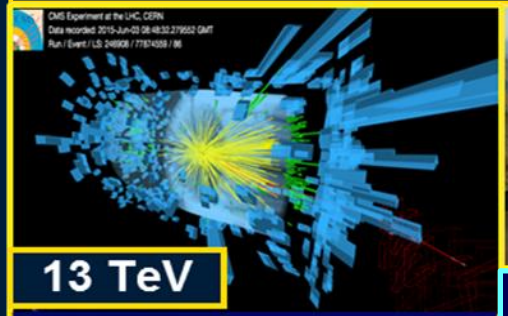
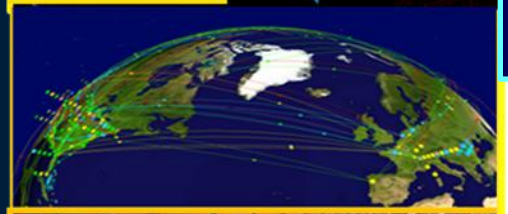
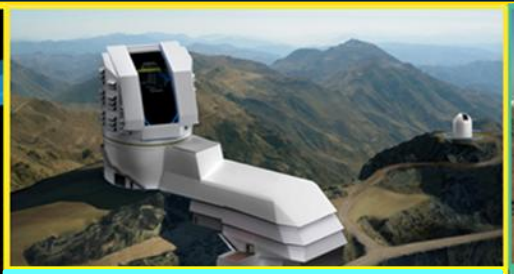
Global Network Advancement Group
Next Generation Network-Integrated System for Data Intensive Sciences

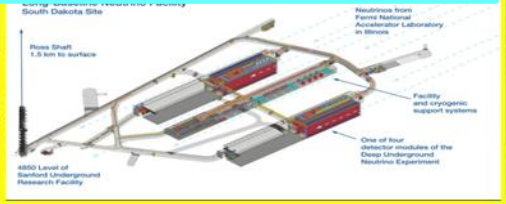


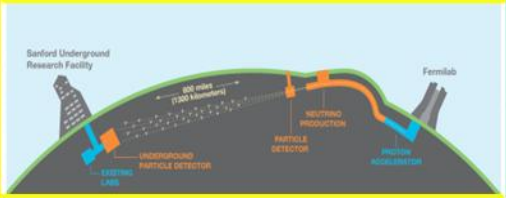


LHC



Rubin Observatory LSST





LBNF/DUNE

SKA

LHC Run 3 and HL-LHC

Rubin Observatory

SKA

BioInformatics

Earth Observation

Gateways to a New Era

Harvey Newman (Caltech)

6th Global Research Platform Workshop, Chicago

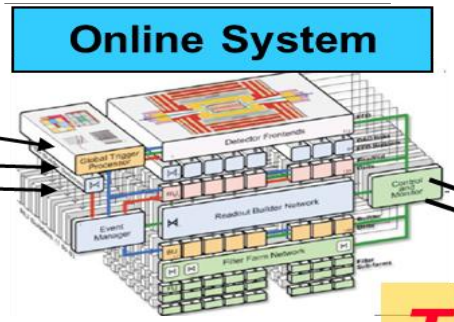
on behalf of the GNA-G and its DIS and AutoGOLE Working Groups

Background Slides Follow

Global Data Flow: LHC Grid Hierarchy
A Worldwide System Invented by Caltech (1999)

Experiment

~PBytes/sec



~1000 MBytes/sec

13 Tier1, *170 Tier2*
+ 300 Tier3 Centers:
1k to N X 10k Cores

Tier 0

CERN Center
100 PBs of Disk;
Tape Robot

Tier 1

N X100 Gbps

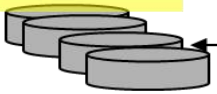
London | Paris | Bologna | Chicago | ...

10 to 100 Gbps

Tier 2

CACR | Center | Center | Center

Tier 3



Physics data cache

Institute | Institute | Institute | Institute

10 to 100 Gbps

Tier 4

Workstations

Tier0: Real Time Data Processing
Tier1 and Tier2: Data and Simulation Production Processing
Tier2 and Tier3: Data Analysis

A Global Dynamic System
Increased "Elastic" Use of Additional HPC and Cloud Resources
Fertile Ground for Control with Ai/ML

LHC: Discovery of the Higgs Boson and Beyond; 75 Years of Exploration !

Physicists Find Elusive Particle Seen as Key to Universe

2013 Nobel Prize

Englert	Higgs

	Energy Frontier	Intensity Frontier	Cosmic Frontier
Higgs Boson	●		
Neutrino Mass		●	●
Dark Matter	●	●	●
Cosmic Acceleration			●
Explore the Unknown	●	●	●

48 Year Search; 75 Year Exploration

Theory (1964): 1950s – 1970s;

LHC + Experiments Concept: 1984

Construction: 2001; Operation: 2009

Run1: Higgs Boson Discovery 2012

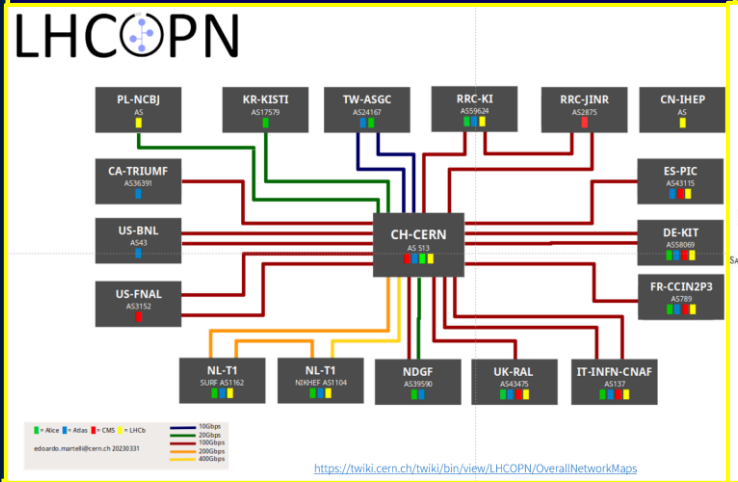
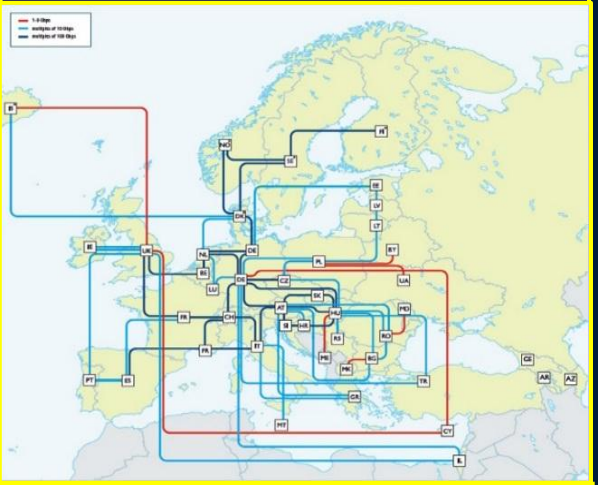
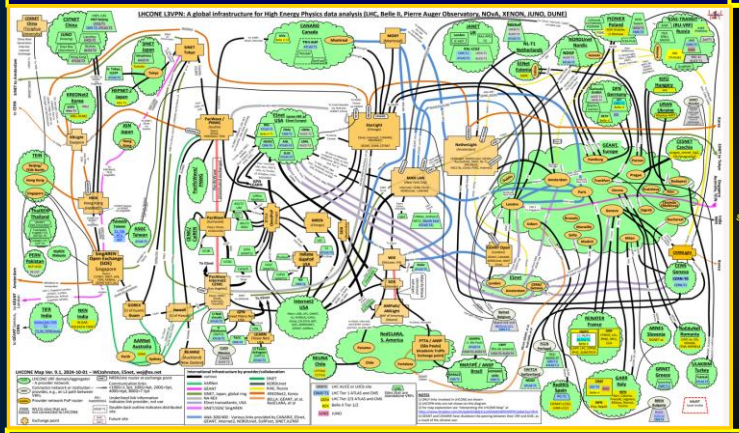
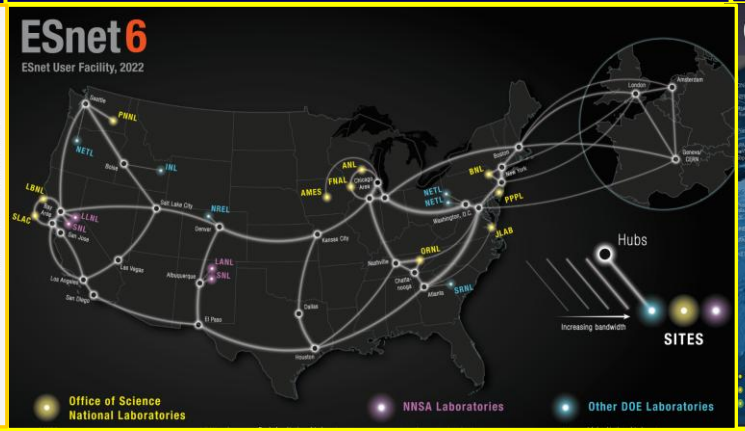
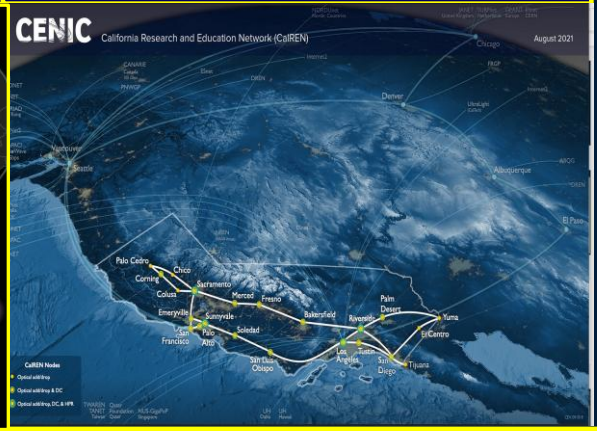
Run2 and Going Forward: Precision Measurements and BSM Exploration: 2013 - 2042

Advanced Networks Were Essential to Higgs Discovery and Every Ph.D Thesis; They will be Essential to All Future Discoveries

- **NOTE: ~85% of Data Still to be Taken**
- **Greater Intensity: Upgraded detectors for more complex events**
- ***To 5X Data Taking Rate in 2029-40***

The Core of LHC Networking

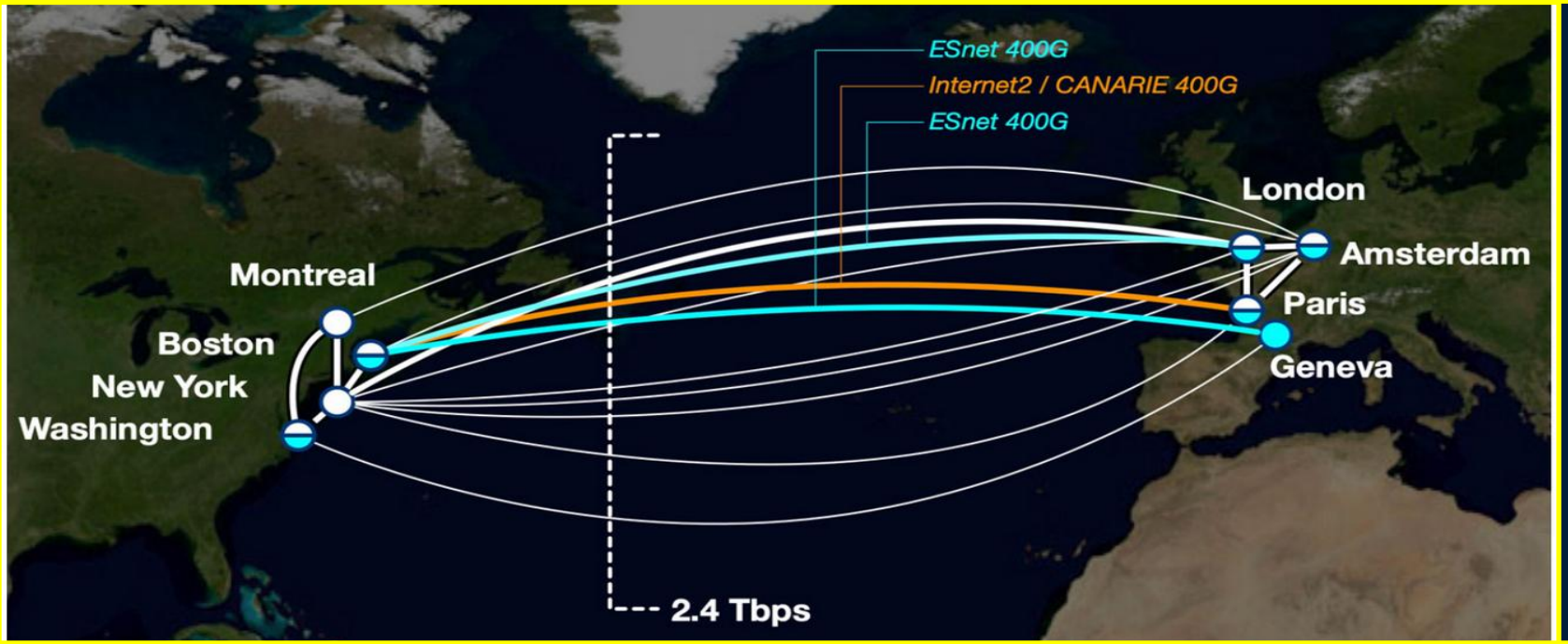
LHCOPN, LHCONE, GEANT, ESnet, Internet2, CENIC…

LHCOPN: Simple & Reliable Tier0+1 Ops	Internet2n	GEANT
		
LHCONE VRF: 170 Tier2s ++	**ESnet6**	***CENIC and NRP***
		

+ NRENs in Europe, Asia, Latin America (RNP, Rednesp), Au/NZ; *US State Networks*

The Upgraded ANA (to 2.4 Terabits/sec): ESnet, Internet2, GEANT, CANARIE Transoceanic Links and Intercontinental Partnerships

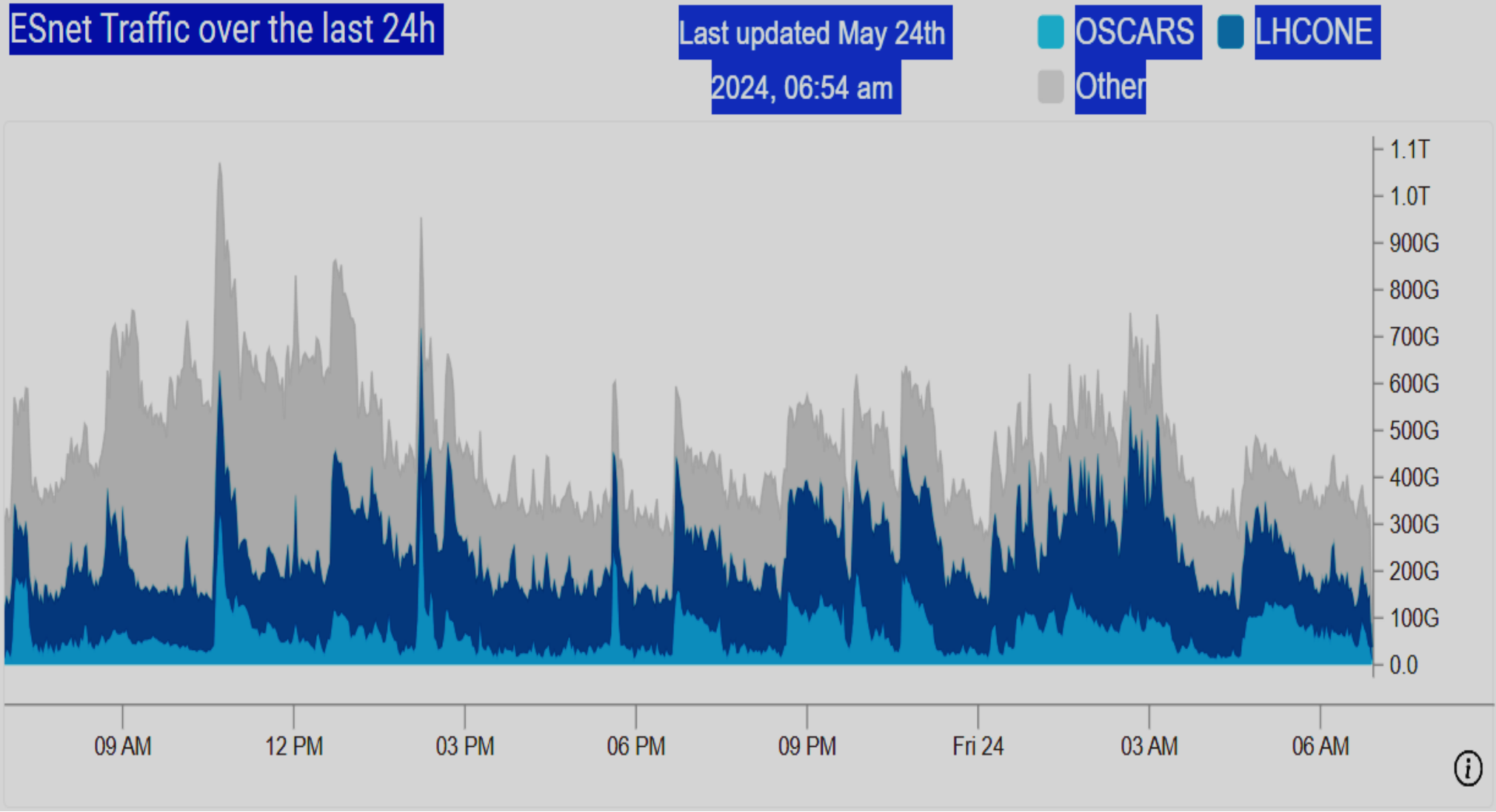
ANA's network expansion supports multinational, data-intensive science collaborations, including the Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator, and the Square Kilometer Array (SKA), the ongoing effort to build the world's largest radio astronomy observatory. It adds much-needed capacity for transmitting instrument findings to researchers globally, enabling ground-breaking discoveries.



The joint effort adds three new 400 Gbps spectrum circuits between exchange points in the U.S., U.K., and France. The connections utilize the record-breaking 400 terabits per second (Tbps) transAtlantic Amitié subsea cable system spanning 6,783 kilometers. The flexibility and scalability of these spectrum circuits enable significant capacity growth, through future upgrades at exchange points.

https://internet2.edu/canarie-esnet-geant-and-internet2-unveil-highest-capacity-transoceanic-connectivity/

ESnet Today: Traffic 0.3 to 1.1 Terabit/sec
with LHCONE and OSCARs 1/3 to 2/3



Global Network Advancement Group (GNA-G)
Since September 2019

Leadership Team: leadershipteam@lists.gna-g.net

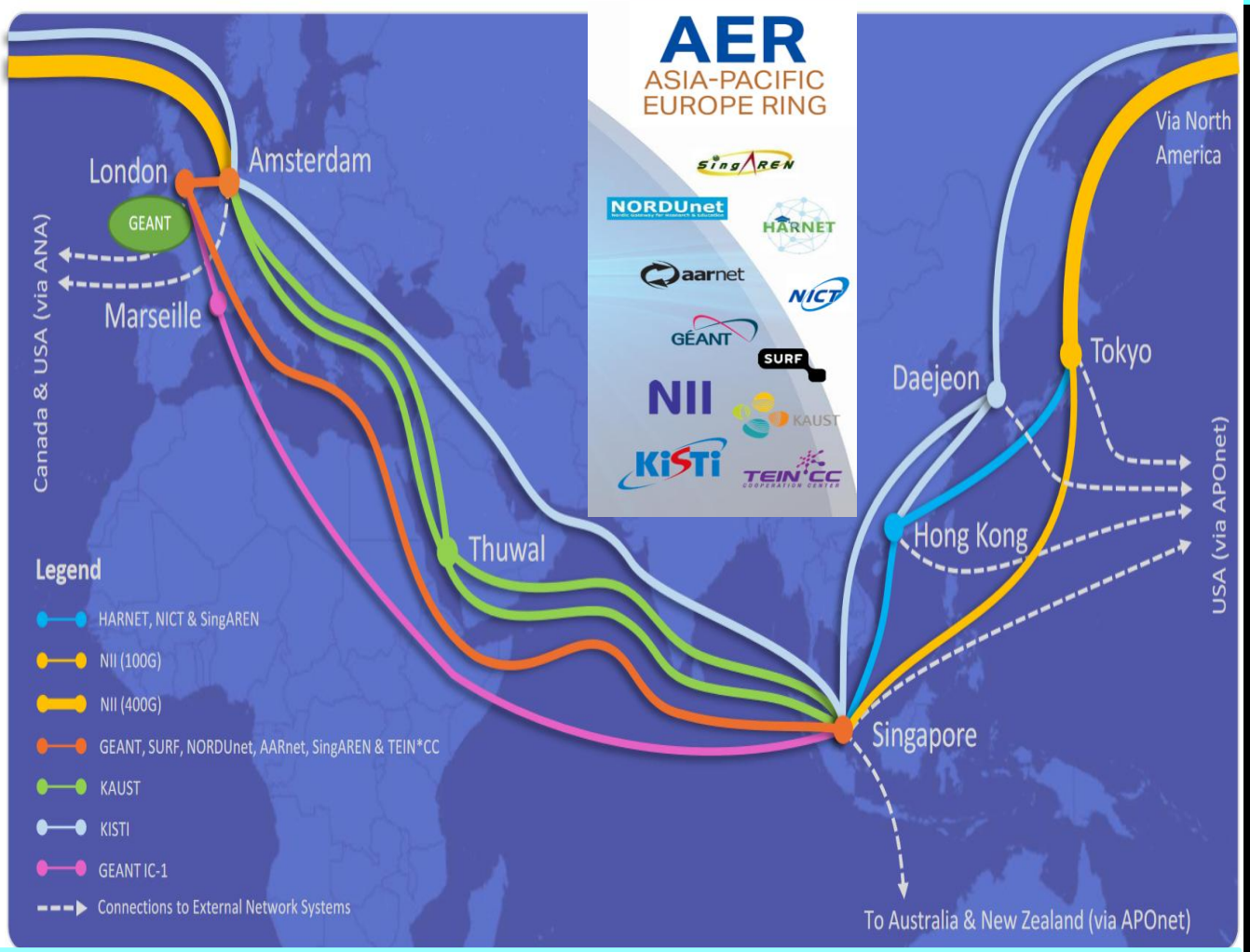
- **Osamu Akashi**, National Institute of Informatics (NII), Japan
- **Dale W. Carder**, ESnet, USA
- **Ivana Golub**, PSNC, Croatia/Poland
- **Renier van Heerden**, SANReN, South Africa
- **Brenna Meade**, Indiana University, USA
- **Alex Moura**, King Abdullah Univ. of Science and Technology (KAUST), Saudi Arabia
- **Marcos Schwarz**, RNP, Brazil
- **David Wilde**, AARNet, Australia (chair)
- **Chris Wilkinson,** Internet2, USA

- **An open volunteer group devoted to developing the blueprint to make using the Global R&E networks both simpler and more effective**
- **Its primary mission is to support global research and education using the technology, infrastructures and investments of its participants.**
- **The GNA-G is a data intensive research & science engager that facilitates and accelerates global-scale projects by (1) enabling high-performance data transfer, and (2) acting as a partner in the development of next generation intelligent network systems that support the workflow of data intensive programs; compatible with at-large A&R requirements**

See https://www.dropbox.com/s/qsh2vn00f6n247a/GNA-G%20Meeting%20slides%20-%20TechEX19%20v0.8.pptx?dl=0

KAUST and the Asia Pacific Oceania Network (APONet): Closing the Global Ring East and West

- Since the AER MoU, KAUST is coordinating with REN partners on shared deployment of spare capacity
- KAUST is supporting the following partners by offering point-to-point circuits for submarine cable backup paths:
  - AARnet
  - GÉANT
  - NetherLight
  - NII/SINET
  - SingAREN
- The SC25 NRE Demonstrations are also supported by KAUST closing the ring from Amsterdam to Singapore and back to the US

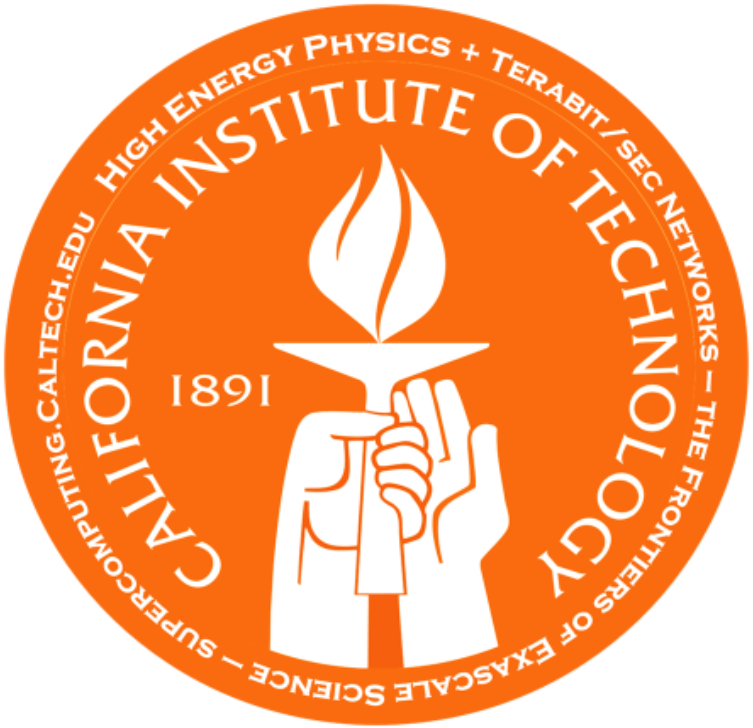


Worldwide Partnership: On the Road to Next Generation Systems and Science Discoveries

GNA-G
Global Network Advancement Group

OFC
The future of optical networking and communications

Global Petascale to Exascale Workflows for Data Intensive Sciences



Accelerated by Next Generation Programmable SDN Architectures and Machine Learning Applications

The GNA-G Data Intensive Sciences WG

Charter: https://www.dropbox.com/s/4my5mjl8xd8a3y9/GNA-G_DataIntensiveSciencesWGCharter.docx?dl=0

- Principal aims of the GNA-G DIS WG:

(1) To meet the needs and address the challenges faced by major data intensive science programs

- In a manner consistent and compatible with support for the needs of individuals and smaller groups in the at large A&R communities

(2) To provide a forum for discussion, a framework and shared tools for short and longer term developments meeting the program and group needs

- To develop a persistent global testbed as a platform, to foster ongoing developments among the science and network communities

- While sharing and advancing the (new) concepts, tools & systems needed
- Members of the WG partner in joint deployments and/or developments of generally useful tools and systems that help operate and manage R&E networks with limited resources across national and regional boundaries
- A special focus of the group is to address the growing demand for
  - Network-integrated workflows
  - Comprehensive cross-institution data management
  - Automation, and
  - Federated infrastructures encompassing networking, compute, and storage
- Working Closely with the AutoGOLE/SENSE WG

AutoGOLE / SENSE Working Group

- **Worldwide collaboration of open exchange points and R&E networks interconnected to deliver network services end-to-end in a fully automated way.** NSI for network connections, SENSE for integration of End Systems and Domain Science Workflow facing APIs.
- **Key Objective:**
  - The AutoGOLE Infrastructure should be persistent and reliable, to allow most of the time to be spent on experiments and research.
- **Key Work areas:**
  - Control Plane Monitoring: Prometheus based, Deployments underway
  - Data Plane Verification and Troubleshooting Service: Study and design group formed
  - AutoGOLE related software: Ongoing enhancements to facilitate deployment and maintenance (Kubernetes, Docker based systems)
  - Experiment, Research, Multiple Activities, Use Case support: Including Fabric, NOTED, Gradient Graph, P4 Topologies, Globus, Named Data Networking (NDN), Data Transfer Systems integration & testing.
- **WG information**
  https://www.gna-g.net/join-working-group/autogole-sense

NRE-125: Global Network System for Data Intensive Sciences with Multi-Tbps Campuses and Regionals (III)

- Demonstrating a paragon network at many sites equipped with P4 programmable devices, including Tofino and Tofino2-based switches, Smart NICs and FPGA-based network interfaces providing packet-by- packet inspection, agile flow management and state tracking, and real-time decisions
- High throughput platform demonstrations in support of workflows for the science programs. Including reference designs of NVMe server systems to match a 400G to Tbps network core, as well as servers with multi-GPUs and programmable smart NICs with FPGAs
- Integration of edge-focused extreme telemetry data (from P4 switches, smart edge devices and end hosts) and end facility/application caching stats and other metrics, to facilitate automated decision-making processes
- Development of dynamic regional caches or “data lakes” that treat nearby as a unified data resource, building on the successful petabyte cache currently in operation between Caltech and UCSD and in ESNet based on the XRootD federated access protocol; extension of the cache concept to more remote sites such as Fermilab, Nebraska and Vanderbilt
- Demonstrate the interoperation in NRP of the SoCal and in-network ES caches and NRP nodes with 50+ clients situated throughout the Internet2 footprint, as a system development platform

NRE-102: Middle-East 800Gbps Transcontinental eScience Bridge

IMPACT

- Building on recent analysis of Red Sea submarine cable vulnerabilities that disrupted up to 90% of Europe-Asia communications on several occasions since February 2024, this demonstration validates KAUST's strategic approach to infrastructure diversification through high-capacity, geographically diverse routing
- This would establish an alternative path to the vulnerable Red Sea corridor where both the AAE-1 and TGN-EA submarine cables create a critical single point of failure
- The new trans-Pacific circuit would also become an important asset for partners and collaborators in Europe and in Asia, in particular the members of the Asia-Pacific Europe Ring (AER) cooperative, including CSTNET, GEANT, NORDUnet, SINET/NII, and SURF together with KAUST
- Critically, this work positions KAUST and Saudi Arabia as an emerging hub for multi-hundred-gigabit research networking while demonstrating practical solutions to the infrastructure vulnerabilities that threaten global research collaboration, and offering a reproducible blueprint for other National Research and Education Networks seeking to enhance both capacity and resilience

✸ This project aligns with the Saudi Vision 2030

NRE-102: Middle-East 800Gbps Transcontinental eScience Bridge

Technical GOALS in Steps

- The objective is to achieve disk-to-disk transfers approaching 800Gbps between KAUST's man campus and its Points of Presence at NetherLight (Amsterdam) and SingaREN's SOX (Singapore).

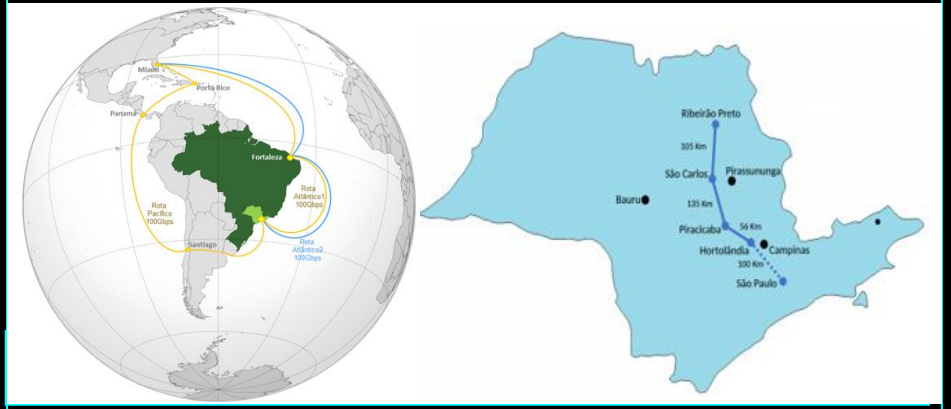
1. Sustain Proven Performance (Baseline): Achieve ≥95 % line-rate disk-to-disk transfers over KAUST's production 2 X 200 Gbps circuits, establishing a repeatable 400 Gbps benchmark and validating DTS stability in a multi-domain environment
2. Double Capacity (Primary Goal) – Demonstrate continuous ≥700 Gbps throughput on the temporary 800 Gbps optical circuit from KAUST to Caltech servers connected to SCInet, confirming that the architecture scales without architectural changes
3. Reusable Tuning Blueprint – Publish the DTN configurations set that enables high-efficiency transfers over ≥12,000 km, offering a plug-and-play playbook for GREN peers

- PATHFINDER: if KAUST manages to get a partner provider the temporary international circuit with capacity from 400 to 800 Gbps in a joint effort with Ciena, it will potentially be one of the first universities in the world, and the first in the Middle East region connected at > 400 Gbps capacity

NRE-129: University of Sao Paulo (USP) High Performance Dark Fiber Infrastructure

- The University of Sao Paulo, in Brazil, maintained by the State of Sao Paulo has consistently been among the 100 best universities in the world and the best in Latin America according to several international ranking organizations
- In 2016, USP decided to interconnect these 4 campuses via a dark fiber infrastructure. This was made possible by a state law that guarantees its right to use state highways for deploying their own ducts and fibers
- All inter-campus connections use fiber optic cables, each with 36 fiber pairs, in ducts buried more than one meter deep in the soil
- Currently, the inter-campus bandwidth is 100 Gbps, with projections to reach 600 Gbps by the end of 2025 using a limited number of fiber pairs
- This demonstrates USP's capacity to readily scale its bandwidth higher as required. This robust infrastructure facilitates a wide range of experiments in high-performance networking and fiber optic data communication technology

The State of Sao Paulo and its international connections (left). USP dark fiber backbone (right)



USP has 10 campuses. The primary USP campus located in São Paulo is currently connected to Rednesp at 100 Gbps, with an anticipated upgrade to 400 Gbps in the near future

The goal of this NRE is to showcase the intercampus bandwidth within USP and the potential for research and development utilizing its dark fiber optic infrastructure

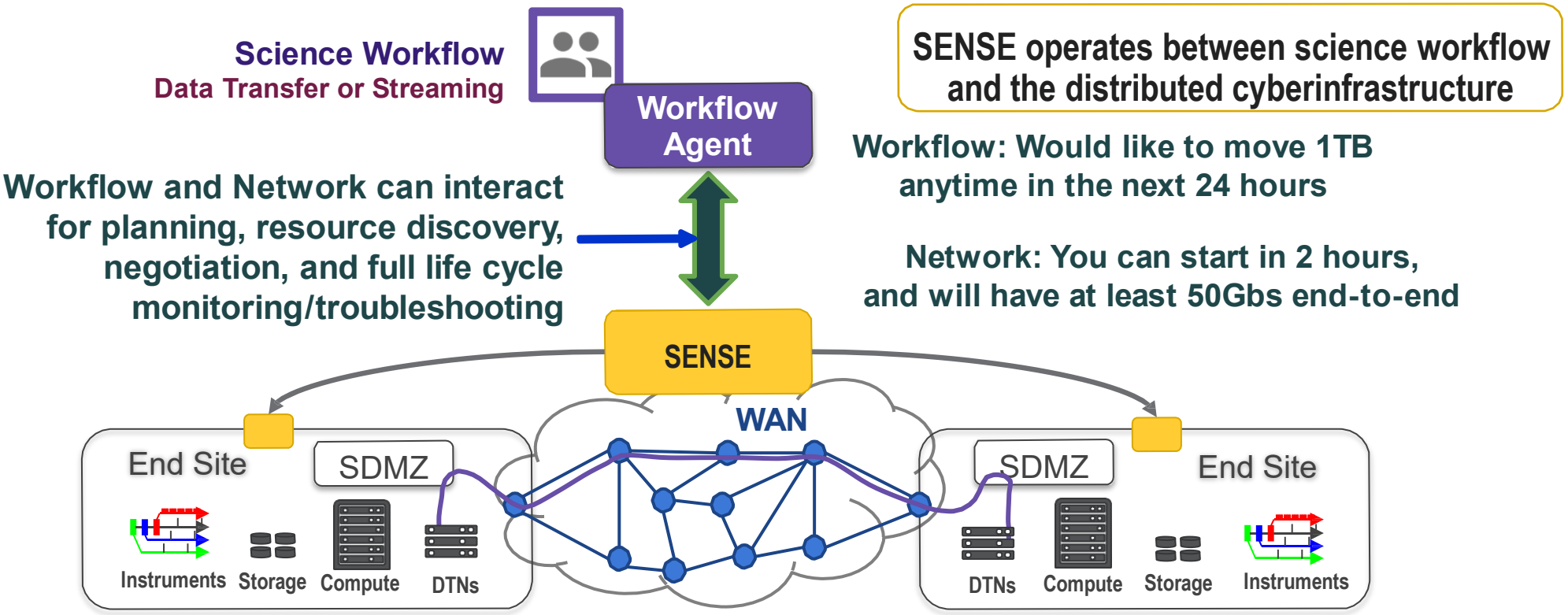
SENSE: SDN Enabled Networks for Science at the Exascale

ESnet, Caltech, Fermilab, LBNL https://arxiv.org/abs/2004.05953

Creates Virtual Circuit Overlays. Site and Network RMs, Orchestrator

Elevate Network to First Class Resource

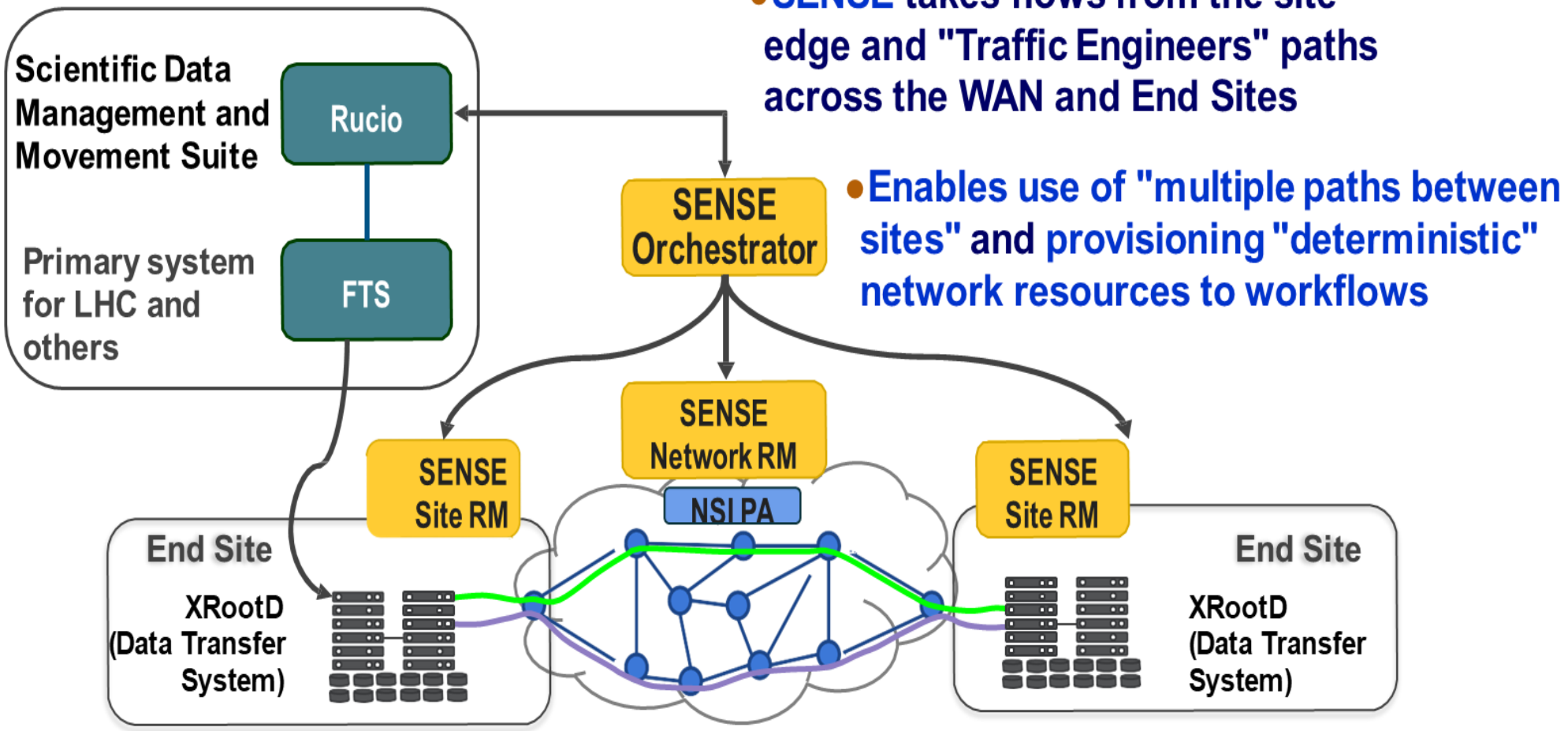
API driven Automation and Orchestration



- Allows workflows to identify data flows which are higher priority
- Allows the network to traffic engineer to fully utilize all network paths

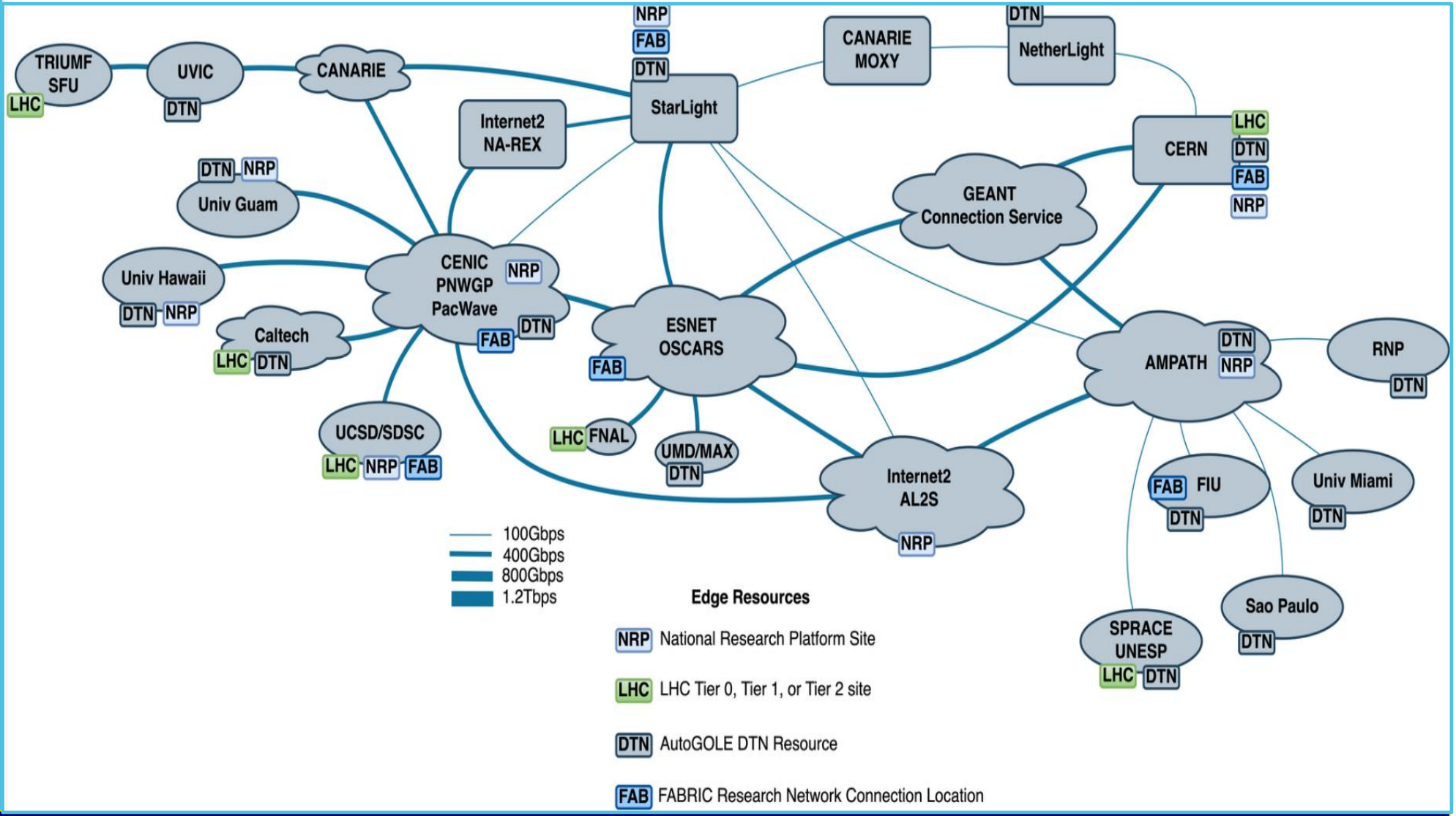
SENSE and Rucio/FTS/XRootD Interoperation

- **Rucio** identifies groups of data flows (IPv6 subnets) which are "high priority"



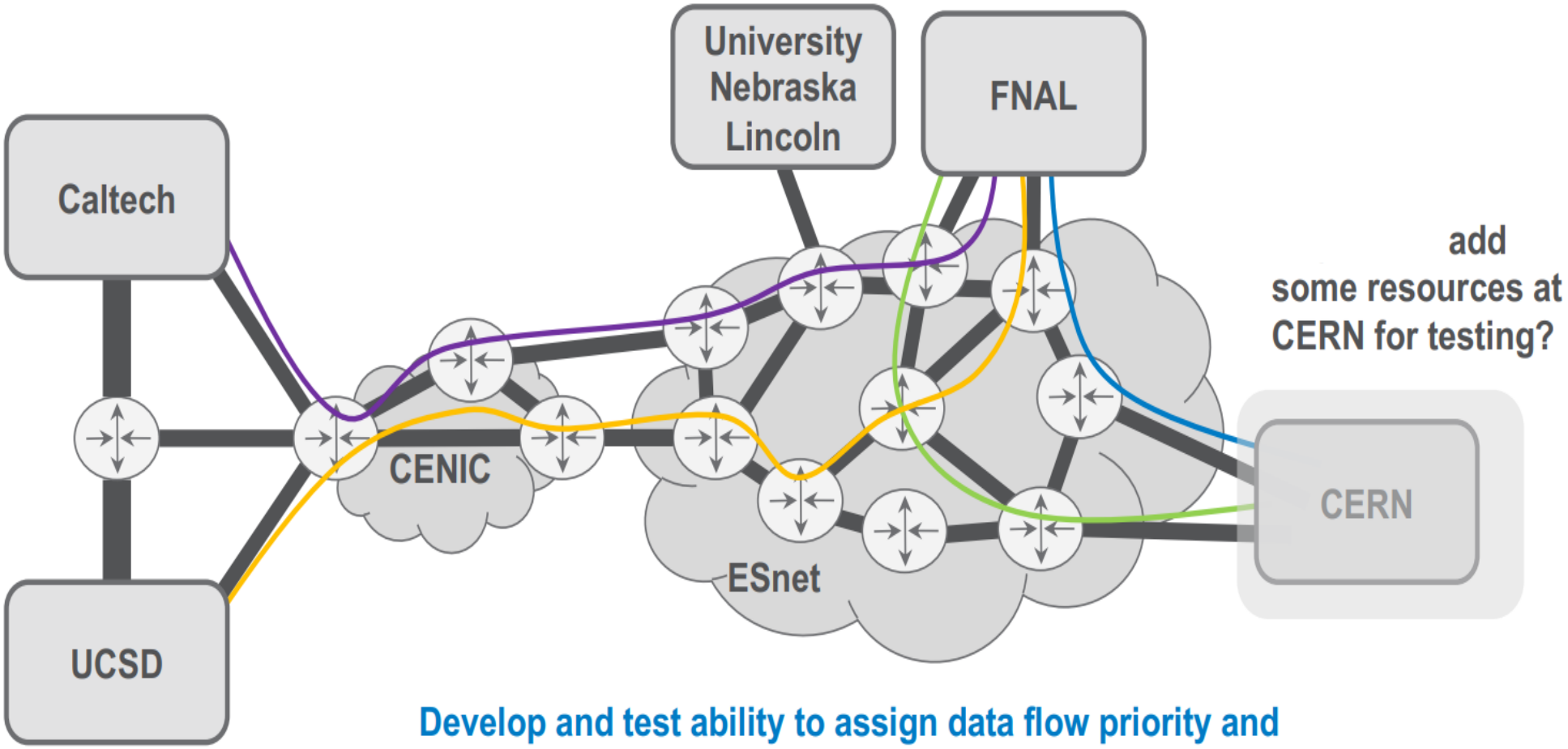
GNA-G AutoGOLE/SENSE WG Global Persistent Testbed

Tom Lehman, ESnet



Software Driven Network OSes; Programmability at the Edges and in the Core

SENSE Rucio/FTS/XRootD Interoperation System Deployment

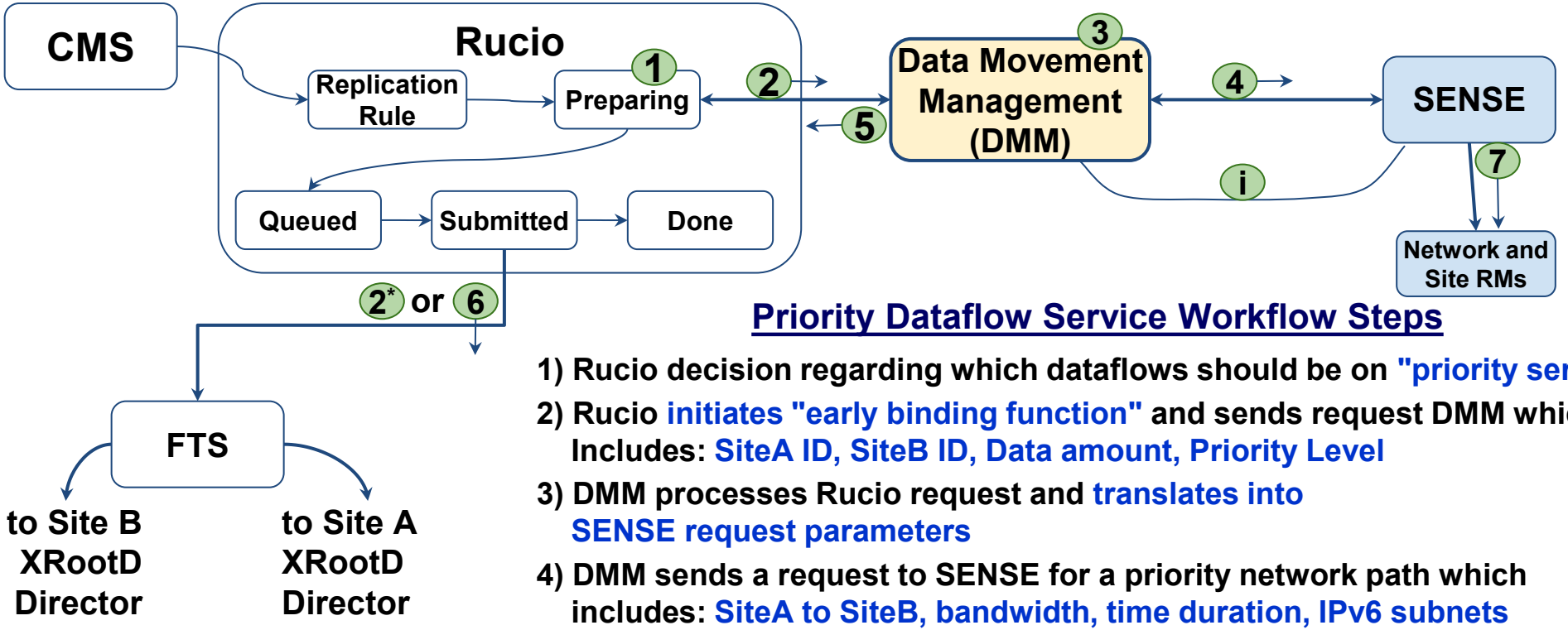


Develop and test ability to assign data flow priority and traffic engineer different end-to-end paths

May add other sites: CERN, Vanderbilt, SPRACE

T. Lehman, ESnet

Rucio, DMM, SENSE Workflow



*Rucio to FTS and DMM interactions can be asynchronous

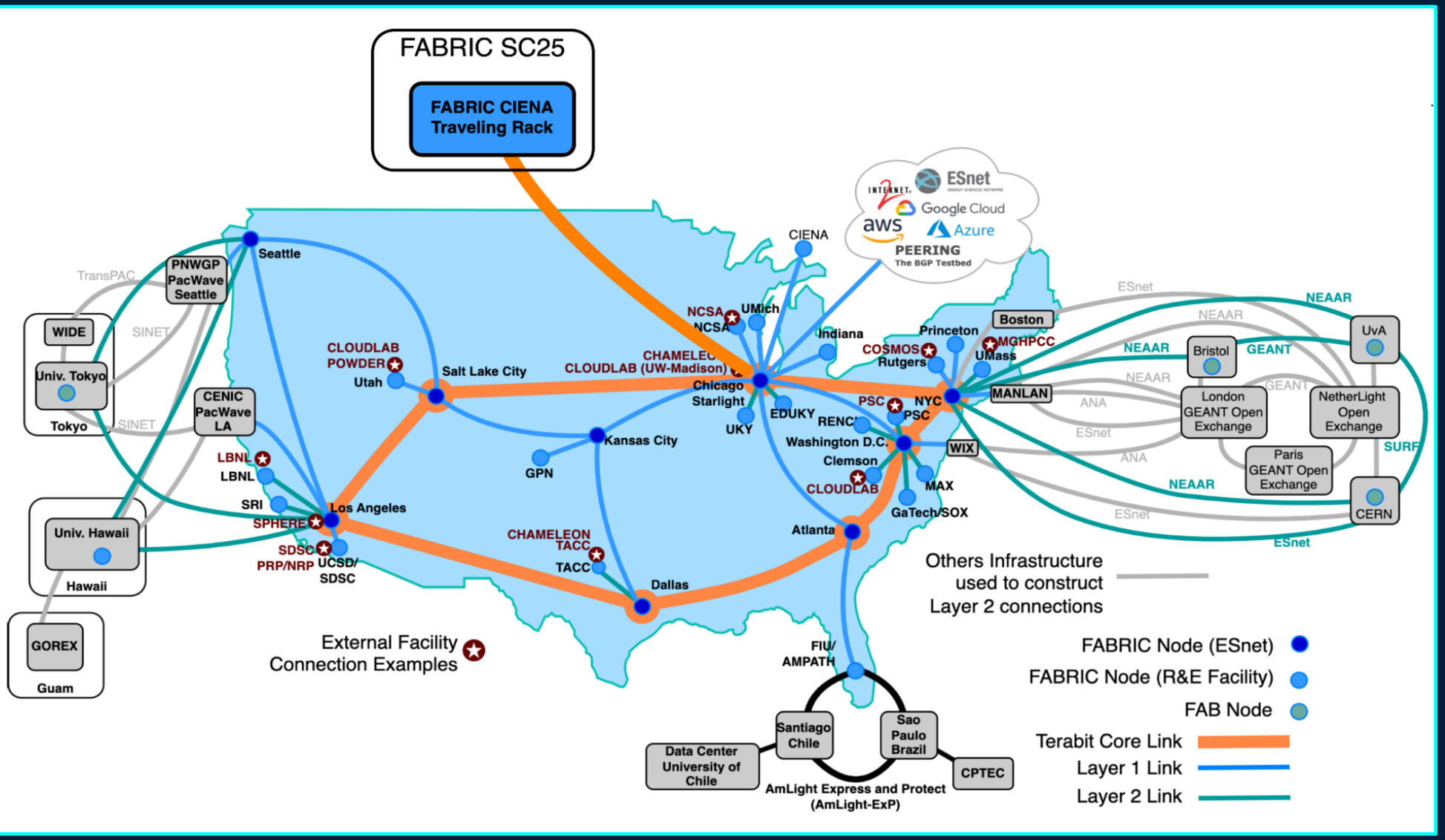
Priority Dataflow Service Workflow Steps

1) Rucio decision regarding which dataflows should be on "priority service"
2) Rucio initiates "early binding function" and sends request DMM which Includes: SiteA ID, SiteB ID, Data amount, Priority Level
3) DMM processes Rucio request and translates into SENSE request parameters
4) DMM sends a request to SENSE for a priority network path which includes: SiteA to SiteB, bandwidth, time duration, IPv6 subnets to be used for transfer.
5) DMM sends a response to Rucio which includes: IPv6 addresses to use for this priority data transfer
6) Rucio sends information to FTS to initiate data transfer, using proper IPv6 addresses
7) SENSE sends request to Network and Site Resource Managers to instantiate priority network service

i) DMM to SENSE "discovery services" (one time at DMM startup):
This is the mechanism for DMM to discover information about sites including: Sites available for service, IPv6 subnets available, site network connection speed. BW sharing according to priority.

NRE-117: FABRIC: An Adaptive ProgrammaBle Research Infrastructure for Computer Science and Science Applications

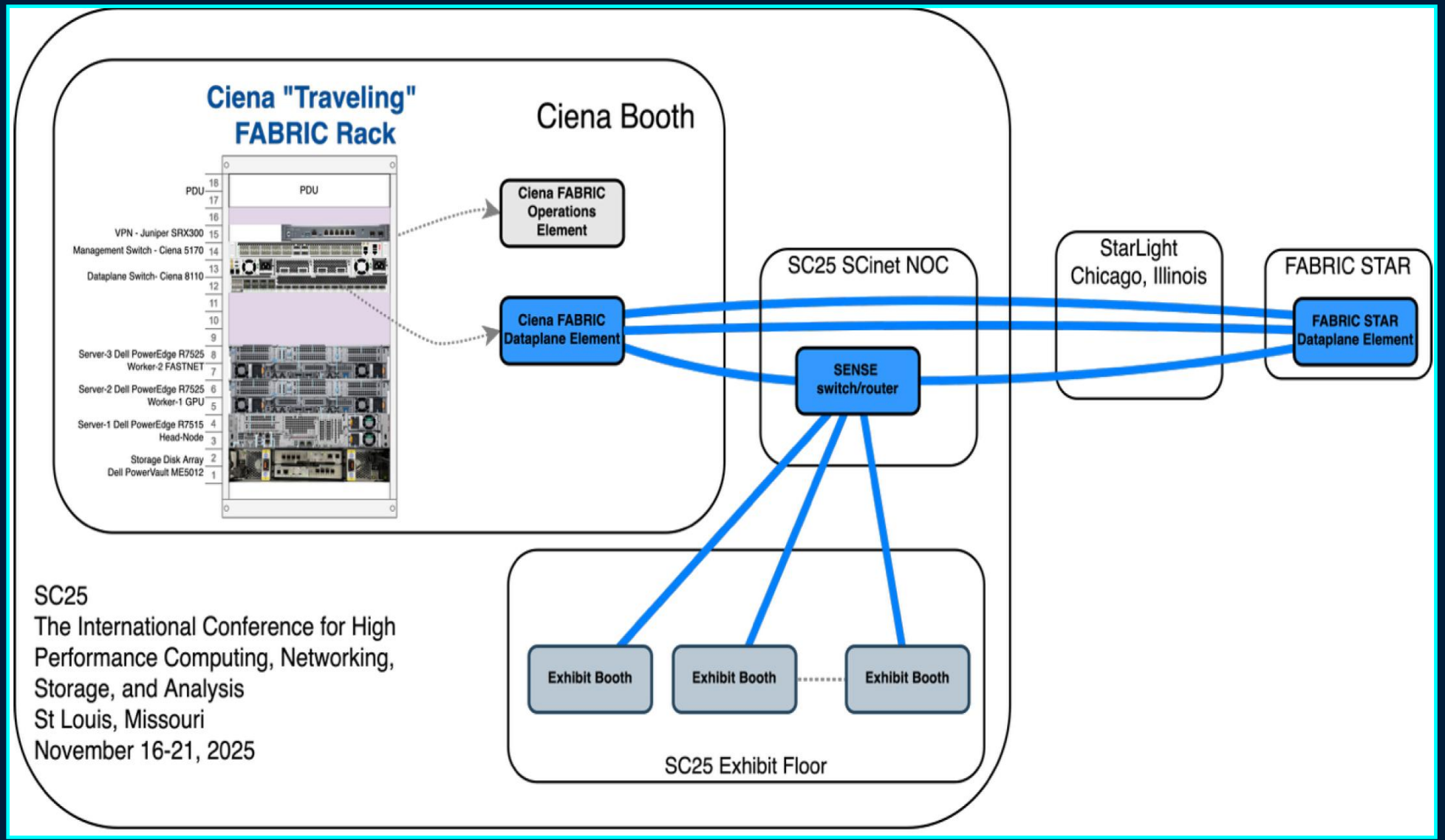
https://portal.fabric-testbed.net



FABRIC Topology with SC25 Connection

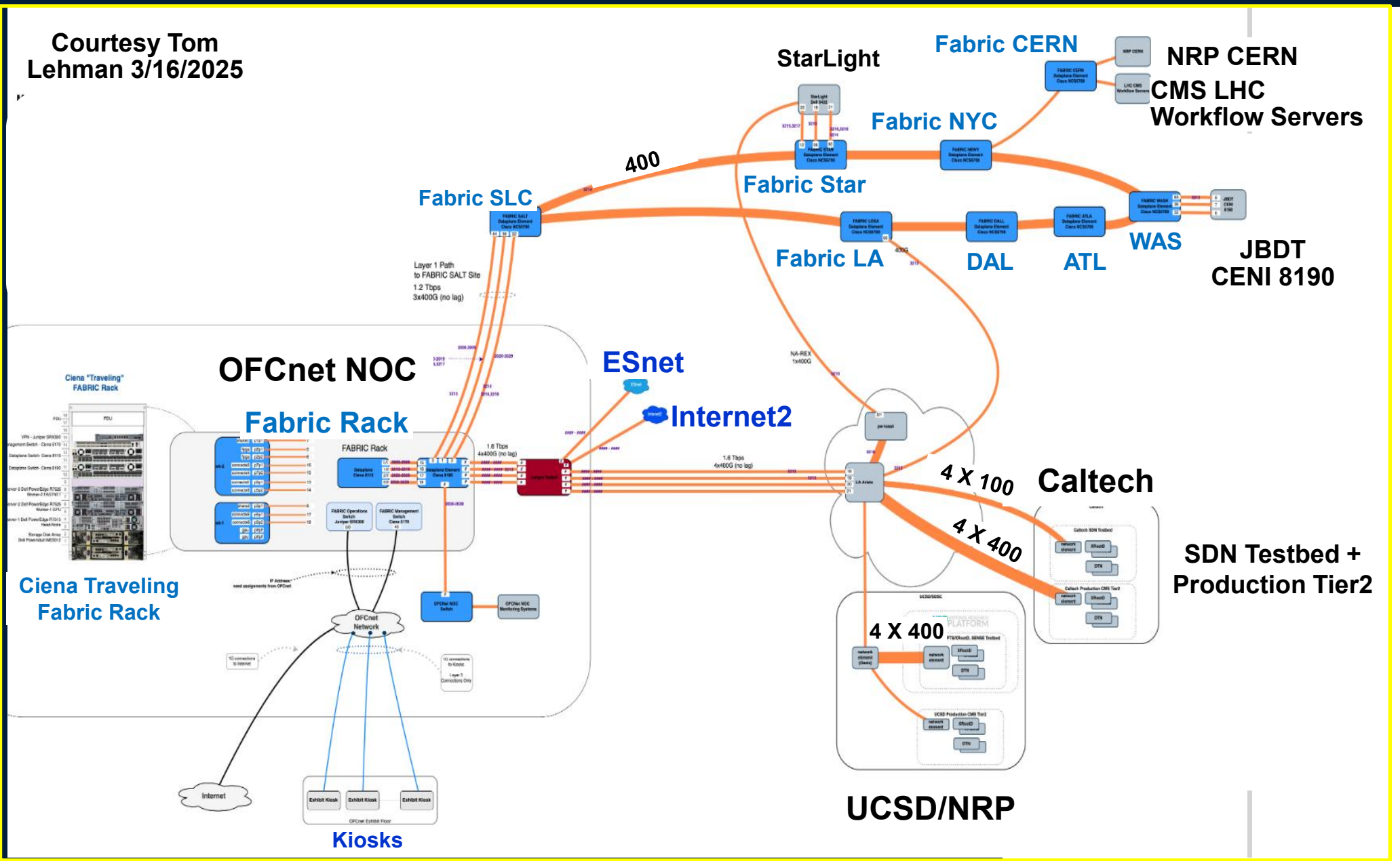
NRE-117: FABRIC: An Adaptive ProgrammaBle Research Infrastructure for Computer Science and Science Applications

https://portal.fabric-testbed.net



The Ciena “Traveling” Rack at SC25

OFC: Advanced Intercontinental Orchestration & Workflow
Testbed: GNA-G/Fabric/NRP/Caltech/NRP/Starlight/Ciena/Arista/Juniper



Advanced Network Services for Domain Science Workflow Innovation

CURRENT STATE

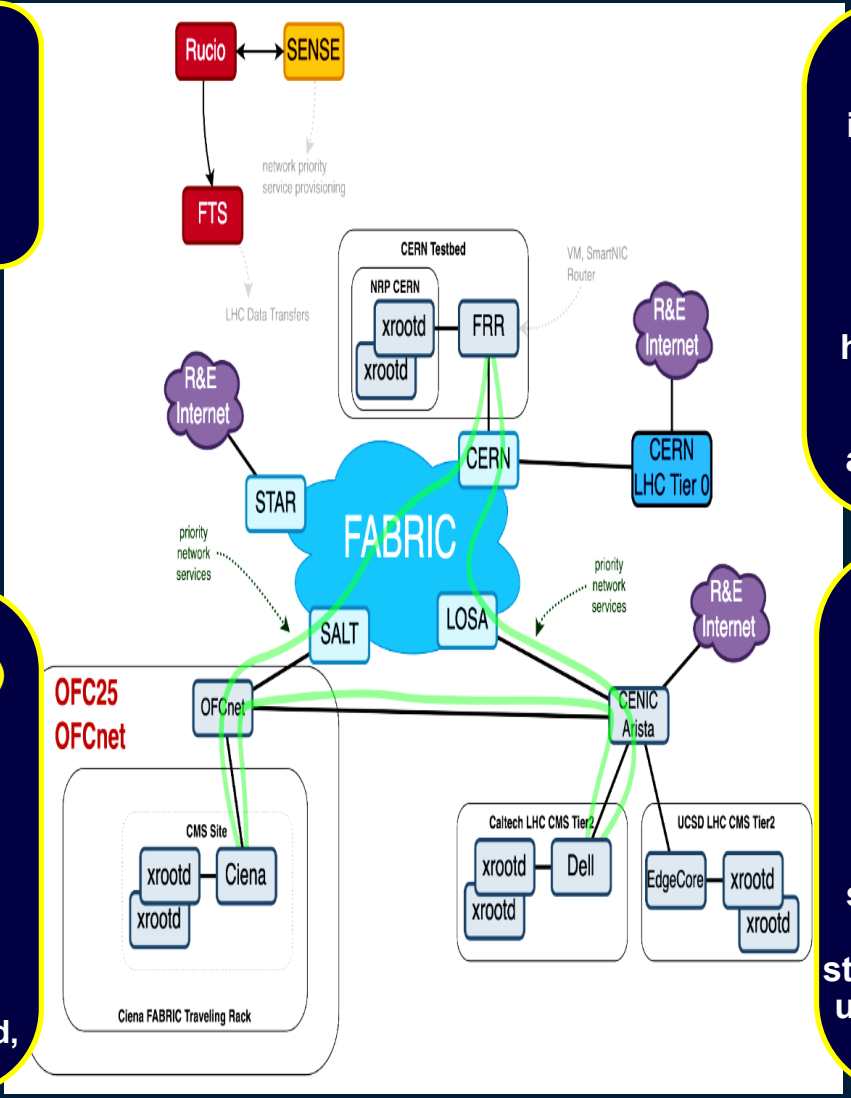
LHC data transfers can be highly variable due to lack of cyberinfrastructure visibility, and services to optimize network paths, server and storage systems.

CHALLENGES

Software Defined Networking (SDN) introduces programmability, but also complexity.

Need to manage policies, software provisioning systems, and capabilities **across multiple domains and world regions.**

Orchestration intelligence is needed, but not commercially available.



APPROACH

Developed solution allows **Rucio** to identify high priority data transfers. **SENSE** ensures minimum network + data mover BW & resources.

FABRIC is the experimental infra-structure enabling cost effective high performance edge site storage and data mover systems, based on server side **FRR, VPP, DPDK** and associated tools and technologies.

FOCUS and IMPACT

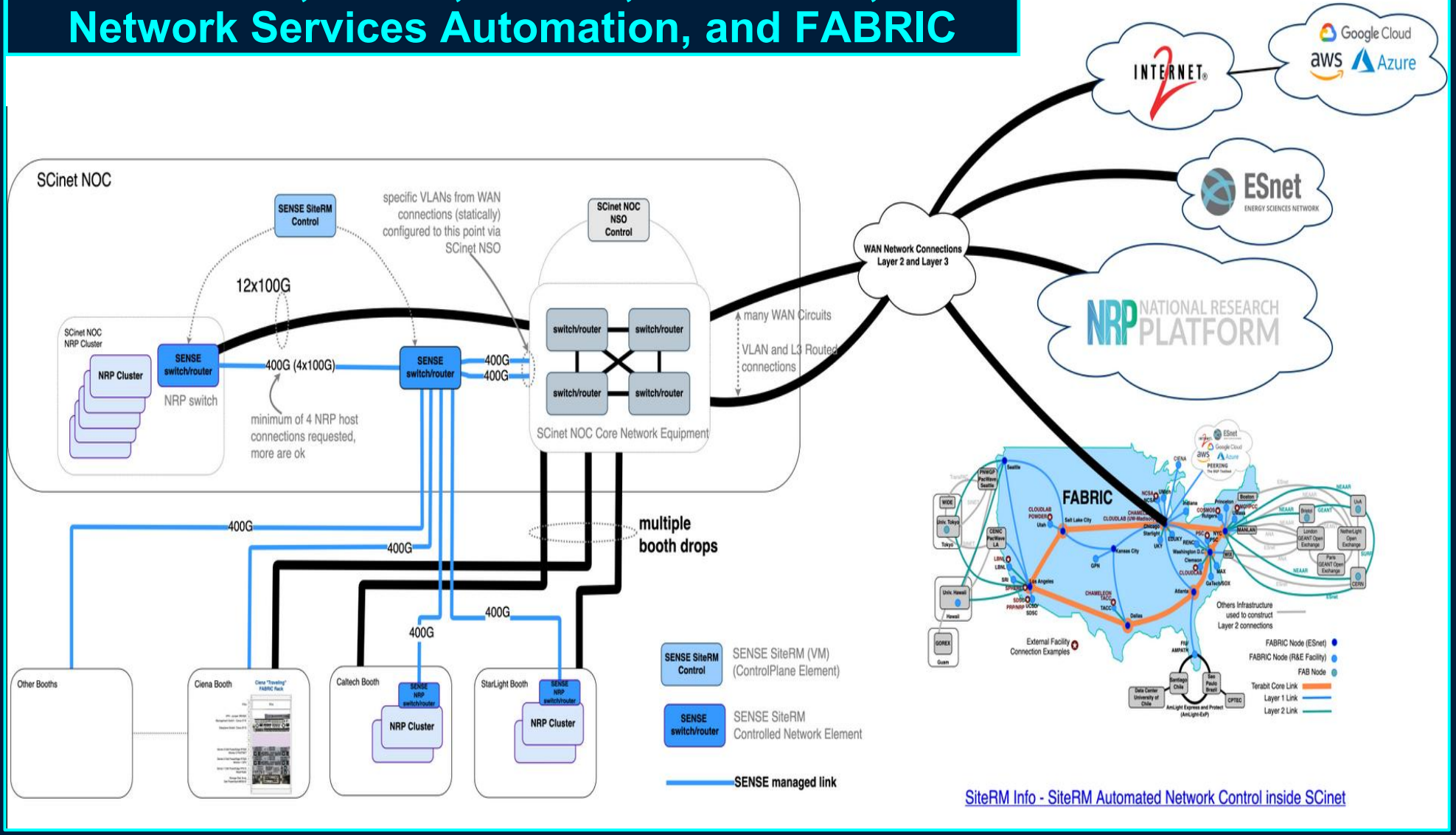
Initial focus for the developed solution is the next-generation **Large Hadron Collider CMS and ATLAS** workflows + **VRO** and others soon.

The **SENSE** and **Rucio** based sol'n enables **full lifecycle workflow services** to maximize data transfer/ streaming efficiency, optimize network utilization, and provide cost-effective cluster deployments.

Courtesy Tom Lehman, ESnet 3/2025

NRE-106+ 117: Multi-Resource Cyberinfrastructure Services for Science Domain Workflows via SENSE and FABRIC

SC25 NREs, Scinet, SENSE, AutoGOLE, NRP Network Services Automation, and FABRIC



SiteRM Info - SiteRM Automated Network Control inside SCinet

NRE-101: DOE Integrated Research Infrastructure (IRI)
IRI Compute Workflow Portability

GOALS

- Demonstrate a "compute portability" workflow across distributed ASCR's computing facilities and testbed resources: FABRIC, NRP, and SC25 Scinet Computing Cluster and data streaming with EJFAT (ESnet/JLab FPGA Accelerated Transport).
  Simulating an application workflow where compute jobs can move between facilities depending on resource availability, with integrated data access and transfers
- Show portable compute jobs running across different sites and facilities, using different technologies (HTCondor, Slurm, Kubernetes) for basic job execution and data placement
- With data movement using Globus, XRootd, HTTP, or streaming data services while providing network guarantees via SENSE, and then expanding into an iterative compute-move-compute pattern, while
  - Enabling tightly integrated workflows across experimental facilities, HPC systems, and high-speed networks
  - Establishing unified interfaces/APIs for allocations, data/job movement, scheduling
  - Supporting data-intensive and time-sensitive applications, such as experiment steering and real-time analytics, together with EJFAT.
  - Fostering modularity and portability, allowing seamless transitions of data and computation jobs across different facilities
- ✸ Pathway to a unified infrastructure + system scaling across agencies & institutions

NRE-122: Agentic AI with QAIC Ultra Cards

for HPC Cluster Management and Resource Provisioning

- NRP, supported by 70+ institutions, provides access to cutting-edge technologies in AI, high-performance computing, data storage, and networking using the Nautilus Kubernetes Cluster.
- NRP is providing a node with Qualcomm Cloud AI 100 Ultra (aka QAIC) inference and fine-tuning cards to advance autonomous, agentic AI-driven cluster management.
- These enable hosting AI-powered provisional alarms and control agents directly within the HPC environment, reducing latency and increasing operational responsiveness
- This NRE will demonstrate the integration of LLMs running on these AI accelerators embedded within the Nautilus cluster, to manage cluster admin operations and provisioning/debugging tasks
- This NRE also will demonstrate the ESnet SENSE Orchestrator + NRP's SENSE Operator for provisioning via the Orchestrator, while focusing on LLM-based agentic effort in reporting and debugging

NRE-127: Onboard AI and Federated Learning for Cosmic Ray Event Detection

Caltech, U. Delaware, MIT, Oligo Collaboration

Next Steps & Longer-Term Plans

- **Immediate (+1-2 years): Scale to +8 institutions:**
  Operational federated learning system for scientific collaboration;
  Int'l Collaboration though a multi-country scientific federated learning network
- **Medium-Term (2026-2028): Space Deployment and Standardization:**
  Autonomous cosmic ray monitoring satellites using federated learning;
  Adopt protocols for application-specific scientific collaboration

Broader Applications

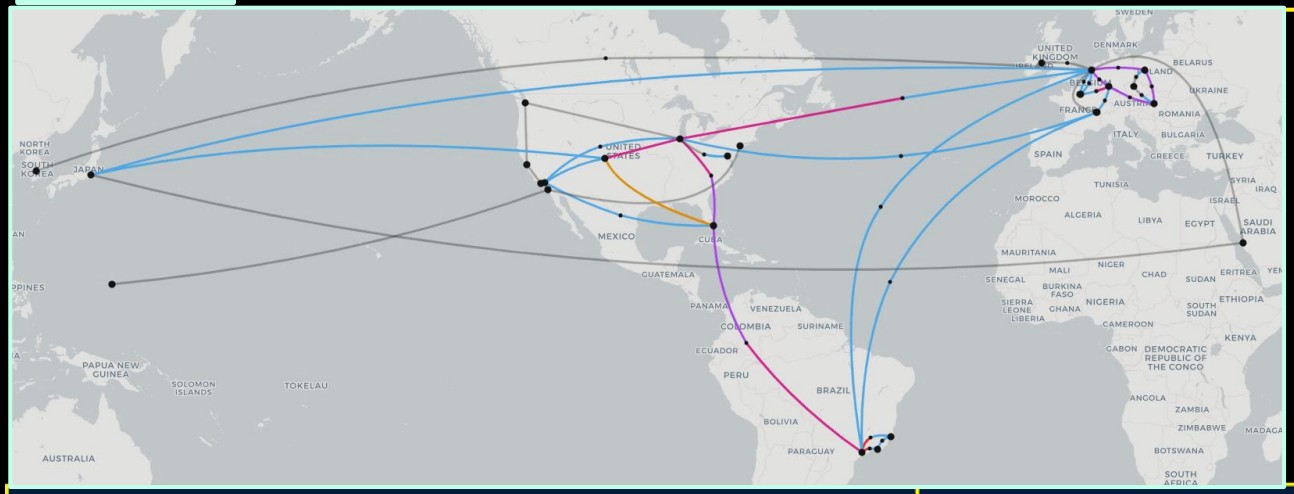
- The CREDO smartphone application, for example, enables citizen science participation, collecting cosmic ray detection data from users with cell phones worldwide

Future Vision

- **Scientific Collaboration Revolution:** CREDO represents a paradigm shift in how scientific institutions collaborate, enabling new levels of distributed processing and knowledge sharing
- **Space Science Applications:** Autonomous satellites with federated learning capabilities could revolutionize space weather prediction and cosmic ray monitoring
- **Broader Impact:** Transform scientific collaboration across disciplines, enabling breakthrough discoveries through distributed processing and collaborative learning

Global Platform for Labs (GP4L)

Prototyping Current/Next-Generation R&E Network Architectures at Scale



RNP + GEANT
Marcos Schwarz
Frederic Loui et al.
2023-5: From 26 to 38 Sites

A new worldwide platform for

- New agile and flexible feature development
- New use case development corresponding to research programs' requirements
- Multiple research network overlay slices

A global playing field for

- Next generation network monitoring tools and systems
- Development of fully automated network deployment
- New network operations paradigm development

VISION: Federate and integrate multiple testbeds and toolsets such as AutoGOLE / SENSE

Rapidly Deployable Network Digital Twin
with high fidelity (Container Lab) for complex network topologies
Plan: Smooth Transition from Simulation to Real-World Global Operations

Global Platform for Labs (GP4L)
Status and Ongoing Progress

Marcos Schwarz RNP
Frederic Loui RENATER

Achieved by SC24; En Route to SC25:

- Persistent global L3 overlay network based on P4 switches, DPUs and DPDK
- Core network based on RARE/FreeRtr and edge networks based also on SONiC and Holo
- Automated generation of a real time world map + dashboards of GP4L
- VPP as software dataplane to achieve 100 Gbps forwarding capacity per CPU
- Intercontinental high capacity transfers (100G and up) exploring multiple source routing solutions and next generation protocols (e.g. PolKA)
- Creation of an on-demand digital twin of the GP4L and production networks
- Interconnections with other testbeds (FABRIC)

In Progress

- Capability to support multiple virtual networks that implement different choices of routing stacks on the same devices: traditional and SDN based
- Integration with initiatives for visibility, controllability and intelligence
- *Working as a reference state of the art / next generation R&E network*

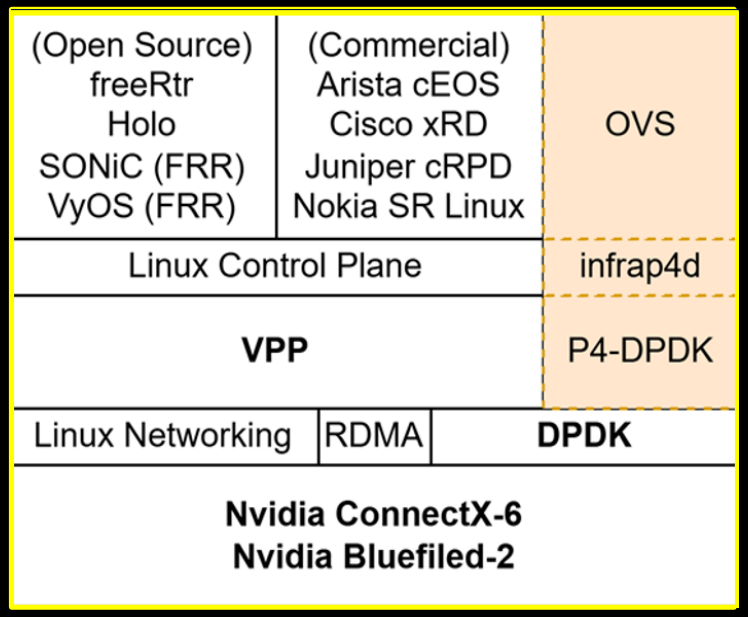
Global Platform for Labs (GP4L)

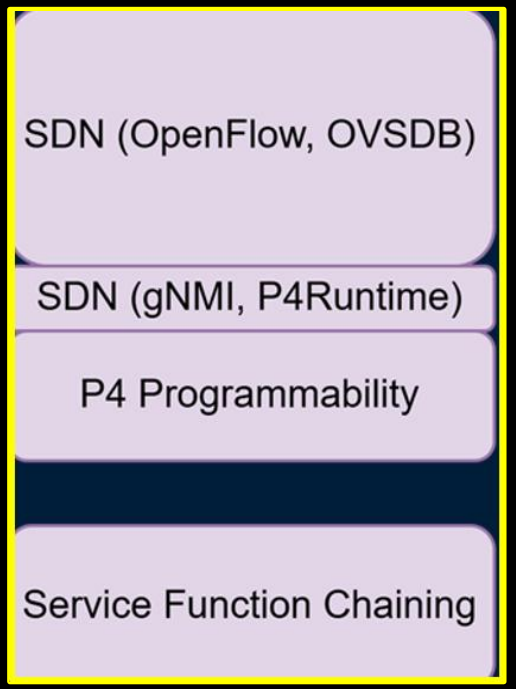
SC25 NRE-126 Next Steps

Marcos Schwarz RNP

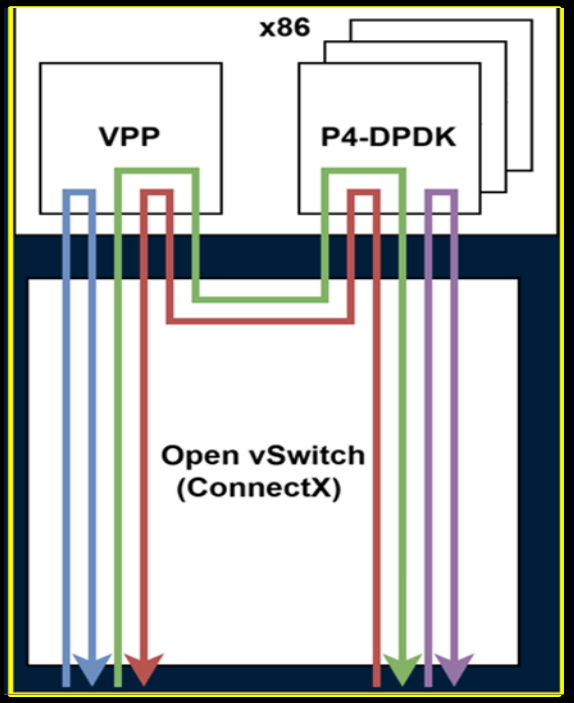
1. Publish Notebook and results on generating single flow 100G using iperf3
2. Control Plane Integration: FRR, Holo, Vendors (Arista, Juniper, Nokia…)
3. Multiple 100G WAN Multipath + Traffic Engineering Scenarios
4. Hardware Accelerated QoS
5. P4-DPDK Integration

VPP + P4-DPDK Integration (Draft proposal)





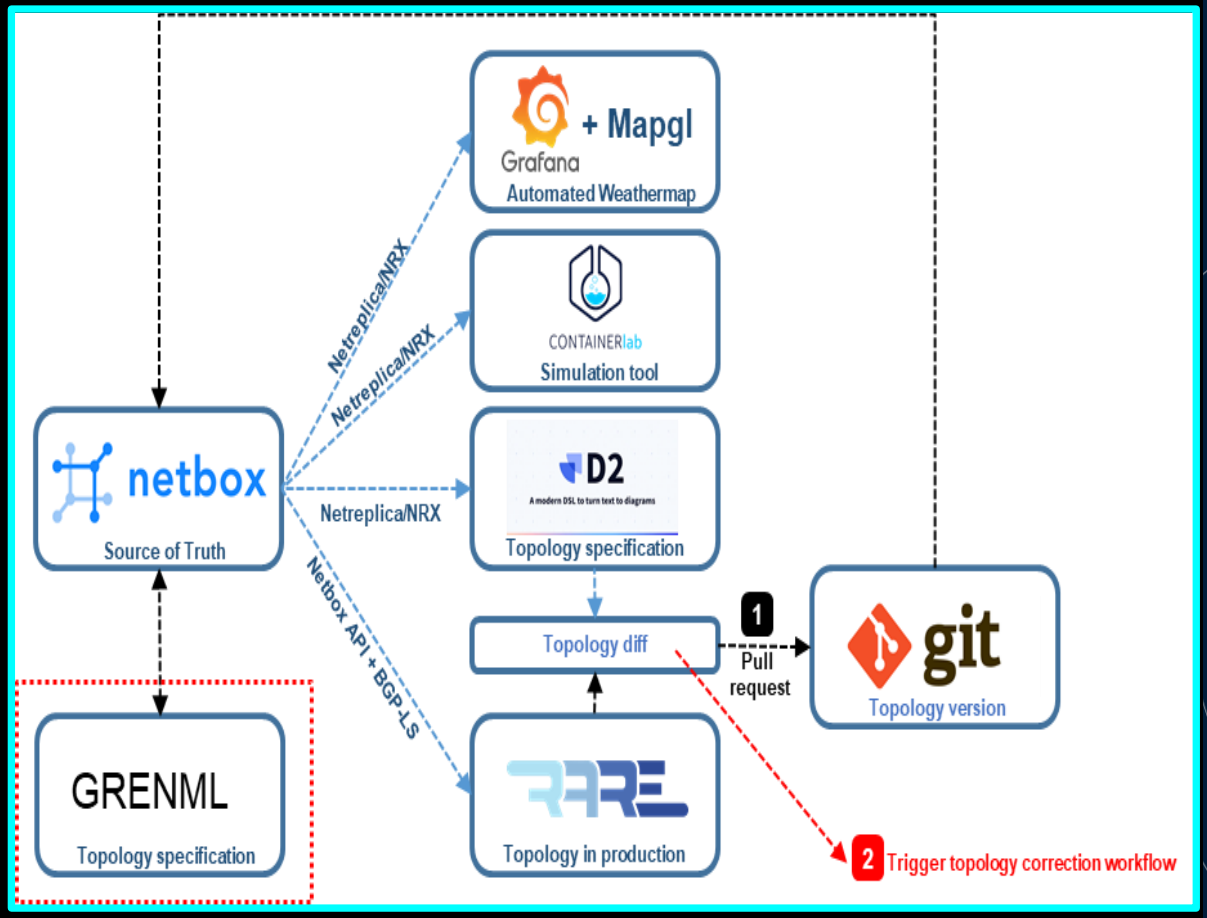
PCIe



Global Platform for Labs (GP4L)
and GRENML Integration

Marcos Schwarz RNP

- Support import/export of topology data from/to Netbox and GRENML
- Allow exporting GP4L data to generate GREN Map based visualizations
- Allow the creation of diagrams, weathermap and digital twins of GREN data



Global Platform for Labs (GP4L)
Context

Marcos Schwarz RNP

- State of P4 hardware, since Tofino was discontinued in 2023:
  - IPUs and DPUs from Intel and AMD/Xilinx are currently the norm
  - Next-Generation P4 ASIC - Xsight Labs X2 + Oxide P4 compiler
  - Tofino compiler and SDE were Open Sourced as part of p4lang
- We are constantly looking for future candidates that can provide production grade features and deep programmability
  - freeRtr + Tofino: Base architecture
  - SONiC/PINS/DASH + SAI:
    - Future proof with the biggest community and ecosystem of partners
    - Progress is slow but steady and high barrier to contribute
  - VPP + P4-DPDK + Linux Control Plane
    - Everything is in software and easier/faster prototyping
  - Flexible control plane options: FRR, Holo, vendors (Arista, Juniper, Nokia…)

Vector Packet Processing (VPP)

Marcos Schwarz RNP

Rich Feature set

- 14+ MPPS, single core
- Multipath - ECMP and Unequal Cost
- Network Programming using SR-MPLS and SRv6, Traffic Steering

Device Drivers

- DPDK, eBPF/XDP, RDMA, device specific (Intel, Marvell)

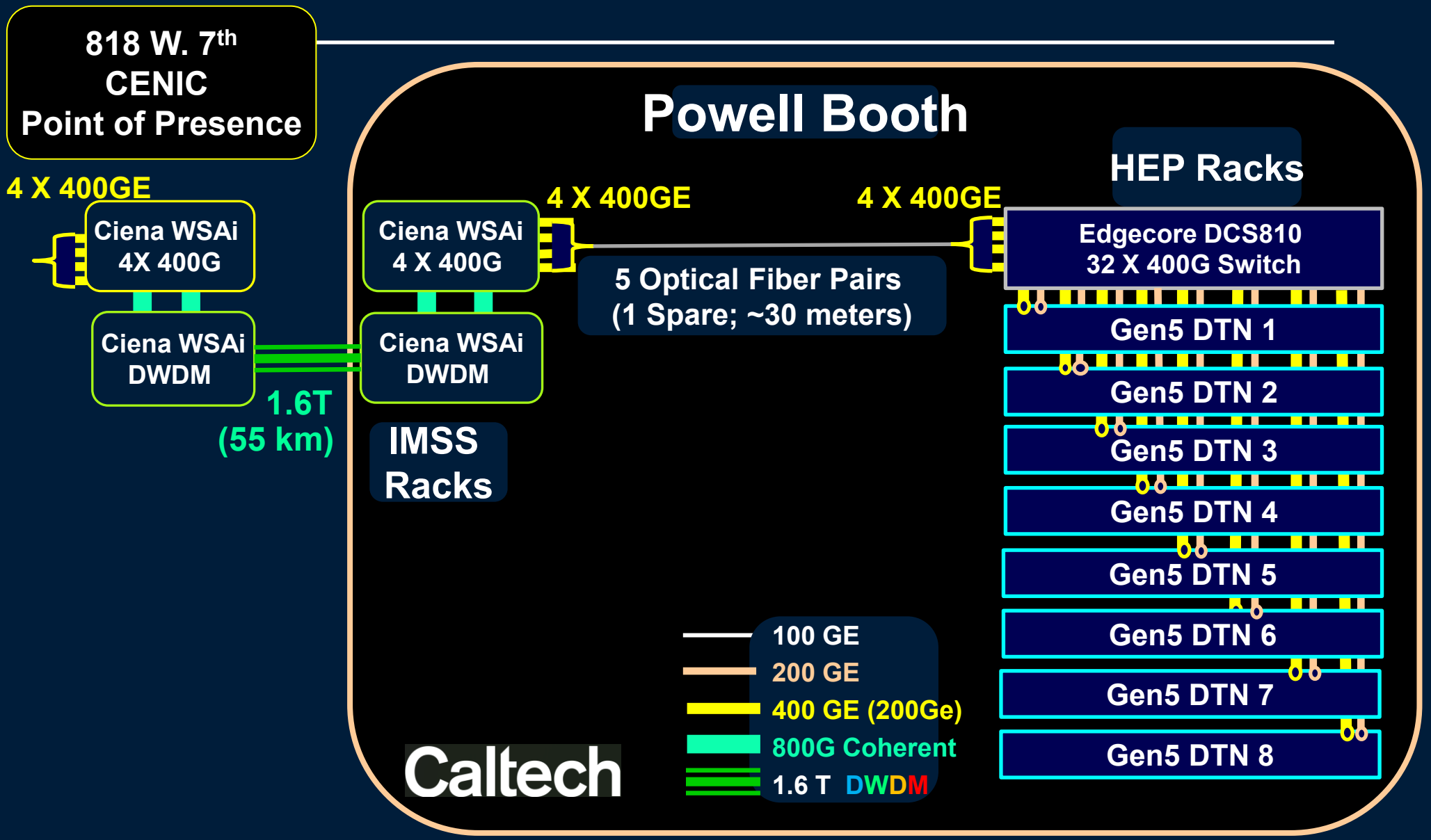
Linux Control Plane

- Any control plane

Adopted by many high performance routing projects

- SONiC, VyOS, Netgate TNSR, Calico-VPP

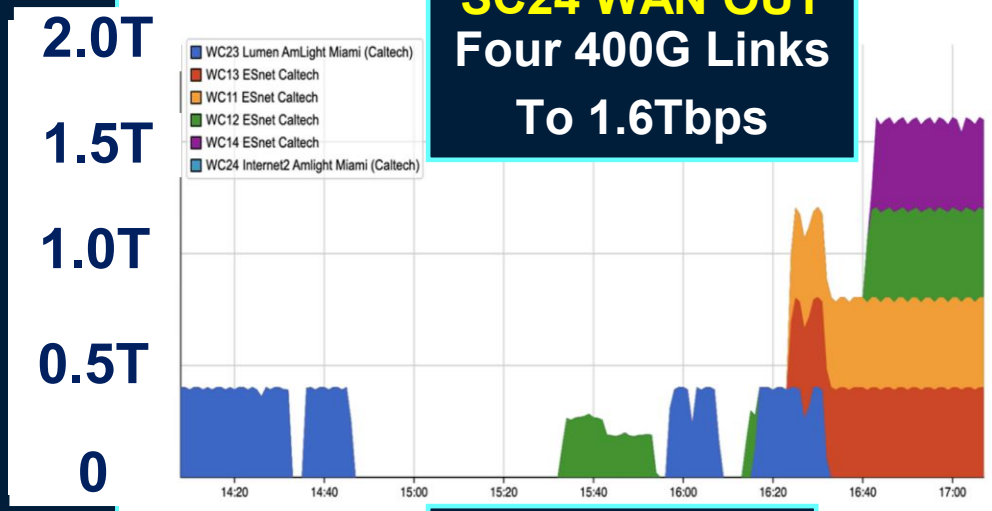
Simplified Caltech – LA Layout



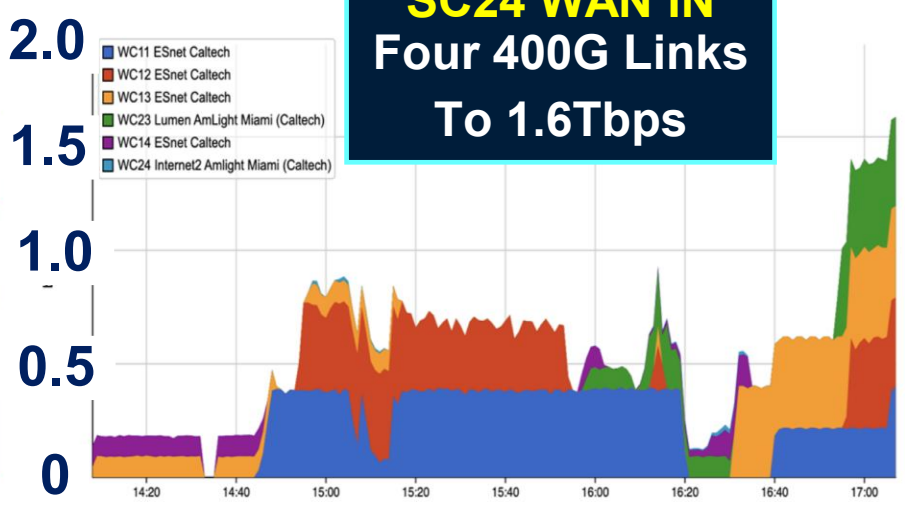
NRE-13 First Results: To 3.2 Tbps on 4 of 6 X 400G Circuits: 4 LA-ATL, 2 Miami-ATL

FDT++ 11/19/24

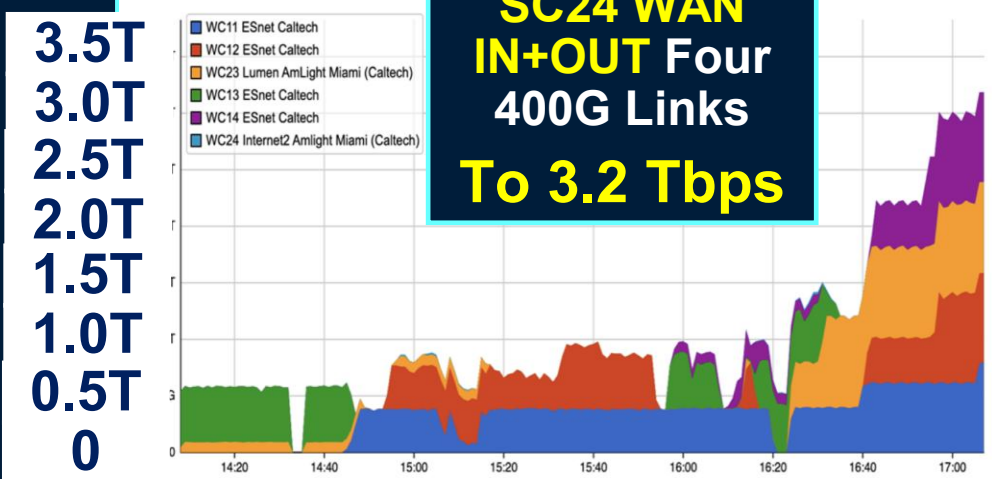
SC24 WAN OUT
Four 400G Links
To 1.6Tbps



SC24 WAN IN
Four 400G Links
To 1.6Tbps



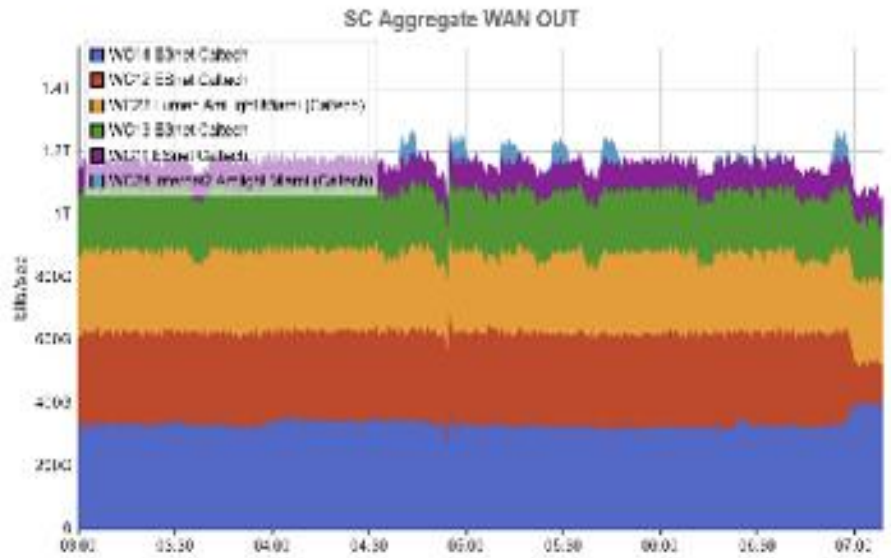
SC24 WAN IN+OUT Four 400G Links
To 3.2 Tbps

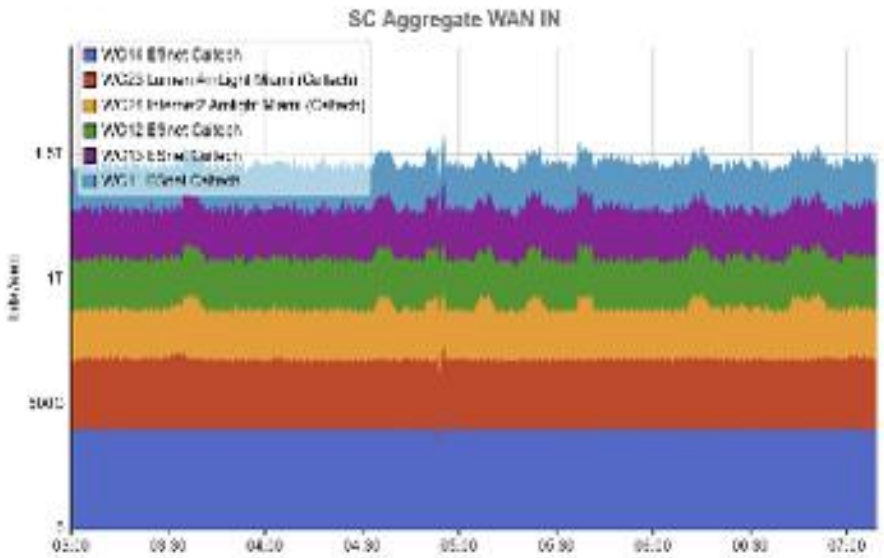


Just Getting Started (on 4 of 6 Links)

+ Studying the limits in depth of a single 32 core CPU and server with 1.1 Tbps of NIC ports

SC24 WAN Circuits Caltech (Booth 845)





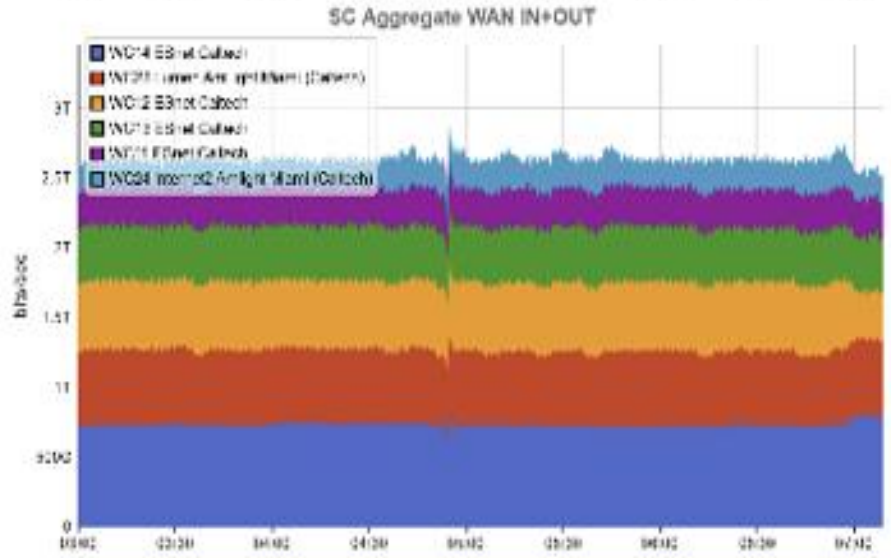
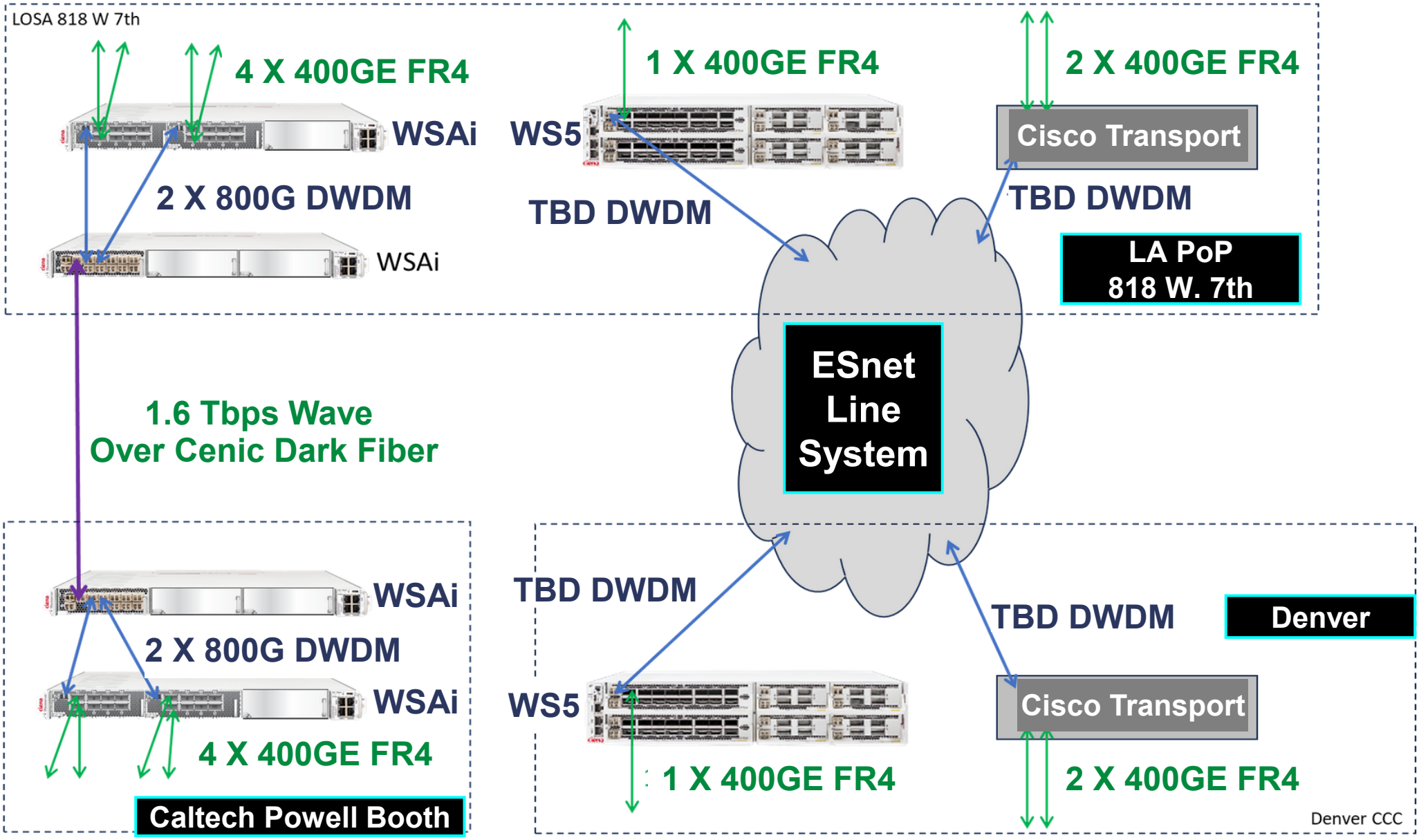


Figure 3 Inmon monitoring of the traffic flows over SCinet between the Caltech booth, the Caltech campus, Miami, Sao Paulo and South Africa before and during the bandwidth tsunami, up to the conclusion of the SC24 Network Research Exhibition. The aggregate in+out rate reached was 4.3 Tbps.

Ciena WaveServer Ais and Waveserver 5s: Site Connections at the SC24 (Atlanta), the CENIC PoP (LA), and Caltech (Pasadena)



General rules for better throughput

Raimondas Sirvinskas and Marcos Schwarz

Caltech

Lessons learned from previous Supercomputing conferences:

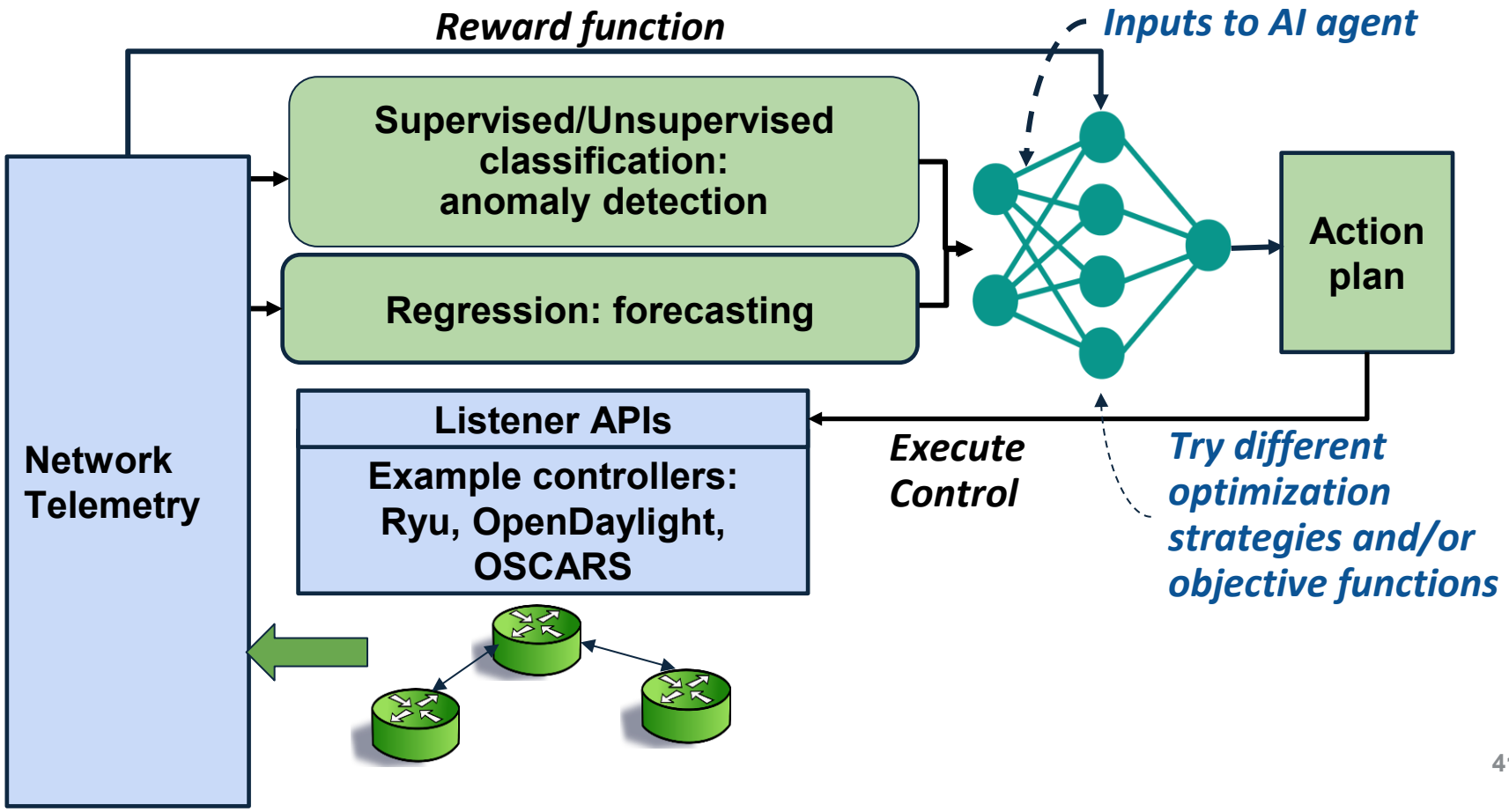
- We should not trust any kernel version, except the one(s) we have tested and confirmed to work well.

  For example kernel 4.18.0 comes with AlmaLinux 8.8 by default:
  It tested well on two 200Gbit links using single direction transfers
  but it failed when we started a third transfer on a third interface
  at the same time, or 2x 200Gbit bidirectional transfers

- Recommended kernel version was 6.5.10
- Ensure Jumbo frames is set on each interface and VLAN
- Ensure the CPU governor is set to performance
- Turn Adaptive RX off
- Set txqueuelen to 10000
- Set the network interface RX and TX buffers to the maximum supported values
- Set the interface to use BIG TCP (newer kernel feature available since 6.3)

Self-Driving Networks
"Networks that self-manage"

Can take simple actions to achieve key goals. Such as:
improving availability, attack resilience and dealing with scale.
Our argument is AI is needed for mission critical actions.



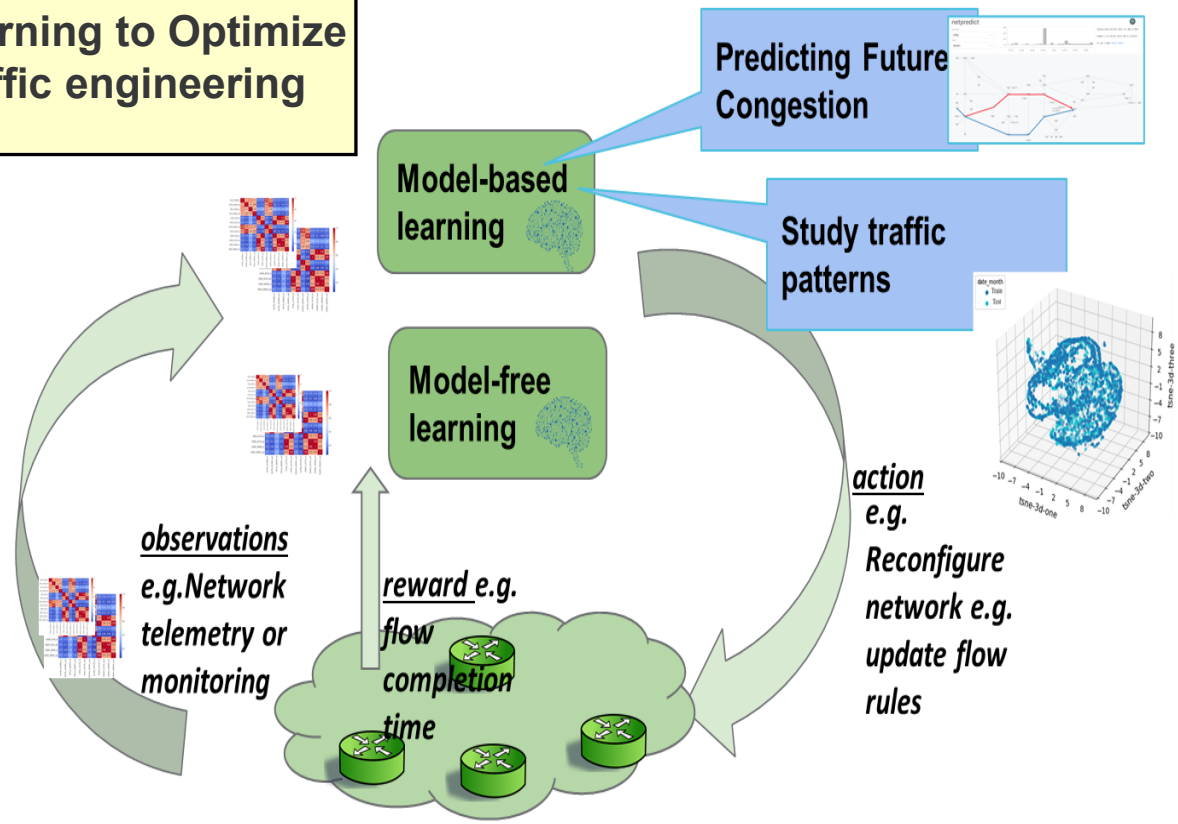
Kiran, Engineering Autonomous Self-Driving Networks, ACM SIGMETRICS Performance Evaluation Review 51 (2), 97-99

Self Driving Network
Adaptive Routing (e.g. Real time data for routing decisions)
Learns to Avoid Congestion
Congestion Free ➔ Loss Free
Towards 100% Utilization
Proactive Fault Repair

Mariam Kiran (ORNL) et al. Intelligent Networks DOE Project

Self-Driving Network for Science

Use Deep Learning to Optimize network traffic engineering



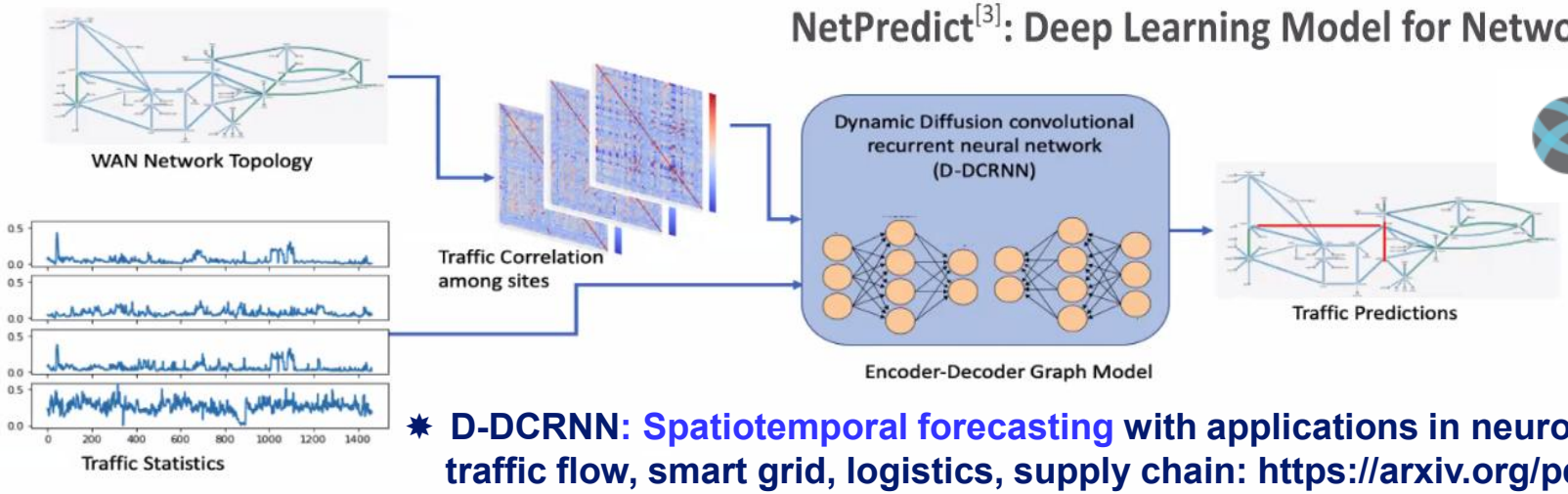
Case Studies:

1. **Model free:** Path selection for large data transfers: better load balancing **HECATE**
2. Forwarding decisions for complex network topologies: Deep RL to learn optimal packet delivery policies vs. network load level **HECATE+POLKA**
3. **Model Based:** Predicting network patterns with **Netpredict**

OAK RIDGE National Laboratory

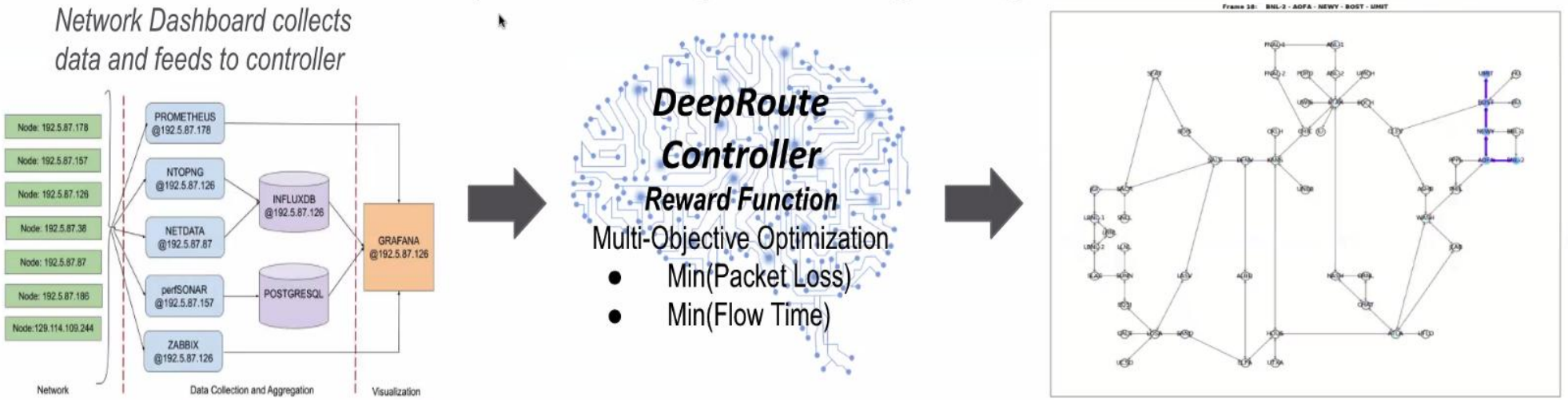
DAPHNE[2] - Deep and Autonomous High Performance Networks

NetPredict[3]: Deep Learning Model for Network Congestion



- **D-DCRNN: Spatiotemporal forecasting with applications in neuroscience, climate, traffic flow, smart grid, logistics, supply chain: https://arxiv.org/pdf/1707.01926.pdf**

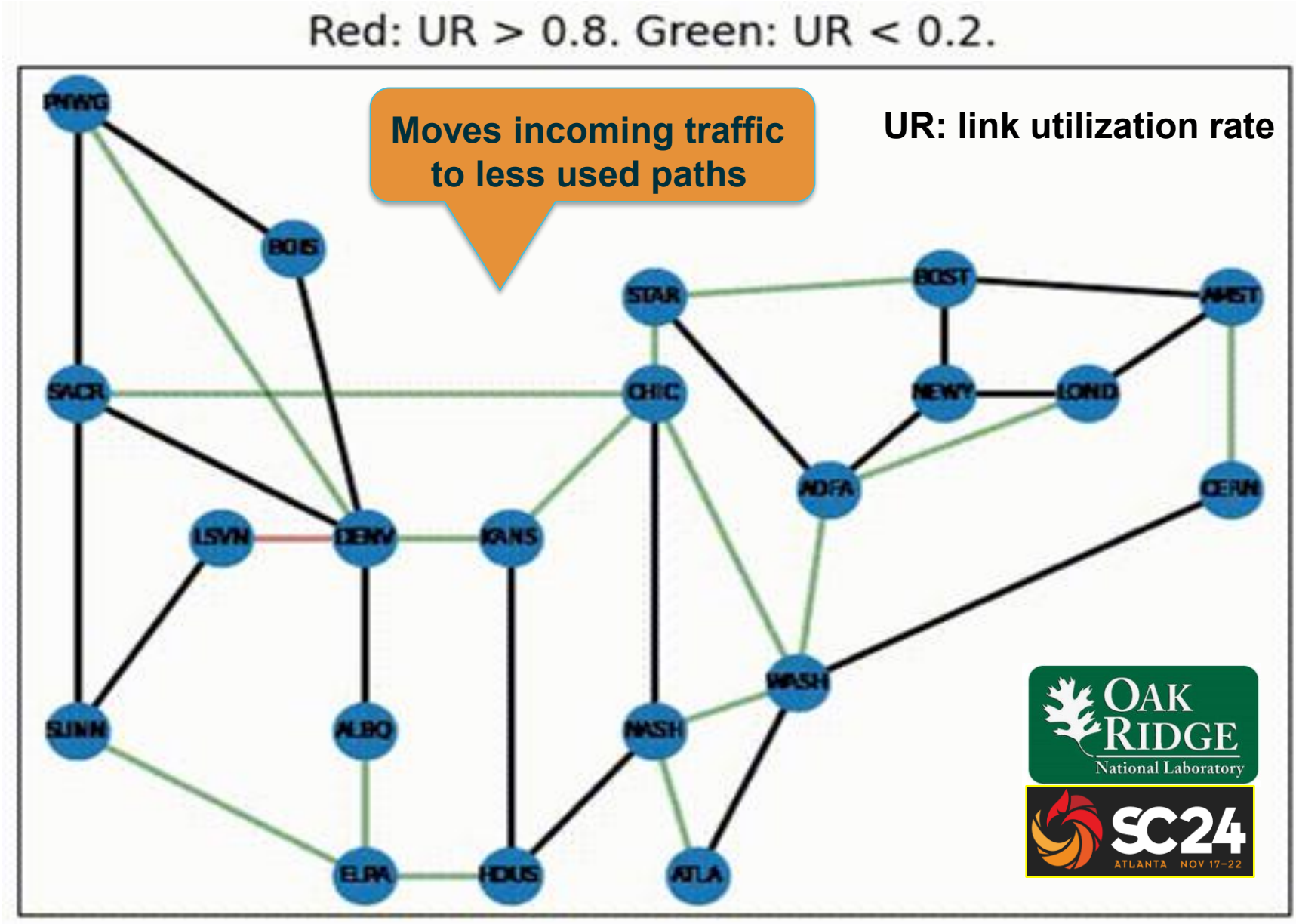
DeepRoute[4]: Intelligent Traffic Engineering



- **Future: Hooks for a Richer, More Stateful Objective Function (Policy, Priority, Deadlines, Path Quality, Co-Flows...); Event Response; Applications to 5G and Quantum Networks**

Mariam Kiran (ORNL) et al

HECATE Simulation



Next Generation System for Data Intensive Sciences

- The development of effective ML optimization methods, and multidimensional, real-world metrics
  - Are themselves challenging, groundbreaking activities
  - A new area of application in multidomain distributed systems
- Strategic Aim: Compatible coexistence of programmable goal-oriented networks, and production networks
  - Simultaneously meeting the needs of the leading edge science programs and the at-large A&R communities

Acknowledgements

This ongoing work has been partially supported by the US National Science Foundation (NSF) Grants OAC-2030508, OAC1841530, OAC-1836650, MPS-1148698, and PHY-1624356, along with research grants from many international funding agencies and direct support from the many regional, national, and continental network and industry partners mentioned. The development of SENSE is supported by the US Department of Energy (DOE) Grants DE-SC0015527, DESC0015528, DE-SC0016585, and FP-00002494.

Finally, this work would not be possible without the significant contributions and the collaboration of the many HEP, network and computer and research teams partnering in the Global Network Advancement Group, in particular the GNA-G Data Intensive Sciences and AutoGOLE/SENSE Working Groups, CENIC, Internet2, ESnet, STARLIGHT, NRP, Rednesp, USP, the GP4L and the RNP Brazilian National Network, together with the LHC experiments and many industry partners, most notably Ciena, Arista and Dell.