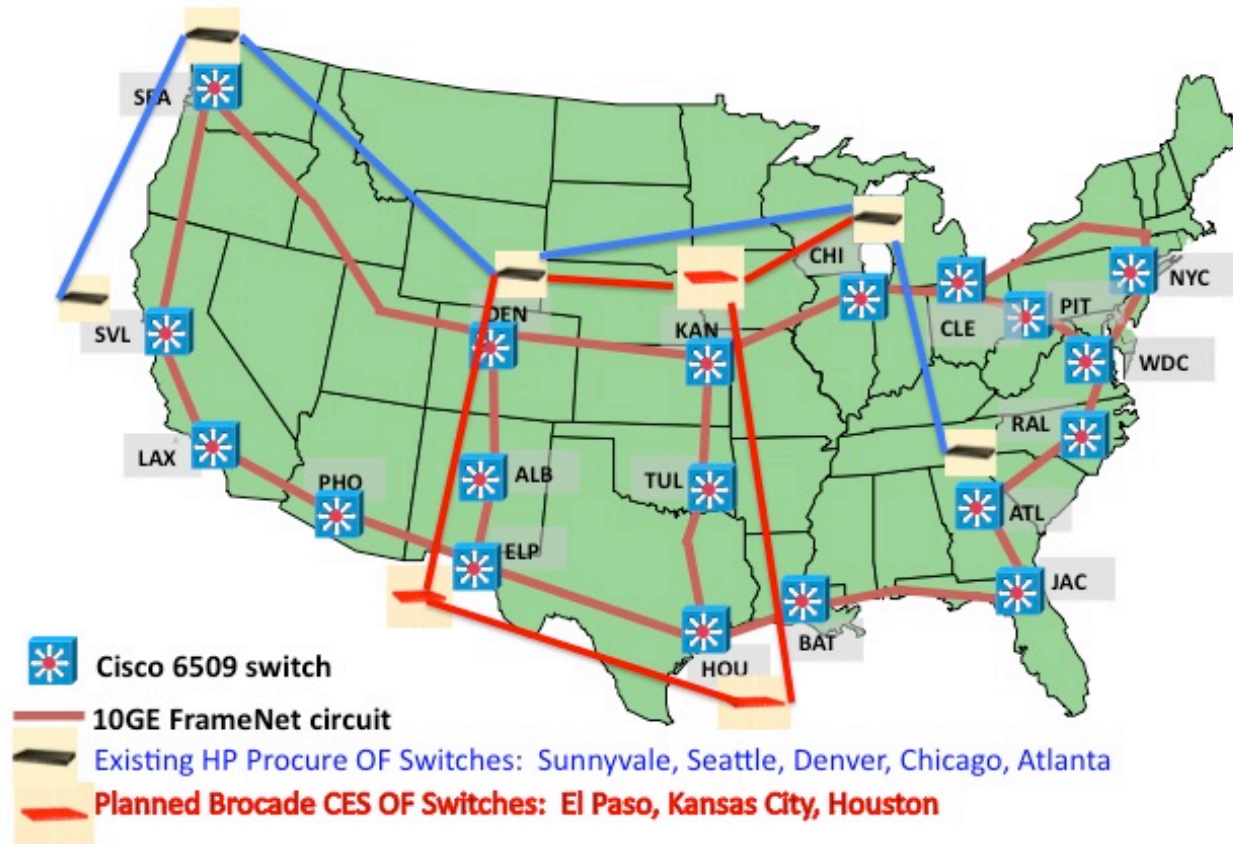
The OpenFlow[®] protocol separates the data and control paths from operating on the same device and prevents exposure of the internal workings of network devices [5].



Graphic used with permission from OpenNetworking.org

The bigger picture

NLR OpenFlow Backbone



graphic used with permission from H. Dempsey (BBN Technologies) and dissemination by geni.net user group [6a]

How will we use OFFN?

The initial domain Science, Technology, Engineering & Mathematics (STEM) research projects include numerical weather prediction, high energy physics, bioinformatics and weather radar, and research slated for OFFN has already begun expanding to other research disciplines.

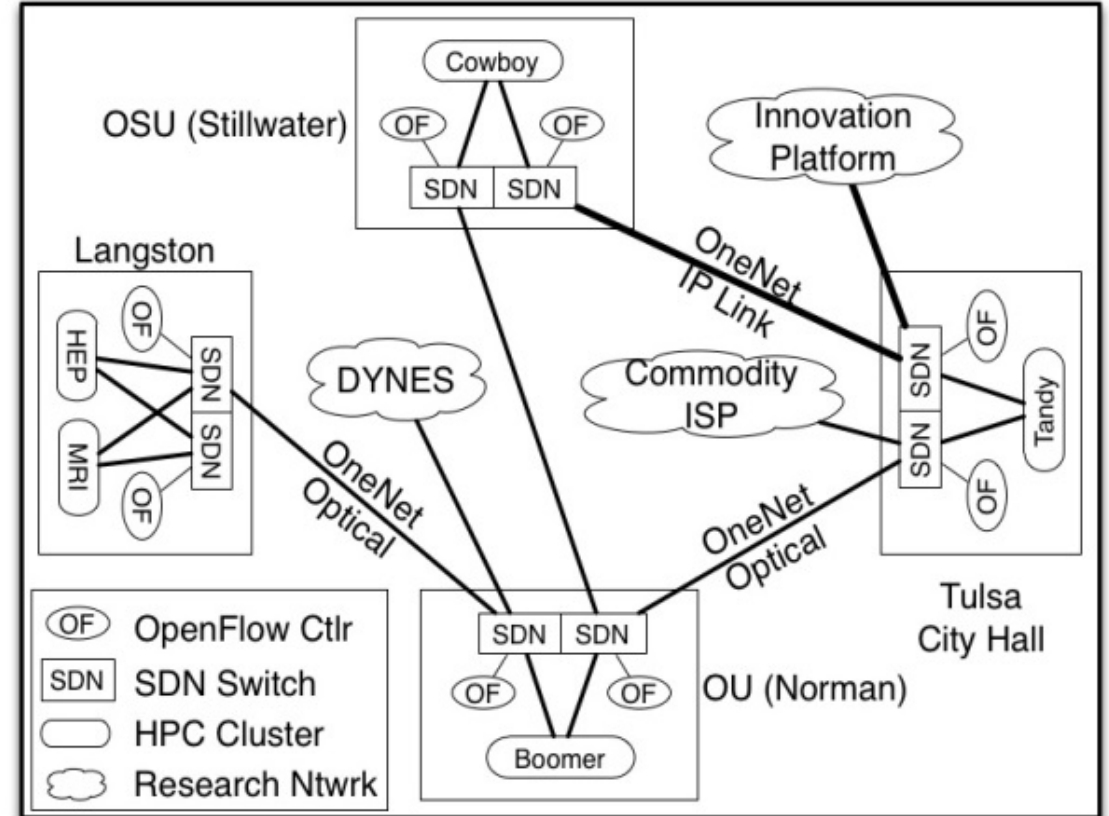


Photo courtesy of Oklahoma State University

- **ATLAS Tier 2 and DØ High Energy Physics:** One of the largest physical science collaborations ever, with 2000+ physicists (including ~800 students) from 150+ universities and laboratories in 34 nations, ATLAS is a High Energy Physics (HEP) experiment at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) that explores the fundamental nature of matter and the basic forces that shape our universe, by examining head-on collisions of protons of enormously high energy. OUHEP personnel have contributed significantly to network monitoring research to quickly pinpoint network-related problems [8].
- **Real-time Numerical Weather Prediction:** Located at OU's National Weather Center (NWC), CAPS is a leader in storm scale Numerical Weather Prediction (NWP) and Data Assimilation (DA) research. Since 2007, CAPS has led a Spring Real-time Storm Forecast Experiment using systems both on XSEDE/TeraGrid and at OSCER, as part of an ongoing collaboration with other academic researchers and research and operational centers in the National Oceanic and Atmospheric Administration (NOAA), e.g., the Storm Prediction Center. Forecast suites have included 25-50 member ensembles, using multiple NWP models (WRF-NMM, WRF-ARW, ARPS, COAMPS) at 4 km resolution over the Continental US (CONUS); hourly 2D fields and images are generated at NICS and sent back to OU for real time use and evaluation by meteorologists from research and operations communities in the Hazardous Weather Testbed (HWT). Experimental forecasts of convection initiation and severe thunderstorms are created from the ensemble forecasts, to evaluate their utility [9,10, 12, 13].
- **Weather Radar:** The societal impact of weather-related hazards such as tornadoes, high winds, snow, and flash floods is substantial, but can be mitigated by timely, accurate prediction via adequate observations and advanced scientific understanding. Weather radar is one of the most crucial instruments to improve the lead time for hazardous weather, benefitting operational and research communities. Weather radar advancements such as polarimetric capability and phased array technology provide valuable information to transform our understanding of radar meteorology [9, 12].
- **Eco-informatics and Geo-Informatics:** EOMF leads efforts in mapping land use and land cover change at various spatial scales, modeling terrestrial ecosystem gross and net primary production, and predicting risk of highly pathogenic avian influenza (subtype H5N1) in Asia, using optical sensors (e.g., Landsat, SPOT-VEGETATION, MODIS, IKONOS, WorldView), synthetic aperture radar images (e.g., PALSAR), and LiDAR images across spatial domains from sites worldwide. Datasets from the US Geological Survey EDC data portal, including MODIS and Landsat require ~320 GB/day for surface reflectance products, and ~80 GB/day for surface temperature products [13-15].
- **Data Networks Research:** Remote visualization with interactive steering needs real-time multicasting, to ensure that (a) data reaches all destination nodes at roughly the same time, and (b) at any given time, only one host can control the visualization, to avoid visualization or data inconsistency. Floor control is the problem of providing exclusive access in a communication session; control mechanism is to use a circulating token: at any given time, a controller holds the token so any node requesting the floor needs the controller for the token. Once the requesting node gets the token, all other requests for the token are queued or processed according to application demands and features. The node relinquishing the floor releases the token to the controller, and the process continues. This project has proposed distributed floor control protocols that are effective on overlay networks., and plans include extending floor control protocols to enable them to work in OpenFlow networks, with testing on a real-time simulation and visualization experiment [16-17].

OFFN Goals

- Provide a proven, commercial off-the-shelf hardware platform backed with vendor support.
- Realize the Science DMZ goals through the use of a truly independent network at each campus site. The network deployment will consist of dedicated optical pathways to the optical transport provider (OneNet), as well as to the local campus backbone where desired.
- Deploy a fully virtualized infrastructure, to be used simultaneously by multiple research entities, presented to each entity as a dedicated “slice” of the overall resource.
- Leverage federation to provide oversight and visibility into the operations of the virtualized platform.
- Realize the full potential of OFFN through awareness, training, site-specific hand-off, and communities of support for OFFN adopters.



OFFN implementation

Each of OFFN's client site deployments will consist of the following resources, two of each component per institution for redundancy and failover:

- SDN switches (Dell Force10 S4810, each 1U, 14.39 lbs, 1194 BTU/hr, 350 W)
- Platform support switches (Dell PowerConnect 7024, each 1U, 14 lb, 300.5 BTU/hr, 88 W)
- Servers (Dell PowerEdge R720xd, each 2U, 64.9 lbs for 2.5" drives, 4100 BTU/hr, 485 W)

Total physical requirements per institution are 8U rack space, 187 lbs, 11,189 BTUs/hr and 1846 W of power.

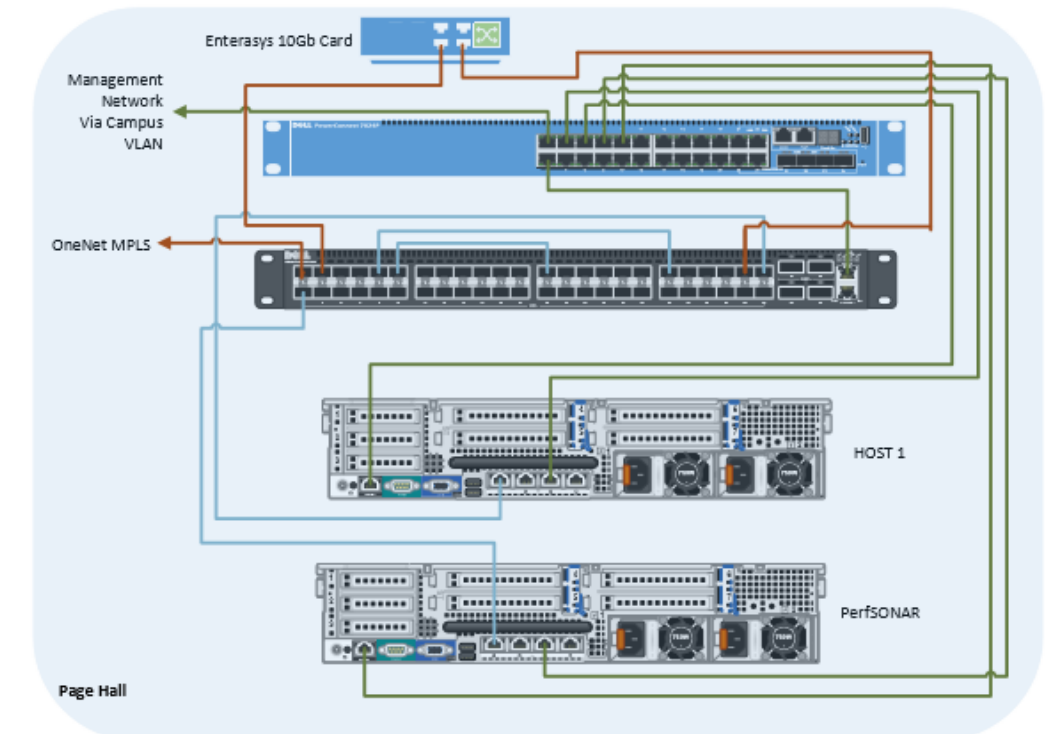
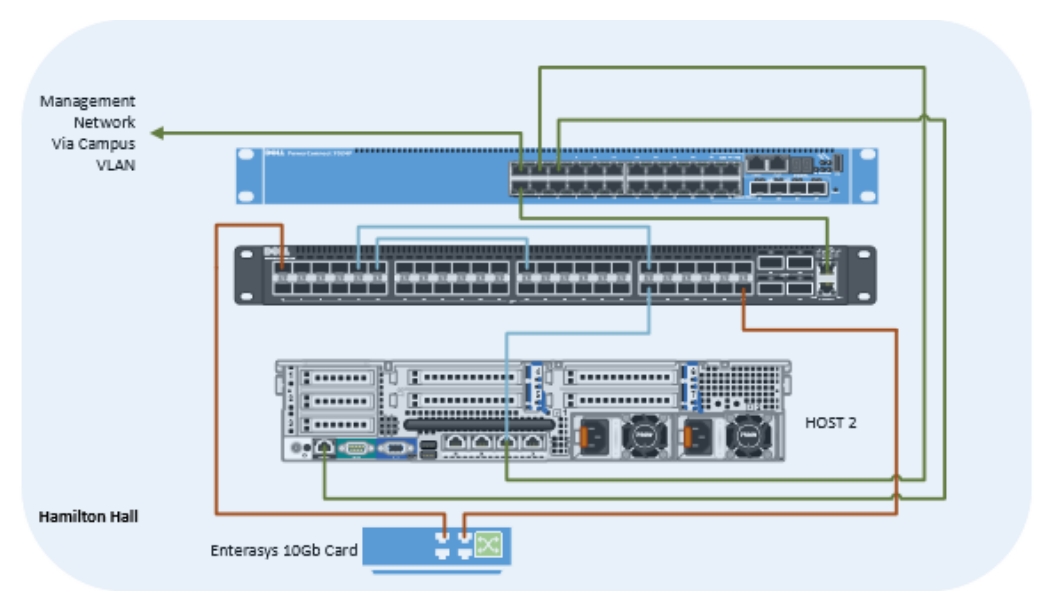
Collectively, these institutional assets provide a hybridized foundation both for production data and for multiple instances of SDN with associated OpenFlow control, allowing OFFN to be both flexible for rapid deployment needs and stable for always-on requirements.

- Software (all Open Source and/or free): the OS virtualization platform, Linux host and guest OS, SDN controller, performance testing, as well as monitoring.
- As part of the toolset for statistics gathering, performance monitoring, and health checking, perfSonar is provided via physical devices at each site, already deployed by OneNet (ps.onenet.net), connecting to other perfSonar instances at all OFFN sites, to provide a true end-to-end analysis of network performance.

Example site implementation-Langston University

Additional Equipment needs:

- 4ea 3m jumpers SC to LC (Single mode)
- 2ea Single Mode SR Optics for S4810
- 4ea Multi Mode SR Optics for S4810
- 2ea 4 port 10Gb Cards for Enterasys K6
- 8ea 3m Multi Mode SR Optics for Enterasys K6
- 8ea 1m TwinAx cables
- 2ea C13 power cable for 7024
- 2ea R720 power cables (standard plug)
- 2ea 7024 power cables (standard plug)
- 10ea 3M Cat6 network cables
- 2ea Single Mode SR Optics for OneNet equipment
- Copper ports for OFFN Management Network
- Copper Ethernet ports for OFFN Management Network
- Interconnecting Fiber



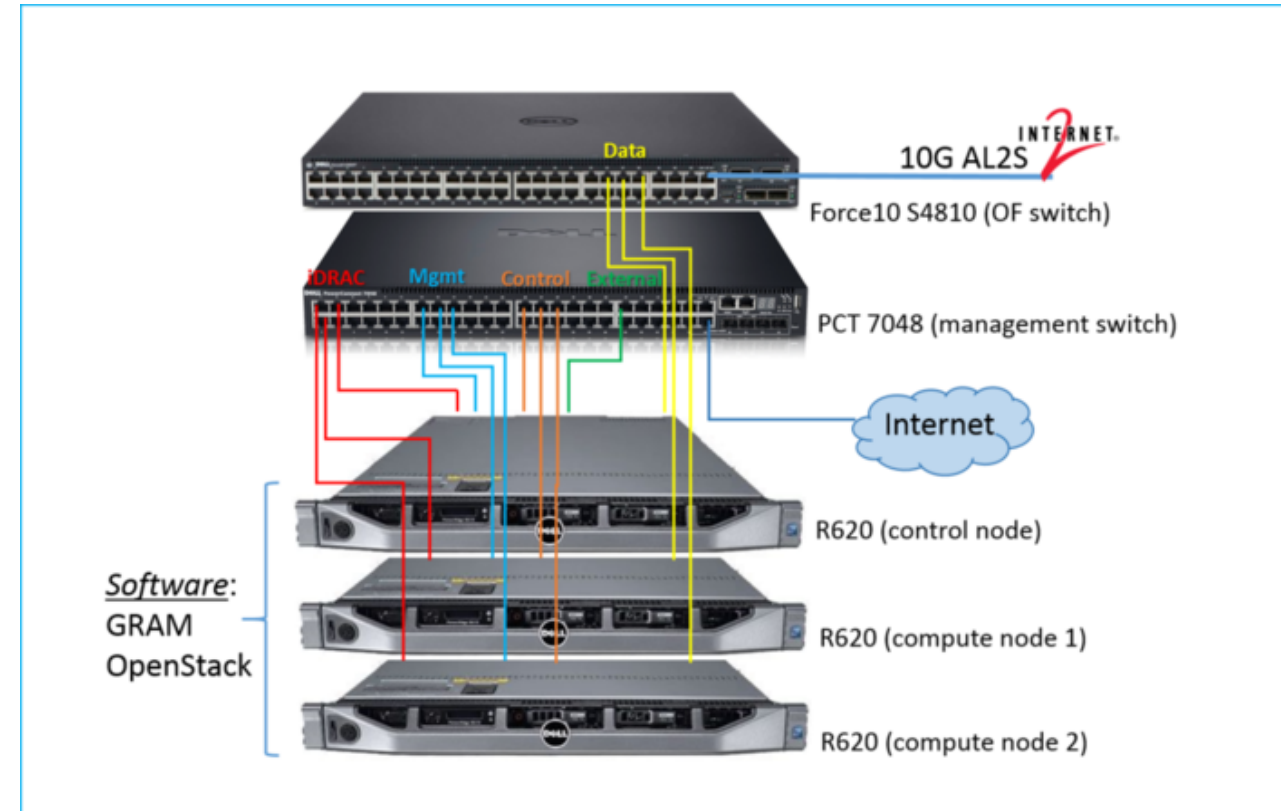
drawing by Matt Younkins

— Cat 6
— Twinax
— Fiber

Example GENI implementation –OSU site

- GENI racks are being implemented to meet the goals specified by rack requirements developed by GENI [7] which include requirements for: software, integration, monitoring, production aggregate, local aggregate owner, and experimenter.

<http://groups.geni.net/geni/wiki/GeniRacks>



Matt will send citation or use correct Rack Config diagram

Our checklist for deploying an OpenGENI site?

- Does the connection between OpenFlow switches appear as a layer 1 connection?
 - using dedicated fiber, existing pseudo-wire technologies, OF switches to pass traffic, or applying special configurations to traditional layer 2 VLANs such as disabling mac learning and STP.
- Additional hardware for preparing OSU's network environment for hosting a GENI rack?
- Identification of transceiver connecting the GENI rack to the 10Gb uplink connection?
- Post-Purchase ****Important:** If the following information is not or cannot be provided prior to the rack being sent to your site, you will receive the hardware with only a bare OS installed on all servers and the switch un-configured. In order to have your rack delivered as a ready to be deployed product, you **MUST** provide the following:
- Identification of Hostname, GENI identifier, and external plane VLAN number
- How do we check conflicts?
- DNS server
- GRAM software Range and subnet of public IP addresses for virtual machines:
- Define a range of contiguous IP addresses to be assigned to the Virtual Machines.
- External network configuration on control nodes and compute nodes
- Assign the following for the control node attached to the system. (DHCP is not recommended as a change in IP address can disrupt the software utilized by GENI)
- Static IP address:
- Subnet:
- Gateway:
- DNS Server:
- Will the following ports be allowed through your campus/network firewall, to the entire rack subnet:
- 22 - SSH
- 25 - SMTP (outbound connections only, from control node)
- 80 - HTTP (must also allow outbound connections from control node)
- 443 - HTTPS (must also allow outbound connections from control node)
- 843 - Flash Policy Server
- 3000-3300 - SSH access for experimenter resources

Expected OFFN Metrics

In general, metrics describing effectiveness at sites, at deployment connections, from the OFFN commercial provider and within specific research projects will be used to evaluate the project as a whole as well as a sum of its parts.

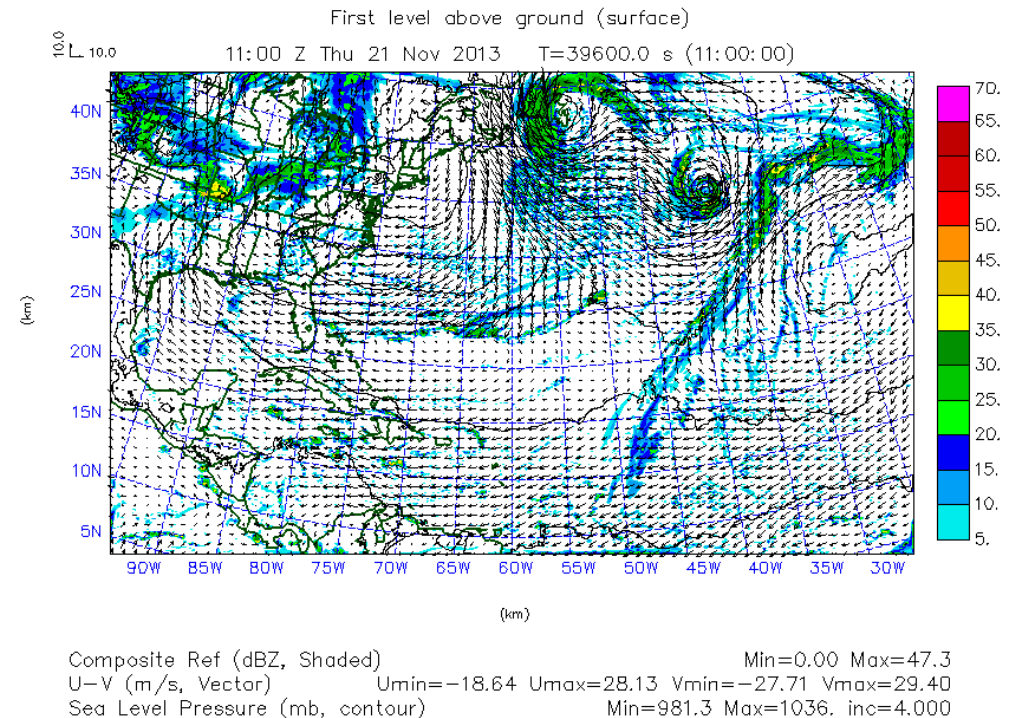
Specifically:

Triangulation of the below as a measurement of signal delivery:

- Bitstream reliability
 - Payload
 - Overhead calculations will be captured as a Switch reliability, capacity, and availability metrics
- Increased bandwidth, traffic metrics including:
- Last mile metrics
 - OneNet
 - local campus backbone
 - Server and switch hardware performance metrics
 - Metrics related to software-defined networking
 - Metrics related to specific research
 - presumably increase in: storage capacity, processing, search time, etc.,
 - user satisfaction with hardware and software resources
 - Information on statewide collaborations (research teams, IT teams) (presumably an increase in research collaboration)

How will we gather effectiveness data?

- The scaling behavior of selected applications will be averaged across multiple runs (5 executions) performed at different times.
- For sequential performance compare each unit (Schooner, Cowboy, Lucille, Buddy, Tandy, etc.) in terms of processor performance
- Microbenchmarking may reveal that network performance related to the 100G OneNet connection is a dominant factor
- Investigation of failure to reach performance benchmarks
- Possible performance limitations:
 - short ssh sessions
 - data transfer node underperformance
 - multiple cores
 - virtual instances\communication
 - variability in execution time across runs
 - shared resource performance
- A representative, test case for benchmarking activity is an interacting Spring Storms simulation
 - A next-generation mesoscale numerical weather prediction system of two dynamical cores, a data assimilation system, and a software architecture facilitating parallel computation and system extensibility
 - Typical problem sizes are in the range of [60 raw timesteps on disk at a time, with allotments of 2 TB for the raw data and 2 TB for the metadata].



How do researchers work on a Science DMZ?

Emerging paradigms in the field include:

- **Data Center:** RN aggregation switches at 10-Gb/s, providing direct connections
- **Lab-based:** A researcher, department, or other entity could purchase additional Brocade switches to expand the network in a manner similar to an edge connection.
- **Ethernet Fabric:** 10Gb/s Ethernet Fabric Switch. This option provides one (1), 10Gb/s connection, an additional 10Gb/s fiber edge port, and either 24 or 48 1Gb/s connections to individual research workstations.
- **Compliance Port:** An individual “research or compliance port” on an existing ITS managed, converged network switch, Such as a virtualized wall jack network connection.



Learn More

- <http://fasterdata.es.net/science-dmz/>

References

1. H. Neeman, D. Akin, J. Alexander, D. Brunson, S. P. Calhoun, J. Deaton, F. Fondjo Fotou, B. George, D. Gentis, Z. Gray, E. Huebsch, G. Louthan, M. Runion, J. Snow, B. Zimmerman, 2014: "The OneOklahoma Friction Free Network: Towards a Multi-Institutional Science DMZ in an EPSCoR State". XSEDE '14: Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment, 49. DOI:[10.1145/2616498.2616542](https://doi.org/10.1145/2616498.2616542).
2. National Science Foundation Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging Final Report, March 2011, p.78. http://www.nsf.gov/od/oci/taskforces/TaskForceReport_CampusBridging.pdf
3. D. Smith, 2014: "Campus Network Design Science DMZ". Retrieved from http://www.interlab.ait.ac.th/training/2014/PPT/Friday/01.1.1_Science_DMZ.pdf.
- 3a. The perfSONAR project. "Performance Beacon Deployment Use Case" graphic. Retrieved from <https://www.perfsonar.net/deploy/deployment-use-cases/>
4. <https://www.opennetworking.org/sdn-resources/sdn-definition>
- 4a. <https://fasterdata.es.net/science-dmz/>
- 4b <http://www.es.net/science-engagement/case-studies/science-dmz-case-studies/>
- 4c. <https://www.perfsonar.net/deploy/hardware-selection/deployment-examples/>
5. <http://archive.openflow.org/wp/learnmore/>
- 6a. H. Dempsey, "NLR Open Flow1" graphic file NLR OpenFlow1.jpeg. Retrieved from <http://groups.geni.net/geni/wiki/NLROverview>
7. GENI, 2016: "GENI Rack Requirements". Retrieved from <http://groups.geni.net/geni/wiki/GeniRacks>
8. S. McKee, A. Lake, P. Laurens, H. Severini, T. Wlodek, S. Wolff and J. Zurawski, 2012: "Monitoring the US ATLAS Network Infrastructure with perfSONAR-PS." *Journal of Physics: Conference Series*, 396 042038. DOI:[10.1088/1742-6596/396/4/042038](https://doi.org/10.1088/1742-6596/396/4/042038).
9. F. Kong, M. Xue, K. W. Thomas, Y. Wang, K. Brewster, A. J. Clark, M. C. Coniglio, J. Correia Jr., J. S. Kain, and S. J. Weiss, 2014: "CAPS Storm-Scale Ensemble Forecasting System: Impact of IC and LBC perturbations". *Extended Abstracts, 26th Conf. on Weather Analysis and Forecasting / 22nd Conference on Numerical Weather Prediction, American Meteorological Society*, Paper 119. Retrieved from http://ams.confex.com/ams/94Annual/webprogram/Manuscript/Paper234762/Kong_26WAF22NWP-extendedAbstract.pdf.
10. A. J. Clark, S. J. Weiss, J. S. Kain, I. L. Jirak, M. Coniglio, C. J. Melick, C. Siewert, R. A. Sobash, P. T. Marsh, A. R. Dean, M. Xue, F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, J. Du, D. R. Novak, F. E. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, T. L. Jensen and J. Correia Jr., 2012: "An Overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment." *Bulletin of the American Meteorological Society*, 93, 55-74. DOI: [10.1175/BAMS-D-11-00040.1](https://doi.org/10.1175/BAMS-D-11-00040.1).
11. J. S. Kain, M. Xue, M. C. Coniglio, S. J. Weiss, F. Kong, T. L. Jensen, B. G. Brown, J. Gao, K. Brewster, K. W. Thomas, Y. Wang, C. S. Schwartz and J. J. Levit, 2010: "Assessing advances in the assimilation of radar data within a collaborative forecasting-research environment." *Weather Forecasting*, 25, 1510-1521. DOI: [10.1175/2010WAF2222405.1](https://doi.org/10.1175/2010WAF2222405.1).
12. D. R. Stratman, M. C. Coniglio and M. Xue, 2011: "Using Traditional and Spatial Verification Methods to Evaluate Real-Time Model Forecasts of Convection." *Proceedings of the 24th Conference on Weather Forecasting/20th Conference on Numerical Weather Prediction of the American Meteorological Society*, Paper 11B.2.
13. National Research Council, Committee on Progress and Priorities of U.S. Weather Research and Research-to-Operations Activities, 2010: "When Weather Matters: Science and Service to Meet Critical Societal Needs." Washington, DC: National Academy Press. DOI: [10.17226/12888](https://doi.org/10.17226/12888).
14. R. J. Serafin and J. W. Wilson, 2000: "Operational weather radar in the United States: Progress and opportunity." *Bulletin of the American Meteorological Society*, 81, 501-518. DOI: [10.1175/1520-0477](https://doi.org/10.1175/1520-0477).
15. J. W. Dong, X. M. Xiao, B. Q. Chen, N. Torbick C. Jin, G. L. Zhang and C. Biradar, 2013: "Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery." *Remote Sensing of Environment*, 134, 11. DOI: [10.1016/j.rse.2013.03.014](https://doi.org/10.1016/j.rse.2013.03.014).
16. C. Jin, X. M. Xiao, L. Merbold, A. Ameth, E. Veenendaal and W. Kutsch, 2013: "Phenology and gross primary production of two dominant savanna woodland ecosystems in South Africa." *Remote Sensing of Environment*, 135, 12. DOI: [10.1016/j.rse.2013.03.033](https://doi.org/10.1016/j.rse.2013.03.033).
17. V. Martin, D. U. Pfeiffer, X. Y. Zhou, X. M. Xiao, D. J. Prosser, F. S. Guo and M. Gilbert, 2011: "Spatial Distribution and Risk Factors of Highly Pathogenic Avian Influenza (HPAI) H5N1 in China." *Plos Pathogens*, 7. DOI: [10.1371/journal.ppat.1001308](https://doi.org/10.1371/journal.ppat.1001308).
18. S. M. Banik, S. Radhakrishnan, and C. N. Sekharan, 2007: "Multicast routing with delay and delay variation constraints for collaborative applications on overlay networks." *IEEE Transactions on Parallel and Distributed Systems*, 18(3), 421-431. Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4079539>.
19. S. M. Banik, S. Radhakrishnan, V. Sarangan and C. N. Sekharan, 2008: "Implementation of distributed floor control protocols on overlay networks." *IEEE Transactions on Parallel and Distributed Systems*, 19 (8), 1057-1070. DOI: [10.1109/TPDS.2007.70807](https://doi.org/10.1109/TPDS.2007.70807)

Additional Contributors

David Akin, University of Oklahoma
david.akin@ou.edu

Joshua Alexander, University of Oklahoma
jalexander@ou.edu

S. Patrick Calhoun, University of Oklahoma
phineas@ou.edu

David Brockus, University of Oklahoma
dbrockus@ou.edu

James Deaton, OneNet
jed@onenet.net

Franklin Fondjo Fotou, Langston University
ffondjo@langston.edu

Debi Gentis, University of Oklahoma
dgentis@ou.edu

Brandon George, University of Oklahoma
bcg@ou.edu

Zane Gray, University of Oklahoma
zgray@ou.edu

John Hale, University of Tulsa
john-hale@utulsa.edu

Peter Hawrylak, University of Tulsa
peter-hawrylak@utulsa.edu

Eddie Huebsch, University of Oklahoma
ehuebsch@ou.edu

Evan Lemley, University of Central Oklahoma
elemley@uco.edu

George Louthan, Oklahoma Innovation Institute
george.louthan@tulsahpc.org

Matt Runion, University of Oklahoma
mrunion@ou.edu

Joel Snow, Langston University
jmsnow@lunet.edu

Brett Zimmerman, University of Oklahoma
zim@ou.edu