Workshop on P4 Programmable Switches

Using FABRIC for Cybertraining

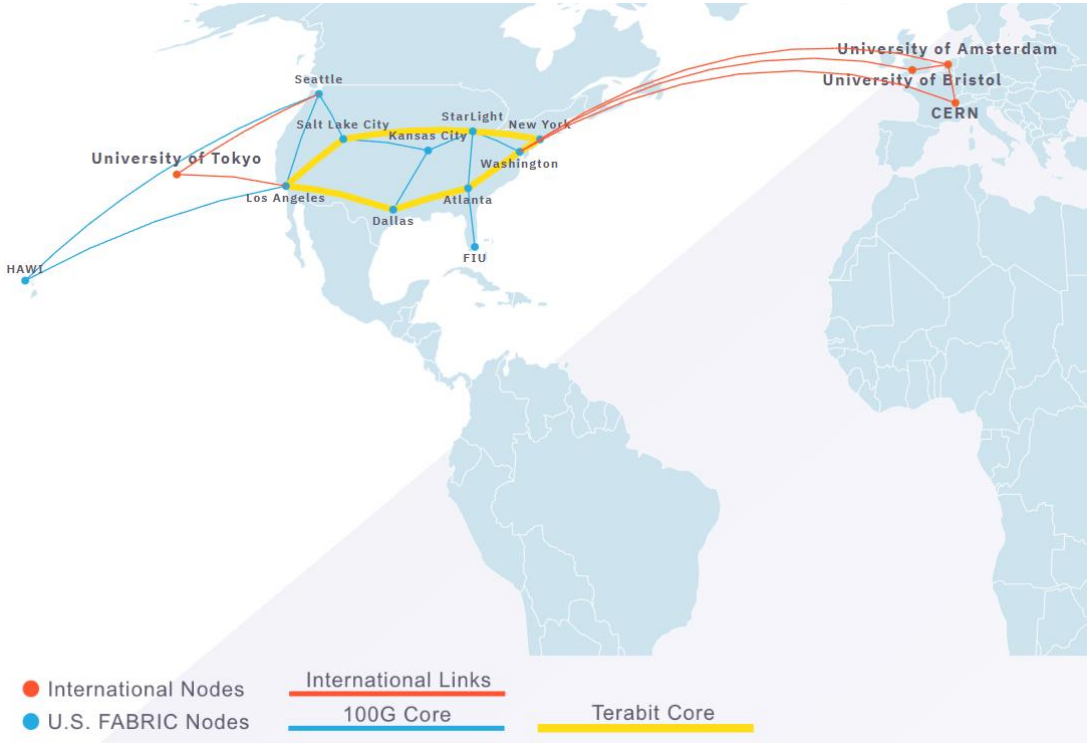
Elie Kfoury, Jorge Crichigno
College of Engineering and Computing, University of South Carolina
http://ce.sc.edu/cyberinfra

University of South Carolina (USC)

August 22, 2023

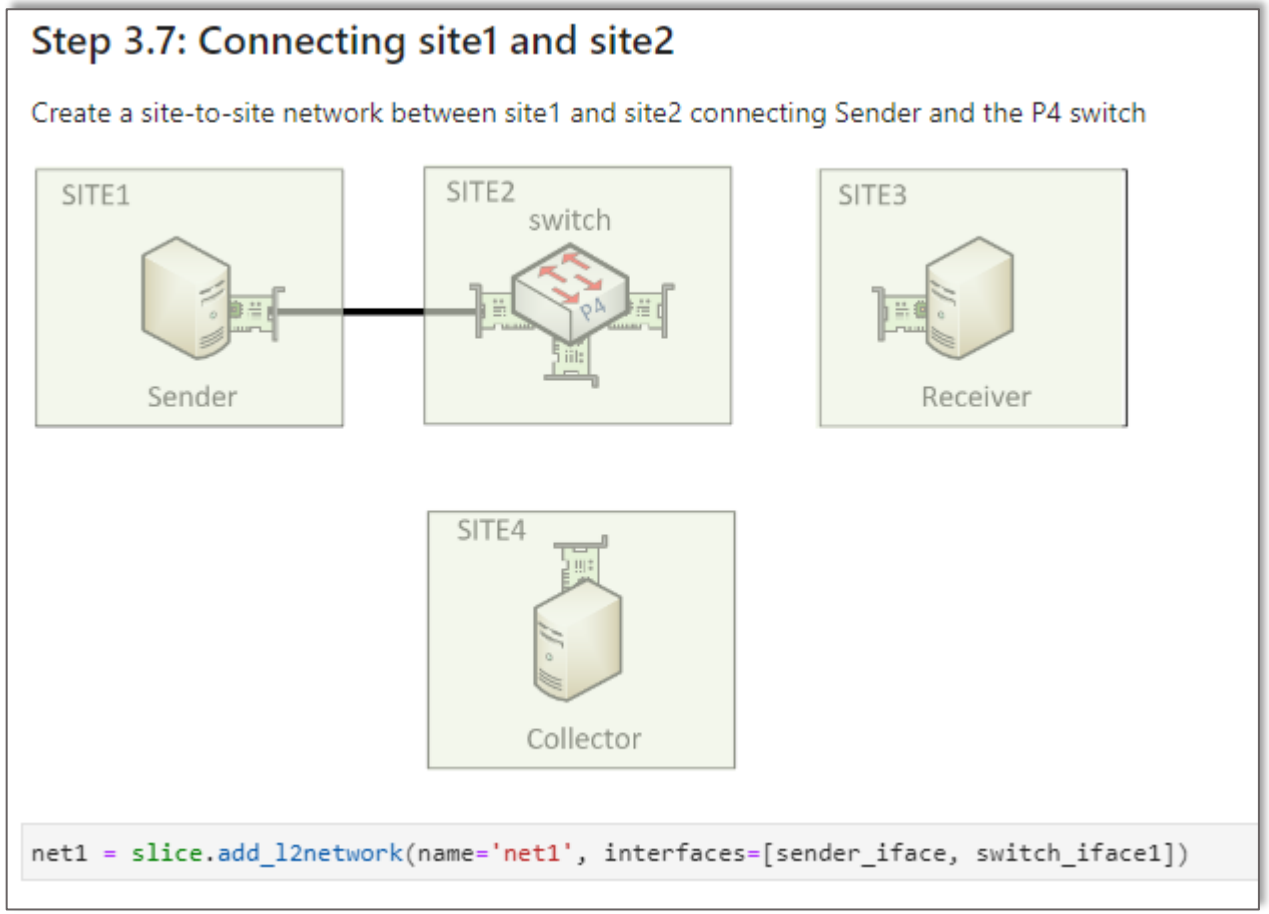
FABRIC Testbed

- FABRIC is an NSF-funded international infrastructure for at-scale experimentation and research
- Areas include networking, cyber, distributed computing, storage, 5G, ML, etc.
- Equipment is located at commercial collocation spaces, U.S. national labs, and campuses – 29 FABRIC sites



Cybertraining on FABRIC

- FABRIC is a real network with physical propagation delays and high-speed links
- With its integrated JupyterHub, it can be ideal for cybertraining:
  - P4 programmable switches/NICs
  - High-speed networks (SDMZ)
  - PerfSONAR
  - Measurement and telemetry
  - Cybersecurity (Zeek, Suricata, etc.)
  - Etc.



Organization of the Labs

Each lab starts with a section *Overview*

- Objectives
- Lab topology
- Roadmap: organization of the lab

Part 1

- Background information of the topic being covered
- Section 1 is optional (i.e., the reader can skip this section and move to lab directions)

Part 2… n

- Step-by-step directions

Labs on P4 Programmable Data Planes over FABRIC

- The following labs have been developed[1]:
  - Lab 1 – Preparing the Environment
  - Lab 2 – P4 Program Building Blocks
  - Lab 3 – Parser Implementation
  - Lab 4 – Introduction to Match-action Tables
  - Lab 5 – Populating Match-action Tables from the Control Plane
  - Lab 6 – Checksum Calculation and Packet Deparsing
- The labs are available to FABRIC users
  - 'Beyond Bleeding Edge' container

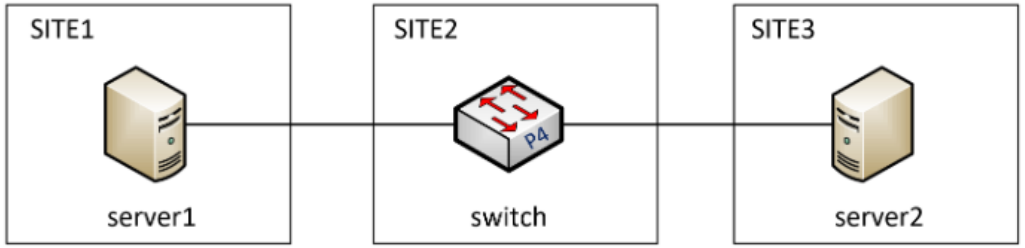
[1] https://learn.fabric-testbed.net/knowledge-base/p4-programmable-data-plane-switches-bmv2-over-fabric/

Labs on P4 Programmable Data Planes over FABRIC

Virtual Labs on P4 Programmable Data Plane Switches (BMv2)

The labs provide a hands-on experience on P4 programmable data plane switches using the Behavioral Model version 2 (BMv2) software switch. The lab series explains topics that include parsing, match-action tables, checksum verification, and others.

The lab series is developed by the Cyberinfrastructure Lab (CILab) at the University of South Carolina (USC).

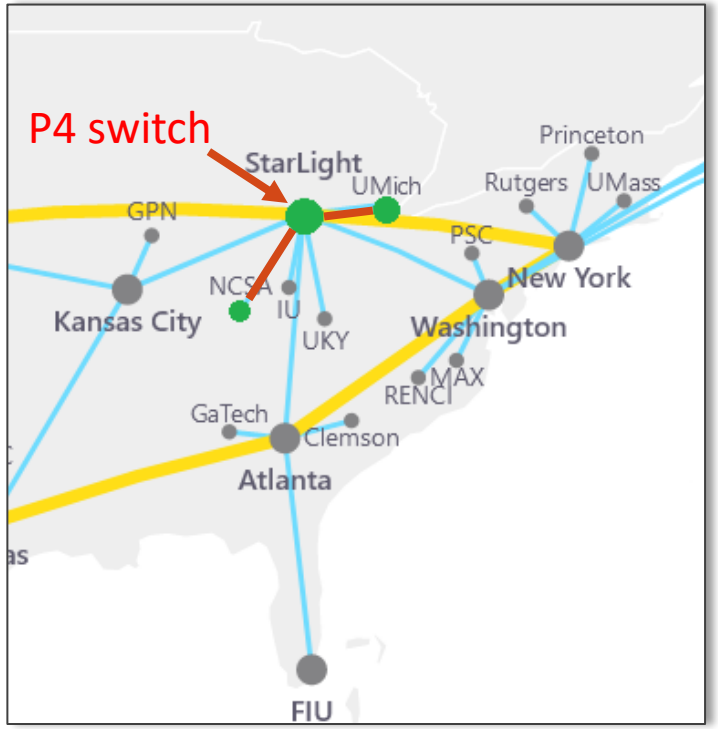


Labs:

- Lab 1 - Creating a Slice with a P4 Switch: This lab describes how to create a slice with a P4 switch. It also shows how to deploy the high-performance BMv2 switch to achieve up to ~1Gbps throughput.
- Lab 2 - P4 Program Building Blocks: This lab describes the building blocks and the general structure of a P4 program. It maps the program's components to the Protocol-Independent Switching Architecture (PISA).
- Lab 3 - Parser Implementation: This lab describes how to define custom headers in a P4 program. It then explains how to implement a simple parser that parses the defined headers.
- Lab 4 - Introduction to Match-action Tables: This lab describes match-action tables and how to define them in a P4 program. It then explains the different types of matching that can be performed on keys.
- Lab 5 - Populating and Managing Match-action Tables at Runtime: This lab describes how to populate and manage match-action tables at runtime. It then explains a tool (simple_switch_CLI) that is used with the software switch (BMv2) to manage the tables.
- Lab 6 - Checksum Recalculation and Packet Deparsing: This lab describes how to recompute the checksum of a header. Recomputing the checksum is necessary if the packet header was modified by the P4 program. The lab also describes how a P4 program performs deparsing to emit headers.

Throughput Test over BMv2

- BMv2 software switch is running on StarLight



Step 8.4: Starting iPerf3 on server2

```
[107]: server2.execute_thread('iperf3 -s')

[107]: <Future at 0x7fee04ab7b50 state=running>
```

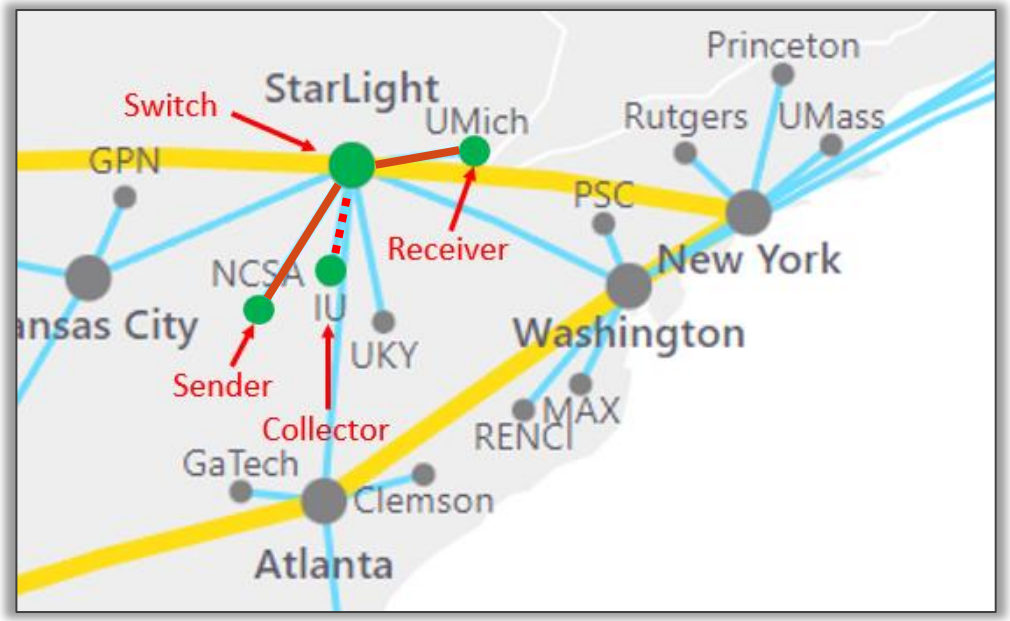
Step 8.5: Starting iPerf3 client on server1

```
[114]: server1.execute('iperf3 -c 192.168.2.10 -P 2')

Connecting to host 192.168.2.10, port 5201
[  5] local 192.168.1.10 port 57904 connected to 192.168.2.10 port 5201
[  7] local 192.168.1.10 port 57908 connected to 192.168.2.10 port 5201
[ ID] Interval           Transfer     Bitrate         Retr  Cwnd
[  5]   0.00-1.00   sec  64.5 MBytes   541 Mbits/sec  427   1.25 MBytes
[  7]   0.00-1.00   sec  72.4 MBytes   607 Mbits/sec  1050   1.47 MBytes
[SUM]   0.00-1.00   sec   137 MBytes  1.15 Gbits/sec  1477
- - - - - - - - - - - - - - - - - - - - - - - - -
[  5]   1.00-2.00   sec  60.0 MBytes   503 Mbits/sec   31    952 KBytes
[  7]   1.00-2.00   sec  70.0 MBytes   587 Mbits/sec   47   1.10 MBytes
[SUM]   1.00-2.00   sec   130 MBytes  1.09 Gbits/sec   78
- - - - - - - - - - - - - - - - - - - - - - - - -
[  5]   2.00-3.00   sec  56.2 MBytes   472 Mbits/sec    0   1024 KBytes
[  7]   2.00-3.00   sec  66.2 MBytes   556 Mbits/sec    0   1.18 MBytes
[SUM]   2.00-3.00   sec   122 MBytes  1.03 Gbits/sec    0
- - - - - - - - - - - - - - - - - - - - - - - - -
```

Queue Measurement Lab

- BMv2 software switch is running on StarLight
- Microseconds granularity



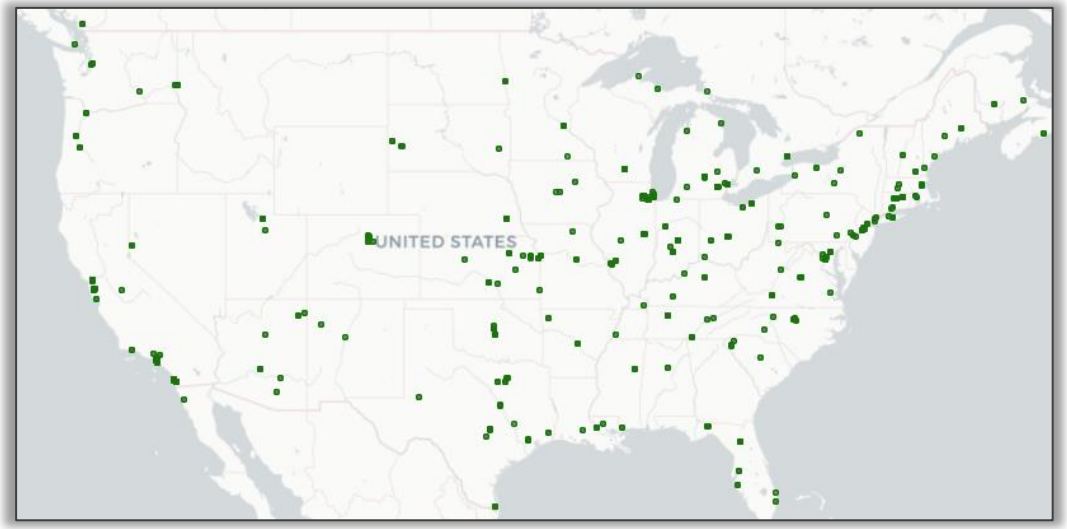
Queue Occupancy

```
384.21 us
380.21 us
395.49 us
389.79 us
383.94 us
386.61 us
375.84 us
5375.27 us
62615.75 us
132152.93 us
```

Upcoming Lab Libraries over FABRIC

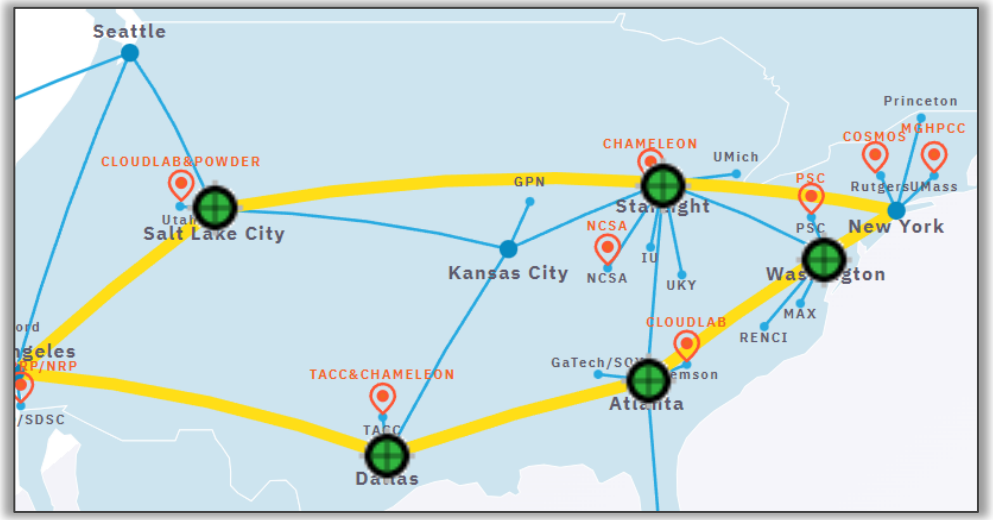
perfSONAR

- perfSONAR is a tool coordinated suite of tools to:
  - Set network performance expectations
  - Find network problems (“soft failures”)
  - Helps coordinate fixing these problems
- perfSONAR provides a standardized way to publish active and passive monitoring data
- Used in Science DMZ networks
- FABRIC can be used to train CI engineers and operators on perfSONAR



perfSONAR Nodes in the US

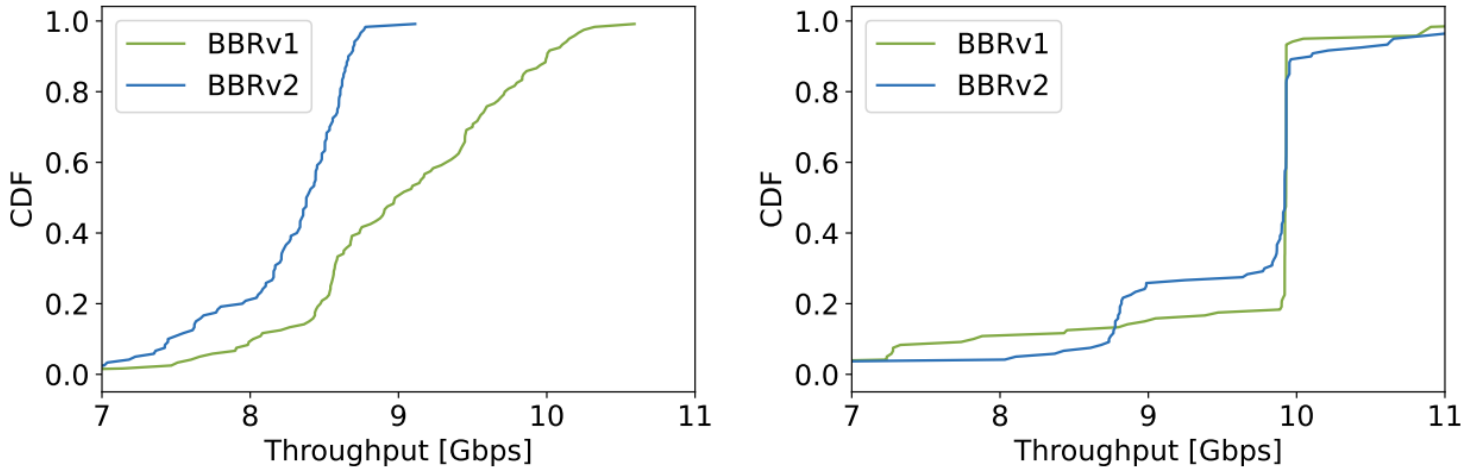
https://stats.perfsonar.net/d/spFwAQi4z/perfsonar-public?orgId=2/



FABRIC with perfSONAR nodes

High Speed Networks (TCP Congestion Control)

- Large scale network emulation with 40/100Gbps does not guarantee high-fidelity
- FABRIC can be used for testing the performance of TCP using real propagation delays and high-speed links
- Preliminary results show that TCP exhibits different behavior when tested in real networks



CDF of the bottleneck bandwidth estimation of BBRv1 and BBRv2.
(a) with 45ms emulated delay. (b) with 45ms propagation delay.

Other Upcoming Lab Libraries over FABRIC

- Advanced P4 Programmable Data Planes: Applications, Stateful Elements, and Custom Packet Processing (Tofino ASIC on hardware switches)
- Plotting Granular P4 Measurements on Grafana
- Writing Cybersecurity Applications on P4 Programmable Data Planes
- Software-defined Networking and Open vSwitch (ONOS + OVS)

Demo: Queue Measurement Lab