



FABRIC Research Infrastructure

Status, Features, Uses Cases

Tom Lehman
FABRIC Team

Presented at Joint Engineering Team (JET) Meeting
May 9, 2023



ESnet
ENERGY SCIENCES NETWORK

CLEMSON
UNIVERSITY



UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

What is FABRIC?

FABRIC enables a completely *new paradigm for distributed applications and Internet protocols and services:*

- A **nation-wide programmable network** testbed with **significant compute and storage at each node**, allowing users to run computationally intensive programs and applications and protocols to maintain a lot of information **in the network**.
- Provides **GPUs, FPGAs, and network processors (NICs)** inside the network.
- Supports **quality of service (QoS)** using dedicated optical 100G links or dedicated capacity
- **Interconnects national facilities:** HPC centers, cloud & wireless testbeds, commercial clouds, the Internet, and edge nodes at universities and labs.
- Allows you to design and test **applications, protocols and services that run at any node in the network**, not just the edge or cloud.

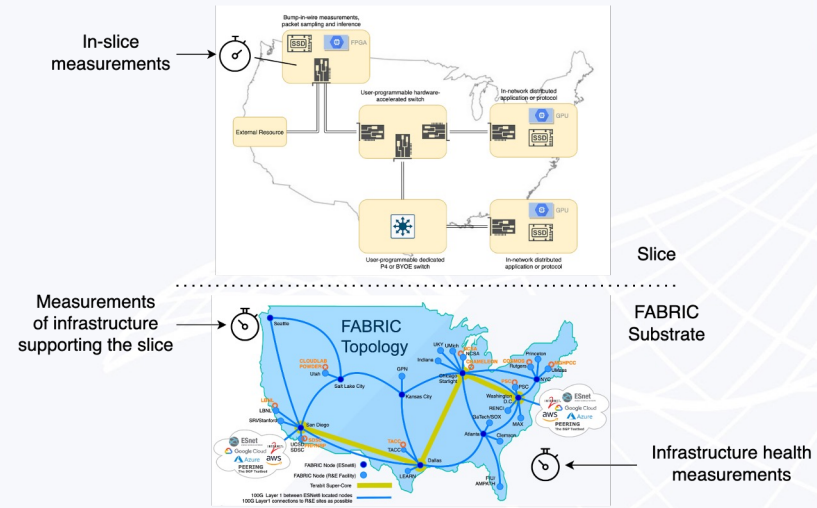
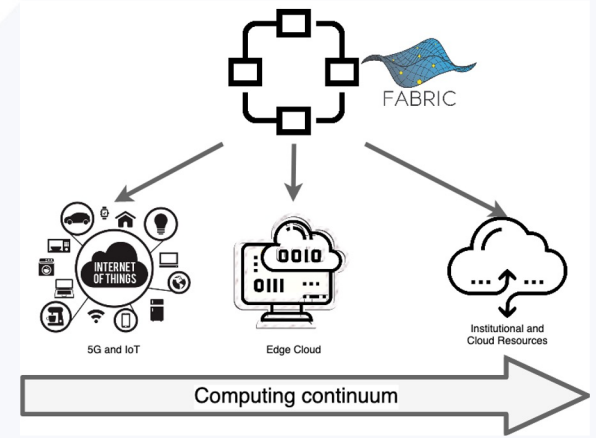


Why FABRIC?

- The mantra of the last 20 years – ‘Internet is showing its age.’
 - Applications designed around discrete points in the solution space
 - Inability to program the core of the network
- What changed?
 - Cheap compute/storage that can be put directly in the network
 - Multiple established methods of programmability (OpenFlow, P4, eBPF, DPDK, BGP flowspec)
 - Advances in Machine Learning/AI
 - Emergence of 5G, IoT, various flavors of cloud technologies
- Opportunity for the community to push the boundaries of distributed, stateful, ‘everywhere’ programmable infrastructure
 - More control or dataplane state, or some combination? Multiple architectures (co)exist in this space.
 - Network as a big-data instrument? Autonomous network control?
 - New protocols and applications that program the network?
 - Security as an integral component

Key FABRIC features

- Network as part of computing continuum
 - ‘Everywhere-programmable’ using different abstractions (P4, OpenFlow, others)
 - Diverse compute, storage capabilities in places where routers typically reside today
 - Dedicated 100G optical links between many sites
 - Support new paradigms in network aware applications and protocols
 - Ability to peer with Internet
- Network as a scientific instrument
 - Pervasive measurement collection capabilities in- and outside the slice available to researchers
 - GPS-disciplined PTP clock sources at every site
- Serve a broad range of scientific domains and applications
 - Concerned with data transport for big-data science, cybersecurity, terrestrial and 5G hybrid network architectures, federated ML/AI, Internet measurements and many more



FABRIC Edge



UMass Amherst

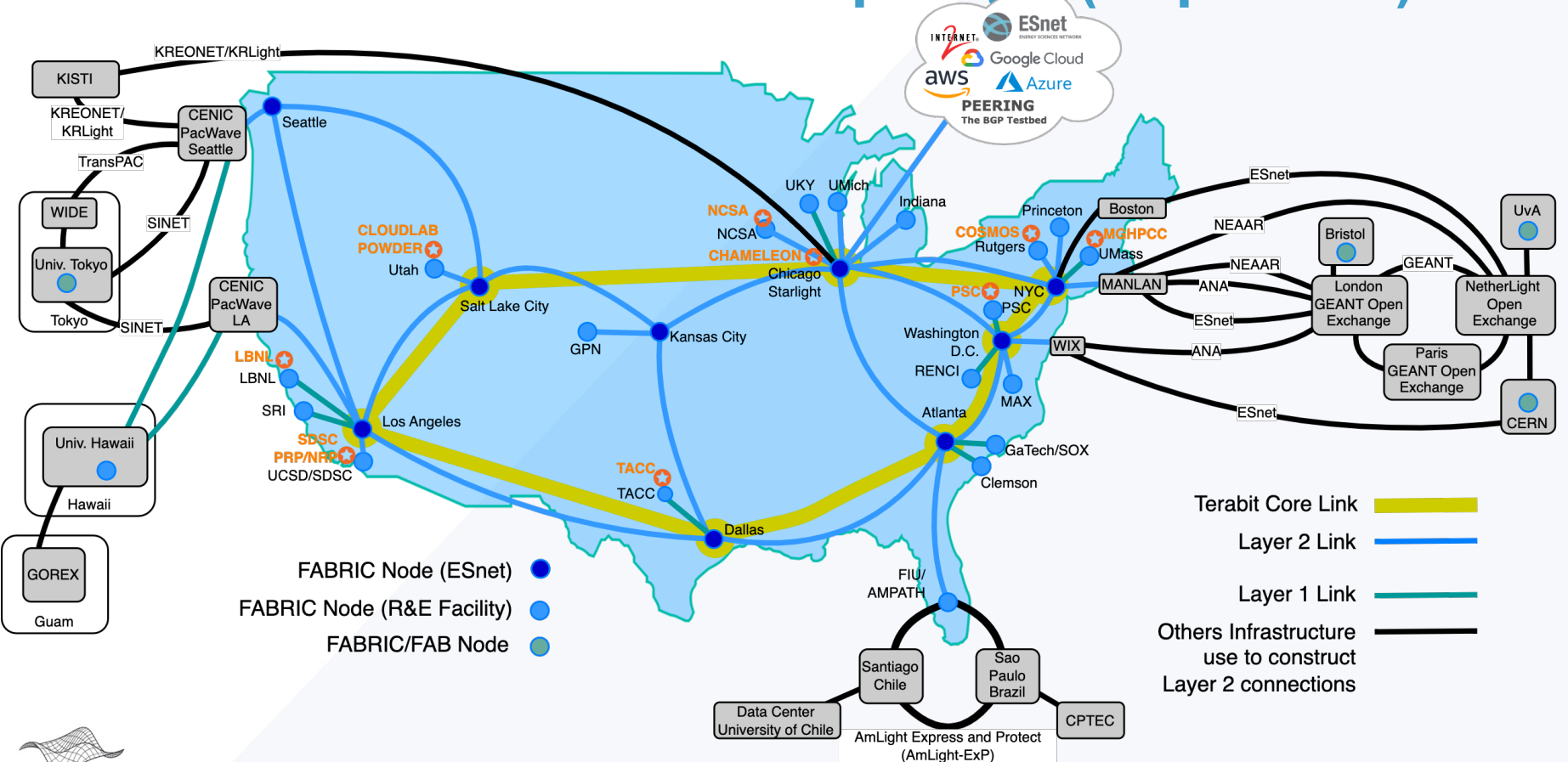


RUTGERS



PRINCETON UNIVERSITY

FABRIC - Production Topology (Sept 2023)



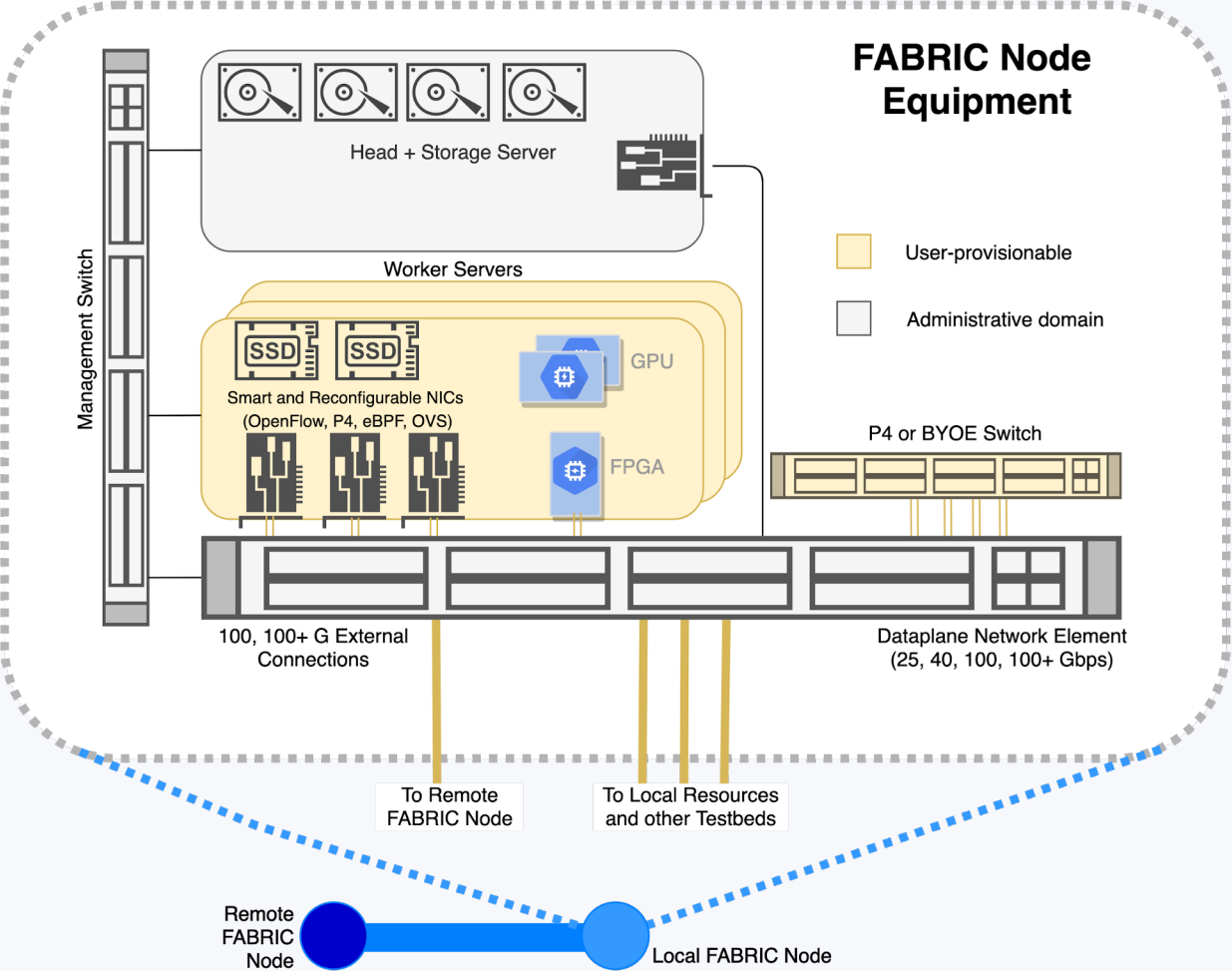
Overview

- 33 FABRIC Nodes
- 9 nodes co-located at ESnet6 Points of Presence
 - Connected via dedicated 100 Gbps DWDM across the ESnet6 openline optical system; some sites to be upgraded to Terabit SuperCore soon
- 20 other nodes distributed across the R&E community at various regional networks, major cyberinfrastructure facilities, and university hosting sites
 - many connected via 100 Gbps Layer 1
 - 4 Nodes deployed at International Locations (CERN, University of Amsterdam, University of Bristol, University of Tokyo)

FABRIC Nodes - Network, Compute, Storage

- Interpose compute and storage into the path of fast packet flows
- Rack of high-performance servers (Dell 7525) with:
 - 2x32-core AMD Rome and Milan with 512G RAM
 - GPUs (NVIDIA RTX 6000, T4, A30), FPGA network/compute accelerators
 - Storage - experimenter provisionable 1TB NVMe drives in servers and a pool of ~250TB rotating storage at each site.
 - Network ports connect to a 100G+ switch, programmable through control software
 - Tofino-based P4 switches (4 or more sites)
- Reconfigurable Network Interface Cards
 - FPGAs (U280 XILINX with P4 support)
 - Mellanox ConnectX-5 and ConnectX-6 with hardware off-load
 - Multiple interface speeds (25G, 100G, 200G+(future))
- Kernel Bypass/Hardware Offload
 - VMs sized to support full-rate DPDK for access to Programmable NICs, FPGA, and GPU resources via PCI pass-through

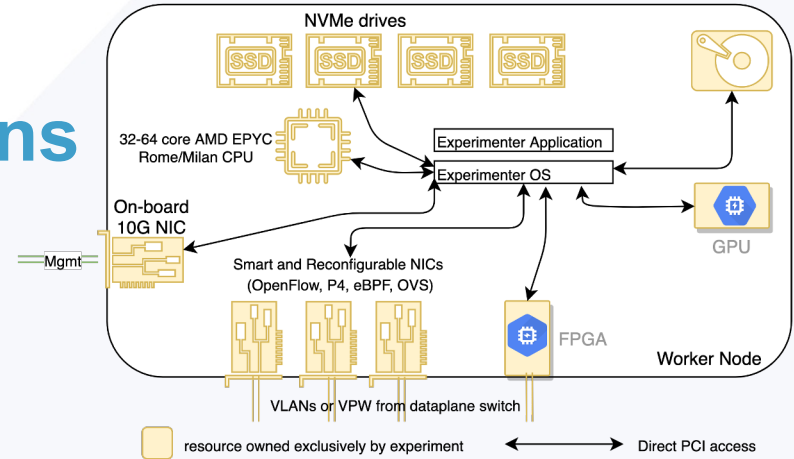
FABRIC Node 'Hank' Overview



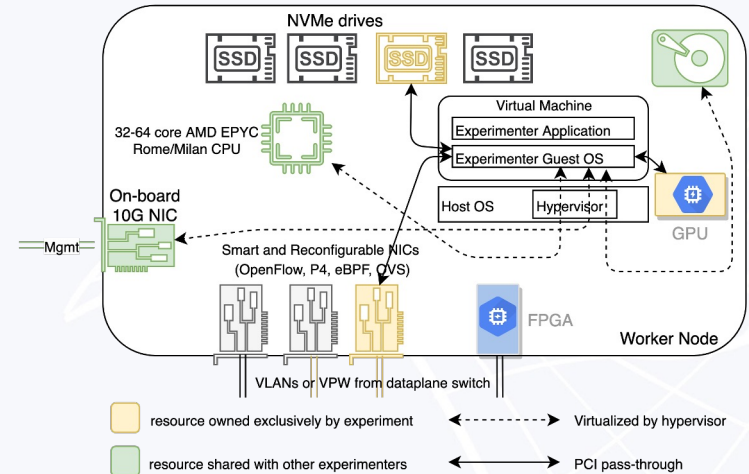
Node Level Programmability Abstractions

- Main capabilities are various PCI cards in individual servers
 - NICs, GPUs, FPGAs
- Additional switches and BYOE hardware
- Depending on experimenter request can be provided as part of a bare metal server or via PCI pass-through for VMs and containers

FABRIC experiment using a bare-metal server



FABRIC experiment using a virtual machine



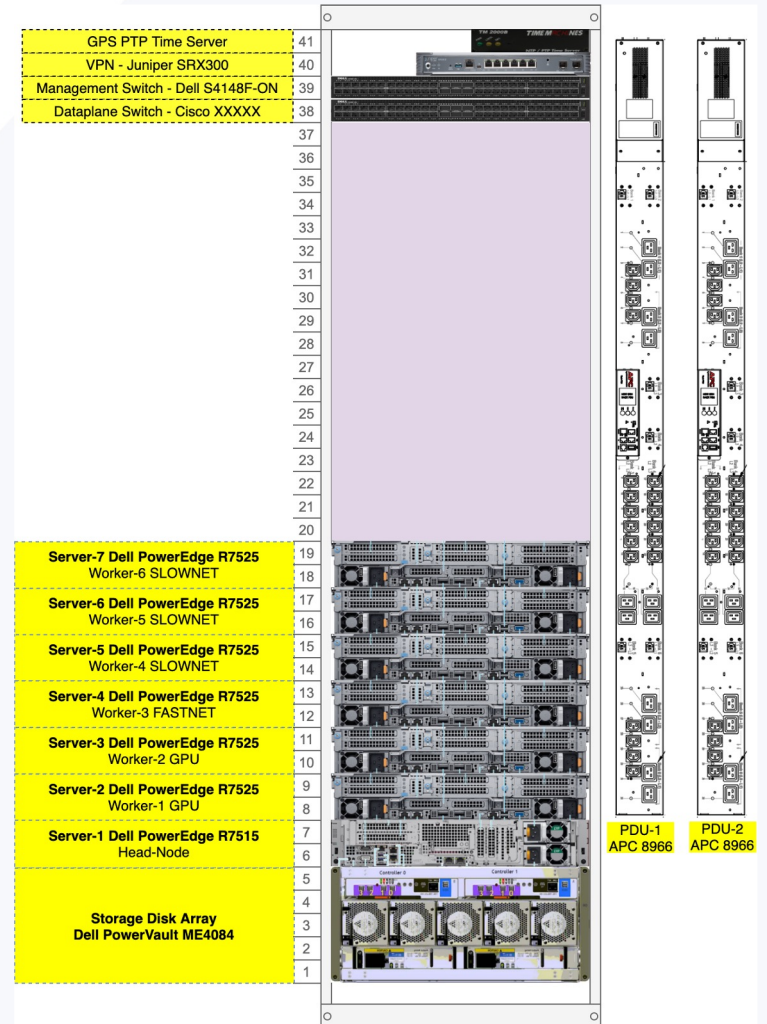
FABRIC Network Services

- A rich set of L2, L3 and specialized services to aid the experimenters
- L2 services provide Ethernet service between experimenter topology interfaces
 - **L2Bridge** - local to individual site
 - **L2STS** - two sites, any number of interfaces
 - **L2PTP** - two sites, two interfaces, QoS guarantees
- L3 services provide IPv4 and IPv6 services with an option to connect to the outside world
 - **FABNetv4** - FABRIC-routed IPv4 service, using RFC1918 addresses
 - **FABNetv6** - FABRIC-routed IPv6 services using FABRIC's IPv6 allocation
 - Both have externally-connectable counterparts (IPv4 variant uses FABRIC's limited IPv4 allocation)
- Specialized services
 - **Port mirroring** - currently mirrors a specific physical port in the topology



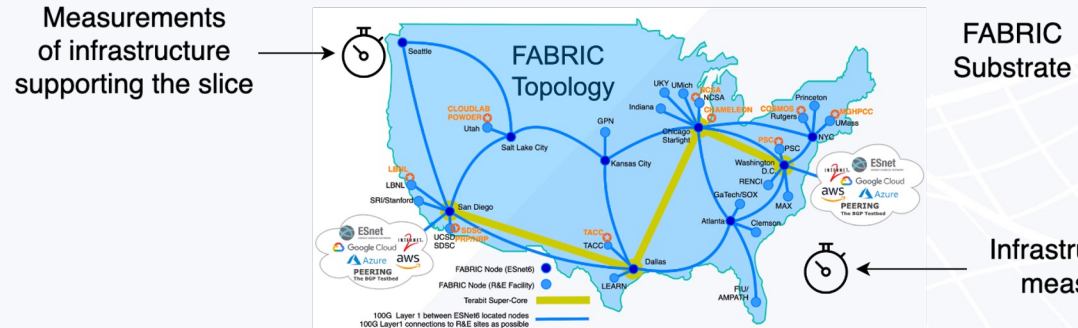
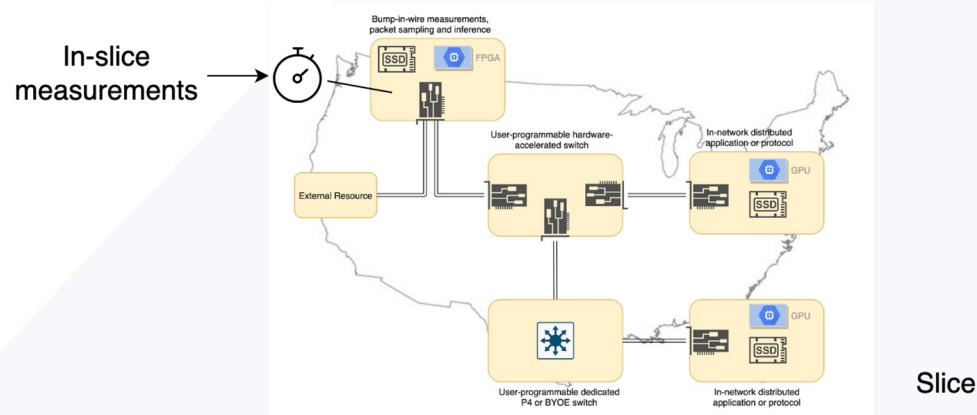
FABRIC Rack Configuration

This is an example FABRIC Rack Configuration. There are multiple configurations which vary the number and type of compute and storage elements.



FABRIC Measurement Capabilities

- Key to FABRIC being a scientific instrument
- Provides measurements
 - Inside the slice
 - Outside the slice



FABRIC Node Design: Measurement Hardware

- GPS-disciplined clock source with PTP at most sites
 - Subject to constraints of the hosting site
- NICs capable of accurate packet sampling/timestamping
- Programmable port mirroring
- Smart PDUs to measure power
- Optical layer measurements (where available)
- CPU, memory, disk, port/interface utilization and other time-series (software)

FABRIC features

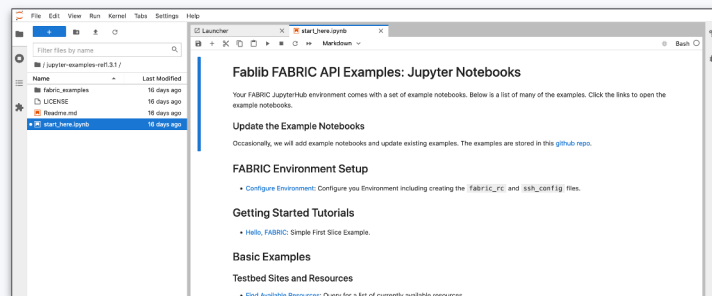
- **Facility Ports** - ability to add external facilities to slices using on-demand L2 connections
- **Mirror Ports** - ability to mirror traffic from the dataplane switch into slice
- Support for **on-demand public connectivity**
 - Slices with L3 IPv4 or IPv6 can connect on-demand with public internet
 - This is in testing
- **L3VPN service + CloudConnect**
- **Persistent storage** for slices - get storage allocation at multiple sites for your project
- **In-slice measurement framework** - instrumentize your slice to get data about its performance
- Support for P4 Tofino switches in topologies (future)
- Support for P4 workflows on top of U280 FPGAs (in collaborations with OCT/NorthEastern and ESnet) (future)
- Inter-testbed federation features, more Facility Ports

FABRIC External Connections Overview

- FABRIC experiments (slices) can run in an isolated manner within FABRIC Infrastructure, and isolated from external networks.
- Slices can also utilize FABRIC's external connections to access a variety of external experimental and production resources.
- These external connections are organized as follows:
 - Layer 3 IPv4 and IPv6 public connectivity on demand (with policy restrictions) via peering points provided by ESnet and Internet2
 - Layer 2 Services Peering (aka Facility ports)
 - Public Cloud Connections via CloudConnect
- Status:
 - Facility ports to
 - Chameleon Chicago and TACC **operational**
 - CloudLab and Powder at Utah **operational**
 - Cryo-EM facility at Rutgers - **operational**
 - CloudLab Clemson and Wisconsin - being configured
 - ASU/Sun Corridor Network - being configured

Interacting with FABRIC

- FABRIC has three ways to interact with it:
 - Portal - focus on creating small topologies
 - Jupyter Hub notebooks via API - don't just create topologies - run experiments
 - API libraries from your laptop/desktop
- We recommend the notebooks as a starting point
- Everything continues to be in active development
 - We keep adding infrastructure
 - We keep improving all of the software
- Infrastructure is updated as it becomes available
 - Generally non-disruptive
- Software updates are on a 3-4 month cadence between releases
 - Backward compatibility is not always assured, but we try



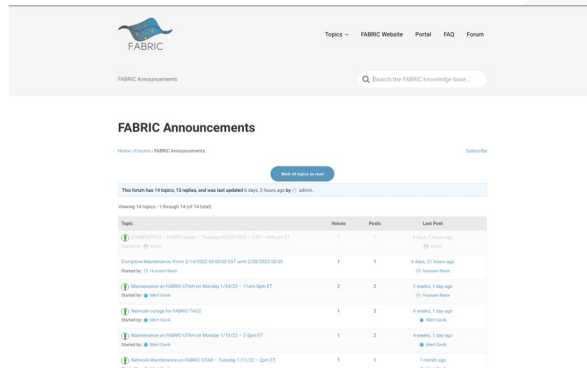
Interacting with FABRIC

- REST-based APIs and a Python library (FABlib) on top to make your experiments easier to run
- FABRIC does not use SSL certificates - instead it uses short-lived API tokens (1hr) that can be refreshed
 - There is a *Credential Manager* service (API and UI front end) to help do that - it requires you to login using your institutional credentials to generate a token
 - Jupyter Notebooks do this automatically - get a new token and refresh as needed
- There is an equivalent of 'RSpec' for topology descriptions based on property graph abstraction, serializable to GraphML and JSON.
 - You can save your topology description
 - Not meant to be human editable - manipulate it using API or Portal
 - Portal does not yet allow you to download topology description, this feature is coming.
 - Topology != Experiment
- Our goal is to make Jupyter Notebooks your reproducible and shareable experiment profile
 - Build a topology (even across multiple testbeds!)
 - Configure it
 - Run the experiment
 - Annotate
 - Display plots and graphs
 - Share and let others reproduce your results
 - All from the comfort of your browser!

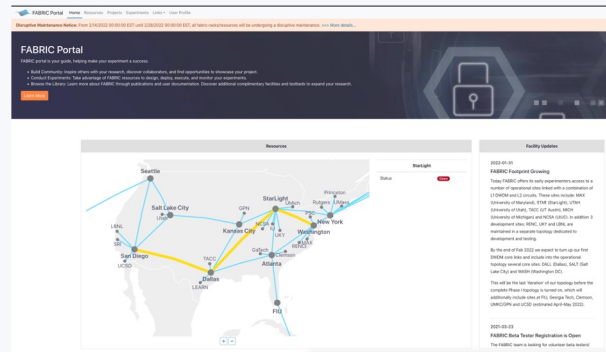
FABRIC UI

- FABRIC Portal, Jupyter Hub, User Knowledge base all operational
- Integrated with CI Logon for federated identity
- Accessible and used by early experimenters

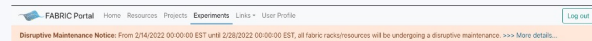
FABRIC Knowledge Base with User Forums



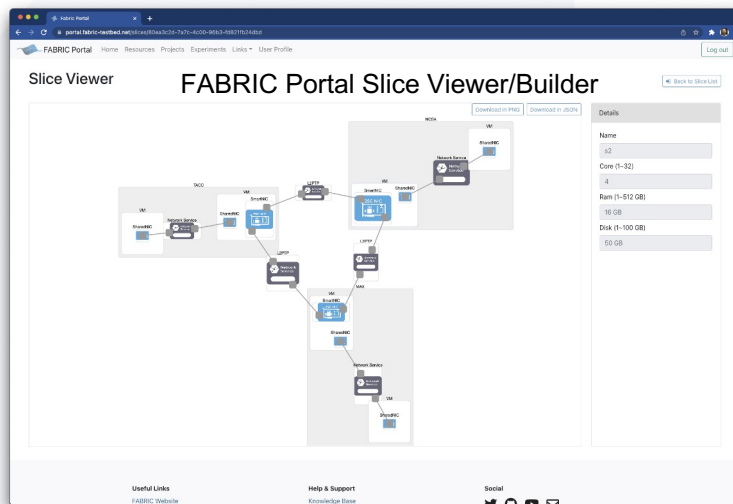
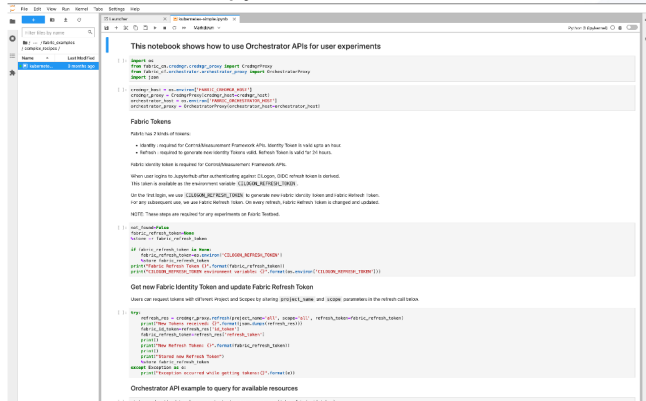
FABRIC Portal Home Page



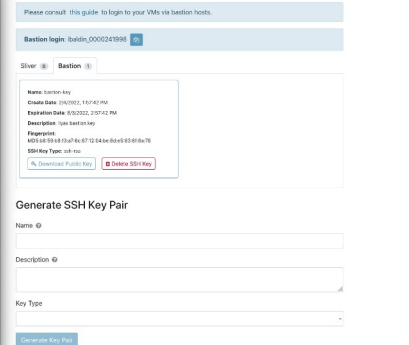
FABRIC Portal Experiments page



FABRIC Jupyter Hub

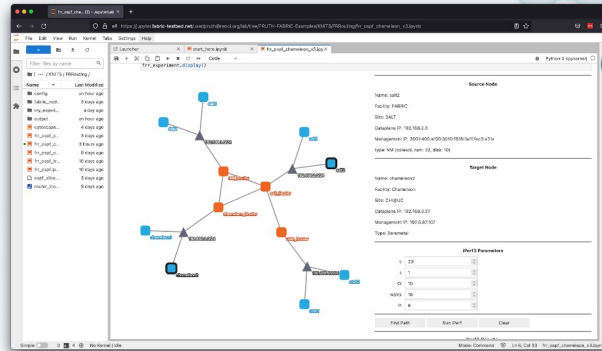


SSH Keys



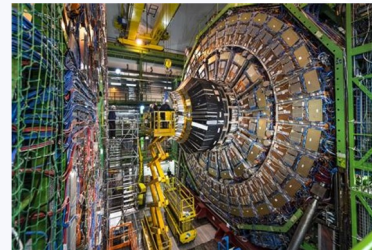
Early Experiments

- Users: ~300
- Projects: 40 (36 research, 4 education)
- Project topics:
 - P4/SDN
 - Honeypots
 - Named Data Networking (NDN)
 - ServiceX
 - BGP Peering
 - Datacenter protocols
 - Scalable Genome Analysis
 - Fast data transfers
 - Internet Privacy



Additional Science Use Cases & Partners

- Astronomy (Vera Rubin Observatory/LSST, Chile)
- Cosmology (CMB-S4)
- Weather (UMiami & CPTEC, Brazil) - Ben Kirtman, Atmospheric Science & Paolo Nobre
- High-Energy Physics (CERN) - Rob Gardner, **FAB Co-PI**, Physicist
- Urban Sensing/IoT/AI at Edge (UBristol) - Dimitra Simeonidou, Prof. of Networking
- 5G across borders, P4/SDN - (UTokyo) Aki Nakao, Prof. of CS; KISTI (Korea Institute of Science and Technology Information)
- Censorship Evasion - Richard Brooks, Prof. of ECE



BRISTOL IS OPEN
open programmable city region



Thank You!

Questions?

Visit <https://whatisfabric.net>

Ask info@fabric-testbed.net

FABRIC Software: <https://github.com/fabric-testbed>

Beta Tester request form: <https://whatisfabric.net/get-involved/beta-testers-request>



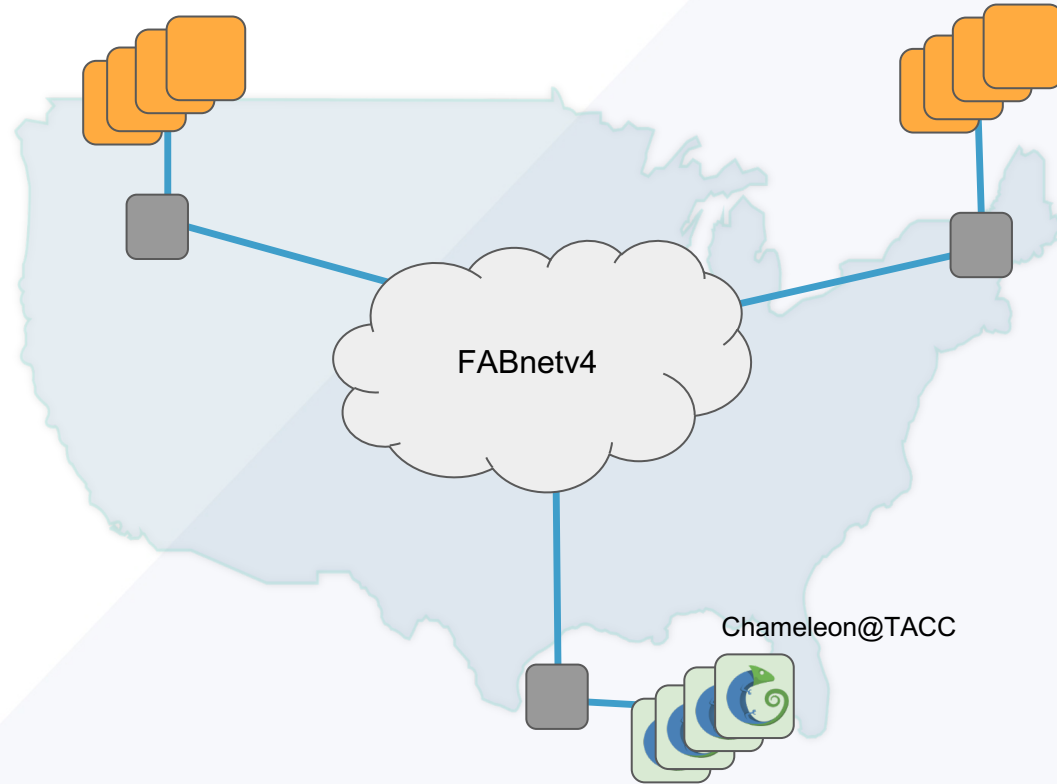
This work is funded by NSF grants CNS-1935966, CNS-2029261, CNS-2029235, CNS-2029200, CNS-2029261, CNS-2029260



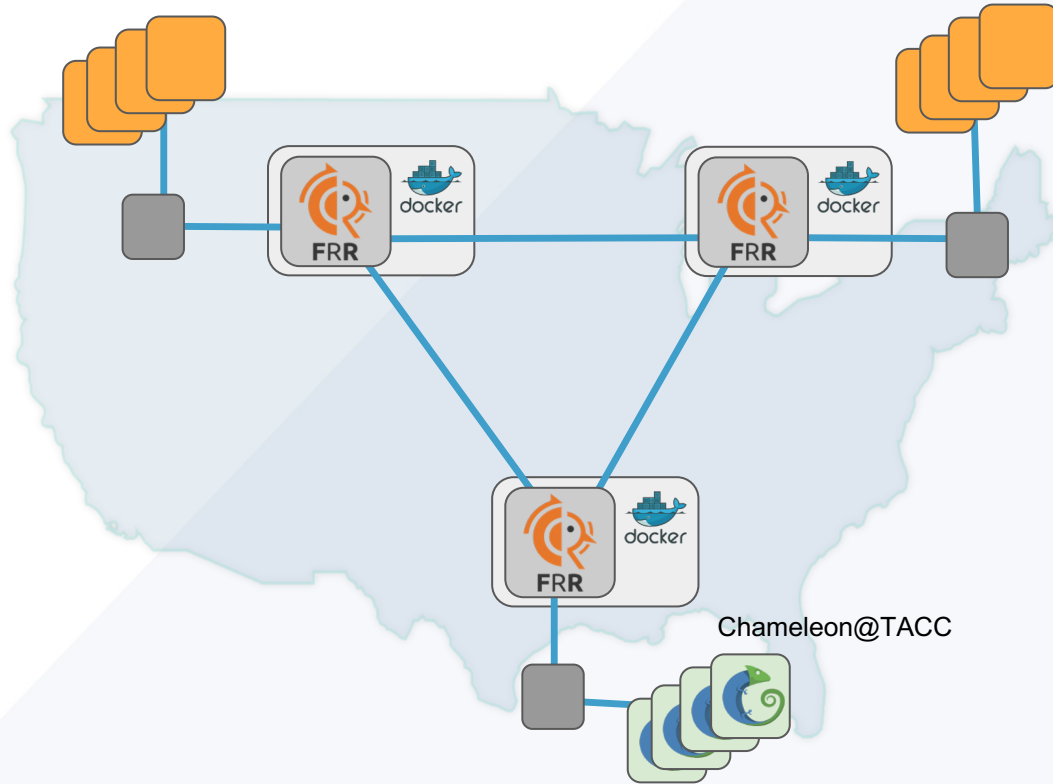
Extra Slides



FABRIC Experiments - Chameleon

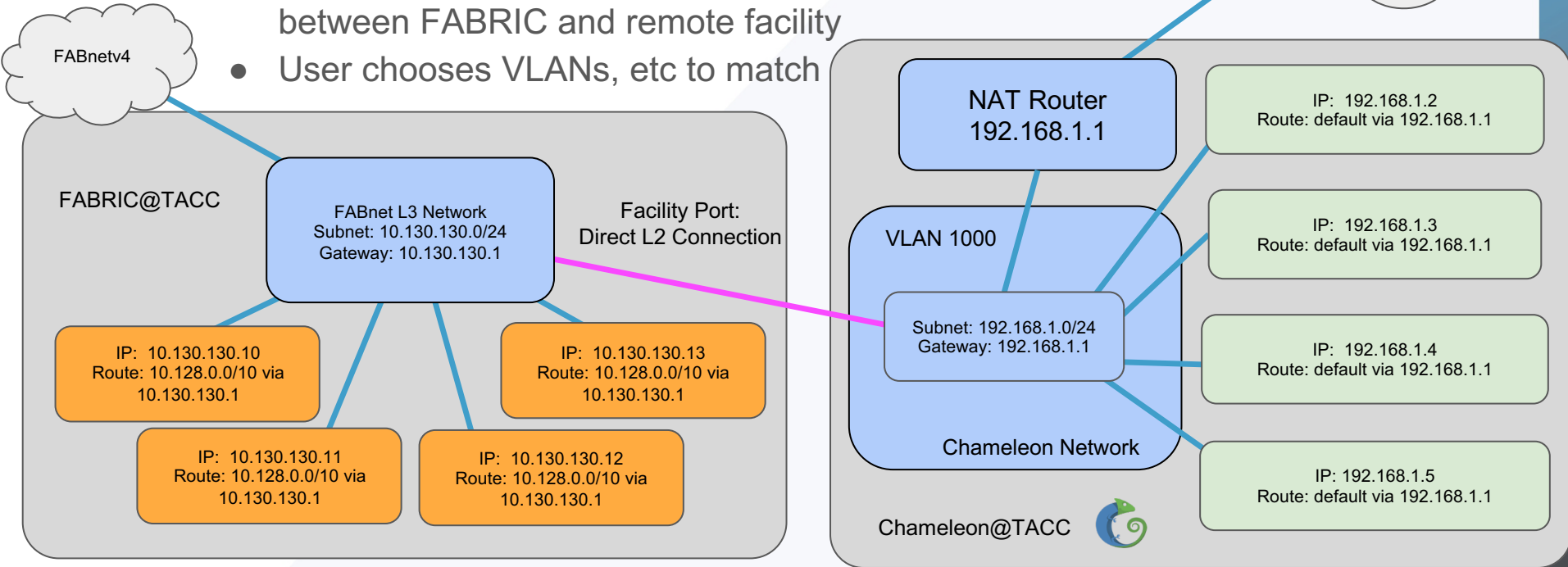


FABRIC Experiments - Chameleon



FABRIC: Chameleon Facility Ports

- Direct L2 Connection (VLAN) between FABRIC and remote facility
- User chooses VLANs, etc to match

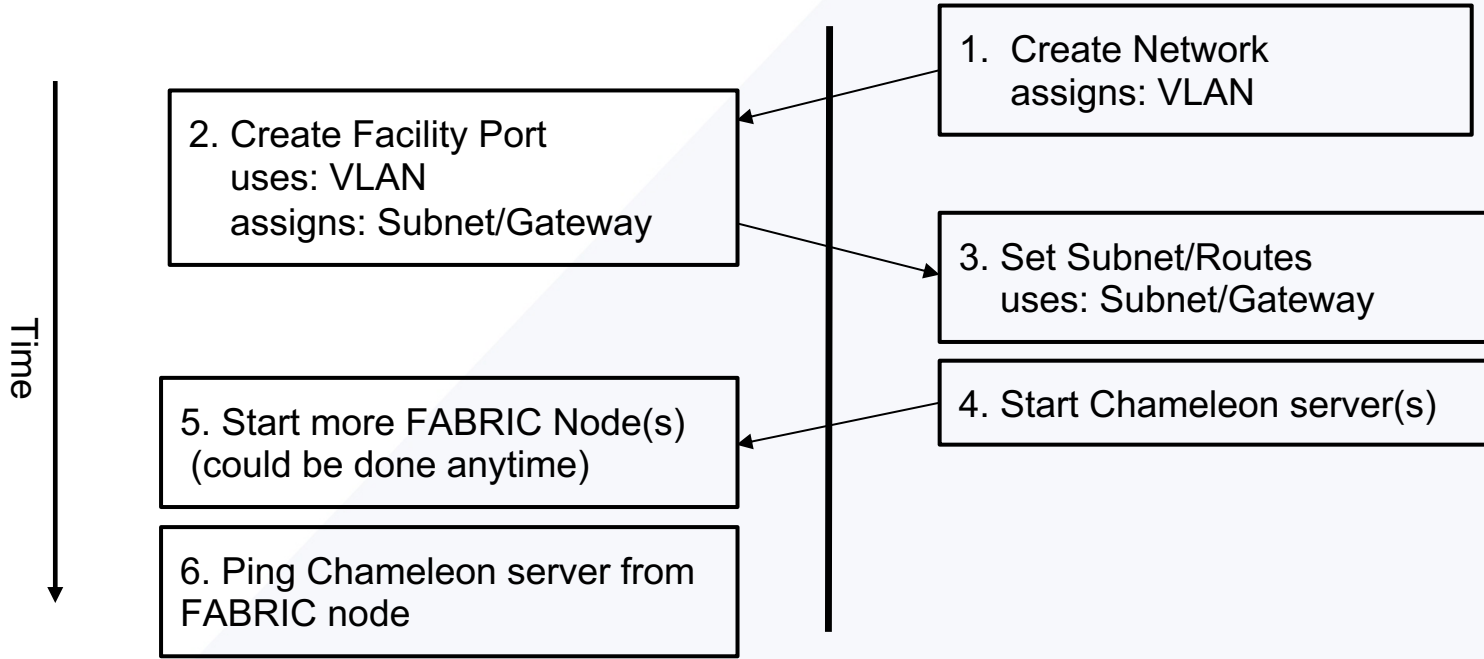


What do we do about the IPs/subnets/routes?

FABRIC: Chameleon Facility Ports

FABRIC@TACC

Chameleon@TACC



What do we do about the IPs/subnets/routes?

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development
(NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674,
Fax: 202-459-9673, Email: nco@nitrd.gov, Website: <https://www.nitrd.gov>

