ESnet
ENERGY SCIENCES NETWORK

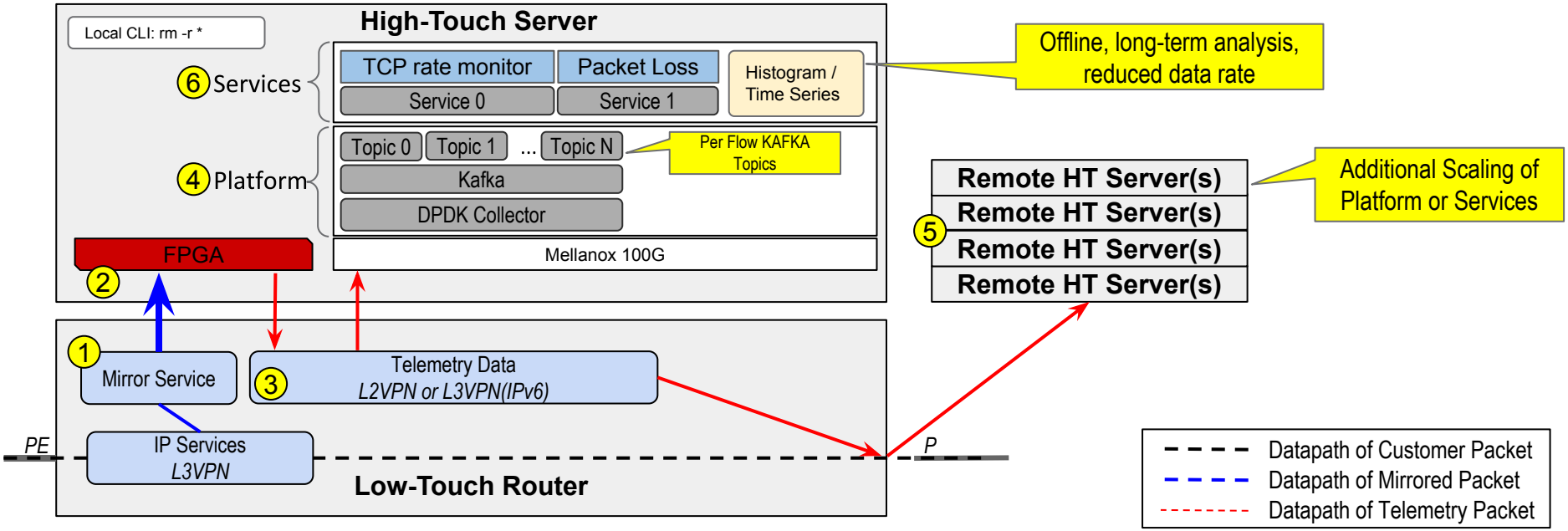
ESnet6 High Touch Services

Precision Streaming Network Telemetry

Bruce A. Mah, Richard Cziva, Chin Guok, Yatish Kumar
Energy Sciences Network
Lawrence Berkeley National Laboratory

October 20th, 2020

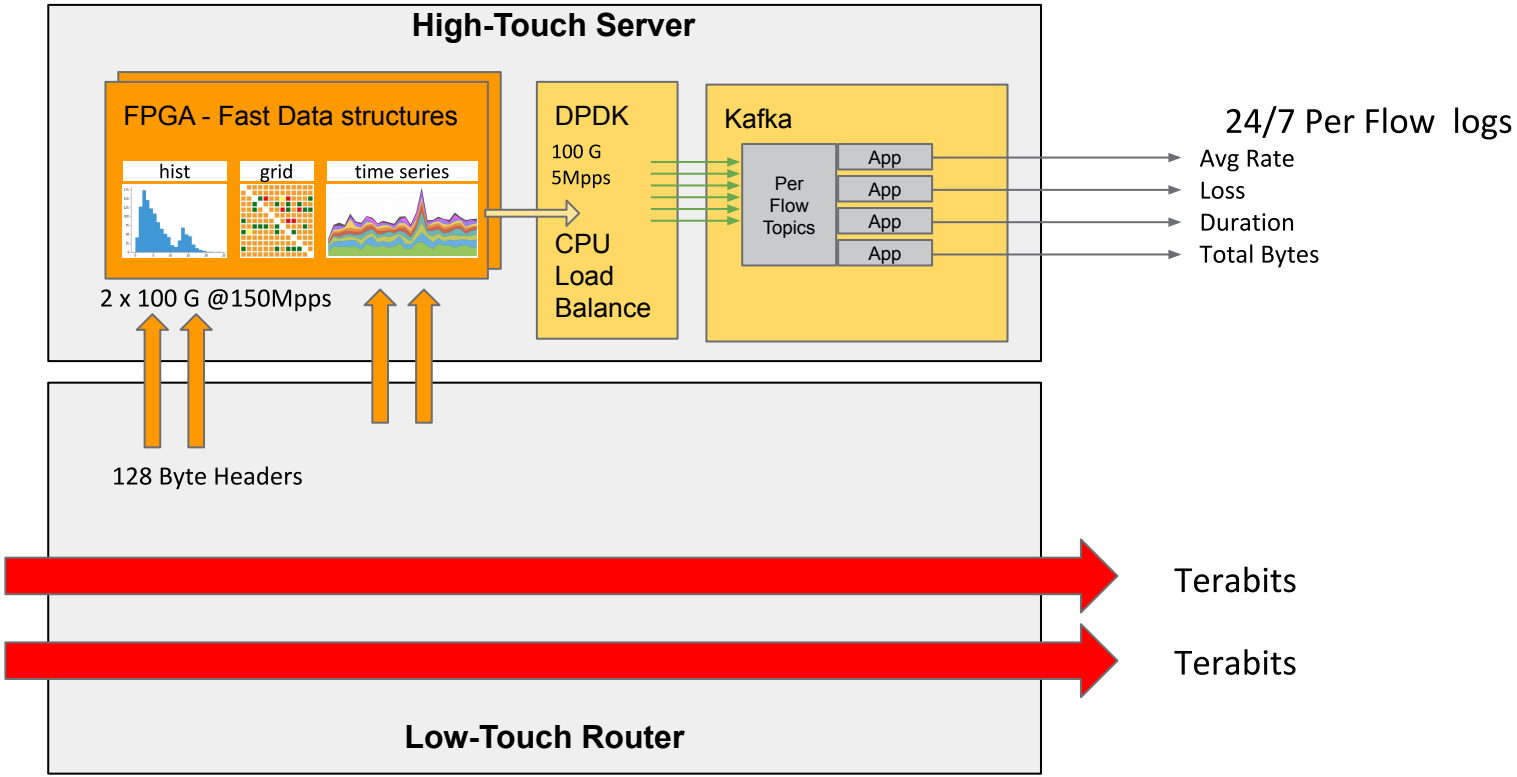
ESnet6 High-Touch Architecture Overview



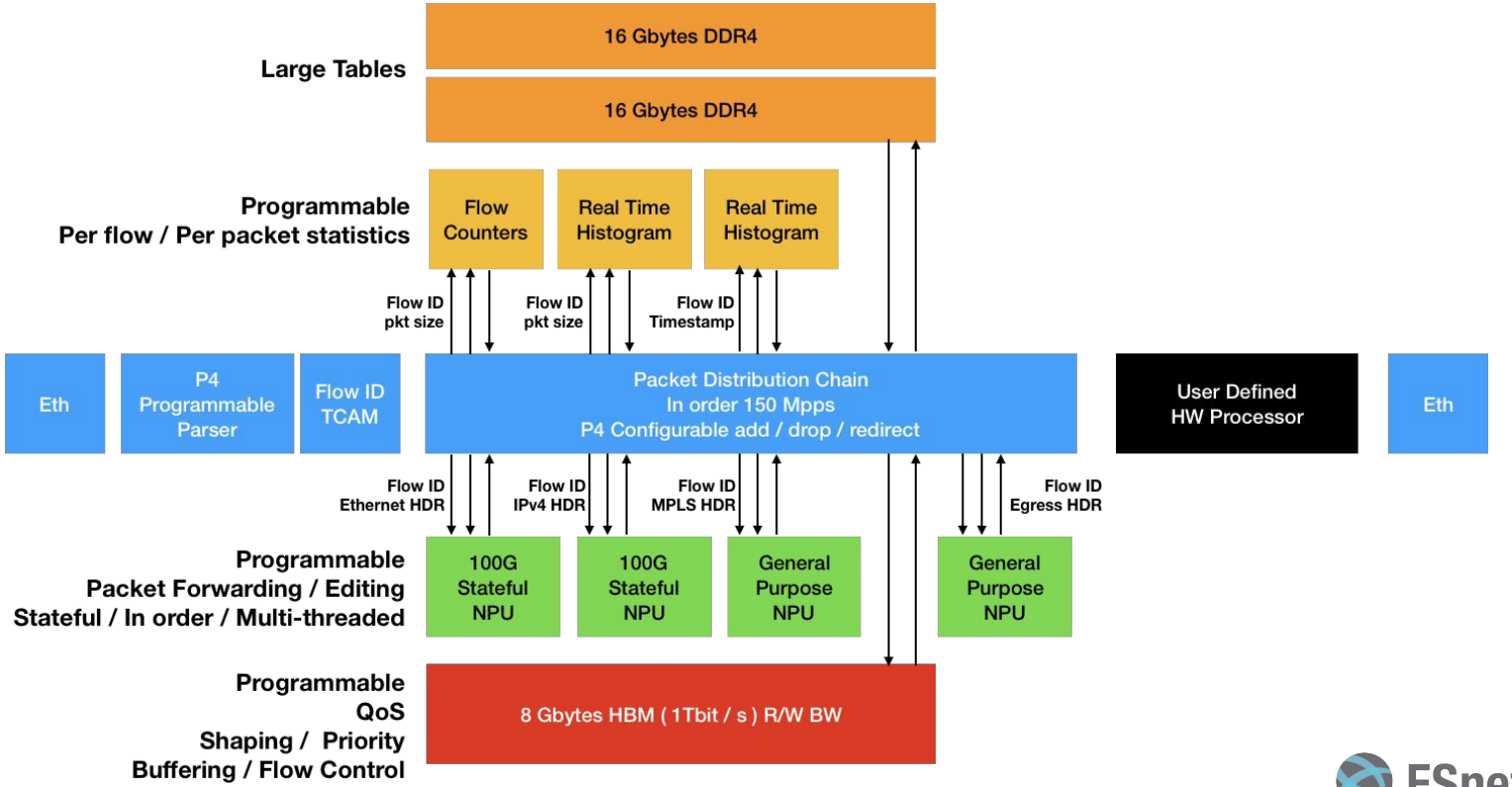
1. Mirror Service - Allows selective flows in the dataplane to be duplicated and sent to the FPGA for processing.
2. Programmable Dataplane (DP) - Appends meta-data, timestamps and repackages packet for transmission to Platform code.
3. Telemetry Data L2VPN or L3VPN(IPv6) - Connect Dataplane and Platform, possibly on different High-Touch (Local and Remote) Servers.
4. Platform - Reads telemetry packets from the network and distributes information to High Touch Services.
5. Remote Server - Hosts Platform components or Services (but not a Dataplane). Telemetry data can be directed to Remote Servers.
6. Services - Reads data from the Platform and performs real-time analysis as well as inserts selected telemetry data into database.

ESnet

Summary Of Rates

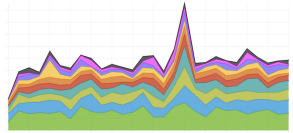


ESnet FPGA Block Diagram - Present and Future

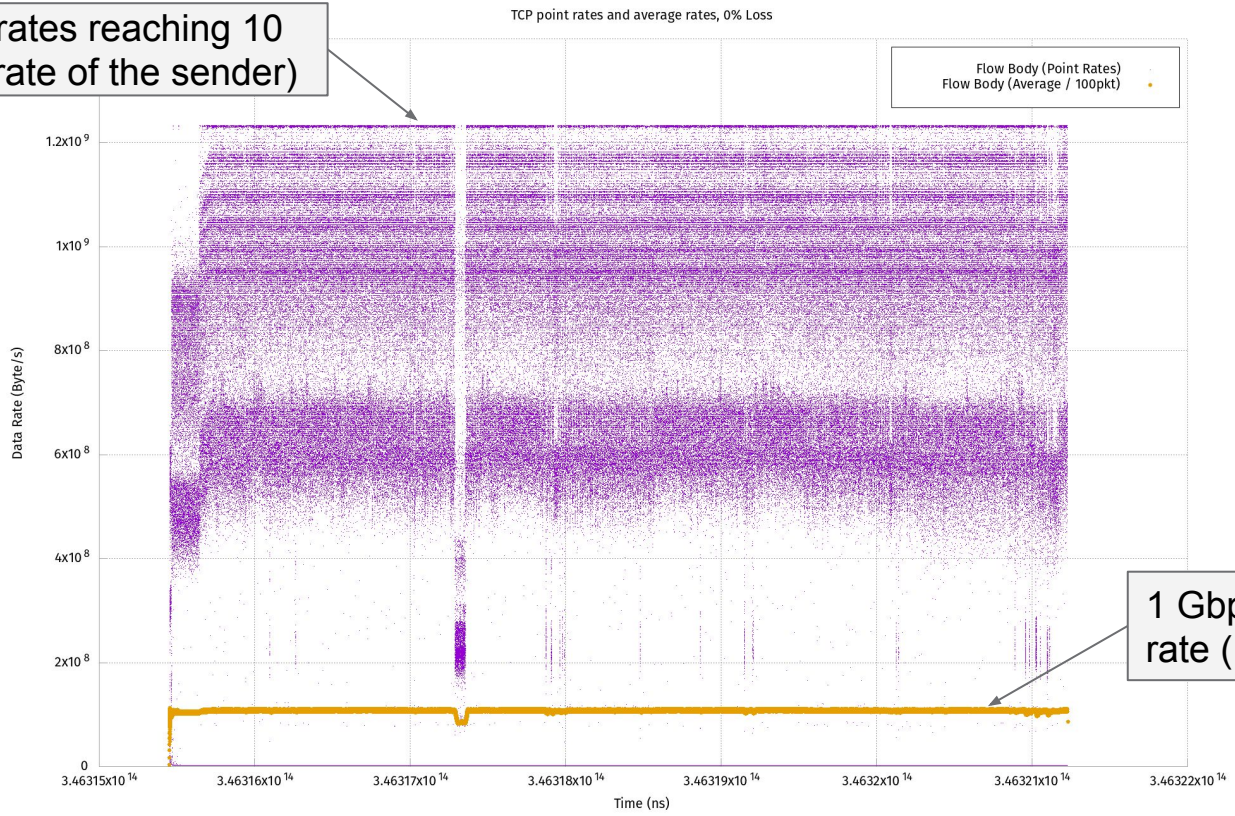


ESnet

1 Gbps iPerf Flow - 600,000 Packets



Per-packet rates reaching 10 Gbit/s (line rate of the sender)

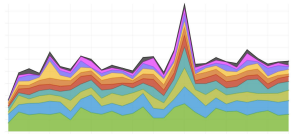


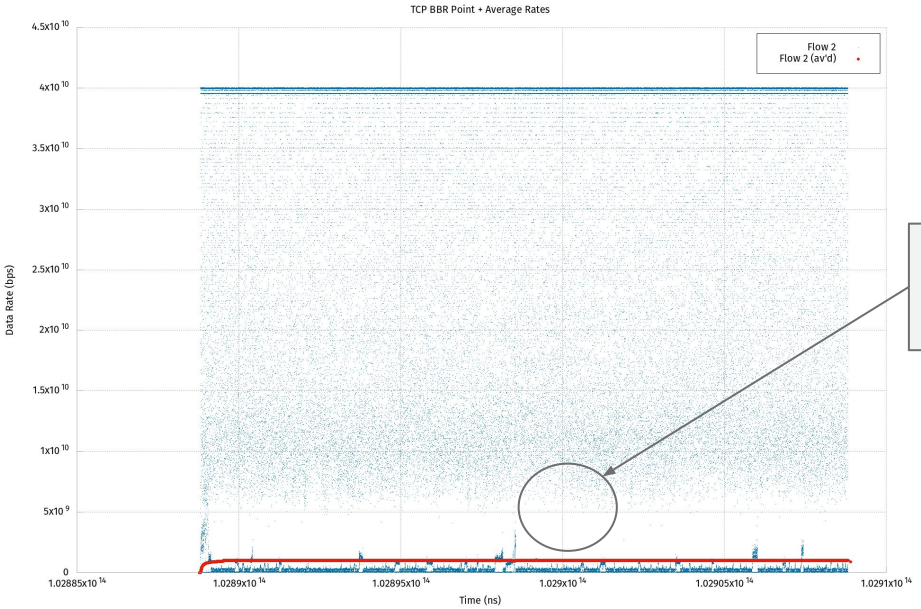
1 Gbp/s average flow rate (100 pkt window)

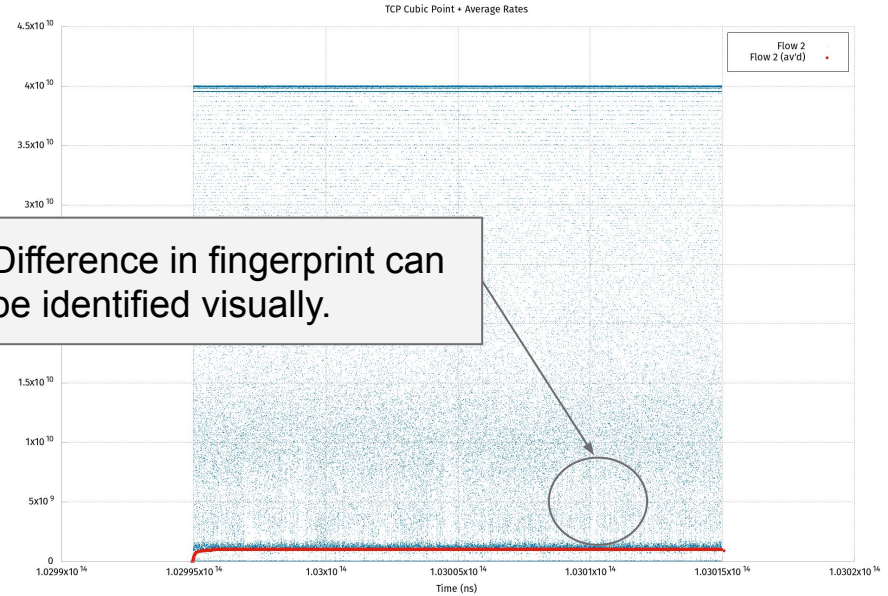
Note: Average rate is calculated using a time-weighted average of per-packet rates.

ESnet

BBR vs Cubic - Point Rates







Difference in fingerprint can be identified visually.

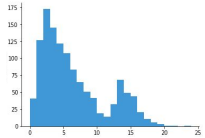
TCP BBR

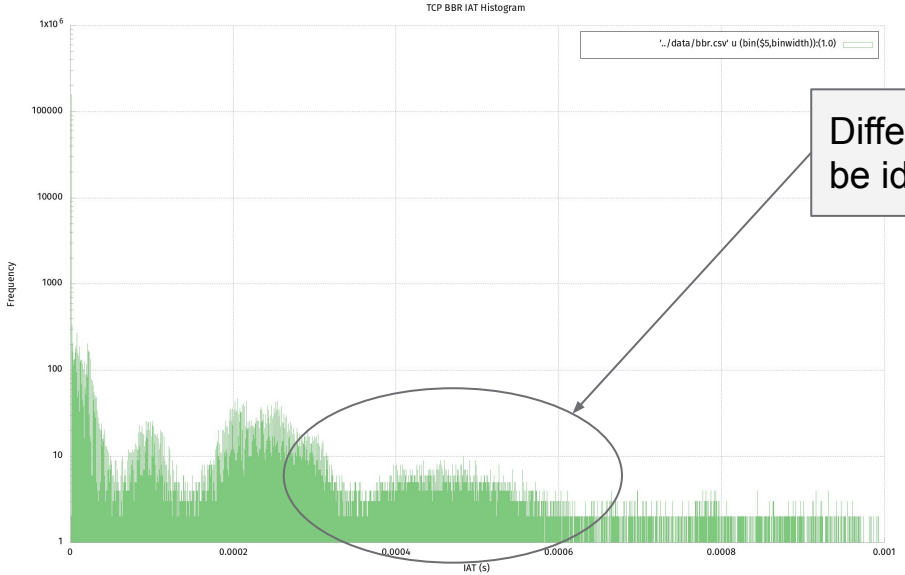
TCP Cubic

2 millions of data points shown (around 600.000 points a second generated)

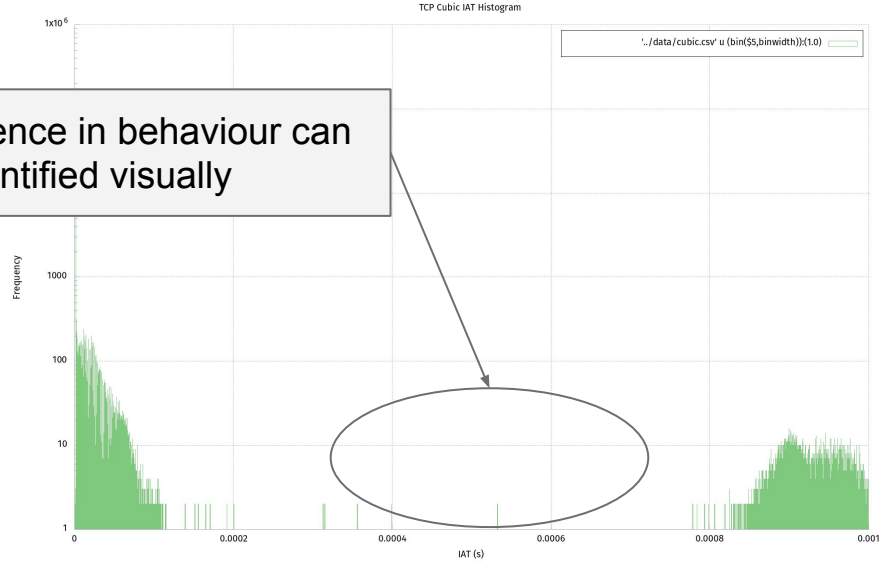
ESnet

BBR vs Cubic - Inter-Arrival Time Histogram





Difference in behaviour can be identified visually



TCP BBR (delay-based)

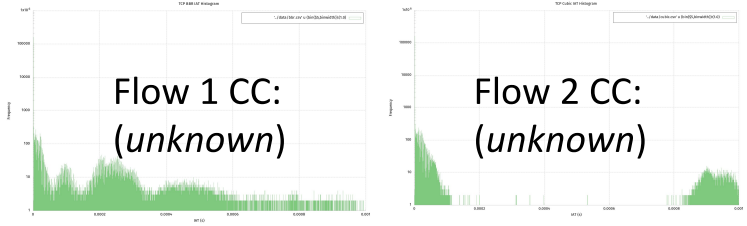
TCP Cubic (loss-based)

BBR: inter-packet timing is more widespread than other congestion control algorithms.

ESnet

Machine Learning on Aggregated Data

- Aggregated data - such as histograms can be used to tell apart congestion control (CC) used by TCP flows
- We are using data plane histograms of inter-arrival times per flow (2000 packets per histogram)
- ML algorithms explored: Convolutional Neural Networks, k-Nearest Neighbors



Input: per-flow histograms of Inter-Arrival Time (IAT)

Machine Learning
(trained with labeled data)

Inference in less than 1 ms in all cases

Flow 1 CC: most likely *TCP BBR*

Flow 2 CC: most likely *TCP RENO*

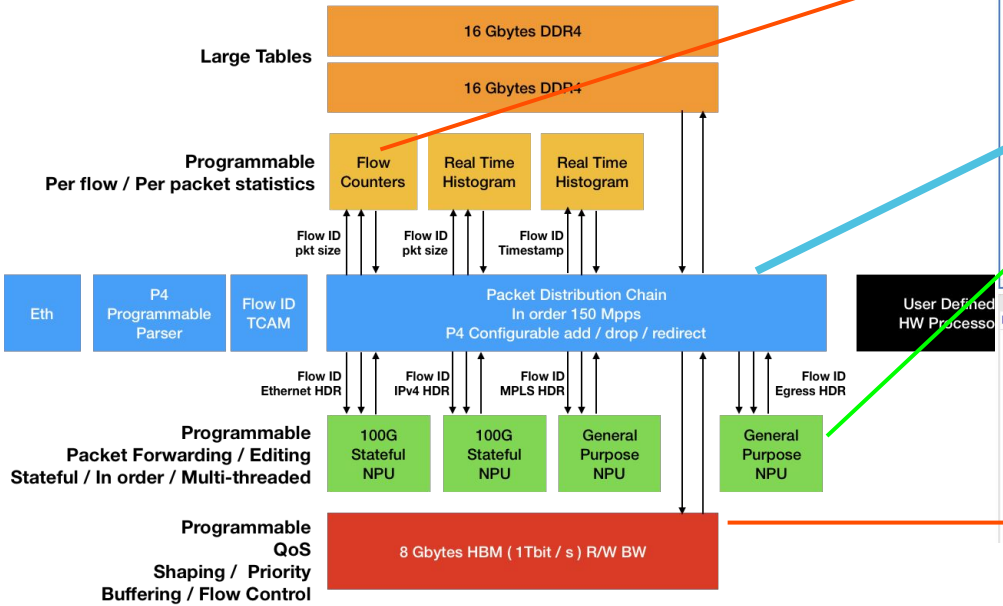
More details, dataplane architecture, ML code in:

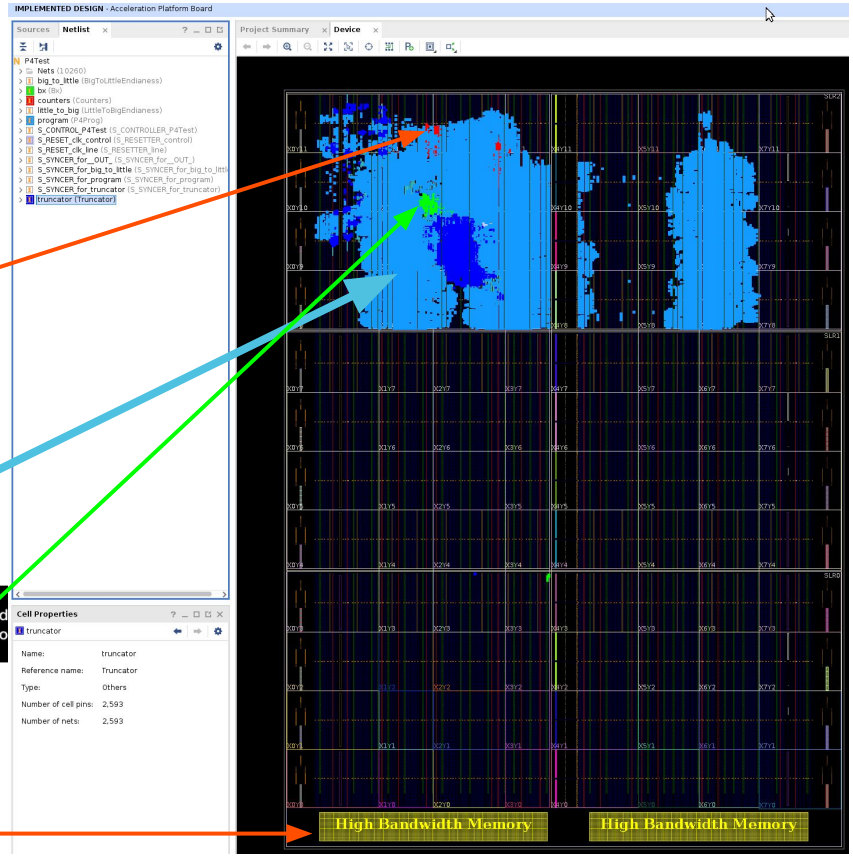
Simpson, Kyle A., Richard Cziva, and Dimitrios P. Pezaros. "Seiðr: Dataplane Assisted Flow Classification Using ML." IEEE GLOBECOM, Taipei, Taiwan (2020).

ESnet

FPGA Utilization

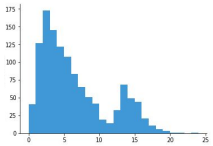
1. Plenty of room for expansion (10 % utilized)
2. P4 pipeline and TCAM are huge compared to NPU and counters.





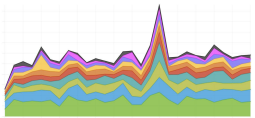
ESnet

Summary of techniques for rate reduction :



Per Flow Histograms - Counters / Sketches -- These are **polled** at Kafka friendly rates

Per Flow Single Facts - Total Bytes / (src,dst pairs) / Duration
Loss / Average Rate -- These are **polled** at Kafka friendly rates



Per Flow Reduced Time Series - Moving Ave Rate / 1:1 Bad Flows / Inter Arrival Time FFT -- These are **streamed** at Kafka friendly rates

All structures implemented in the FPGA at full line rate

ESnet

Questions ?

ESnet