FABRIC

integration of bits, bytes, and xPUs

Inder Monga, Director, ESnet
imonga@es.net

JET meeting, March 17th, 2020

NSF'S 10 BIG IDEAS

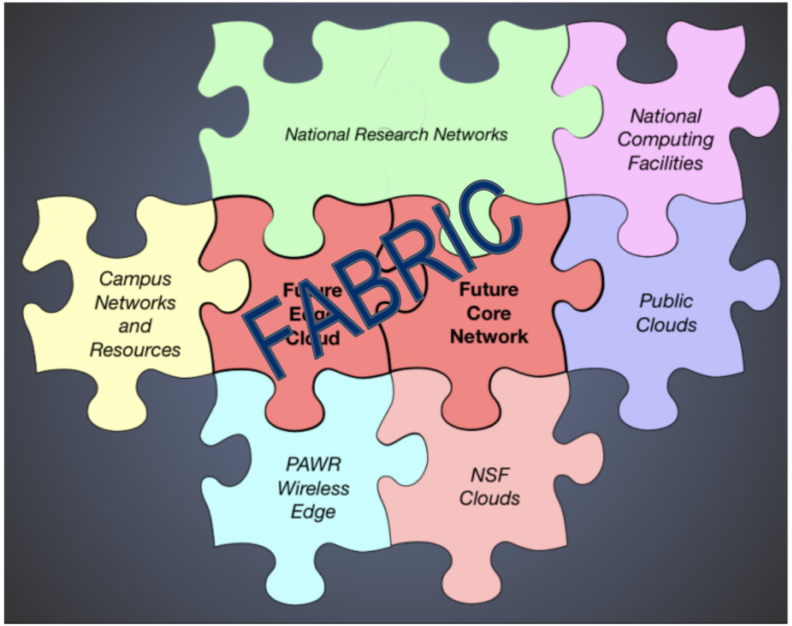
National Science Foundation
WHERE DISCOVERIES BEGIN

Genesis

Big Idea
Mid-scale Research Infrastructure

- Many important potential experiments and facilities fall between the $100K to $4M[1] Major Research Instrumentation (MRI) program and the > $70M Major Research Equipment and Facilities Construction (MREFC) account.
- This gap results in missed opportunities that may leave essential science undone.
- NSF needs a new agile process for funding experimental research capabilities in the mid-scale range.

The Future of CISE Distributed Research Infrastructure



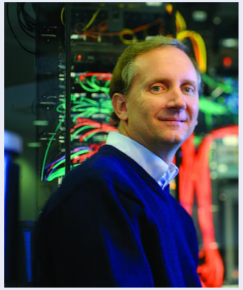
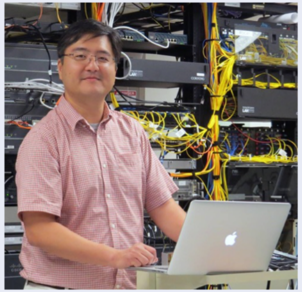
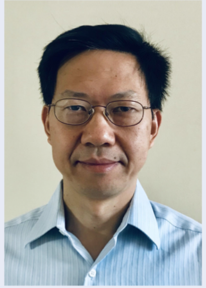
A Community White Paper
03/08/2018

J.Aikat (RENCI/UNC Chapel Hill)
I.Baldin[1] (RENCI/UNC Chapel Hill)
M. Berman (BBN/Raytheon)
J. Breen (Utah)
R.Brooks (Clemson)
P. Calyam (Missouri)
J.Chase (Duke)
W.Chase (Clemson)
R. Clark (Georgia Tech)
C.Elliott (BBN/Raytheon)
J.Griffioen[1] (Kentucky)
D. Huang (ASU)
J.Ibarra (FIU)
T. Lehman (Maryland)
I.Monga (ESnet)
A.Matta (Boston University)
C. Papadopoulos (Colorado State)
M.Reiter (UNC Chapel Hill)
D.Raychaudhuri (Rutgers)
G. Ricart (US Ignite)
R. Ricci (Utah)
P. Ruth (RENCI/UNC Chapel Hill)
I.Seskar (Rutgers)
J.Sobieski (NORDUnet/GEANT)
K. Van der Merwe (Utah)
K.-C.Wang[1] (Clemson)
T. Wolf (UMass)
M. Zink (UMass)

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

FABRIC Leadership Team

Ilya Baldin (RENCI)	Anita Nikolich (IIT)	Inder Monga (ESnet)	Jim Griffioen (UKY)	KC Wang (Clemson)
				
Dale Carder (ESnet)	Tom Lehman (Virnao)	Paul Ruth (RENCI)	Zongming Fei (UKY)	
				

FABRIC: Broad research infrastructure

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

FABRIC Core

ASIA
AUS
NZ
Seattle
LBNL
Salt Lake
UEN
CLOUDLAB
POWDER
FRGP
GPN
Kansas
CHAMELEON
Chicago
NCSA
STARLIGHT
MERIT
WISCNET
KYRON
MGHPCC
NYSERNET
NEREN
NJEDGE
PSC
NYC
EUROPE
COSMOS
Washington
3ROX
MAX
WIX
KINBER
NCREN
Internet2
aws
Microsoft
SDSC
PRP/NRP
San Diego
Internet2
aws
Microsoft
CENIC
TACC
Houston
LEARN
Atlanta
AMLIGHT
SOX
FLR
100G Core - Dedicated DWDM
Terabit Super-Core
SOUTH AMERICA

FABRIC Edge

SRI International

Georgia Tech

FIU FLORIDA INTERNATIONAL UNIVERSITY

Colorado State University

UMass Amherst

POWDER

INTERNET 2

TACC

renci

CLEMSON UNIVERSITY

COSMOS

THE UNIVERSITY OF UTAH

LEARN LONESTAR EDUCATION AND RESEARCH NETWORK

MAX MID-ATLANTIC CROSSROADS

UK

RUTGERS

3ROX THREE RIVERS OPTICAL EXCHANGE

PITTSBURGH SUPERCOMPUTING CENTER

CENIC

STARLIGHT℠ The Optical STAR TAP℠

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

PRP PACIFIC RESEARCH PLATFORM

usignite

BERKELEY LAB

M UNIVERSITY OF MICHIGAN

UC San Diego

SDSC SAN DIEGO SUPERCOMPUTER CENTER

Chameleon

I ILLINOIS NCSA

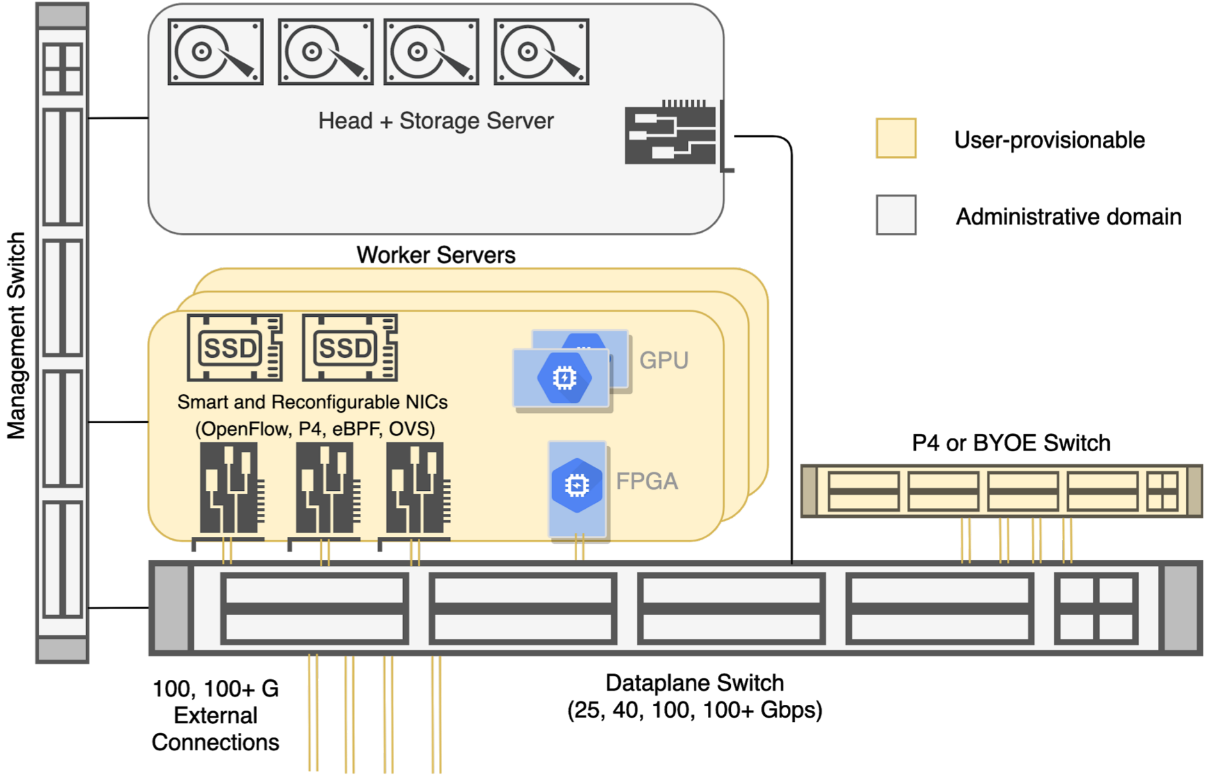
EASTERN REGIONAL NETWORK

THE UNIVERSITY OF CHICAGO

GPN GREAT PLAINS NETWORK

PRINCETON UNIVERSITY

FABRIC Node Concept



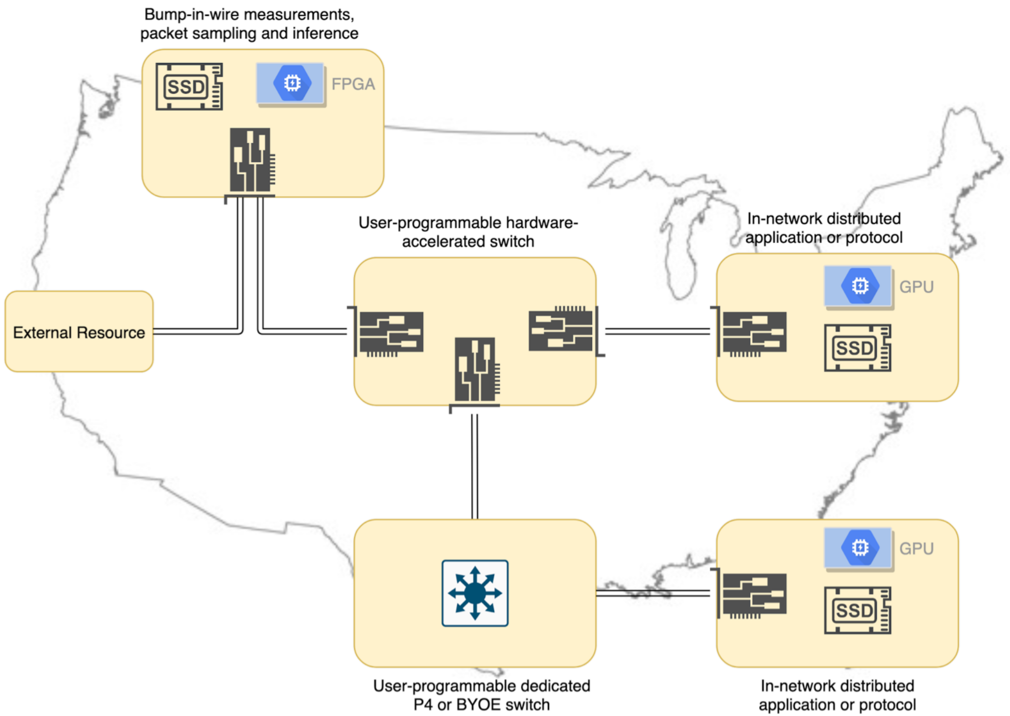
FABRIC Node (‘hank’) Design: Network + Storage + Compute

- We refer to it also as a ‘disaggregated router’
- Network cards with high speed interfaces (25G, 40G, 100G. 200G+ in future)
  - Programmable interface cards (hardware OVS offload + DPDK)
  - Reconfigurable interface cards (FPGA and P4/network processors)
- High-performance servers equipped with
  - GPUs
  - FPGA compute accelerators
  - NVMe drives
  - Storage: User-provisionable short term & shared high volume. Not meant to be persistent.
- All ports interconnected by a 100G+ switch programmable through testbed control software
  - Acts as a ‘patch panel’ connecting various ports in the node together
- Users can fully interact with network, compute, storage
- Nodes are “sliceable” for experimenters to use simultaneously

Potential use-case scenarios

Examples of potential uses:

- ‘Bump-in-wire’ measurements and packet sampling at high bit rates (25, 40, 100, 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
- Connect experiments to external facilities like IoT, 5G, cloud testbeds, public clouds and HPC resources.
- Deploy non-IP protocols on top of wide-area L2 topologies, that may include in-network processing and storage



FABRIC Network Services

- Network services link different elements of requested topologies together and to the outside world
- Examples of FABRIC network services for experimenter topologies:
  - Layer 2 on-demand with bandwidth provisioning or best-effort
    - Layer 2 on-demand services require experimenter to build their own Layer 3 services, possibly from an existing experiment profile
  - Layer 3 (IPv6) best-effort
  - Layer 3 peering between experiment topology and an existing production network (e.g. campus)
  - Layer 2 peering between experiment topology and a cloud provider (Google, AWS or Azure, via Internet2 CloudConnect)
  - VPN from FABRIC node to experimenter desktop or a campus resource

FABRIC Enables Measurement

- Measurement Framework is designed to be Adaptable/Programmable, Scalable, Extensible, and Shareable:
  - Is used to collect, store, and publish measurement data from users and the system
  - Supports a common/shared message bus infrastructure based on pub/sub technology
  - Supports efficient filtering, searching, and (limited) processing of measurement data
  - Interfaces with multiple UIs and alert systems
- Fine-grained Precise Measurements
  - Leverages a highly-accurate PTP timing signal from a node-local GPS receiver
  - Supports precise timestamping of packets using NIC cards (a.k.a., PacketGPS)
- Packet Capture
  - Supports high-speed packet capture and (limited) processing

Early Science Design Drivers and Applications

- Four ‘Science Design Driver’ teams
  - FABRIC-ready experiment use-cases and applications
  - Help formulate design requirements
  - Help validate and commission the facility
  - Leave lasting experimental artifacts - software, experiment profiles, case studies
- Security, IoT, ML in the network, Named Data Networking, advanced transport protocols

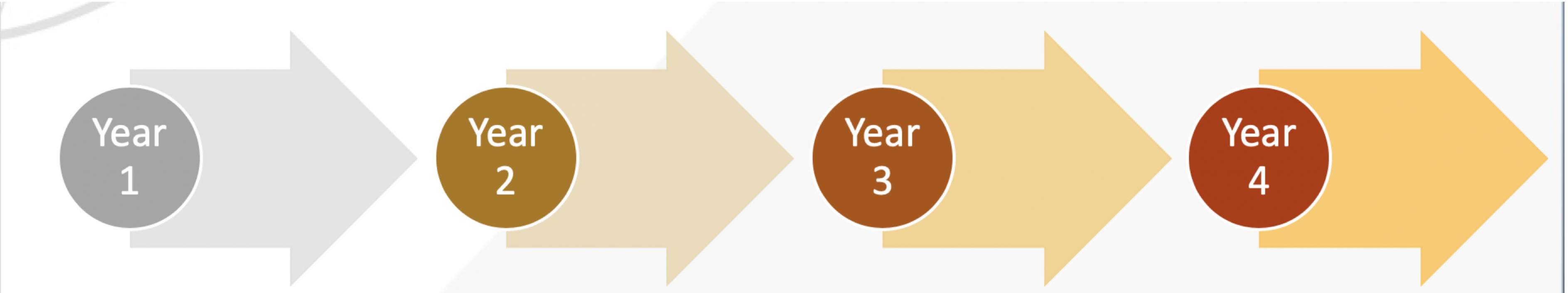
SRI International — ***Security*** — Phil Porras

University of Virginia — ***Machine Learning*** — Malaathi Veeraraghavan

Georgia Tech — ***IoT*** — Russ Clark

FIU Florida International University — ***NDN*** — Alex Afanasyev

Construction Timeline



Year 1

- Planning
- Prototyping
- Software development
- Community building

Year 2

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

Year 3

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

Year 4

- Complete Phase 2 deployment
- Prepare for operations

What FABRIC IS:

- FABRIC is an ‘everywhere-programmable’ network combining *core* and *edge* components that also link to many outside facilities.
- FABRIC is a multi-user facility with support for concurrent experiments of differing scales facilitated through federated authn/authz system with allocation controls.
- FABRIC is a place to experiment on new Internet architectures, protocols and distributed applications using a mix of resources from FABRIC, its facility partners, connected campuses and opt-in users.
- FABRIC is extensible – it will continue to connect new facilities like cloud, networking, other testbeds, computing facilities and scientific instruments. BYOE is also an option.

What FABRIC is NOT:

- FABRIC is not an isolated testbed – it will peer at Layer 2 and Layer 3 with a variety of networks, allowing experiment slices to connect to a wide variety of external resources
- FABRIC is not a place for long-term production workloads - it is intended for CI experiments short- or long-lived.
- FABRIC is not a place for real-world protected (PII or other) data – you can develop such new applications on FABRIC, but the infrastructure cannot support regulated data.
- FABRIC is not a fast new pipe for data between its connected facilities – ESnet, Internet2, and the regional networks provide production capacity, FABRIC provides a place to experiment with new approaches.

FABRIC Community Building

- We are looking to build a vibrant community of stakeholders:
  - Experimenters interested in using FABRIC
  - Facility partners to host equipment
  - Regional and national network providers
  - Government agencies focusing on research
  - Industry looking to test or partner
- Community events & workshops to share the vision, progress and collect feedback
- Virtual Community Visioning Workshop: **April 15-16, 2020**
- Follow on, focused workshops 1-2/ year

FABRIC Scientific Advisory Committee


Sujata Banerjee


Terry Benzel


Kaushik De


Cees de Laat


Phillipa Gill


Abraham Matta


Craig Partridge


Jennifer Rexford


Scott Shenker


Frank Wuerthwein

Thank you!

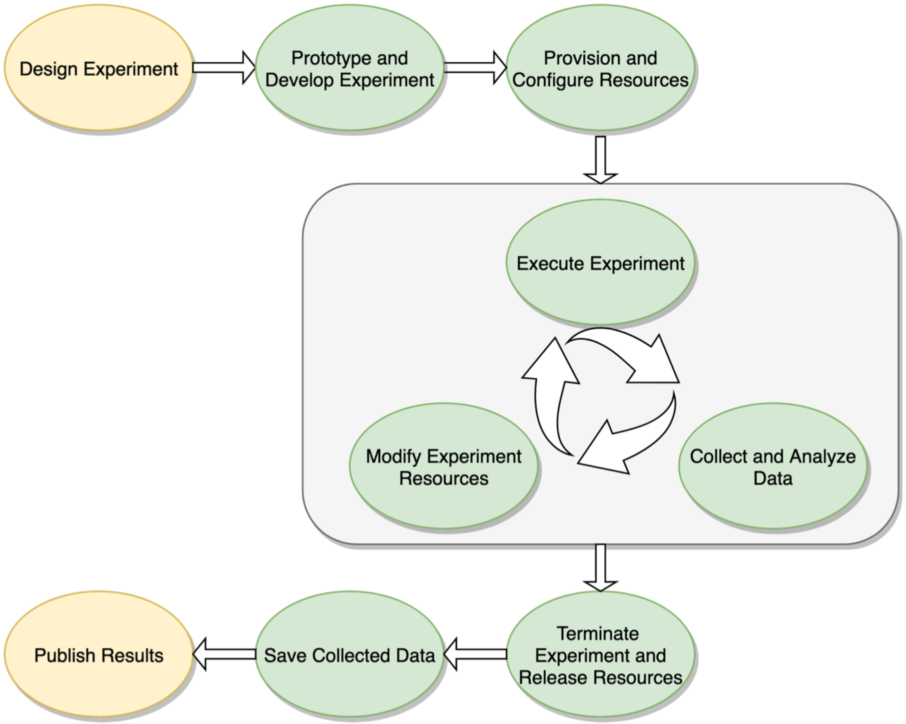
Questions? Ask info@fabric-testbed.net

This work is funded by
NSF grant CNS-1935966

FABRIC Experiment Workflow

Experiment Phases:

- Design - an experiment is imagined and defined
- Prototyping and development - experiment software is written and prototyped (in-house, using FABRIC or other testbed hardware)
- Provision resources - FABRIC and other resources are acquired and configured via APIs or portal
- Experiment is run:
  - Multiple experiment runs include collecting data and modifying resources
- Termination - experiment ends, all resources released
- Saving data - collected data is retrieved from FABRIC storage
- Publish - paper citing FABRIC is prepared, submitted and published



How is FABRIC different from GENI?

- FABRIC has a programmable core infrastructure
- FABRIC interconnects a large number of existing scientific, computational and experimental facilities
- FABRIC provides guaranteed quality of service by utilizing its own dedicated optical 100G infrastructure
- FABRIC experimenter network topologies can peer with production networks on-demand