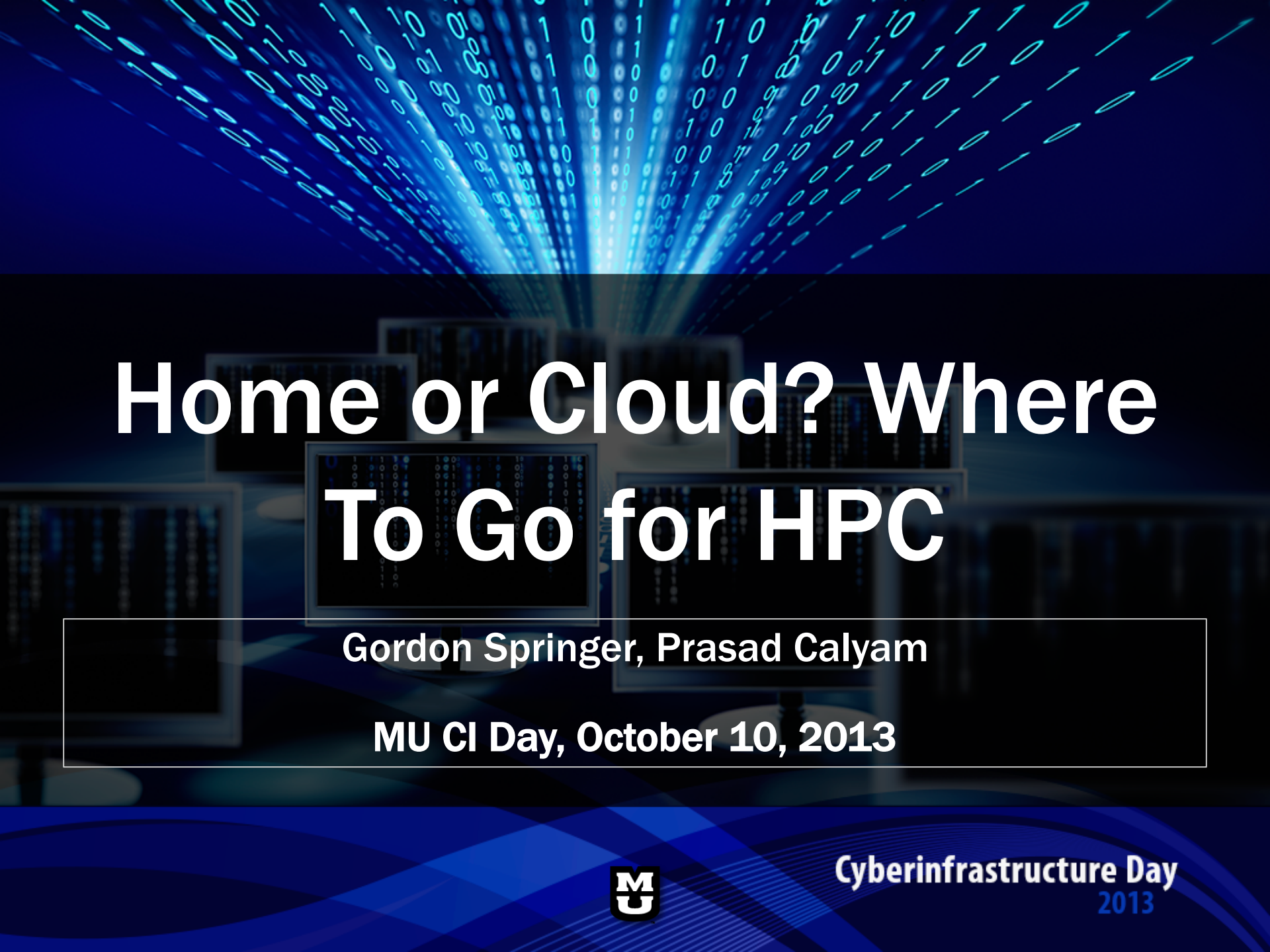


Home or Cloud? Where To Go for HPC

Gordon Springer, Prasad Calyam

MU CI Day, October 10, 2013

Cyberinfrastructure Day 2013

Scaling Science for Performance: Implementing a Cost-Effective "Big Data" Environment for Research Analysis

Gordon K Springer, PhD

Scientific Director, UM Bioinformatics Consortium

Associate Professor, Computer Science

University of Missouri

Cyberinfrastructure Day 2013

Building Infrastructure Support

- The University has built cyberinfrastructure (CI) to support researchers and minimize duplication of expensive resources like computational and storage resources, and advanced networking services.
- Building our own data analysis tools as a strategy has served well up until now.
- **BUT**, the explosive increase in infrastructure and analysis demands has led to a re-evaluation of the strategy and possible modification of the approaches.

The Dawn of a New Era

- “Fabric computing" is enabling the University’s bioinformatics consortium (UMBC) to scale automated pipelines and alleviate the data management burden associated with traditional high-performance computing architectures.
- The environment facilitates research collaborations by enabling scientists to rapidly and cost effectively develop custom pipelines using their preferred analysis tools locally and at other institutions.

What's the Problem?

- Providing research infrastructure to support increasing numbers of scientists with diverse interests is a major issue.
- Data is getting produced at an ever increasing rate and data analysis of the flood of data is increasing faster than the current staff can handle. No additional people have been added to the analysis efforts despite the significant increase in usage and expectations of the researchers.

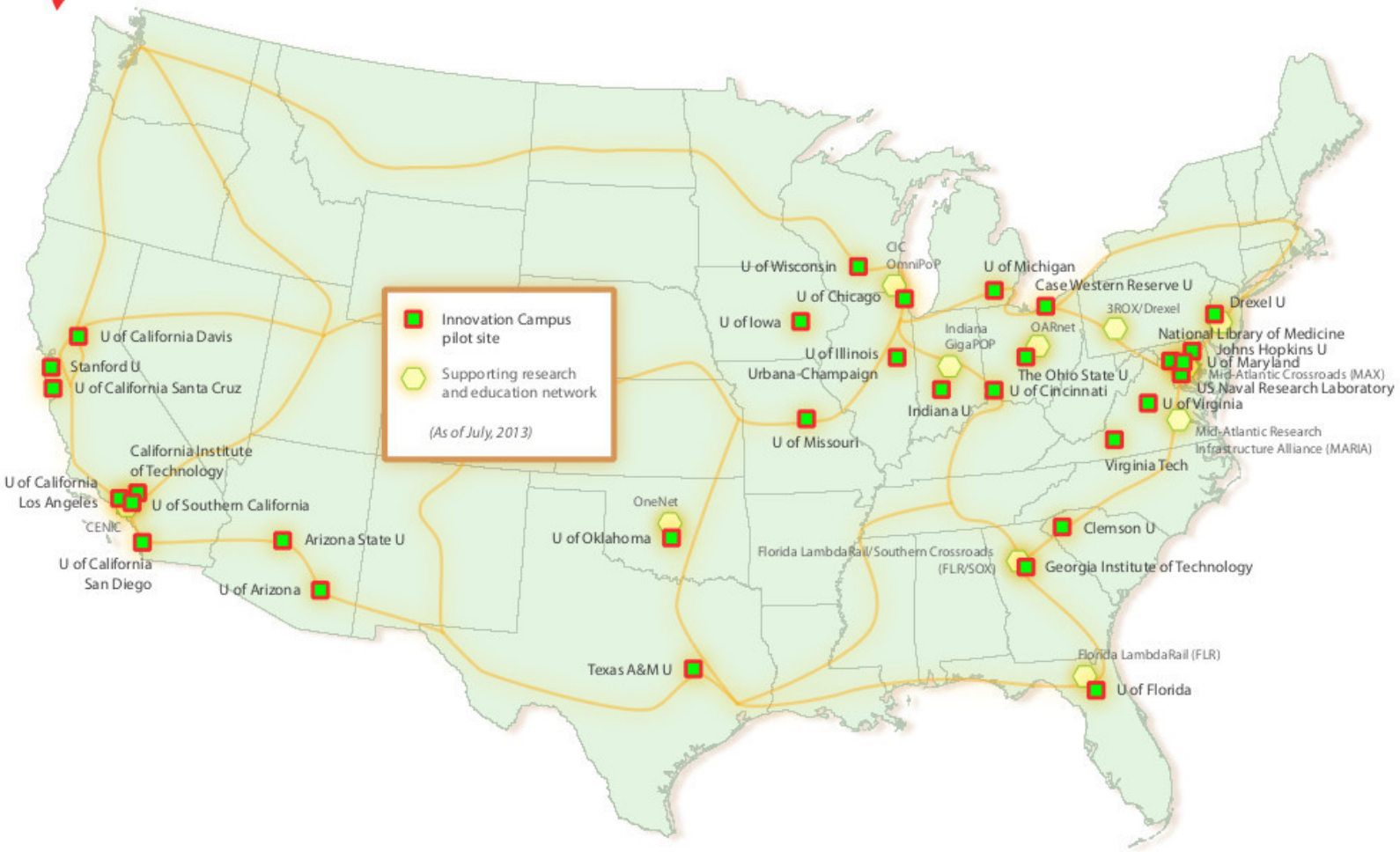
Advantages of Design

The MU/OSU Collaboration at 100G

- A single instance of the system can span multiple data centers, mirroring and distributing files across all cloud storage servers so that the loss of any one data center does not limit access to data.
- We can incrementally scale the capacity of a deployed system by adding new servers and storage, without downtime or loss of file availability. Data replication policies are fully customizable, enabling a high level of fault tolerance, regardless of system configuration.

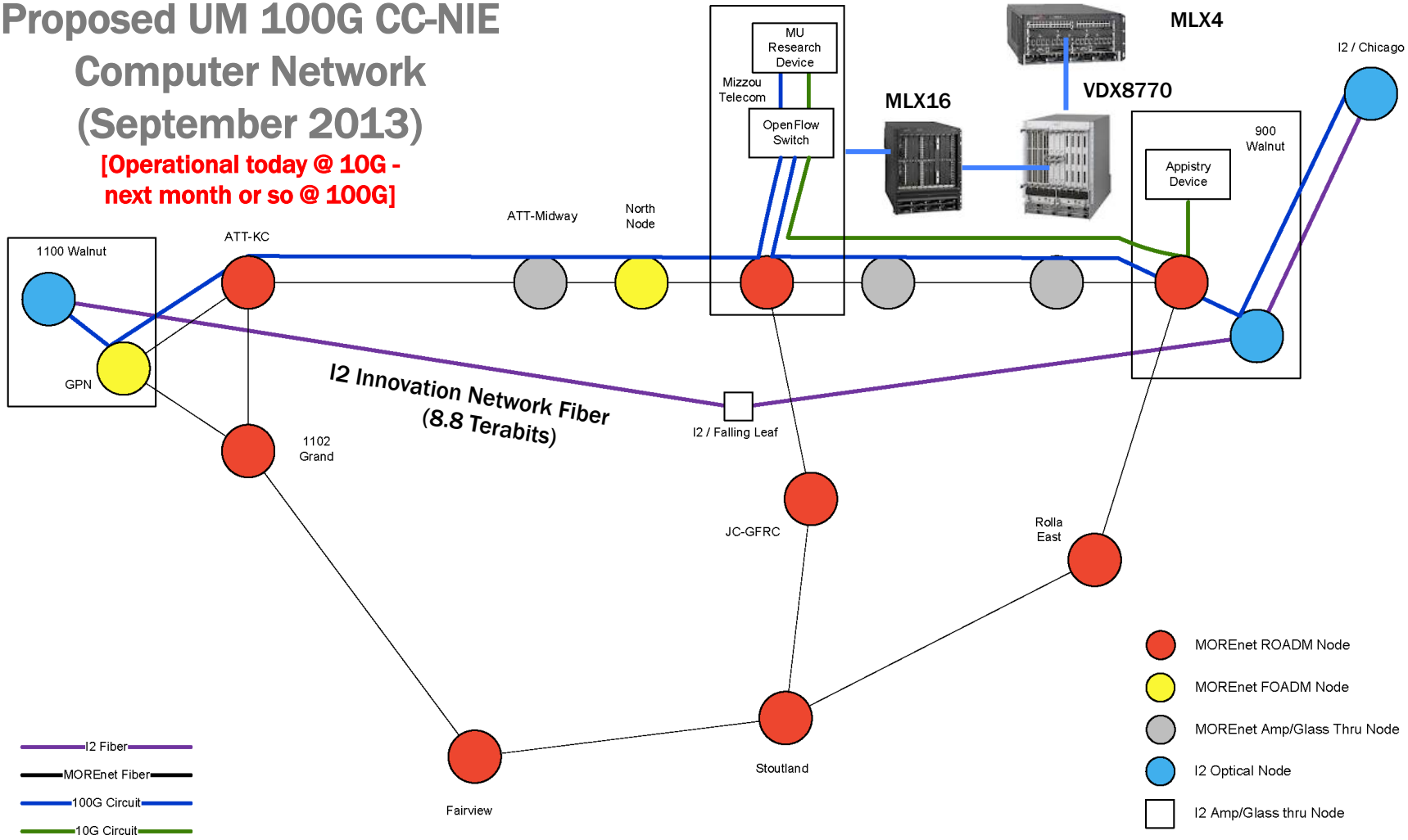
Innovation Campus Pilot Sites

100GE – Science DMZ – Software-Defined Networking



Proposed UM 100G CC-NIE Computer Network (September 2013)

[Operational today @ 10G - next month or so @ 100G]



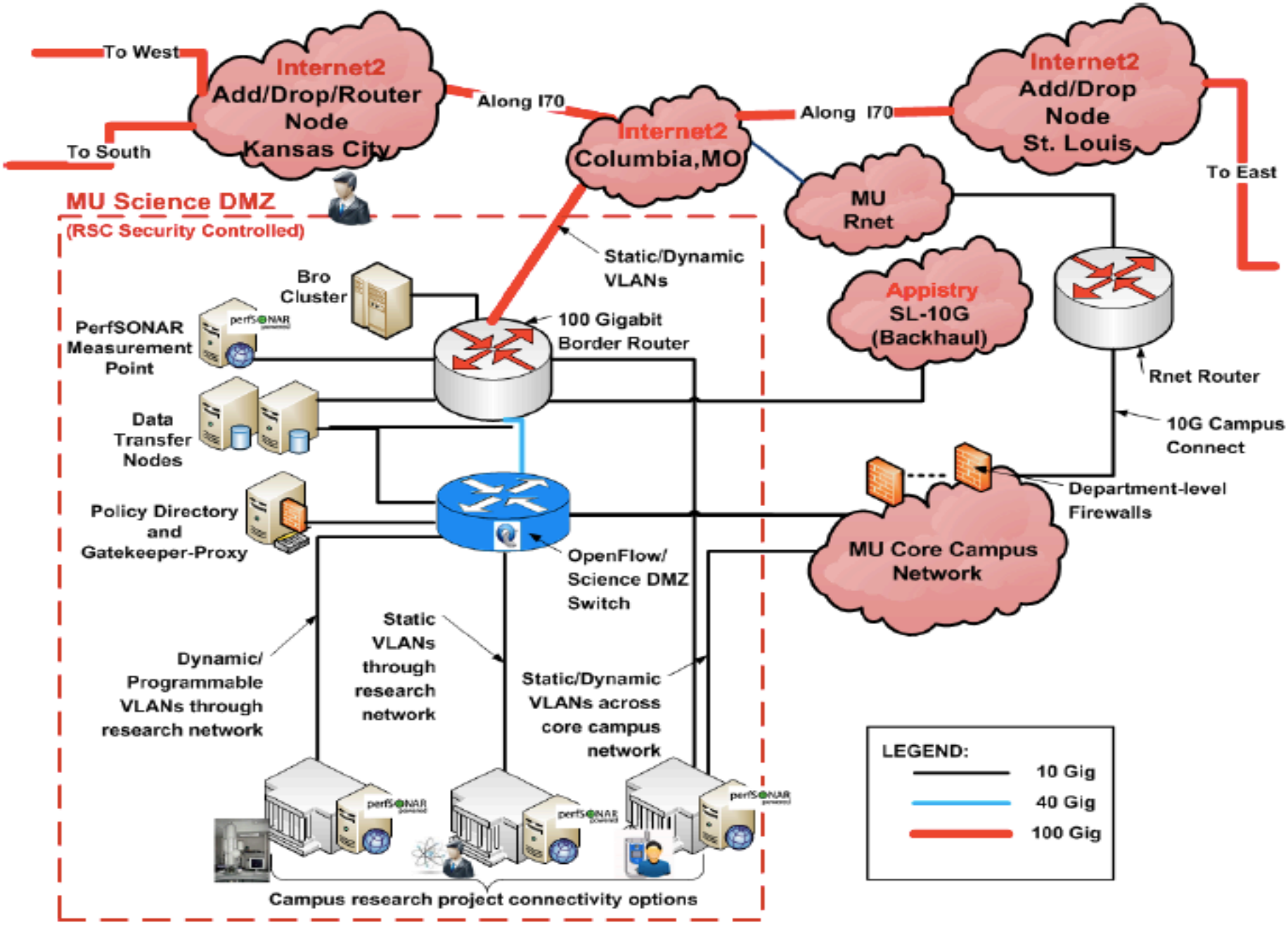


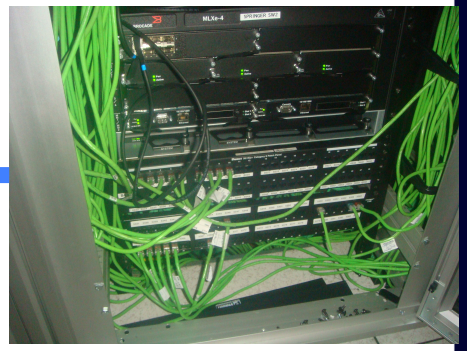
Figure 7 - MU Science DMZ and Network connections

Switches in the Infrastructure

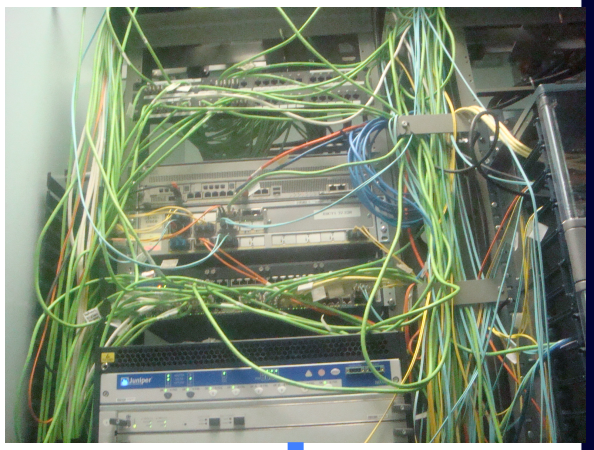
Brocade MLXe-16

Brocade VDX8770
96 10G Ports
SDN/OF

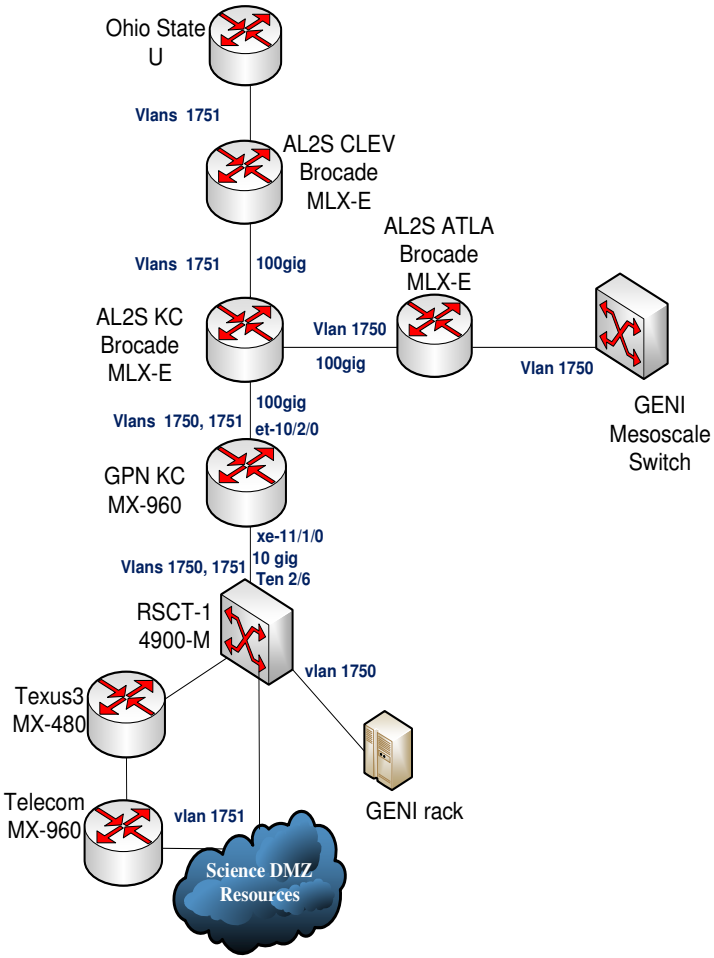
Brocade MLXe-4



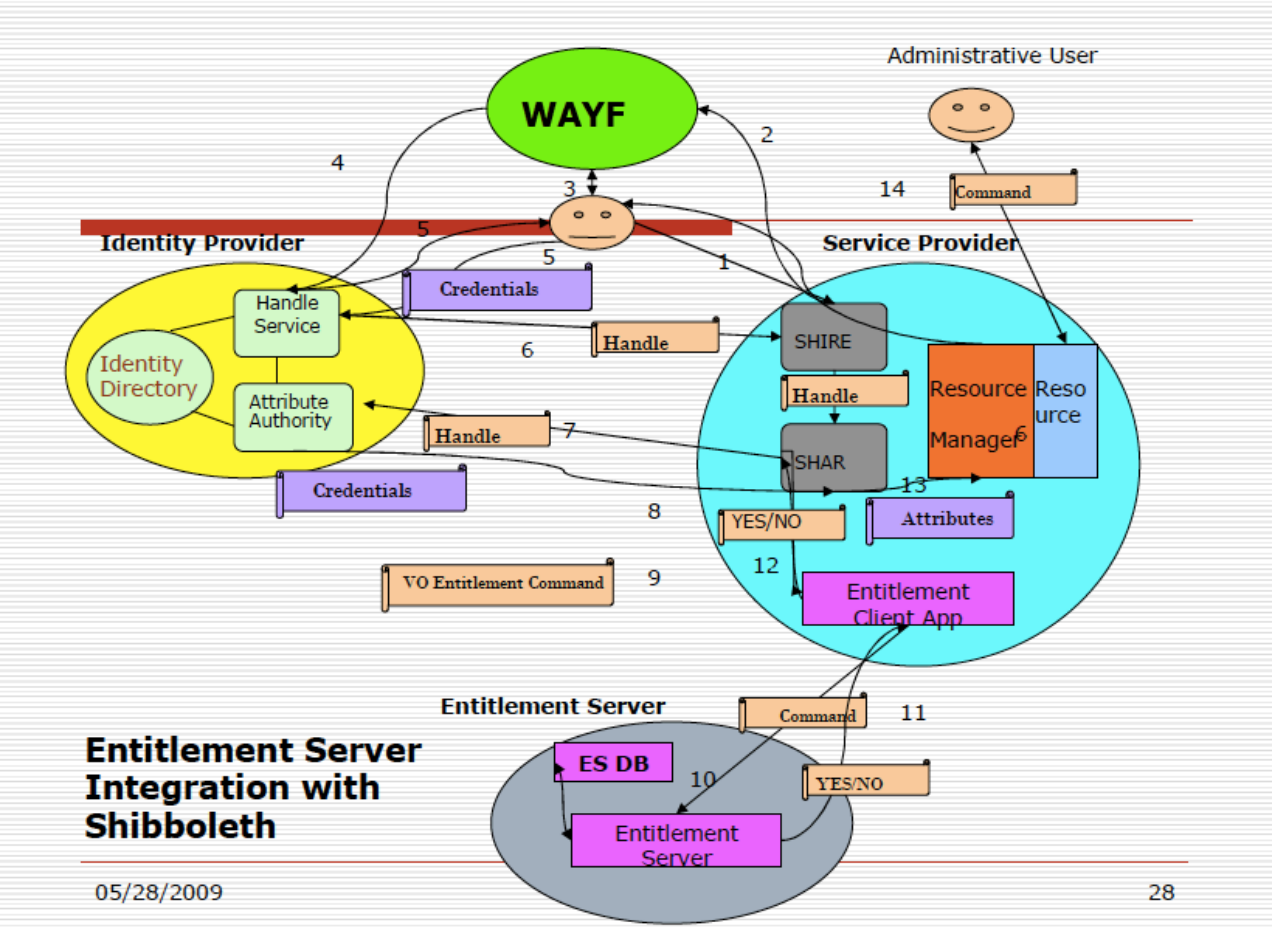
RSCT1
Rnet Router

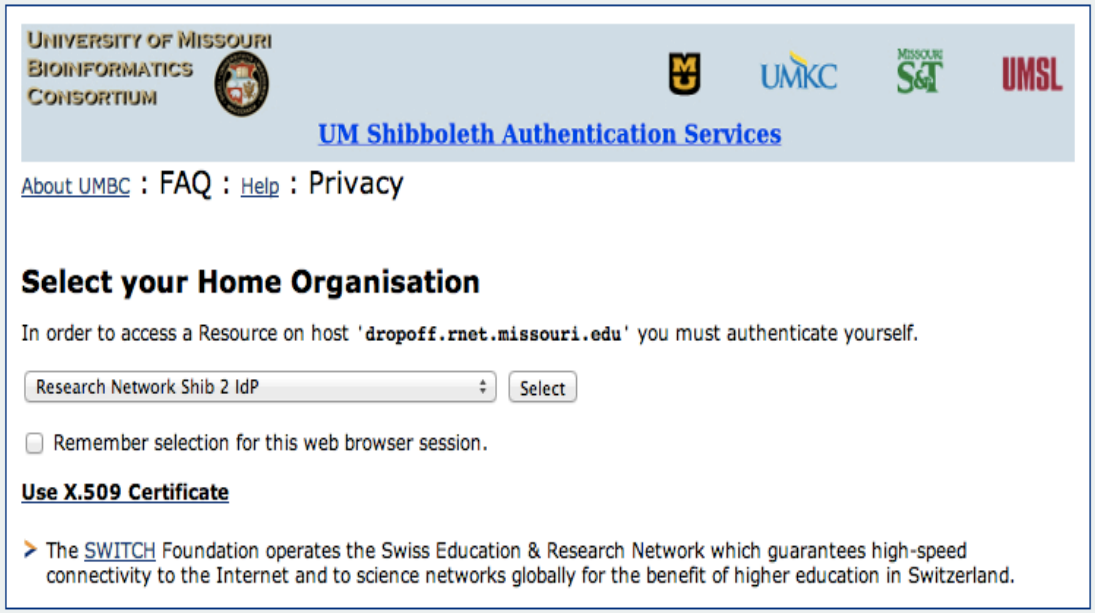


High Level Layer 2/3 Diagram of MU Science DMZ



Federated Identity can Overcome Multi-institutional Collaboration Barriers





SSO Authentication Using Federated Identity Services via InCommon

The Story Continues

University of Missouri

VIMAN Lab
VIRTUALIZATION, MULTIMEDIA AND NETWORKING LAB

Wide-area Overlay Networking to Manage Science DMZ Accelerated Flows

Prasad Calyam, Ph.D.

Assistant Professor, Department of Computer Science

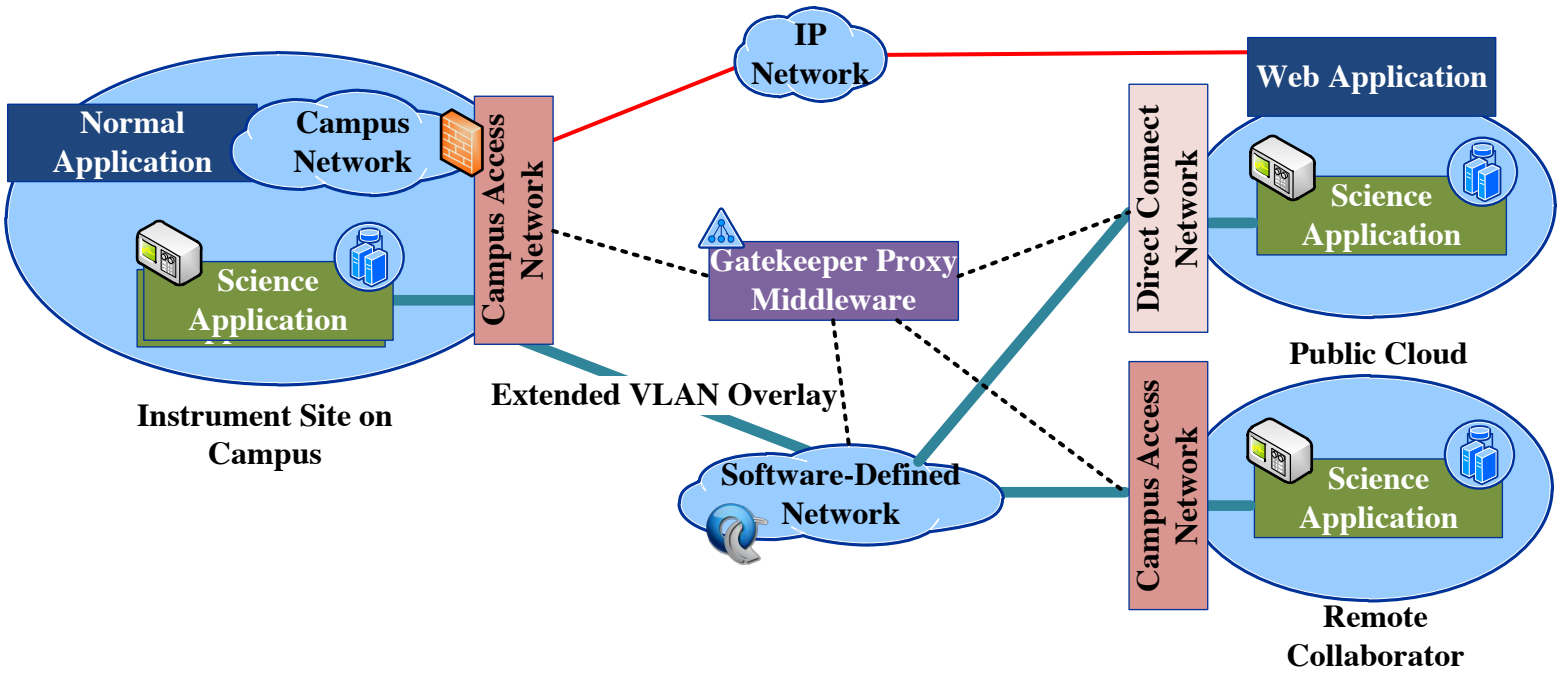
October 10th, 2013

Cyberinfrastructure Day 2013

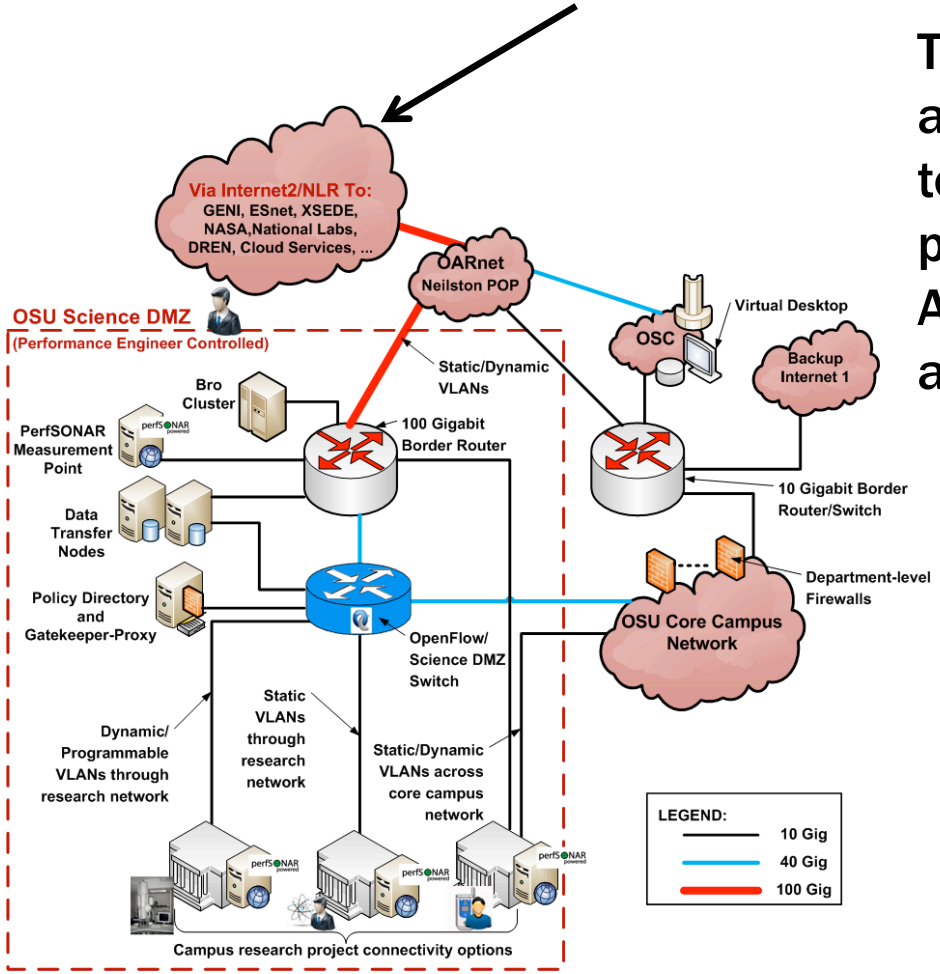
“Network-as-a-Service”

- Need for “Network-as-a-Service” for inter-institutional collaborations
- Network agnostic scientific researcher has to play the role of a “network admin”!
- Design a Science DMZ architecture that can enable:
  - Deep network programmability according to researcher’s requirements with a simple intuitive application dashboard
  - Real-time policy enforcement

Science DMZ Concept

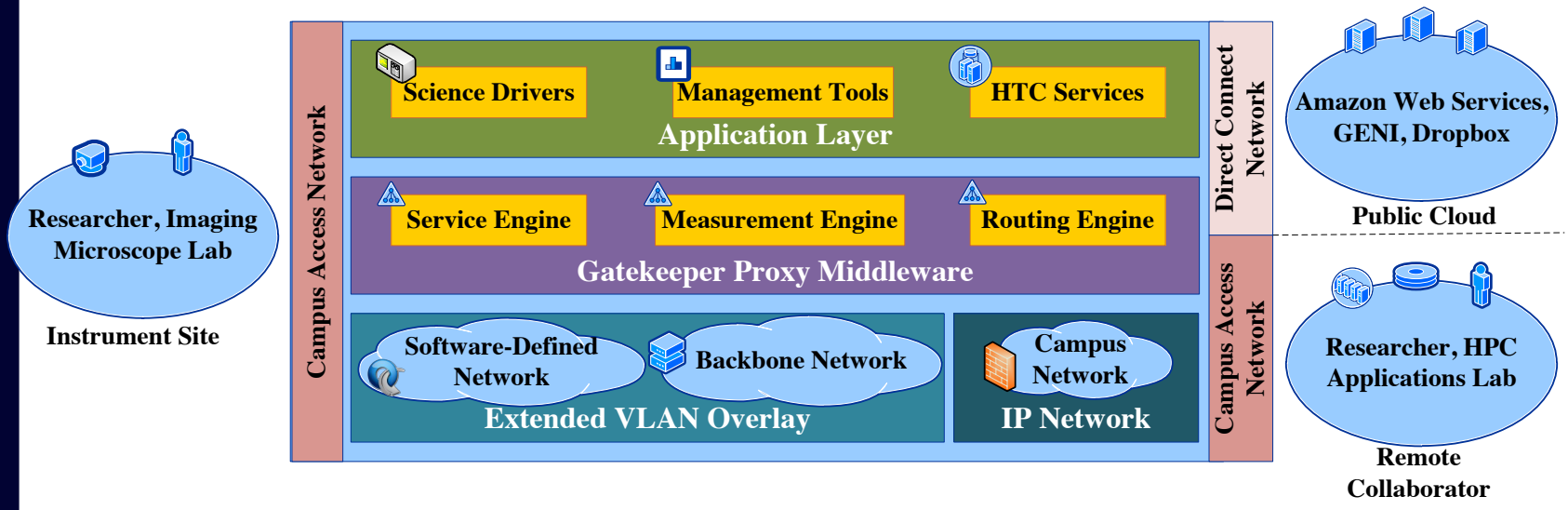


Science DMZ Design at OSU (MU has a mirror!)



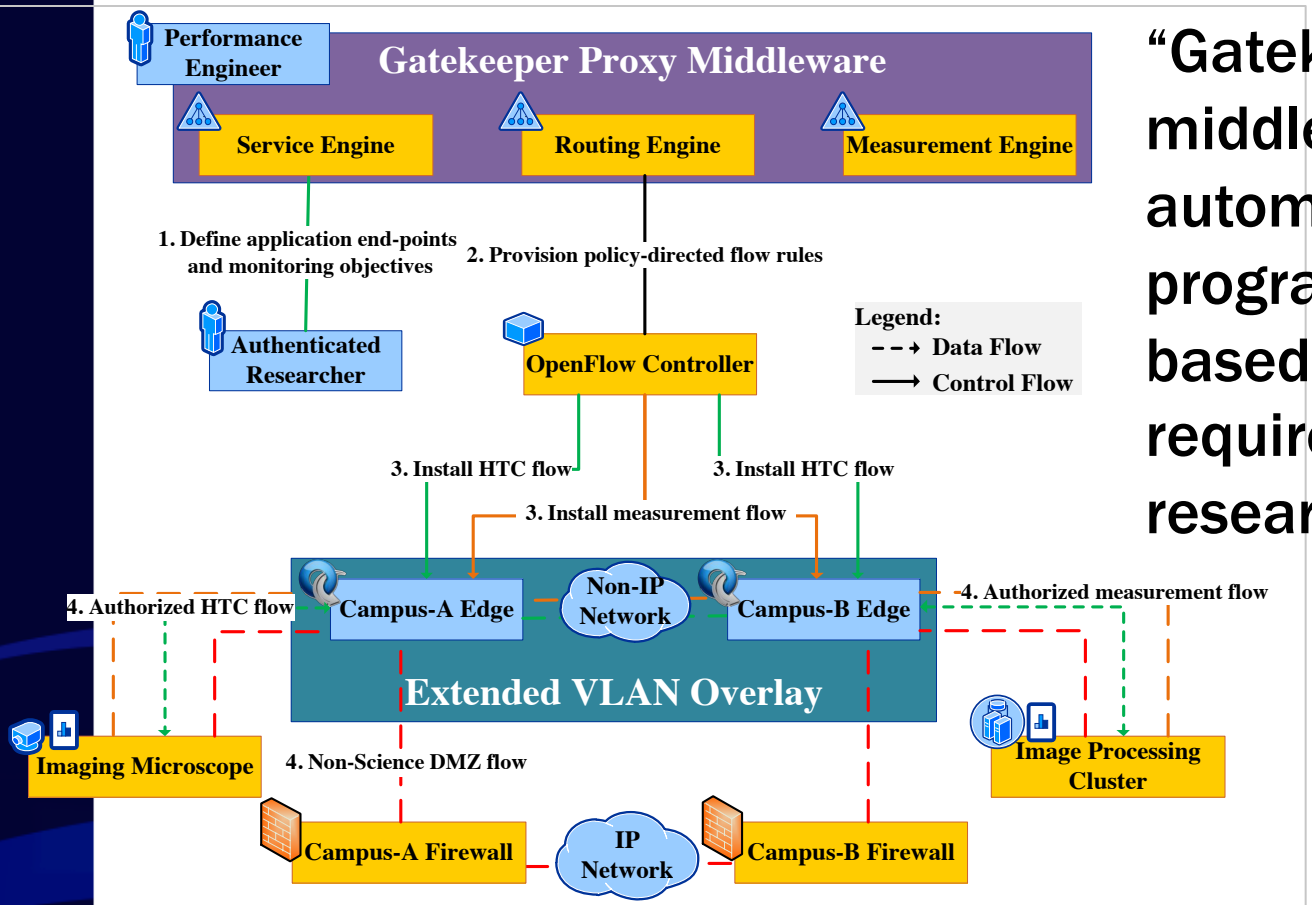
Two layered network architecture with connectivity to both Layer 3 IP network for public data access and Advanced Layer 2 Service for accelerated big data transfers

OSU-MU Dual-ended Science DMZ Case Study



End-to-end overlay network based performance acceleration over wide-area networks and across “dual-ended” Science DMZs

Gatekeeper Proxy Middleware maintained by a "Performance Engineer"



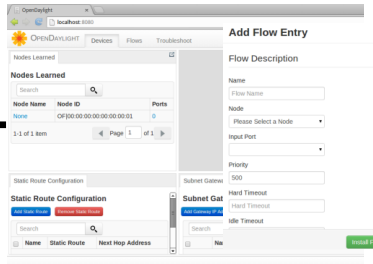
An intelligent "Gatekeeper proxy" middleware that automatically programs network based on application requirements and researchers inputs

Researcher Network-as-a-Service Dashboard

- Intuitive application dashboard as a simple user input form to define end-points and monitoring objectives
- Fine-grain control of application requirements as “Send Data to Host X with Bandwidth 4 Gbps, Priority 1 and No Delay”
- Real-time software defined network monitoring of critical network parameters for bottleneck identification and troubleshooting

Researcher

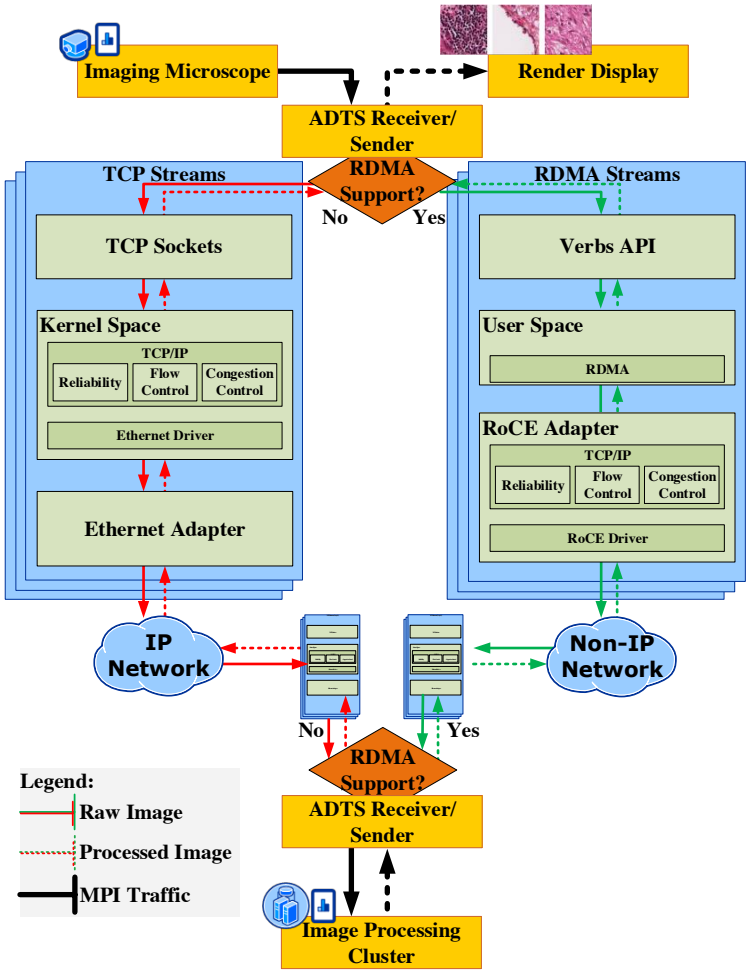
Federated Identity Login



Application Dashboard

Real-time Network Monitoring

Science DMZ Use Case between MU and OSU



Real-time processing of 'Neuroblastma' cell images, has to happen in the order of less than 10-20 seconds per magnification at OSU and sent back to MU Microscope facility!

Questions?

https://umbc.rnet.missouri.edu

https://dropoff.rnet.missouri.edu