SC20

Everywhere we are | more than hpc.

P4 Experimental Networks for the Global Research Platform(GRP)

Jim Chen
International Center for Advanced Internet Research (iCAIR)
Northwestern University
StarLight International/National Communications Exchange Facility

Nov 12 2020 • Chicago

- Overview/ Introduction
- Global Research Platform (GRP)
- Different GRP Services in SC20 NREs
- International P4 Experimental Networks (iP4EN)
- Selected Research Projects in iP4EN
- Q & A

Introduction – International Center for Advanced Internet Research (iCAIR) and StarLight

Accelerating Leading Edge Innovation and Enhanced Global Communications through Advanced Internet Technologies, in Partnership with the Global Community

StarLight – "By Researchers For Researchers"

- Creation and Early Implementation of Advanced Networking Technologies - The Next Generation Internet All Optical Networks, Terascale Networks, Networks for Petascale Science

- Advanced Applications, Middleware, Large-Scale Infrastructure, NG Optical Networks and Testbeds, Public Policy Studies and Forums Related to NG Networks

- Three Major Areas of Activity: a) Basic Research b) Design and Implementation of Prototypes c) Operations of Specialized Communication Facilities (e.g., StarLight)

Global Research Platform: Global Lambda Integrated Facility
Available Advanced Network Resources

NRE03--Global Research Platform (GRP)-Software Distribution

To enable partners to participate in Global Research Platform, a set of software stack is being designed and distributed to GRP participating systems.

For Providers/Operators:

- Kubernetes

  1. GRP-hosted: Enables direct GRP participation for your node

  2. Local-hosted(Federation): Create your own k8s cluster and federate with GRP

  3. NSI Network Control Automation (in progress)

For Users:

- DTN-as-a-Service

- International P4
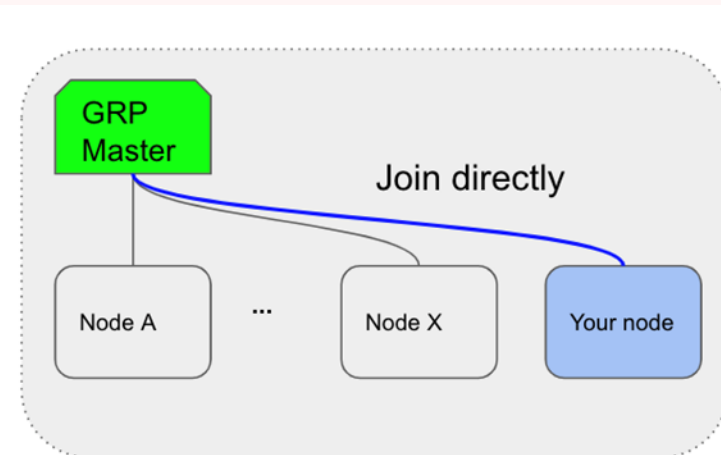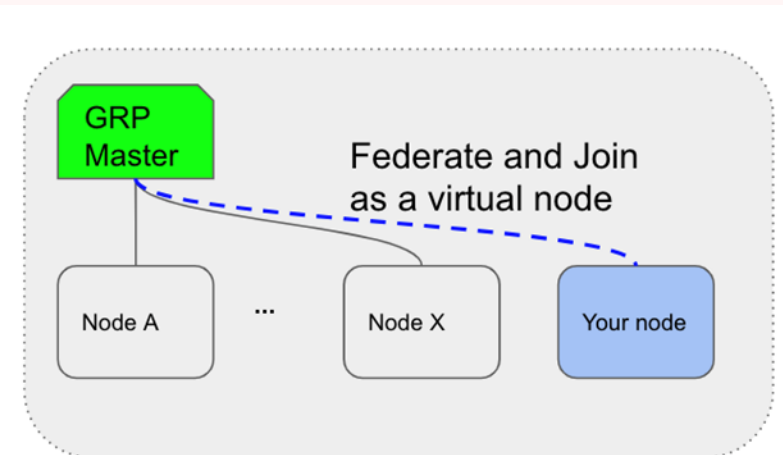
Experimental Networks

- SAGE2

Fig. 1

Fig. 2

NRE11-GRP Service: Research Platforms Federation Demonstration

Goal: Secure multi-domain resource sharing across regional, national and international research platforms
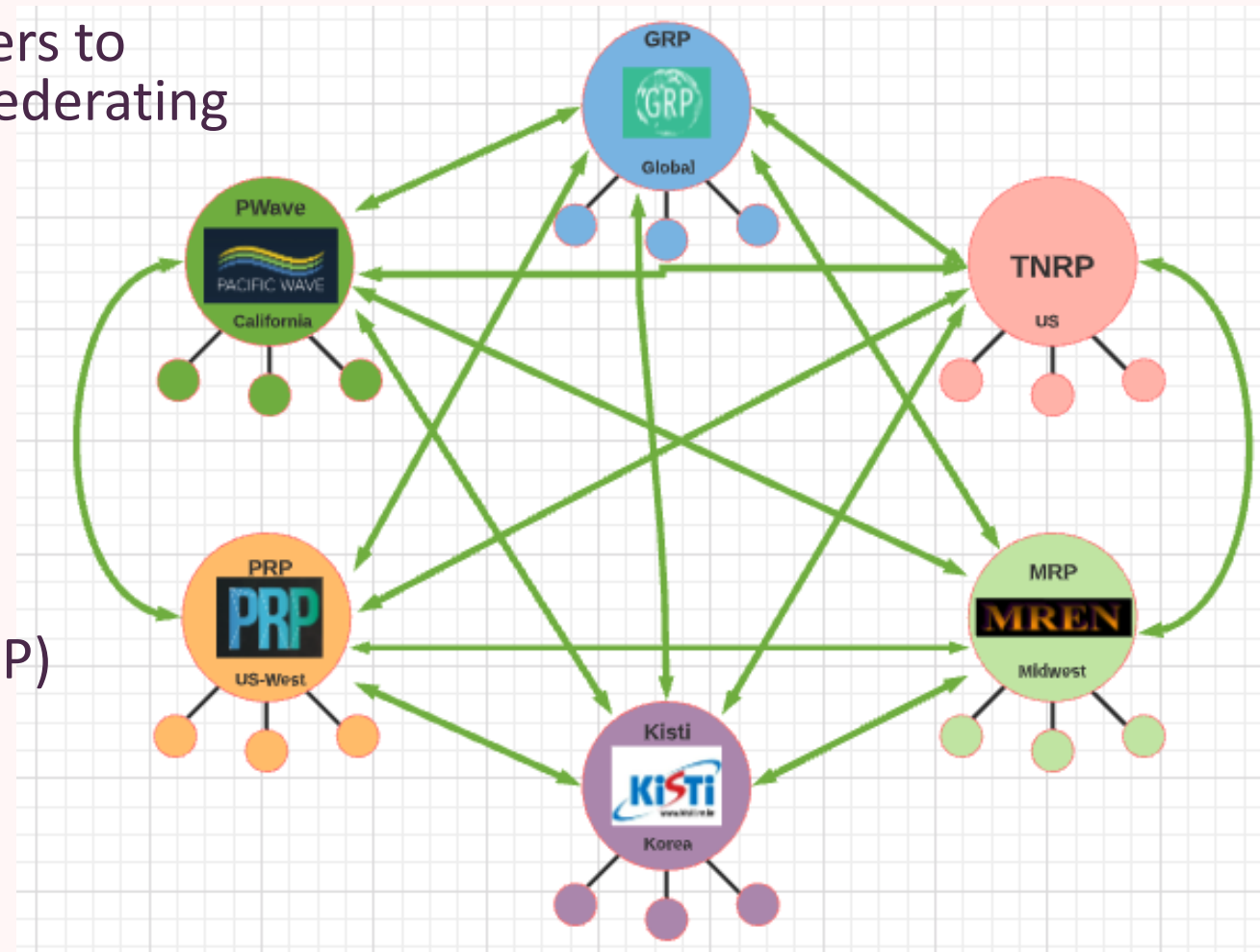
Solution: Admiralty. The software enables users to schedule workloads in a different cluster by federating the source and target clusters.

Participants:

- Global Research Platform (GRP)

- Pacific Research Platform (PRP)

- MREN Research platform (MRP)

MREN:Metropolitan Research and Education Network

- Towards a National Research Platform (TNRP)

- Pacific-Wave

- KISTI

Admiralty Overview

Admiralty is a system of Kubernetes controllers that intelligently schedules workloads across clusters.

- Unilateral: your cluster will be a *source* and you can define the resources from other clusters as *targets*. This process is unilateral, not mutual.

- Decentralized: each cluster has its own control plane without rely on a central clyster owned by a single organization.

- Scheduling: flexibility to schedule pods to a specific namespace on certain nodes from a particular cluster
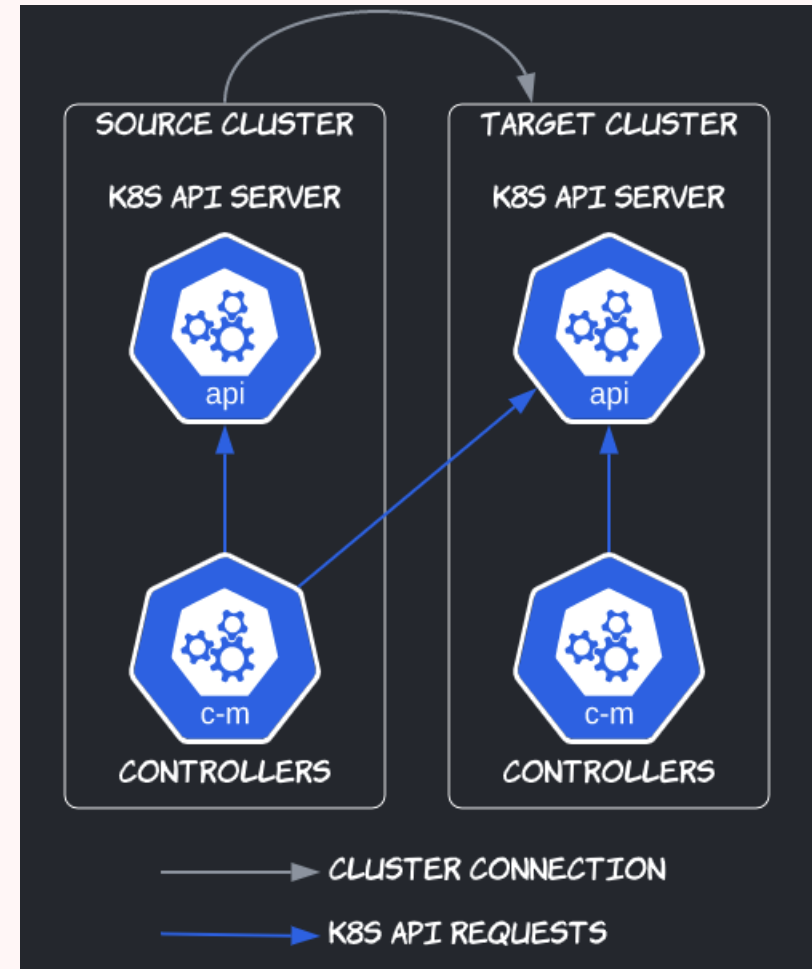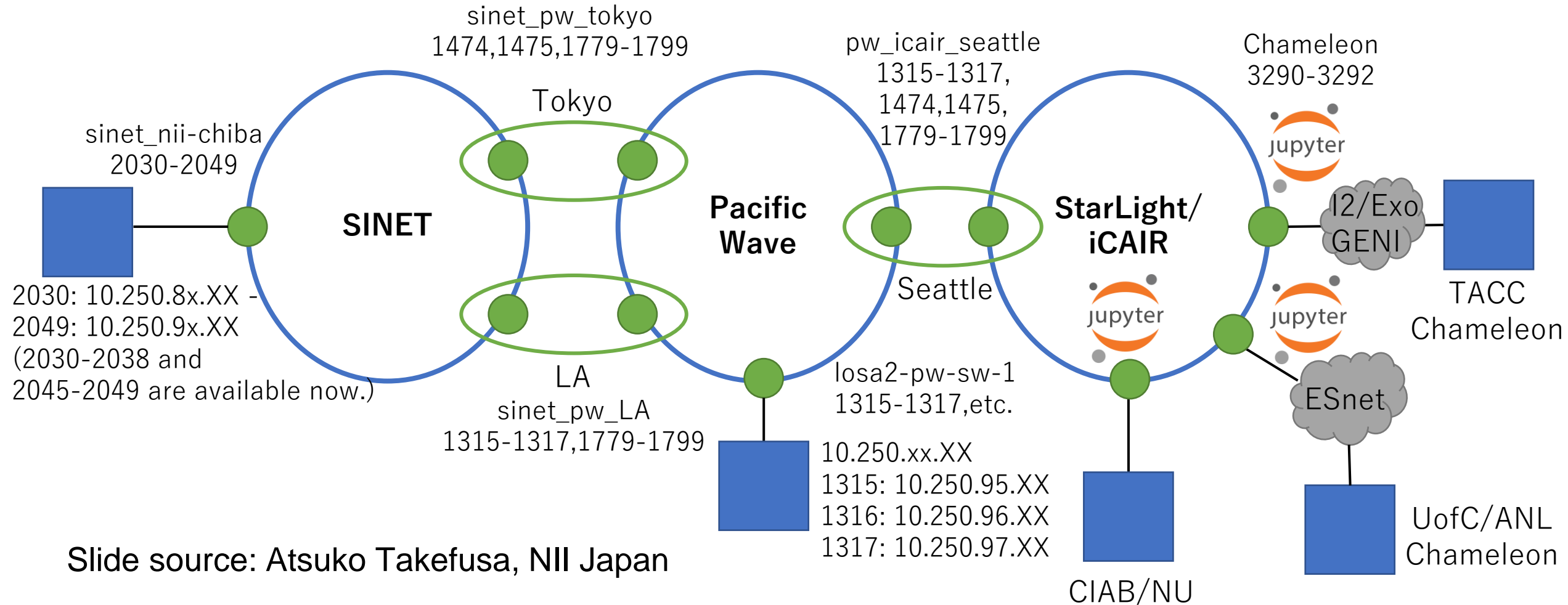
NRE14: AutoGOLE & SENSE

Persistent Multiresource Infrastructure 2020 update 4
Gerben van Malenstein
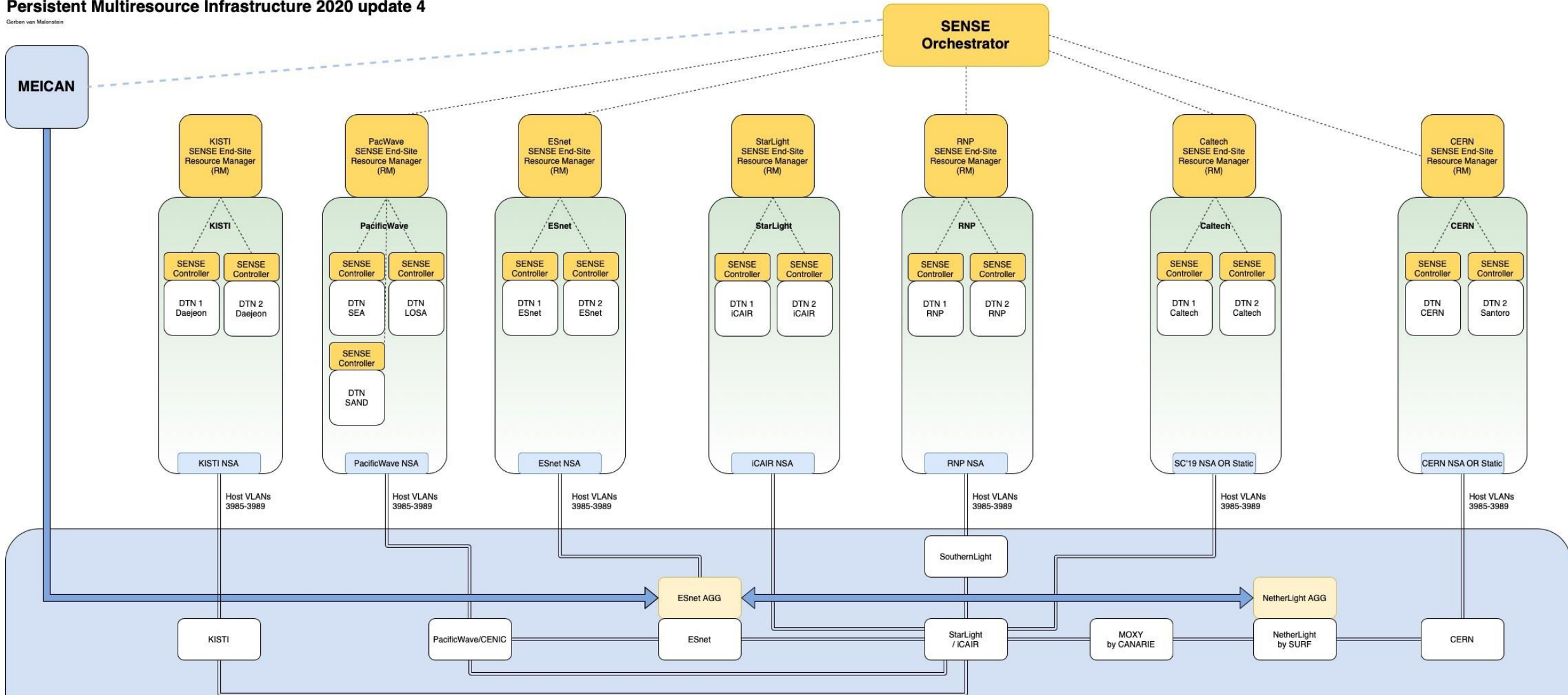
NSI Enabled Dynamic International Multi-Domain Networks for Cloud Research

Virtual Cloud Provider (VCP) : Application-Centric Overlay Cloud
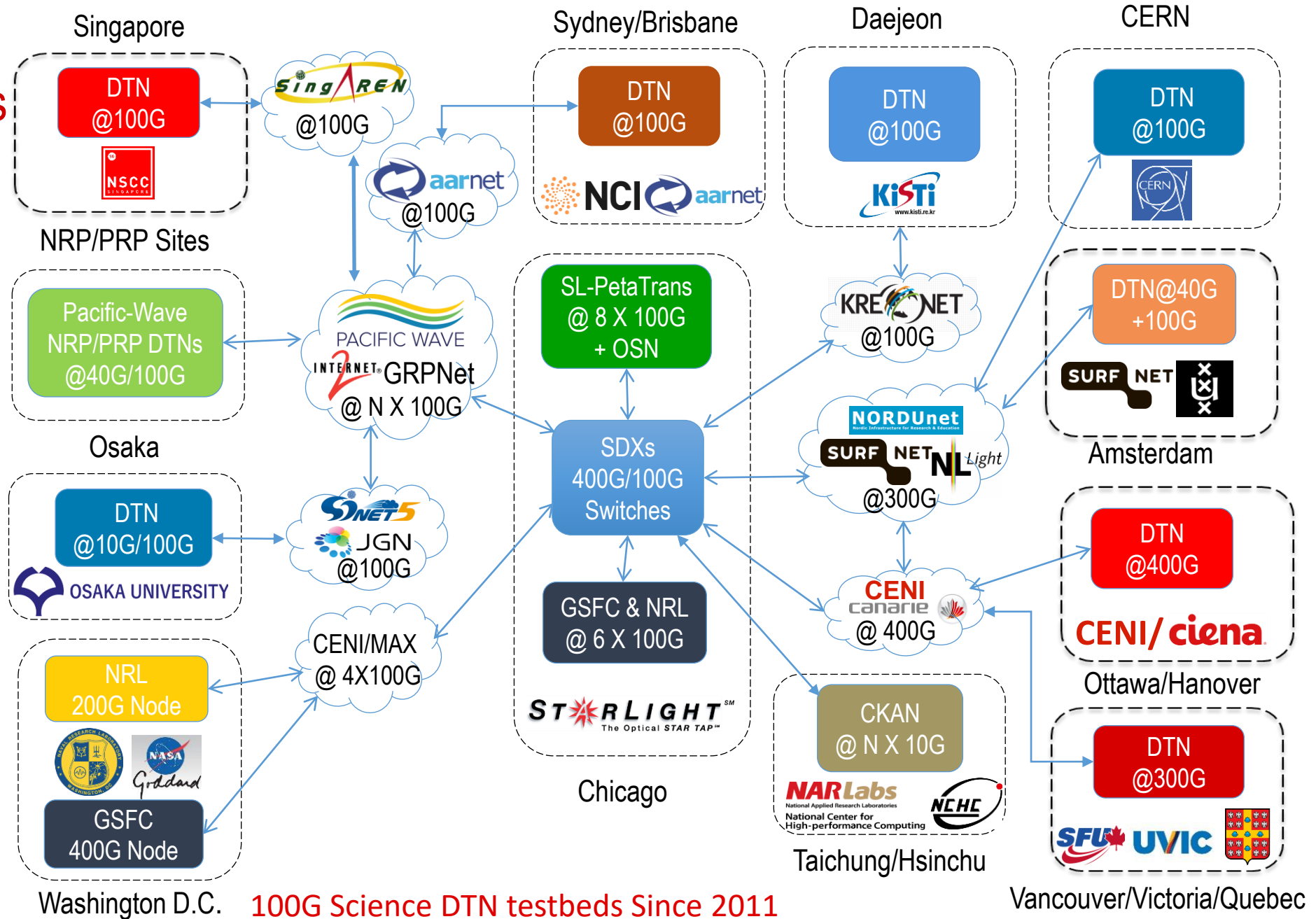Prototyped over NII Cloud-Chiba, Japan and NSF Chameleon Cloud, U.S.

sinet_pw_tokyo
1474,1475,1779-1799

Tokyo

pw_icair_seattle
1315-1317,
1474,1475,
1779-1799

Chameleon
3290-3292

sinet_nii-chiba
2030-2049

SINET

**Pacific
Wave**

**StarLight/
iCAIR**

I2/Exo
GENI

2030: 10.250.8x.XX -
2049: 10.250.9x.XX
(2030-2038 and
2045-2049 are available now.)

Seattle

LA
sinet_pw_LA
1315-1317,1779-1799

losa2-pw-sw-1
1315-1317,etc.

10.250.xx.XX
1315: 10.250.95.XX
1316: 10.250.96.XX
1317: 10.250.97.XX

TACC
Chameleon

ESnet

UofC/ANL
Chameleon

Slide source: Atsuko Takefusa, NII Japan

CIAB/NU

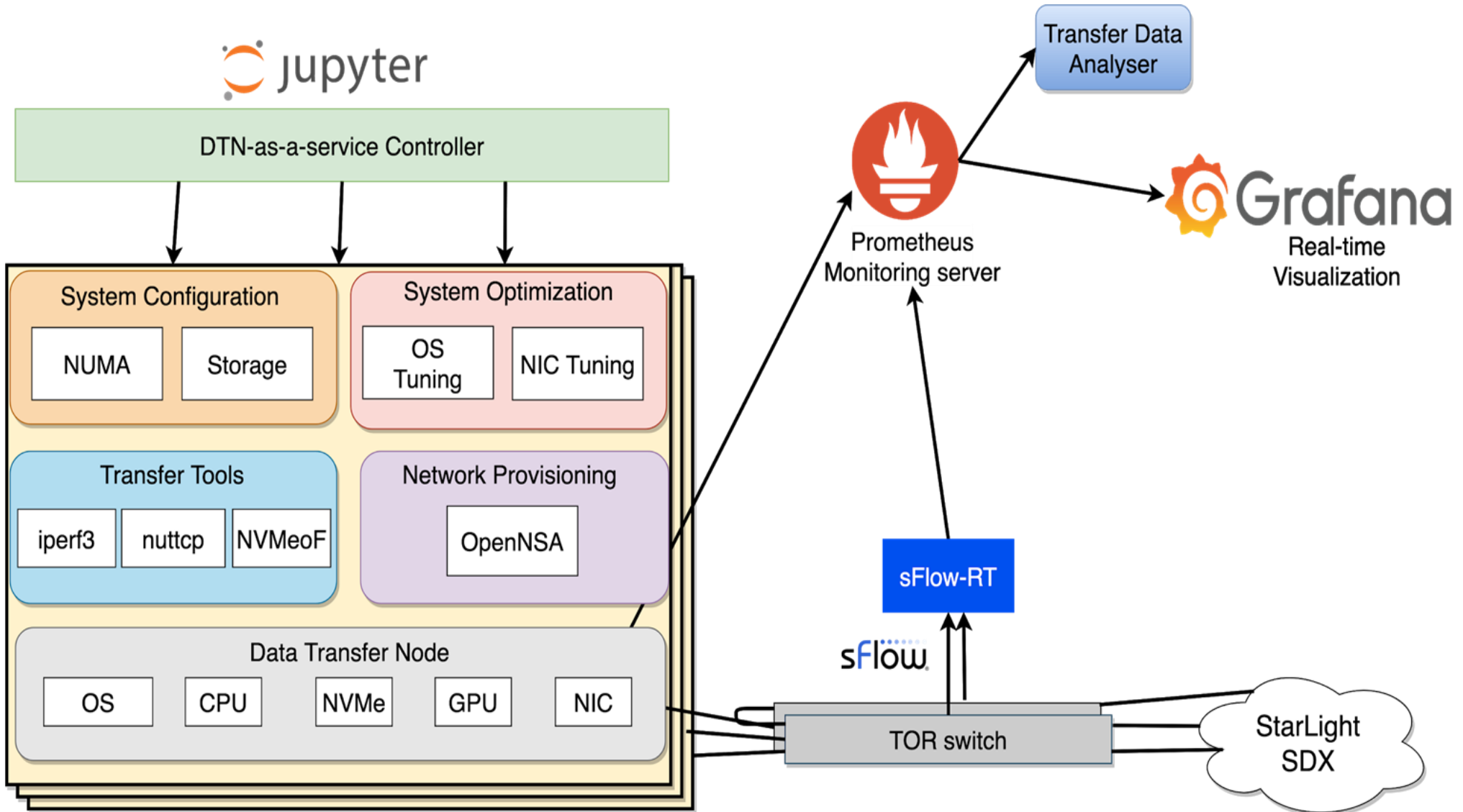# NRE09-GRP Service: DTNaaS for Petascale Sciences Data Transfer

As 10/09 Supports

NRE03, NRE04
NRE05, NRE06
NRE10, NRE11
NRE12, NRE13
NRE14
indis104s1

Singapore

DTN @100G

NSCC SINGAPORE

NRP/PRP Sites

SingAREN @100G

aarnet @100G

Pacific-Wave NRP/PRP DTNs @40G/100G

PACIFIC WAVE

INTERNET2 GRPNet @ N X 100G

Osaka

DTN @10G/100G

OSAKA UNIVERSITY

SINET5 JGN @100G

Sydney/Brisbane

DTN @100G

NCI aarnet

SL-PetaTrans @ 8 X 100G + OSN

SDXs 400G/100G Switches

GSFC & NRL @ 6 X 100G

STARLIGHT The Optical STAR TAP

Chicago

Daejeon

DTN @100G

KiSTi www.kisti.re.kr

KREONET @100G

NORDUnet Nordic Infrastructure for Research & Education

SURF NET NL Light @300G

CERN

DTN @100G

CERN

DTN@40G +100G

SURF NET

Amsterdam

CENI canarie @ 400G

CKAN @ N X 10G

NARLabs National Applied Research Laboratories National Center for High-performance Computing NCHC

Taichung/Hsinchu

DTN @400G

CENI/ciena

Ottawa/Hanover

DTN @300G

SFU UVIC

Vancouver/Victoria/Quebec

NRL 200G Node

NAVAL RESEARCH LABORATORY NASA Goddard

GSFC 400G Node

CENI/MAX @ 4X100G

Washington D.C.

100G Science DTN testbeds Since 2011

# GRP Service: DTN-as-a-Service Software Architecture

GRP Cluster with DTN-as-a-Service

DTN-as-a-Service(DTNaaS) provides a data movement workflow in GRP k8s cluster:

1. Deploy DTNaaS workloads via k8s API server
2. Use Jupyter to optimize and run transfers
3. Observe performance from monitoring service

GRP DTNaaS Components:

- Orchestrator: controller of DTNaaS to manage agent and optimizer pods via REST API.
- Transfer Agent: run transfer jobs
- DTN Optimizer: optimize the DTN resources for workflow
- Jupyter: web interface to run DTNaaS interactively
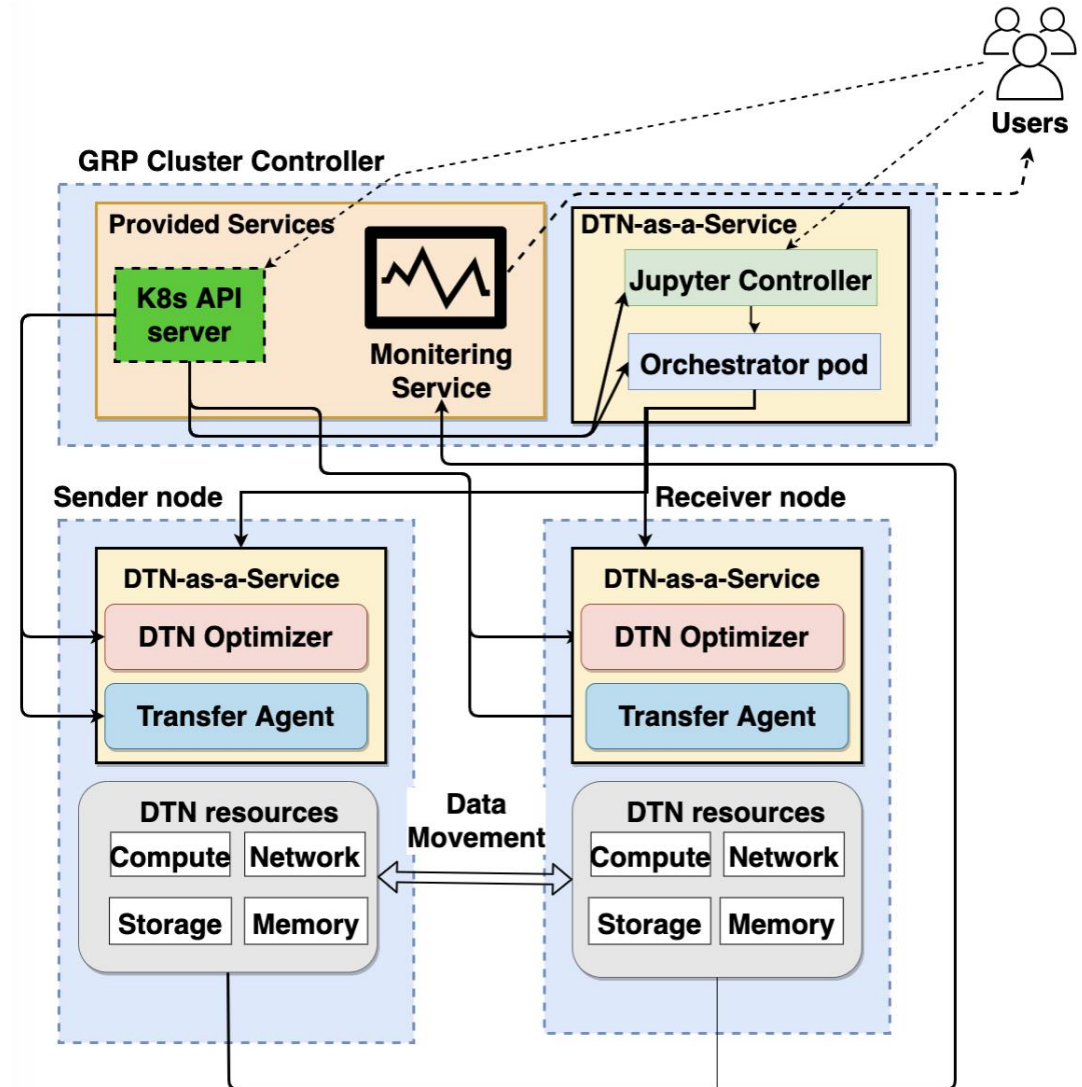
SAGE2 / SAGE3: Integrated Persistent Visualization and Collaboration Services for Global Cyberinfrastructure

International science communities use SAGE2 (and, soon, newly announced SAGE3) to share information, reach conclusions and make decisions with greater speed, accuracy, comprehension and confidence.

SAGE2 is a user-centered platform enabling small groups or large distributed teams to access digital media datasets from various sources and display, juxtapose, share and investigate a variety of related, high-resolution information on large-scale display walls.

NSF #OAC-1441963, #OAC-2003800

SAGE, SAGE2 and SAGE3 are trademarks of University of Illinois Board of Trustees
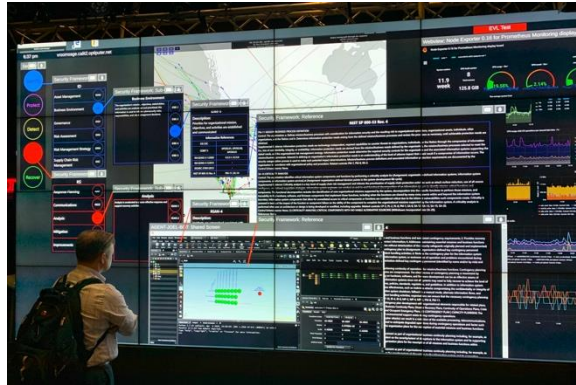
USA Univ. of California San Diego, Qualcomm Institute – Calit2

USA Argonne National Laboratory, Leadership Computing Facility

JAPAN AIST (Nat'l. Institute of Adv. Industrial Science and Technology), Cyber-Physical Cloud Research

POLAND University of Warsaw, Interdisciplinary Centre for Math & Computational Modelling

KOREA KISTI (Korea Institute of Science and Technology Information), KREONET Center

LAVA
Lab for Advanced Visualization & Applications
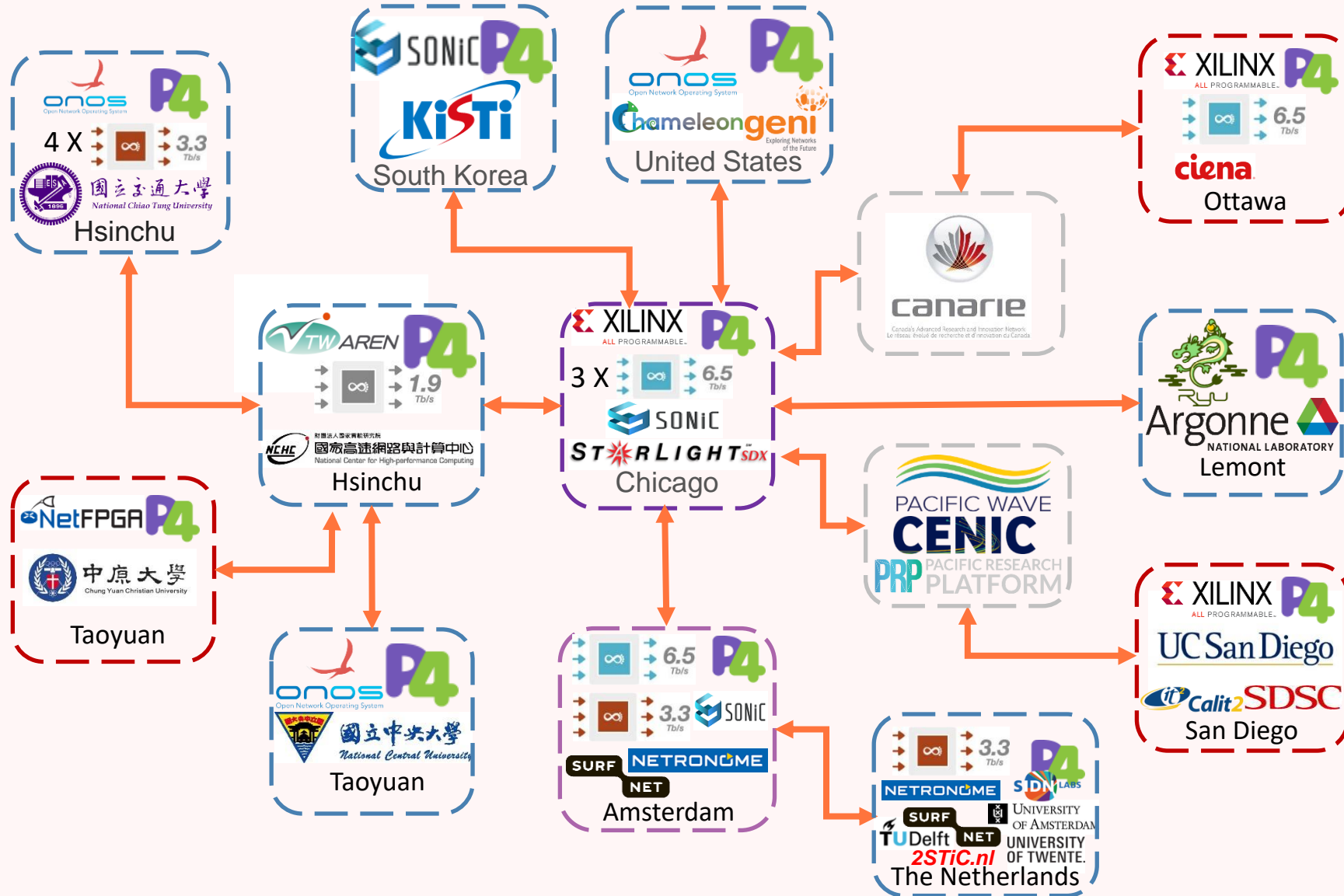**University of Hawai'i at Mānoa**
evl
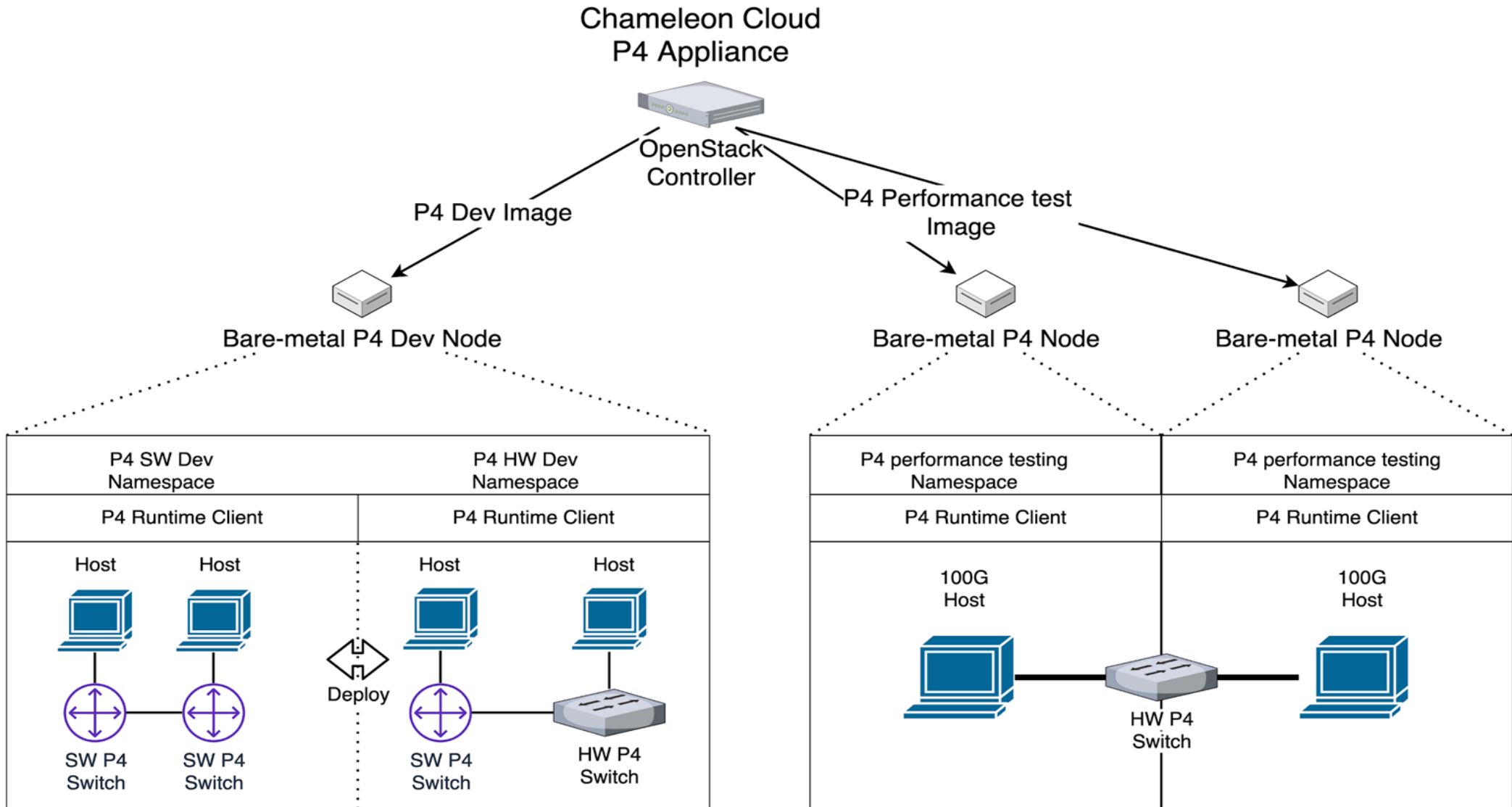Electronic Visualization Laboratory
University of Illinois at Chicago
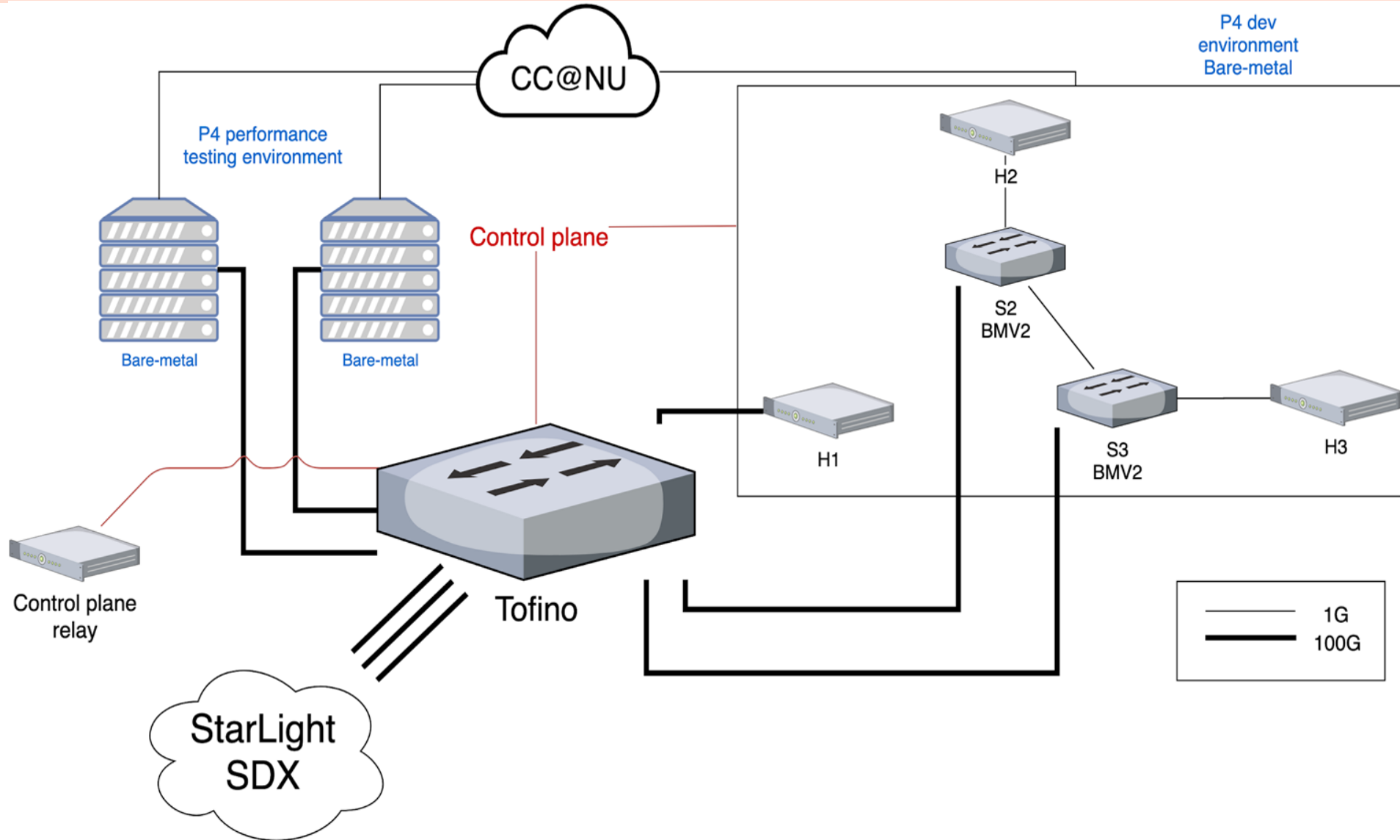VT
Virginia Tech
InfoVis Lab

www.sagecommons.org

Chameleon Cloud P4 Appliance V1

Chameleon Cloud P4 Appliance V1 in StarLight

P4MT: Multi-Tenant Support Prototype for International P4 Testbed

National Chiao Tung University

- **Multiple Tenants** Support
  - Data Plane: **Traffic, Flow rule matching** are isolated for each Tenant
  - Control Plane: **Control message** verification, **Packet I/O** redirection. Based on **P4Runtime**.

- **Dynamic pipeline allocation** for tenants during runtime

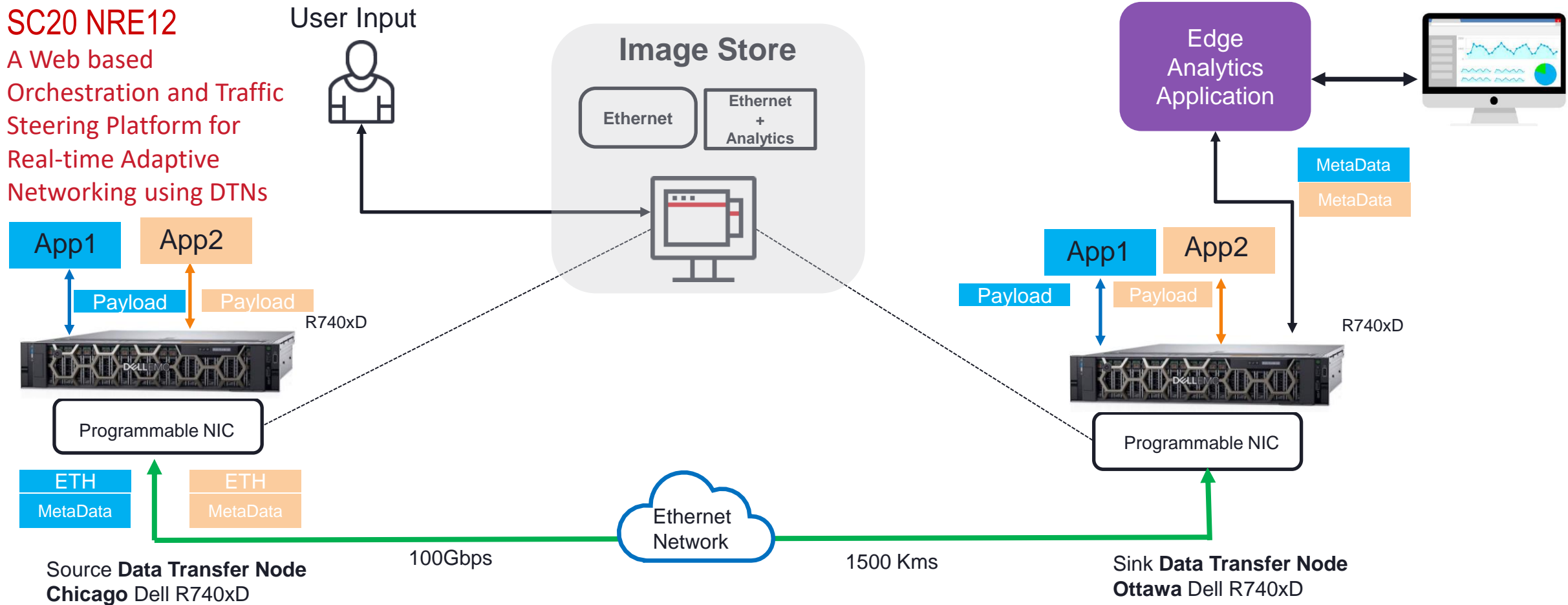- **Multiple packet process method** choosen by tenant (e.g. INT, L2 fwd)
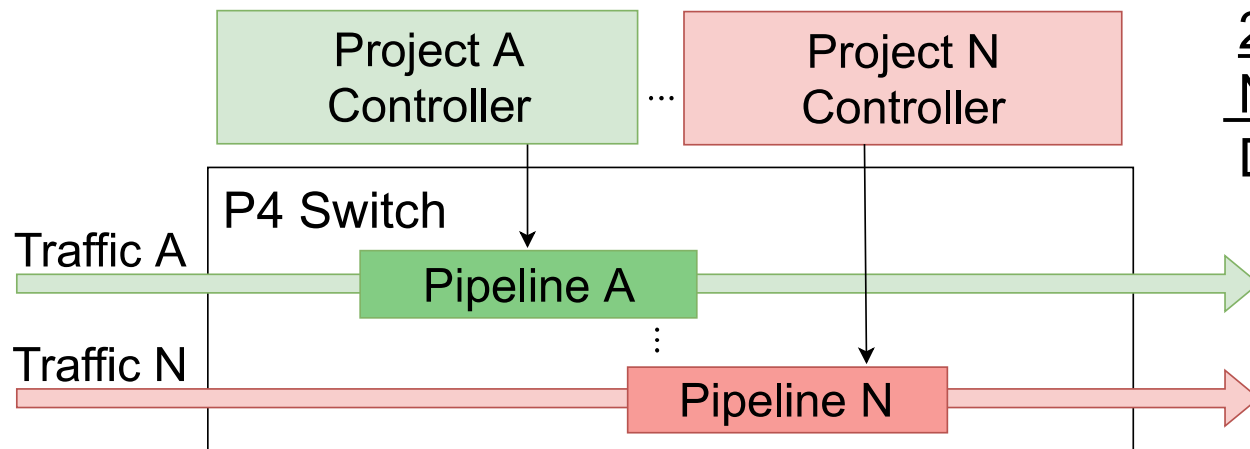
1. Two Data Transfer Nodes on CENI are fitted with Programmable NIC cards, capable at 100Gbps speeds.

2. User interacts with FPGA Image Store to enable In-band Telemetry on Xilinx NICs. The DTNs now act as INT Source and INT Sink nodes adding layers of Metadata into application packet headers.

3. An Edge Analytics application extracts metadata and hands off to visualization engine for live graphing and analysis of key parameters.

SC20 NRE12

A Web based Orchestration and Traffic Steering Platform for Real-time Adaptive Networking using DTNs

User Input

Image Store

Ethernet

Ethernet + Analytics

App1 App2

Payload Payload

R740xD

Edge Analytics Application

MetaData
MetaData

App1 App2

Payload Payload

R740xD

Programmable NIC

ETH ETH
MetaData MetaData

Programmable NIC

Ethernet Network

Source **Data Transfer Node Chicago** Dell R740xD

100Gbps

1500 Kms

Sink **Data Transfer Node Ottawa** Dell R740xD

Unique data package flow produces analytical MetaData, Packet level latency measurements @ 100Gbps

Real-Time DDoS Attack Detection using Sketch-based Entropy Estimation on the NetFPGA SUME

- High-speed DDoS attack traffic detection
  - Shannon Entropy estimation in real-time of selected network traffic headers
  - Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN)

- More detail: "Real-Time DDoS Attack Detection using Sketch-based Entropy Estimation on the NetFPGA SUME Platform" 12th APSIPA, Dec 7-10, 2020, Auckland, New Zealand
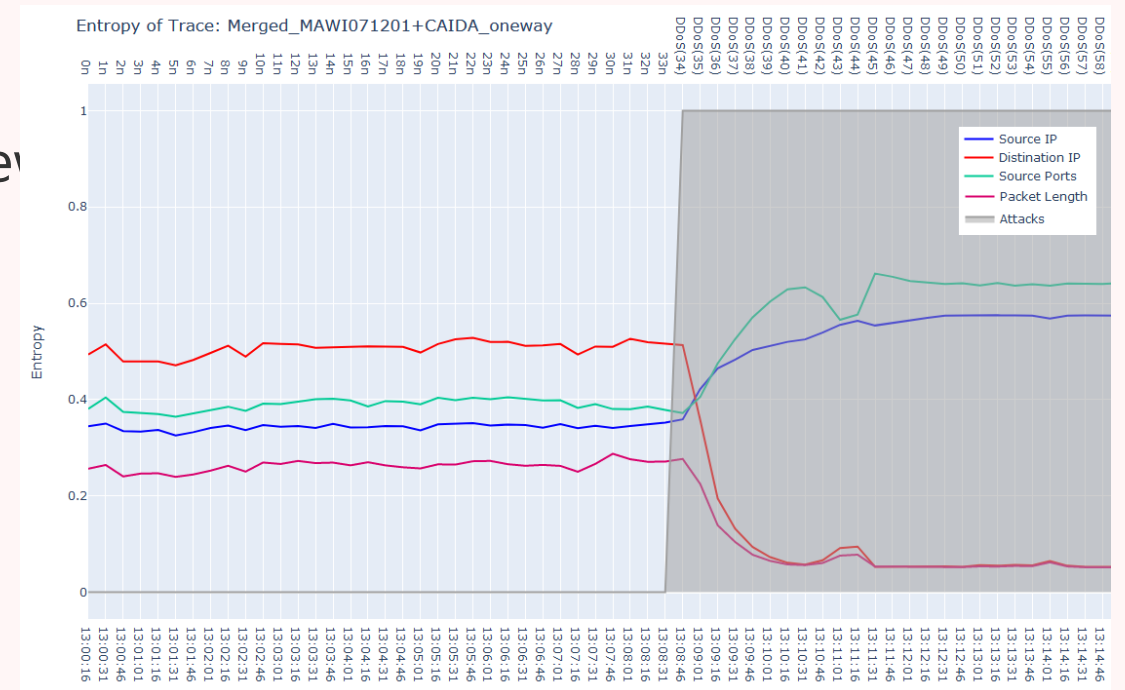
P4 INT Analyzer with Web UI

- An **INT Analyzer** is designed to monitor P4-enabled network
  - **DB Driver Layer** : read INT database supporting several formats (InfluxDB, Prometheus, ...)
  - **Analyze Layer** : parse/analyze data into JSON format
  - **UI Layer** : Configuration and Grafana visualization

- For future work, multi-domain INT analysis / visualization could be implemented for monitoring across P4-enabled NRENs

"網路遙測數據整合系統的設計/Design of an Integrated Analysis System for P4 In-Band Network Telemetry," TANET2020 Taipei, Taiwan, 10/2020
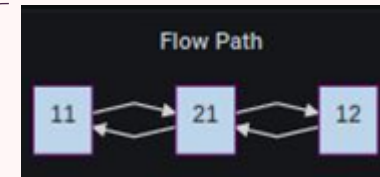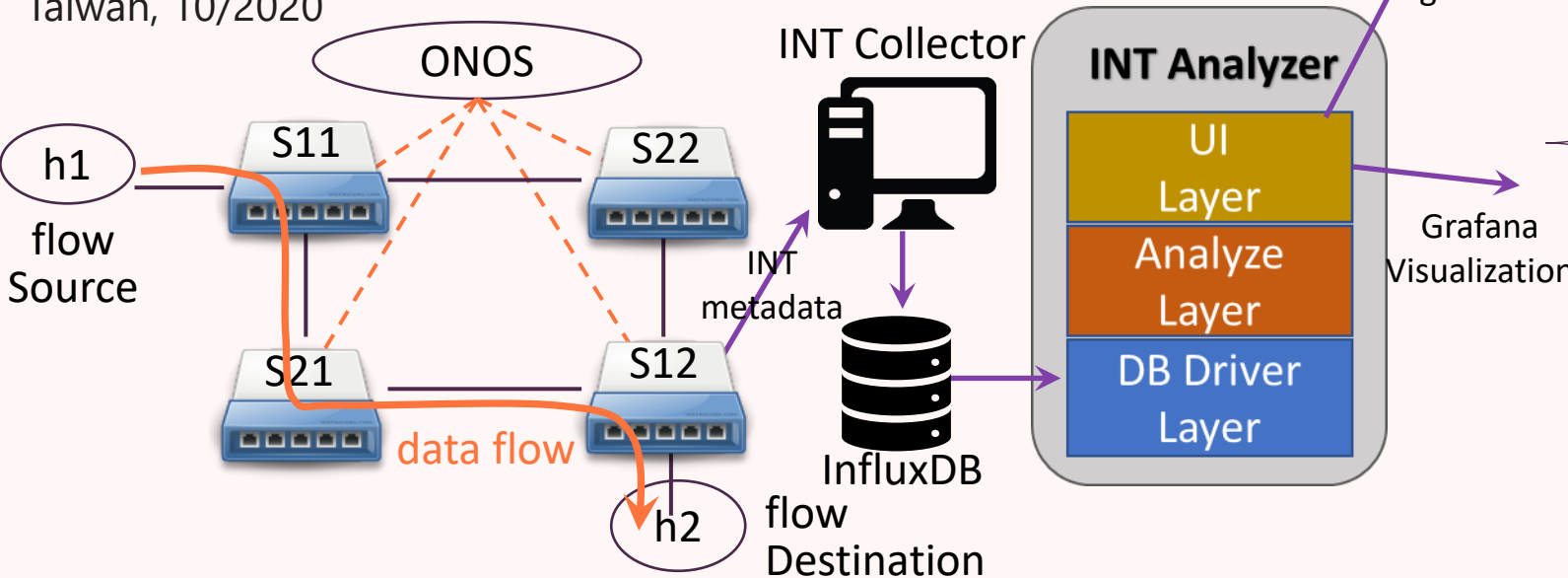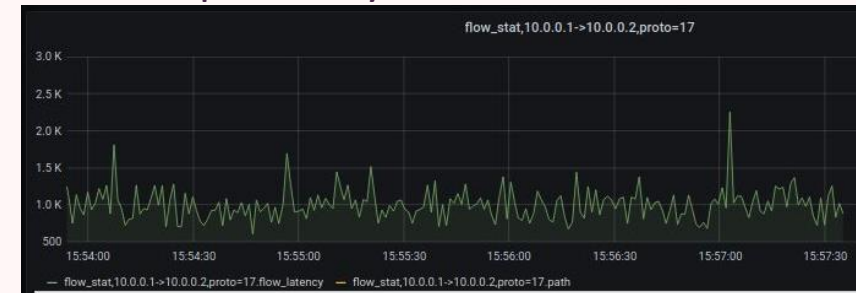

Flow Path Visualization
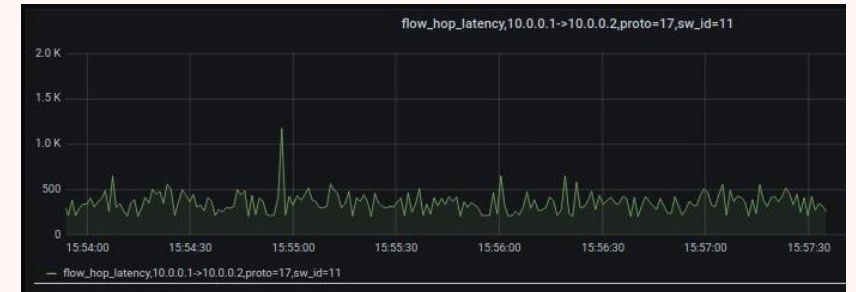
Flow Hop Latency

Flow Statistics

Thanks to the NSF, DOE, NIH, USGS, DARPA
NOAA, Universities, National Labs,
International Partners,
and Other Supporters

Jim-chen@northwestern.edu

Q & A