FABRIC: Tightly Weaving a Better Internet

Anita Nikolich
University of Illinois, Urbana Champaign
&
Cyber Policy Initiative, Harris School, University of Chicago

Topics

- What is FABRIC?
- The Internet Is Awful (short version)
- Testbeds Are Awful
- Build a Better Internet using a Better Testbed!

FABRIC

What is FABRIC?

New ~~testbed~~ paradigm for distributed apps, Internet protocols and services:

- Nation-wide programmable network with significant compute and storage at each node. Run computationally intensive programs, applications and protocols to maintain a lot of information in the network.
- GPUs, FPGAs, and network processors (NICs) inside the network.
- Quality of service (QoS) using dedicated optical 100G links
- Interconnects national & International facilities and production: HPC centers, cloud & wireless testbeds, edge nodes, science instruments, Smart cities testbeds, production internet
- Design and test applications, protocols and services running in any node in the network, not just edge or cloud.

FABRIC

FABRIC for everyone

FABRIC Enables New Internet and Science Applications

- Stateful network architectures, distributed applications that directly program the network

FABRIC Advances Cybersecurity

- At-scale realistic research facilitated by peering with production networks

FABRIC Integrates HPC, Wireless, and IoT

- A diverse environment connecting PAWR testbeds, NSF Clouds, HPC centers and instruments

FABRIC Integrates Machine Learning & Artificial Intelligence

- Support for in-network GPU-accelerated data analysis and control

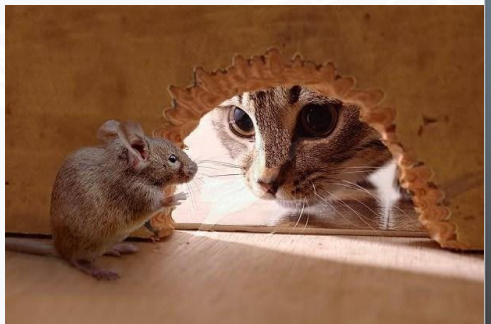
FABRIC helps train the next generation of computer science researchers

Why FABRIC?

- Cheaper compute and storage allows us to re-think the technical Internet
  - “If we had to build a router from scratch today it wouldn’t look like the routers we build today”
  - Explosion of capabilities in augmented computing - GPUs, FPGAs
  - Opportunity to reimagine network architecture as more stateful
  - What functions can we put in the core of the network?
- ML/AI is Everywhere Due to Data
  - Network as a ‘big-data’ instrument: real-time measurements + inferencing control loop
  - ~~Self-driving network~~
- IoT + 5G - new high-speed intelligent network edge
- New science applications and need to process quicker
  - New distributed applications - data distribution, computing, storage
- Continuum of computing capabilities
  - Not just fixed points - “edge” or “public cloud”
  - Network as part of the computing substrate - computing, fusing, processing data on the fly
- Too many testbeds!

Why a New Internet?

- Operators: Workarounds are time consuming and ineffective
- Academics: Proposed solutions often not feasible in real world
- Scientists: Deluge of instrument data for processing
- Industry: Hierarchical model of ‘ownership’ is inefficient
- Average Person: Usability, troubleshooting, access, free from tracking and surveillance



FABRIC

Technically Awful Aspects of the Internet

- Protocols are Fragile (i.e., BGP, DNS, http)
- Systems to workaround Protocols = difficult to implement (i.e. BGPSEC, RPKI, censorship bypassing)
- Troubleshooting workflow is complex
- Storage/Compute/Network model introduces errors
- Access to data is silo’d
- Who ‘does security’? Security – Network – Storage – etc often different teams
- Security research is hampered by CFAA (Computer Fraud and Abuse Act)
- Terrible UX

FABRIC

Designing for All Internet Users

- Activists/Journalists
- Censored Internet users
- Average Users: elderly, disabled, etc
- Underserved Areas

Curb the move towards Internet Balkanization/Splinternet

FABRIC

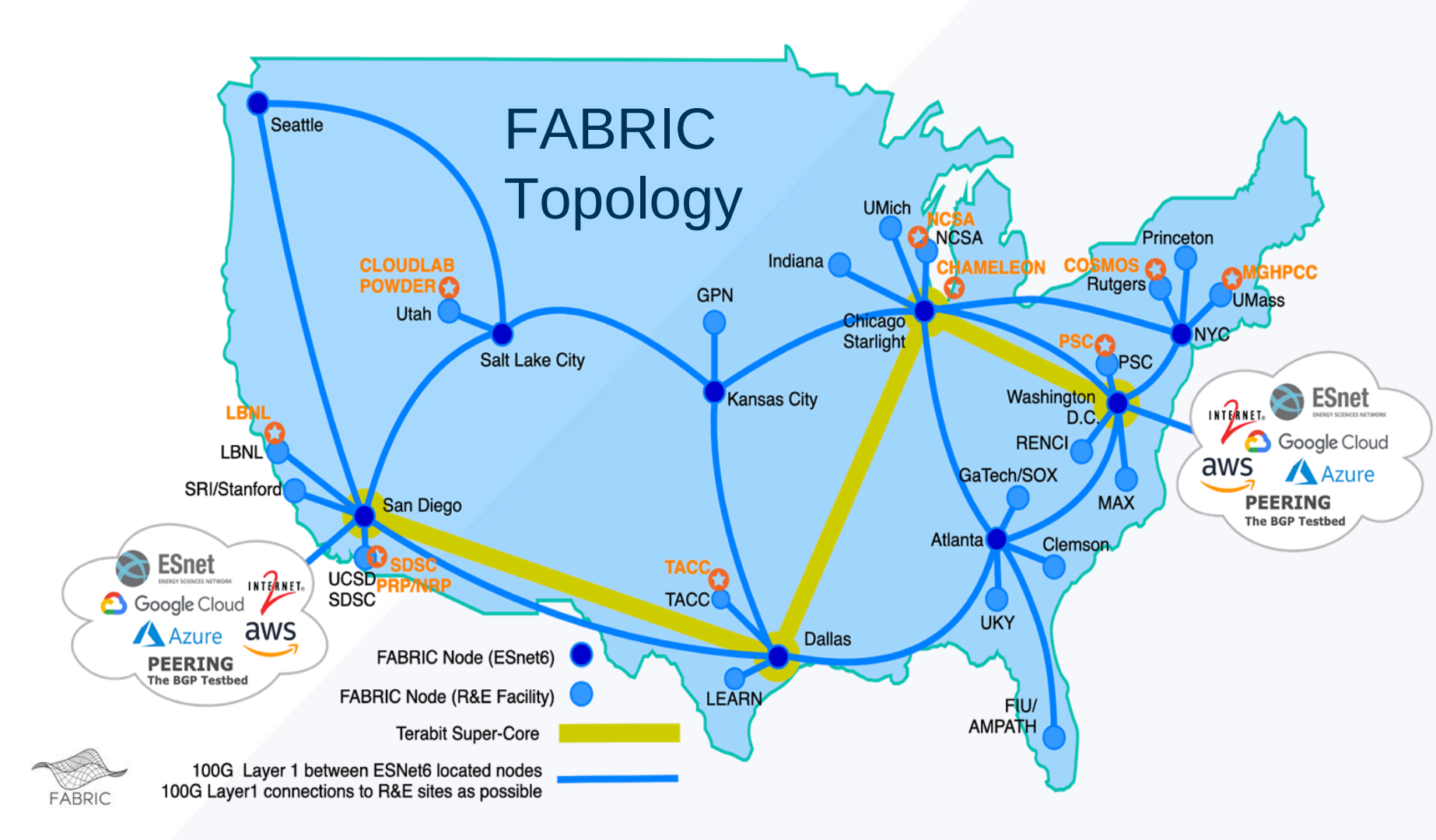
What testbed features will help researchers/industry develop the future Internet?

FABRIC

Basic Capabilities

- Widely distributed
- Highly programmable
- Multi-user/multi-tenant
- Suitable for education and research
- Core programmability
- Hardware support for programmable abstractions
- **Connect to the Internet and/or production facilities!**

FABRIC


FABRIC
Topology
Seattle
CLOUDLAB
POWDER
Utah
Salt Lake City
GPN
Kansas City
Indiana
UMich
NCSA
NCSA
CHAMELEON
Chicago
Starlight
Princeton
COSMOS
Rutgers
MGHPCC
UMass
NYC
PSC
PSC
Washington
D.C.
RENCI
GaTech/SOX
MAX
Atlanta
Clemson
UKY
FIU/
AMPATH
LBNL
LBNL
SRI/Stanford
San Diego
UCSD
SDSC
SDSC
PRP/NRP
TACC
TACC
Dallas
LEARN
ESnet
ENERGY SCIENCES NETWORK
INTERNET2
Google Cloud
Azure
aws
PEERING
The BGP Testbed
FABRIC Node (ESnet6)
FABRIC Node (R&E Facility)
Terabit Super-Core
100G Layer 1 between ESNet6 located nodes
100G Layer1 connections to R&E sites as possible
FABRIC

FABRIC Across Borders (FAB) – New!

- Expansion to:
  - Japan, University of Tokyo
  - UK, University of Bristol
  - University of Amsterdam
  - Geneva: CERN

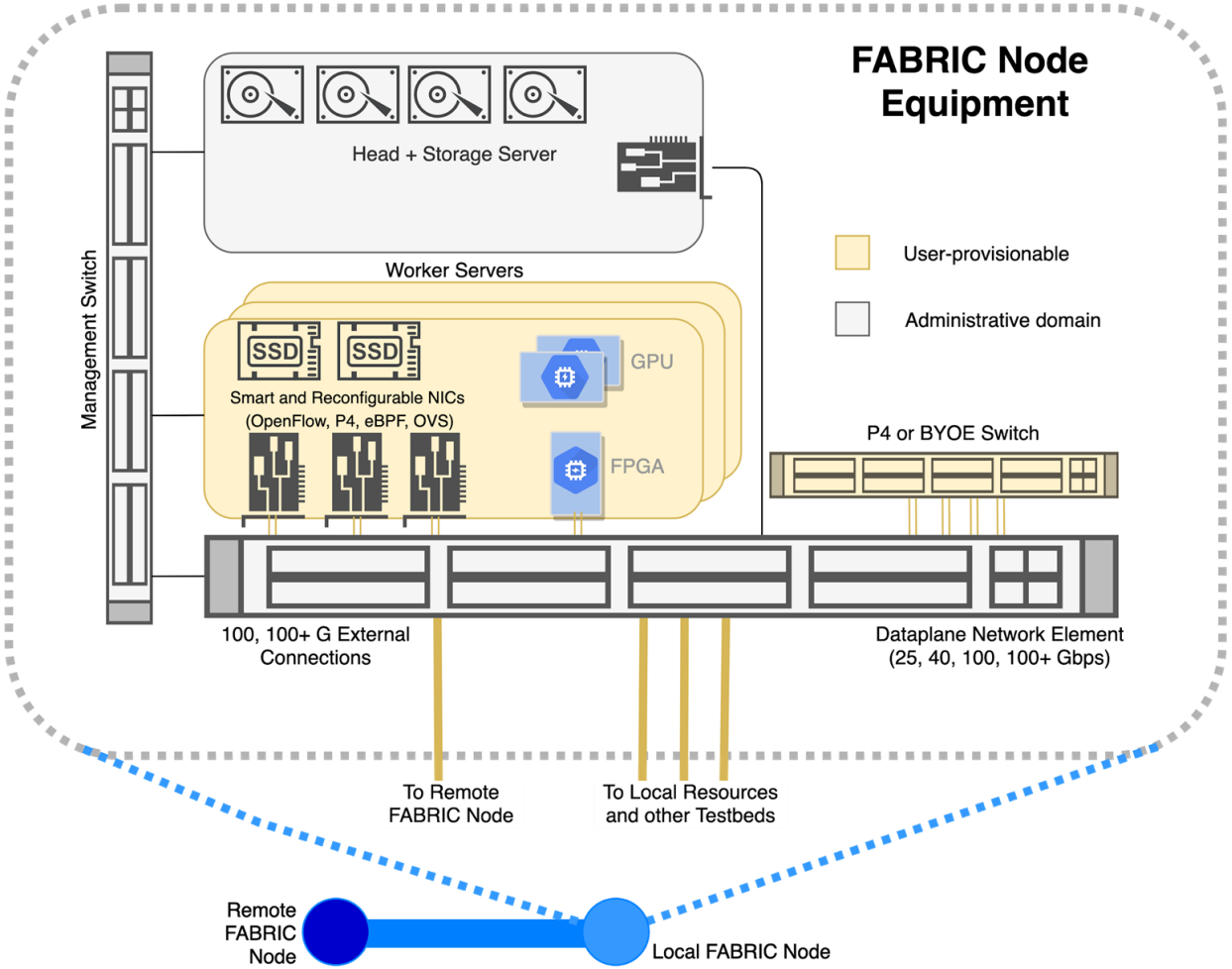
- Additional use-cases:
  - Astronomy/Cosmology (CMB-S4; LSST)
  - Weather
  - High-Energy Physics (LHC)
  - Smart Cities/IoT (Bristol Smart City Lab/Chicago Array of Things)
  - 5G across borders
  - Anti-Censorship protocol development

FABRIC Nodes

- Interpose compute and storage into the path of fast packet flows
- Rack of Dell 7525 servers:
  - 2x32-core AMD 7532 with 512G RAM
  - GPUs (RTX 6000 and T4), FPGA network/compute accelerators
  - Storage - 1TB drives in servers & pool of ~250TB rotating storage
  - Network ports connect to a 100G+ switch, programmable through control software
- Reconfigurable Network Interface Cards
  - FPGAs (with P4 support)
  - Multiple interface speeds (25G, 100G, 200G+(future))
- Kernel Bypass/Hardware Offload
  - VM/Containers support full-rate DPDK for access to Programmable NICs, FPGA, and GPU resources

Conceptual FABRIC Node ‘**Hank**’

‘disaggregated router’



FABRIC Node Design: Measurement Hardware

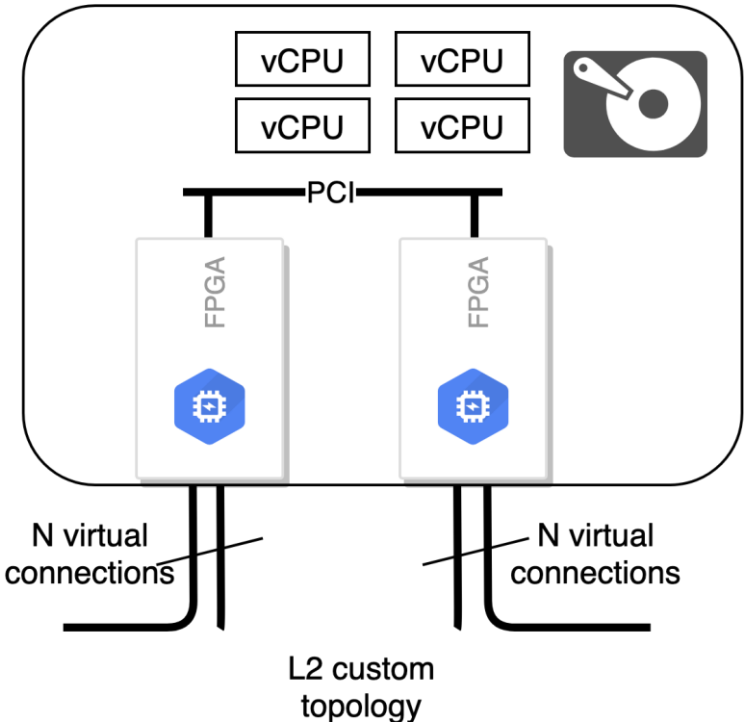
- GPS clock source (at most sites) using PTP
  - Subject to constraints of the hosting site
- NICs capable of accurate packet sampling/timestamping
- Programmable port mirroring
- Smart PDUs to measure power
- Optical layer measurements (where available)
- CPU, memory, disk, port/interface utilization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

FABRIC Experiment building blocks

- Each experiment is encapsulated in a slice - a topology
- Slices consist of slivers
  - Individually programmable or configurable resources
- Slices can change over time
  - Grow or shrink, adding or shedding resources under programmatic control
- Slice topologies can be
  - Custom L2 using underlying MPLS-SR
  - Rely on persistent routable IPv6 layer in FABRIC
- Basic sliver classes
  - Nodes - can include a selection of PCI-passthrough devices
  - Links - L2 or L3 with QoS and without
  - Measurement points - inside and outside the slice

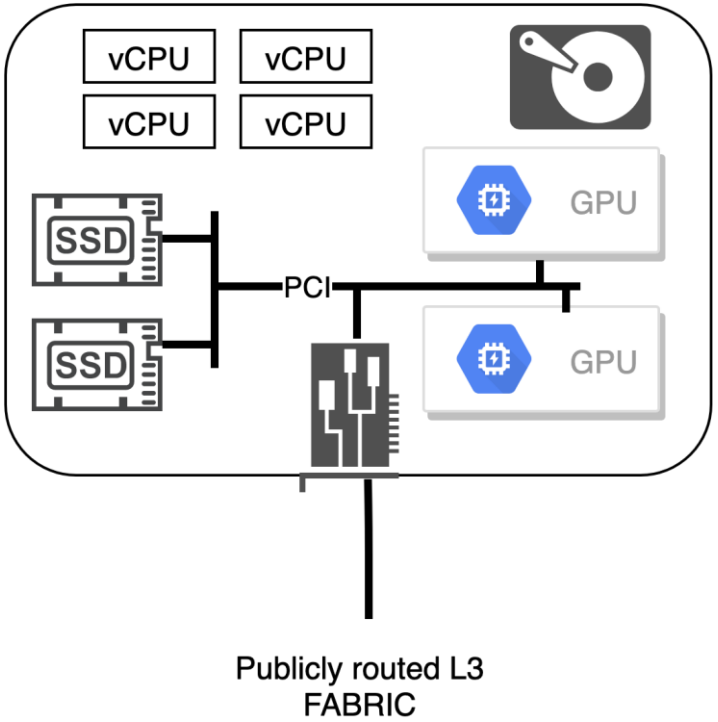
FPGA or P4 router sliver

- Uses Xilinx FPGAs in a node
- Can build a small port-count FPGA router
- With additional tools support can also serve as a P4 router built on top of the FPGA
- Can route between multiple virtual connections based on e.g. VLAN tags or other header information



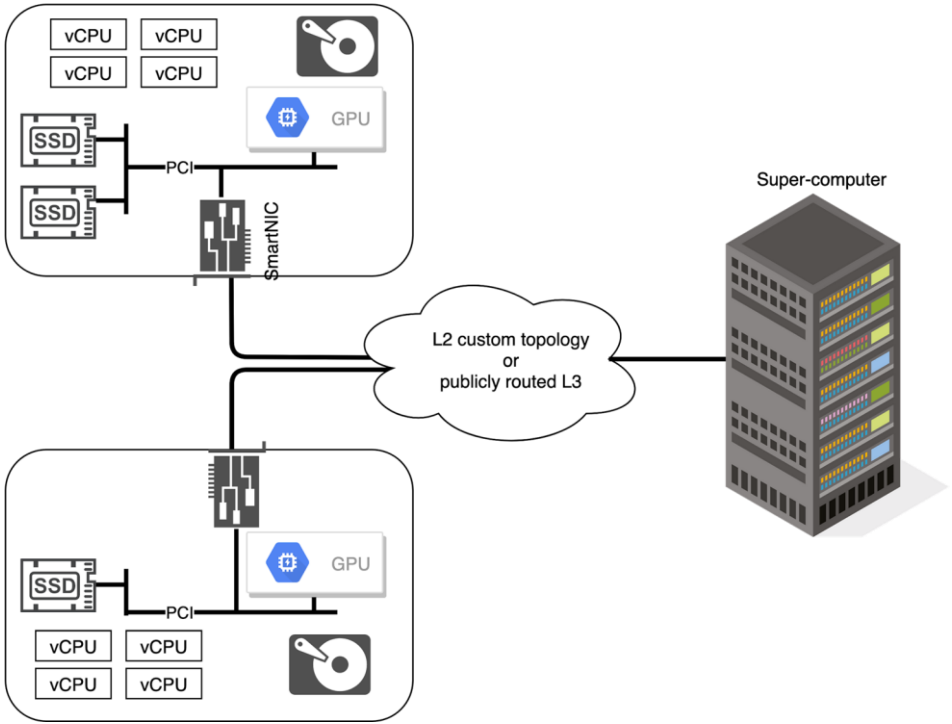
In-network AI/ML

- Investigating autonomous network behavior using in-network GPU support
  - Using RTX6000 for learning and inference using streaming data
- Perform intelligent data fusion/processing in the network
- Implement in-network analytics/security functions



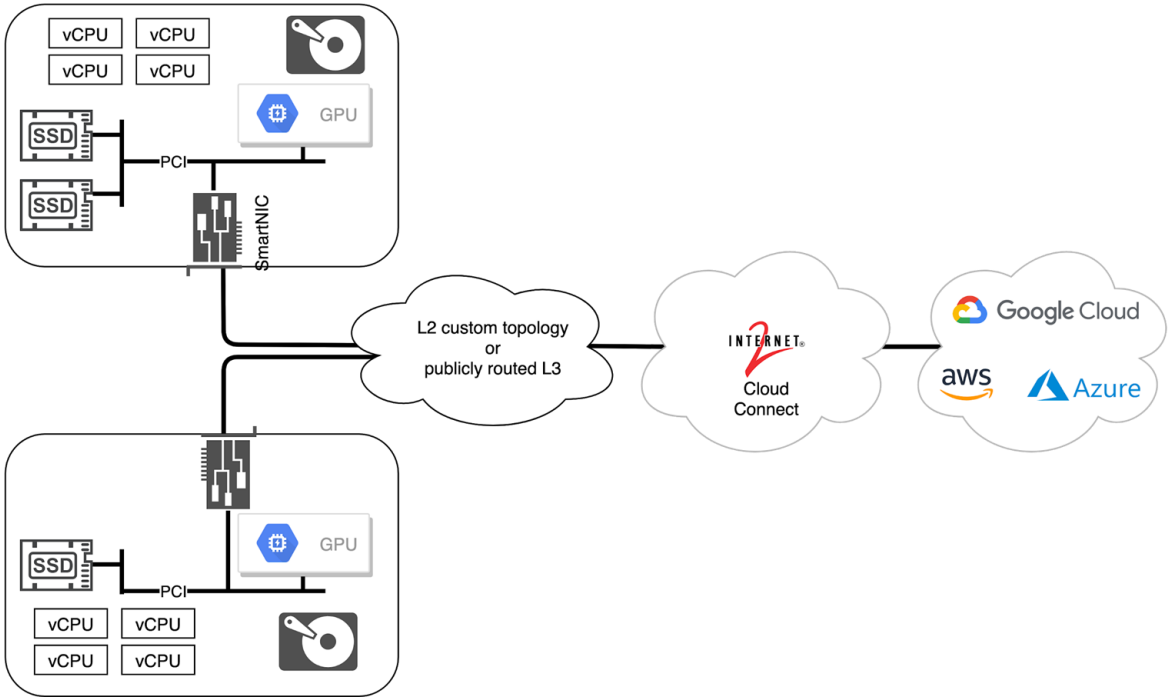
Attaching external facilities

- The US NSF has made significant investments in scientific CI
- Future networks must better support domain science needs
- FABRIC connects to a number of facilities and testbeds to enrich the set of resources that can be used in experiments
  - Supercomputing centers (PSC, NCSA, SDSC, TACC, MGHPCC)
  - Cloud testbeds - CloudLab, Chameleon, Open Cloud Testbed
  - 5G testbeds - COSMOS, Powder
- Through FAB we will also reach
  - University of Bristol, University of Amsterdam, University of Tokyo, CERN



Using public clouds in experiments

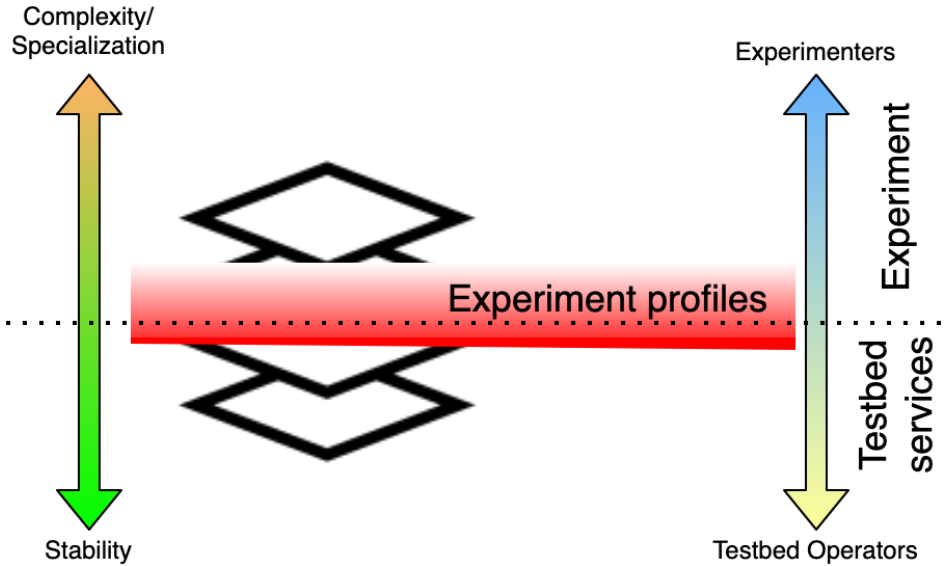
- Future networks will connect clouds and their customers
- 5G+Cloud experiments
- Through partnership with Internet 2 FABRIC will provide connectivity to commercial clouds
    - Utilize I2 CloudConnect system



FABRIC Testbed Services

- Central to FABRIC are ideas of ‘testbed services’ and ‘experiment profiles’
- Testbed services are provided by us

- Experiment profiles can be created by testbed operators or you and shared with others
  - Contain additional configuration, reproducible building blocks to help build experiments faster



FABRIC Network Services

- FABRIC is *not a network,* rather **a testbed** that provides network services
- Variety of options to connect compute slivers into topologies
- L2 point-to-point
  - Built on top of MPLS-SR
  - Support for QoS for individual services
  - **Does not assume the use of IP**
  - Routing must be built into the experiment via experiment profiles (e.g. OSPF instances or NDN forwarded instances)
- L3 routed
  - Relies on FABRIC's allocation of public IPv6 addresses
  - Provides high-performance routing using FABRIC's hardware routers
  - **Peers with production networks**
  - Shared between multiple experiments

FABRIC Software Overview

User Tools allow experimenters to interact with FABRIC resources.

Control Framework (CF) is responsible for allocating and provisioning experiment resources

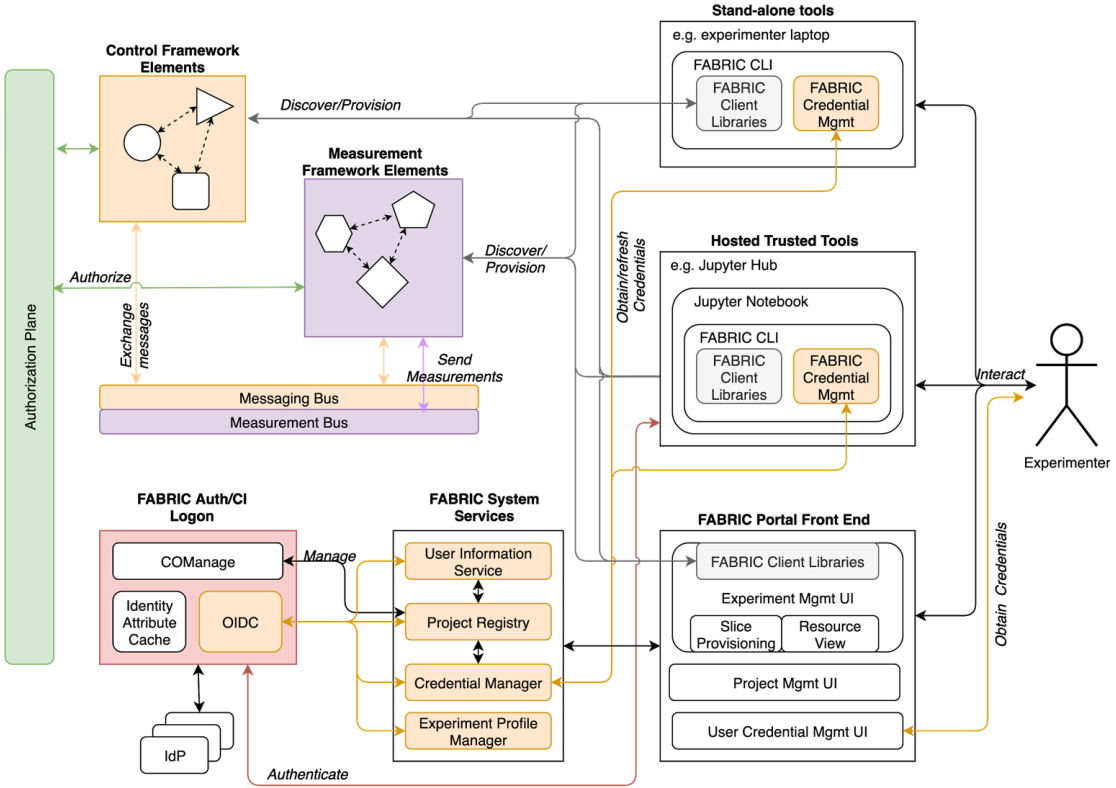
Measurement Framework (MF) is responsible for provisioning and configuring measurement resources and collecting measurement data.

FABRIC Information Model - property- graph-based representation of resources

The Message Bus is used to exchange (and log) messages between components.

Authentication and System Services manages identities, credentials, and registers/manages projects

The Authorization Plane provides a distributed way to express and enforce authorization policies.



FABRIC Measurement Framework (MF)

Manages and provisions measurement resources for experimenters & operators, enabling them to collect, process, store, and visualize measurement data.

- Adaptable/Programmable, Scalable, Extensible, and Shareable:
  - Supports automated instrumentation of experiments
  - Collect, store, and publish measurement data from users and the system (at all levels)
  - Support a common/shared measurement bus infrastructure based on pub/sub technology
  - Support efficient filtering, searching, and (limited) processing of measurement data
  - Interfaces with multiple UIs and alert systems
- Fine-grained Precise Measurements
  - Leverages PTP timing from node-local GPS receiver
  - Precise timestamping of packets using NIC cards (a.k.a., PacketGPS)
- Packet Capture
  - Supports high-speed packet capture and (limited) packet processing/storage

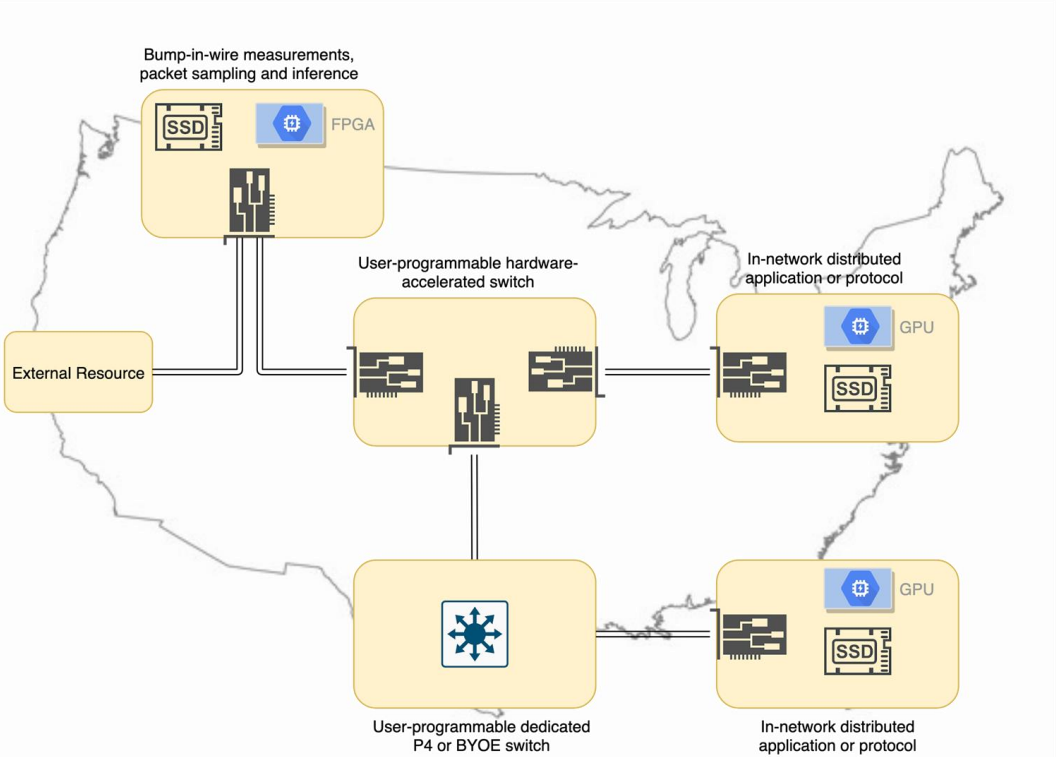
Building a testbed: Data

- Testbeds produce data artifacts
- Multiple questions arise:
  - Where does this data reside in the long run? Does the testbed provide long-term storage?
  - How public is this data? Meta-research can be conducted if data across multiple experiments is available to other researchers
  - If volumes or velocity of measurement data are expected to be high, special attention needs to be paid to the design of the measurement collection and storage system within the testbed

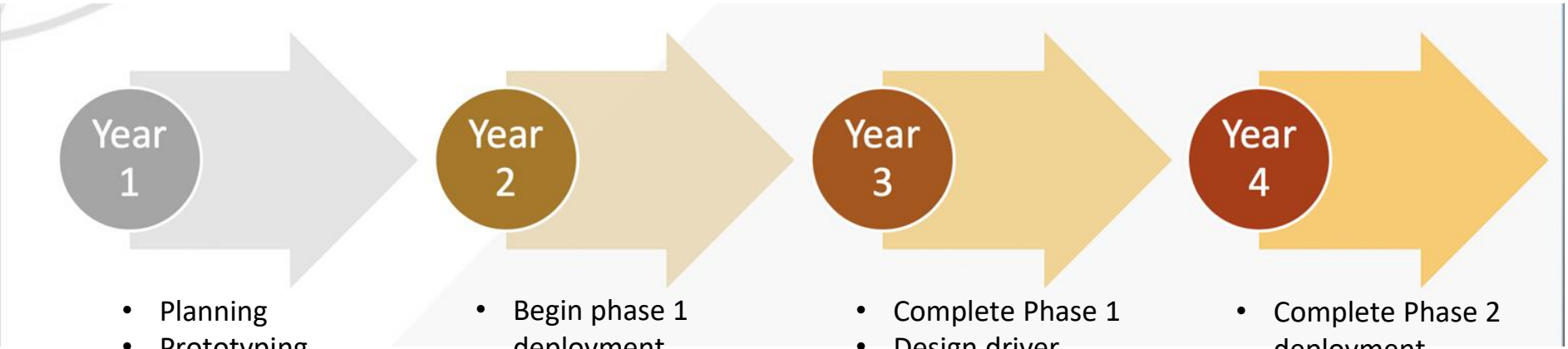
FABRIC

Example FABRIC Use-cases

- Measurements and packet sampling at high bit rates (up to 100+ Gbps)
- Hardware-accelerated switching using Smart NICs, FPGA NICs or P4 switches in individual nodes
- Hosting in-network applications and stateful architectures using a combination of storage and compute resources in individual nodes
- In-network inference, other types of accelerated computing via FPGAs and GPUs
- Experiments between facilities like IoT, 5G, cloud testbeds, Smart Cities testbeds, public clouds and HPC.
- Deploy non-IP protocols on top of wide-area L2 topologies, that may include in-network processing and storage



Construction Timeline



- Planning
- Prototyping
- Software development
- Community building

- Begin phase 1 deployment
- Testing, commissioning
- Design driver on-boarding

- Complete Phase 1
- Design driver experiments and early users
- Begin Phase 2 deployment

- Complete Phase 2 deployment
- Prepare for operations

Thank you!

This work is funded by
NSF grant CNS-1935966

https://whatisfabric.net
info@fabric-testbed.net