DC24 High-Touch Data Analysis

Tristan Sullivan, UVic
Beijing LHCONE Meeting
Oct. 12/24

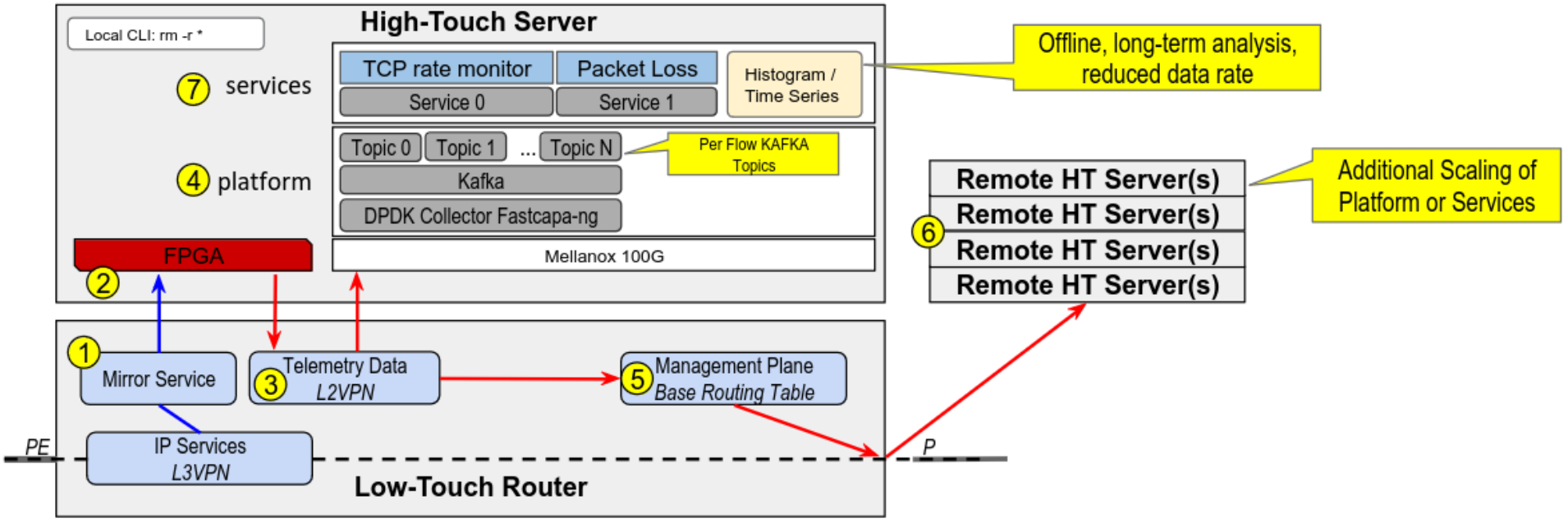
High-Touch Services

- High-precision, real-time visibility into network traffic
  - Process every packet of interest in real-time
  - Accurate, precision timing (ns precision / accuracy)
  - Software-defined functionality
  - Programmatically deployable and customizable
- In contrast to “low touch” services
  - Fixed function services such as IP packet routing, basic statistics
  - Optimized for speed and low cost, but not flexible
- Technology enablers
  - Software-defined networking
  - Programmable network dataplane hardware with accurate timestamps
  - High-speed packet processing libraries (DPDK, etc.)

Slide from ESNet SC20 Presentation

ESnet

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 - Connect Dataplane and Platform, possibly on different High-Touch Servers.
4. Platform - Reads telemetry packets from the network and distributes information to High Touch Services.
5. Management Plane Base Routing Table - Provides connectivity to Remote Servers.
6. Remote Server - Hosts Platform components or Services (but not a Dataplane). Telemetry data can be directed to Remote Servers.
7. Service - Reads data from the Platform and performs real-time analysis as well as inserts selected telemetry data into database.

Datapath of Customer Packet
Datapath of Mirrored Packet
Datapath of Telemetry Packet

ESnet

Slide from ESNet SC20 Presentation

DC24 Analysis

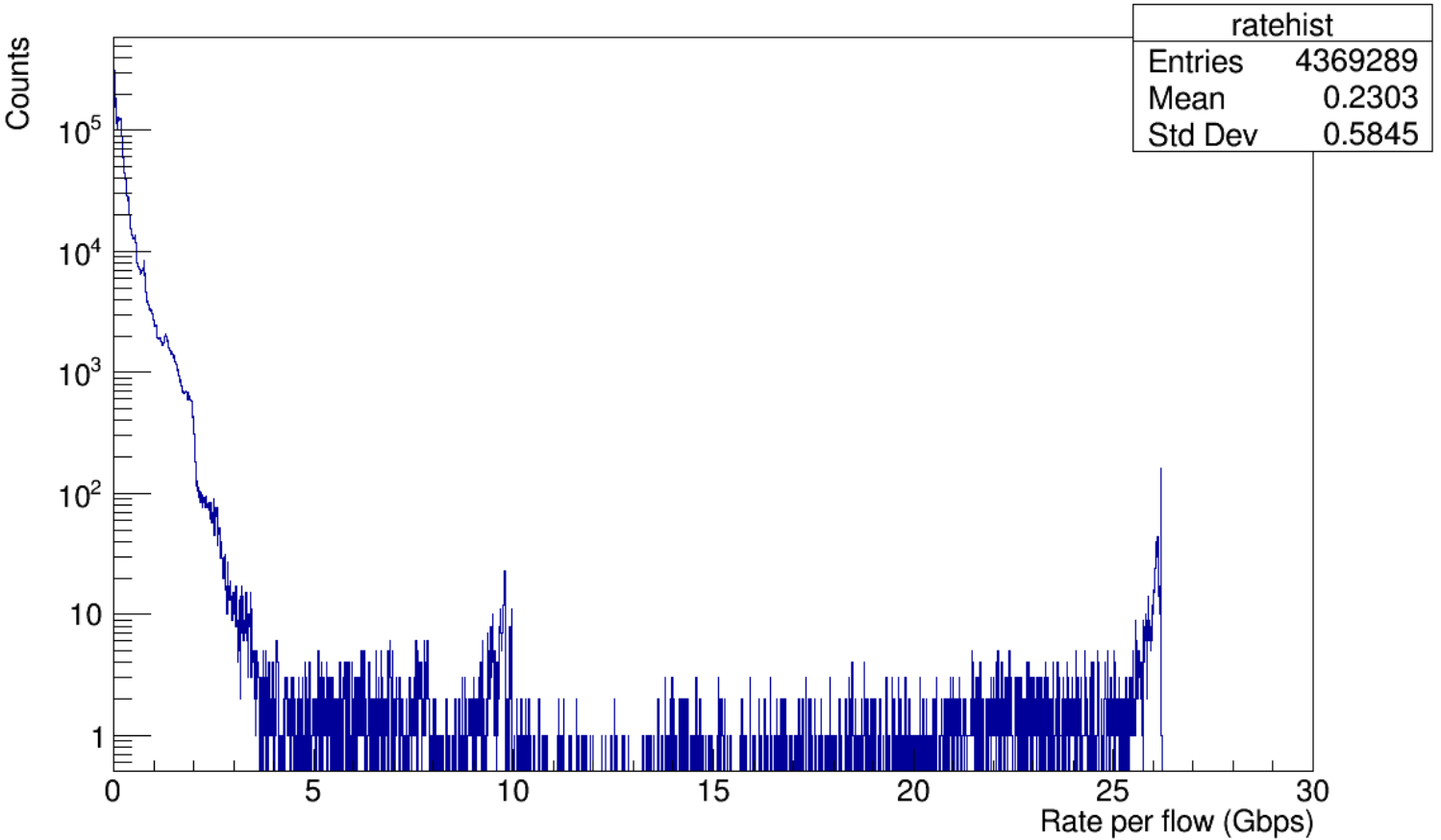
- Have full high-touch dataset from LHCONE interfaces
- Detailed information for all flows with at least one end in the US
- Difficult to isolate actual data transfers: can remove iperf3, very short/long flows, very little data moved

Bandwidth Per Flow

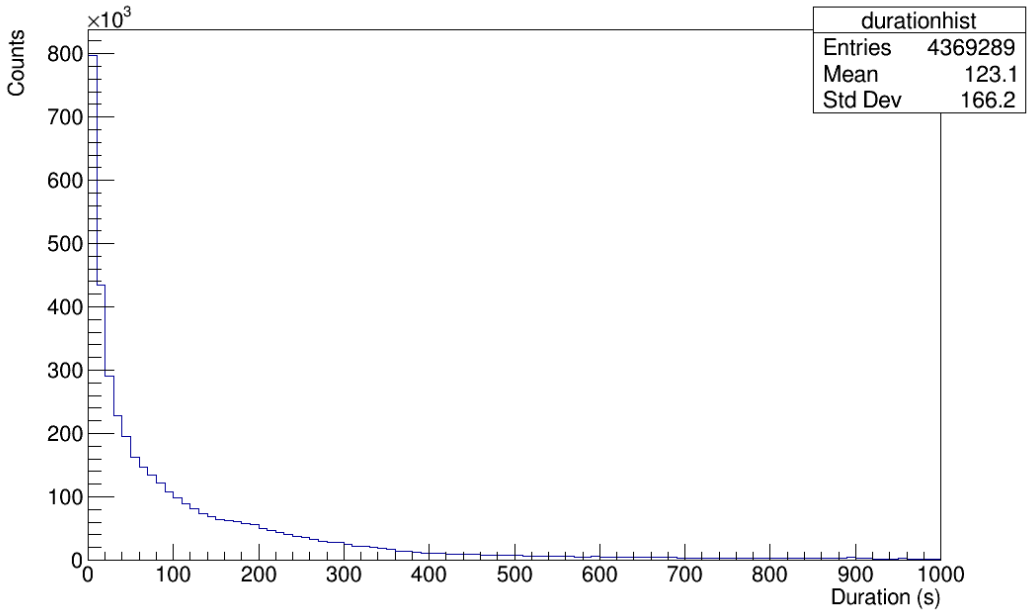
Each entry represents one complete flow

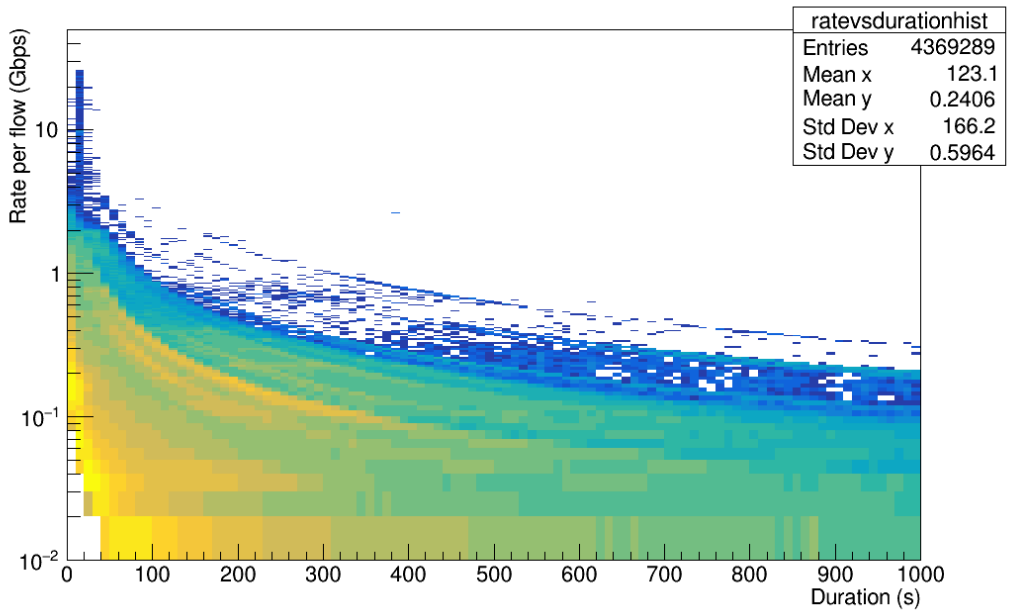
Peaked at very low bandwidth

Average is 230 Mbps



Duration Per Flow



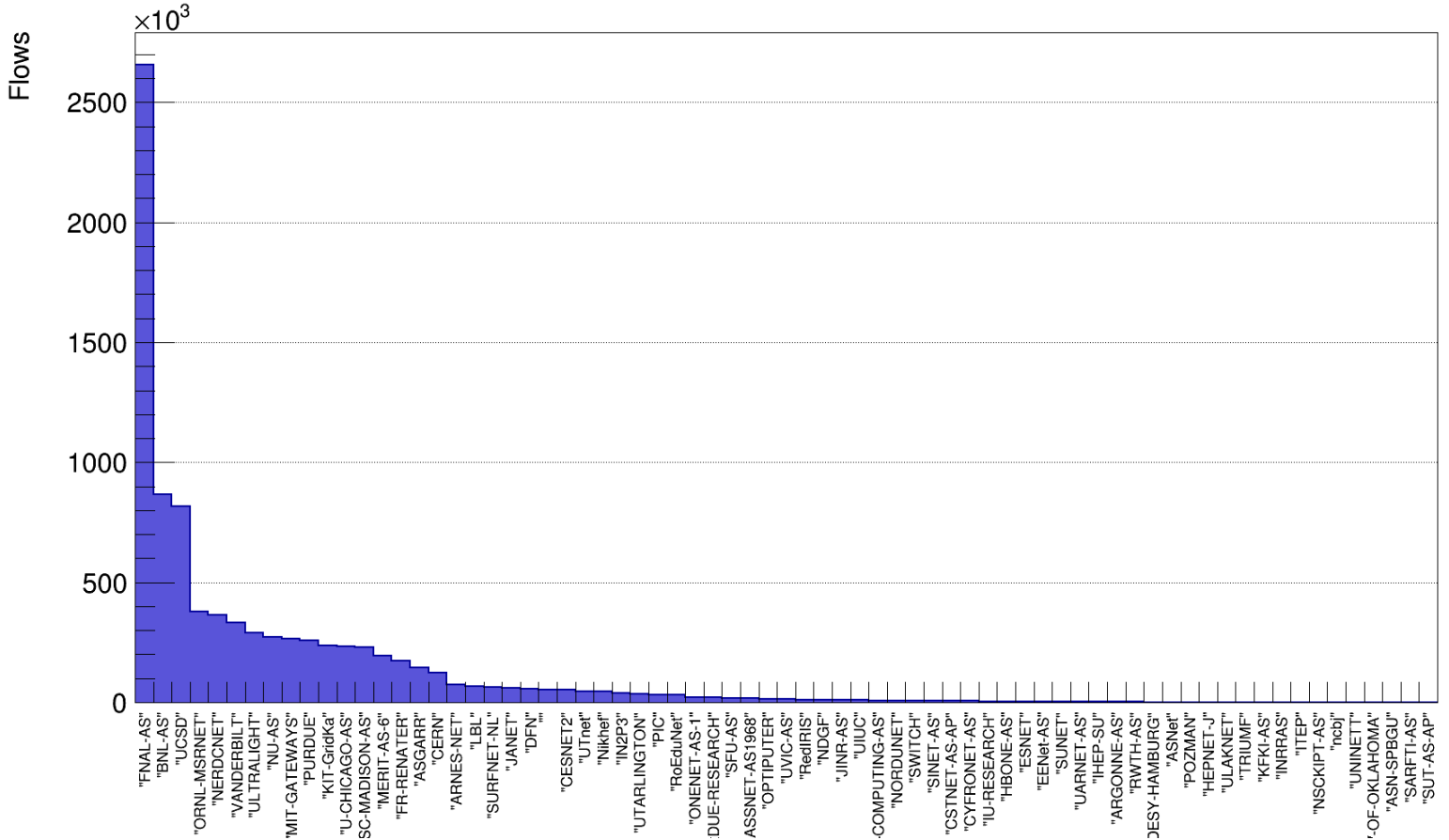


Average flow duration just over two minutes, peaked at very short flows

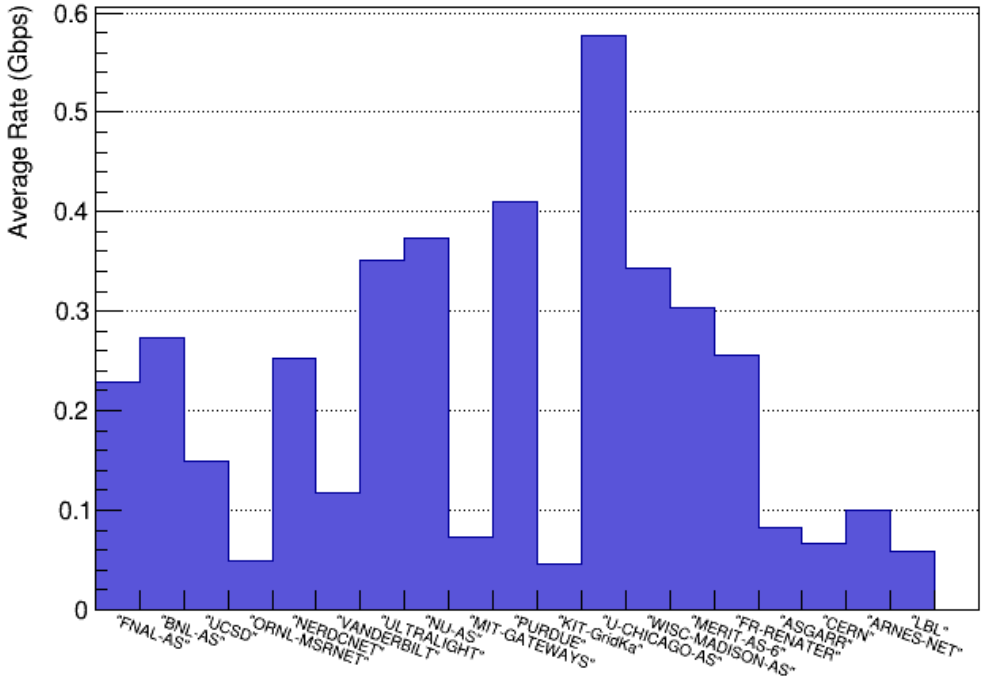
Some interesting structure in rate vs. duration

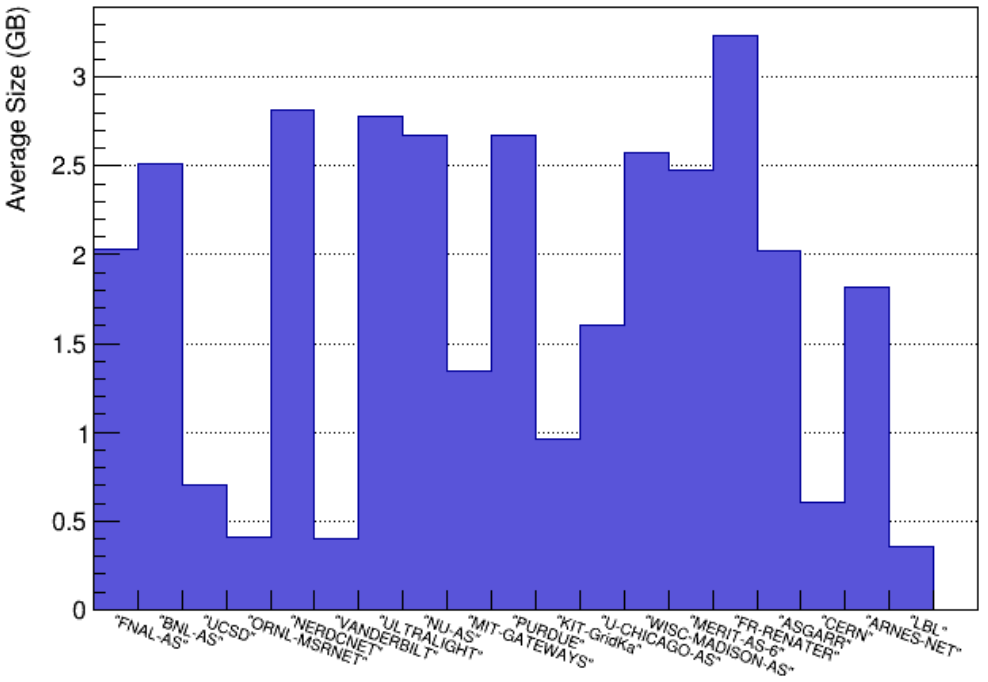
Flows By Site

Site	Flows
Total	8.74e6
FNAL	2.66e6
BNL	0.87e6
UCSD	0.82e6
ORNL	0.38e6
NERDCNET	0.37e6
VANDERBILT	0.33e6
ULTRALIGHT	0.29e6
NIU	0.27e6
MIT	0.26e6
PURDUE	0.26e6



Bandwidth and Flow Size





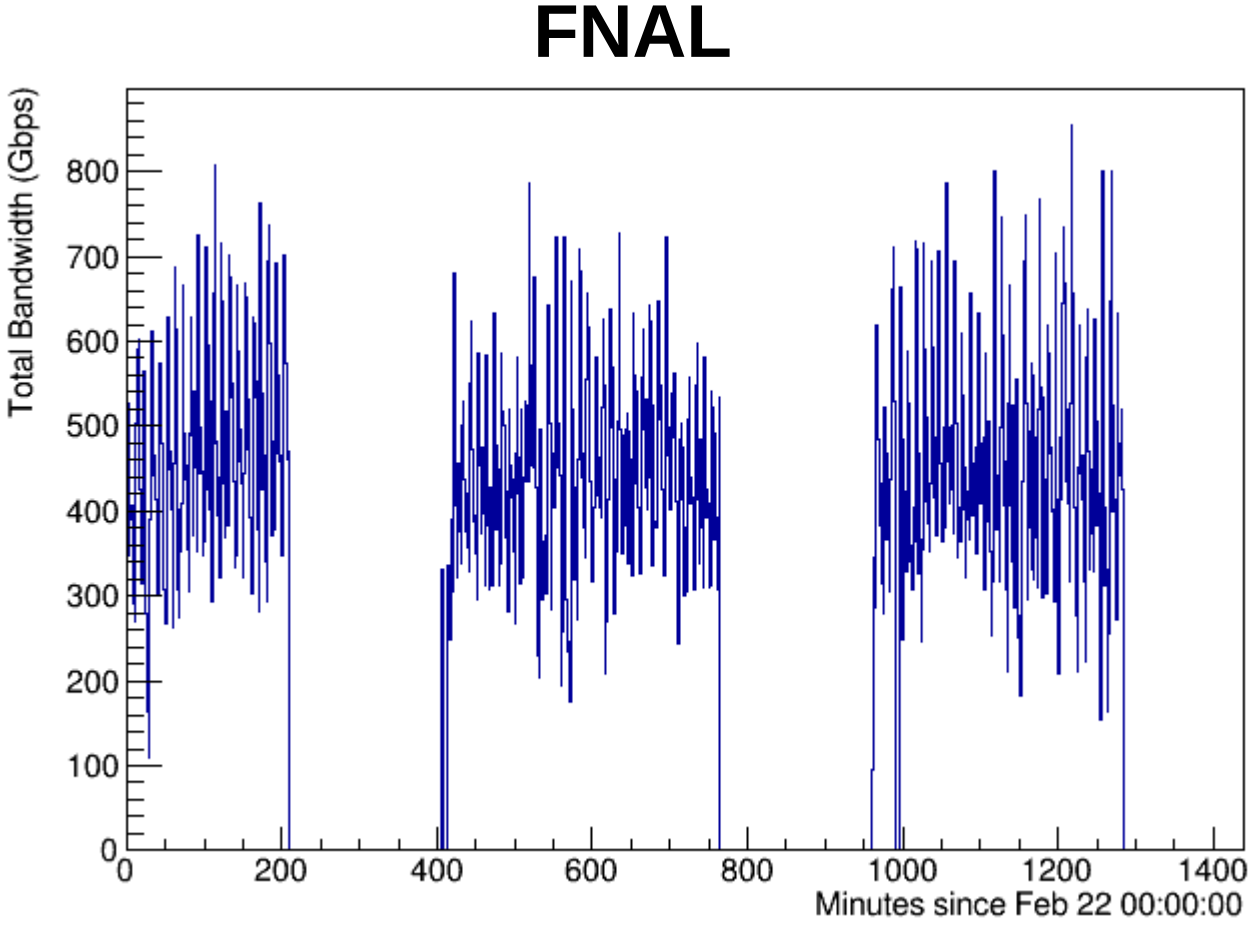
Most sites close to the average for bandwidth, some much lower

Seems to be a rough correlation with average size of flow; smaller flows are slower

Bandwidth By Site

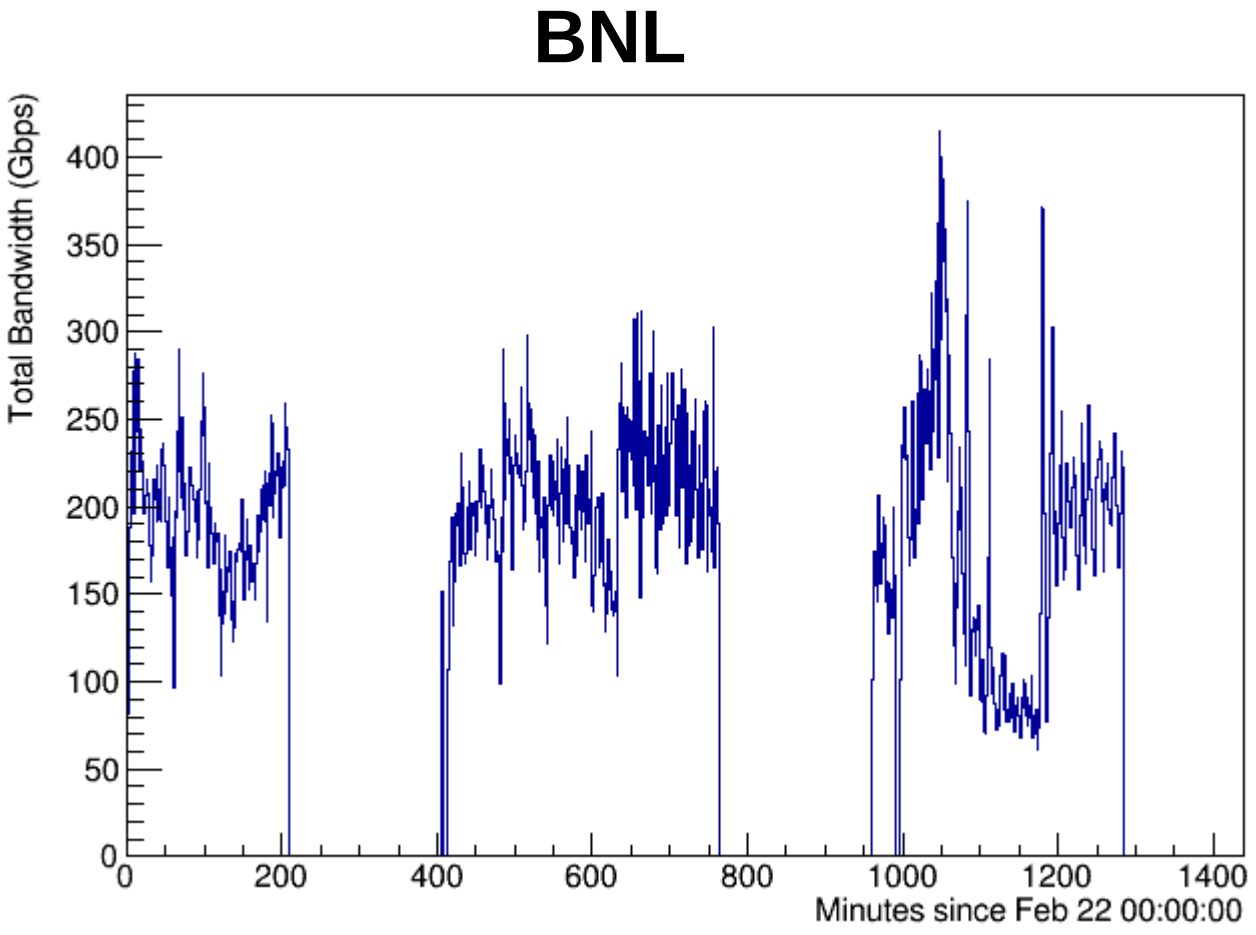
Time series of aggregate bandwidth for Feb. 22/24

Although FNAL average bandwidth is only ~200 Mbps, the aggregate peaks over 800 Gbps



Bandwidth By Site

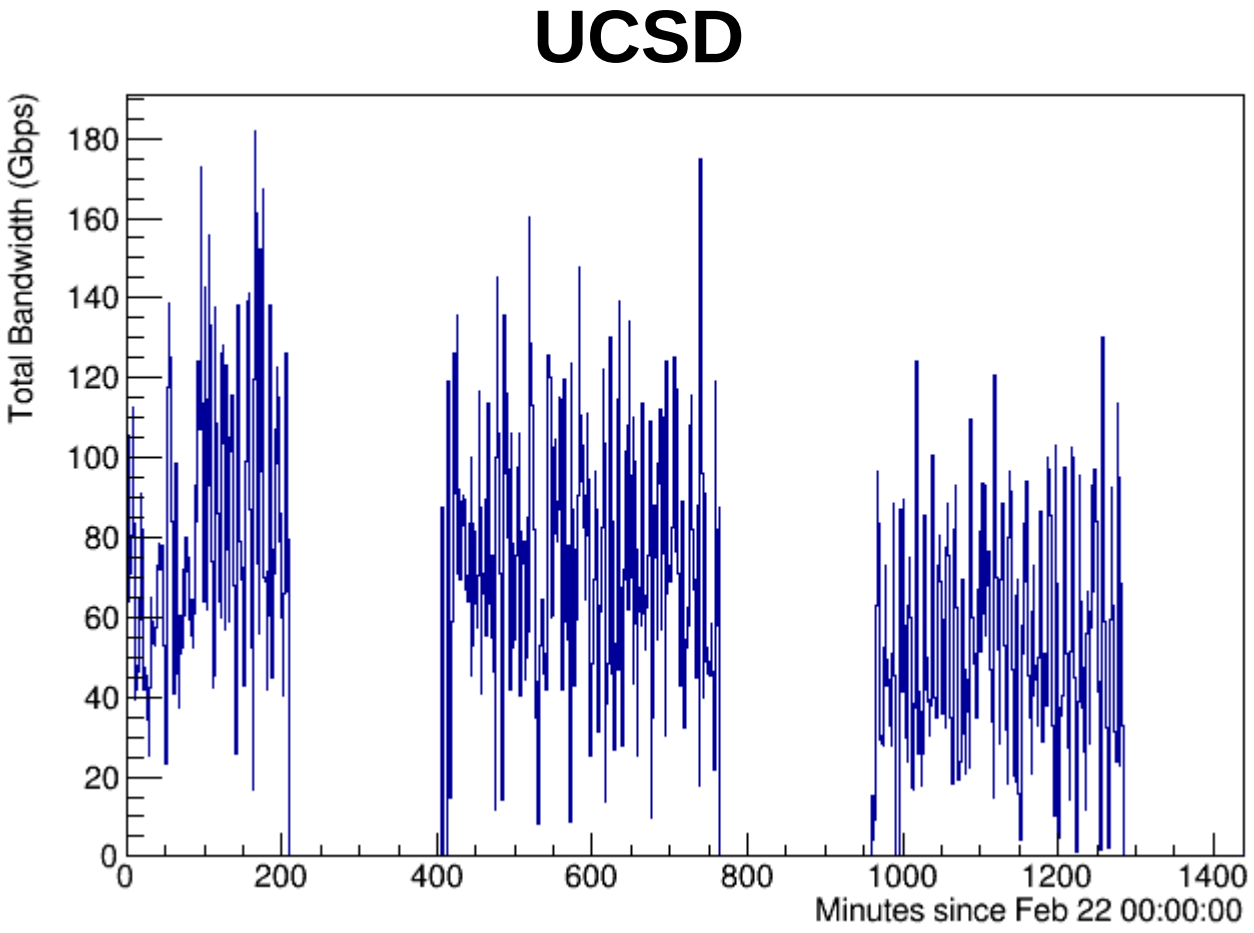
Not as steady as FNAL, but still peaks at 400 Gbps



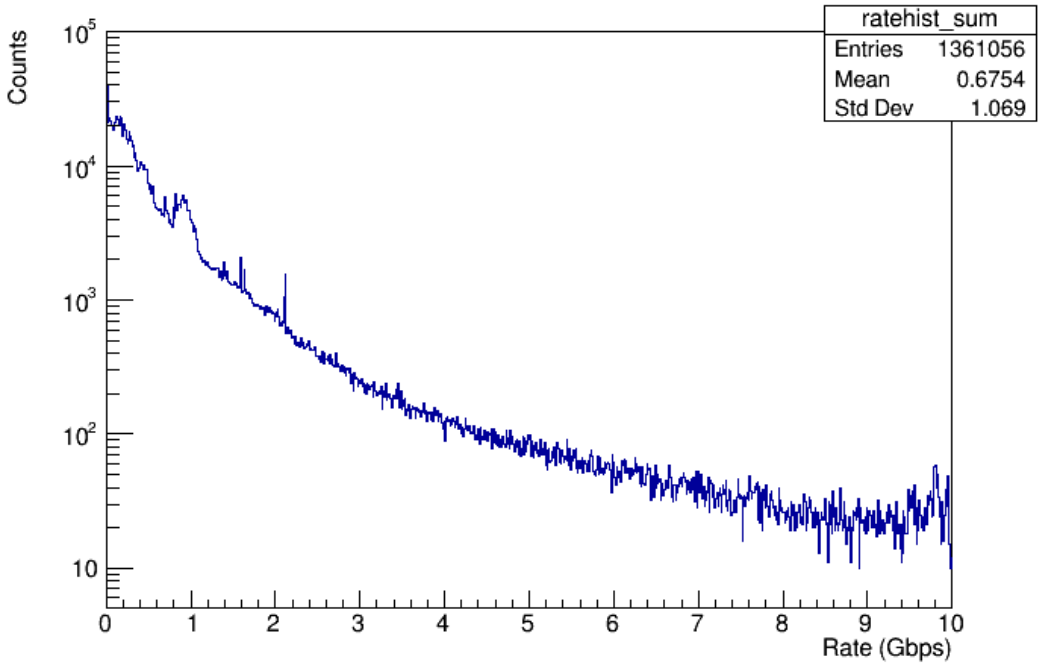
Bandwidth By Site

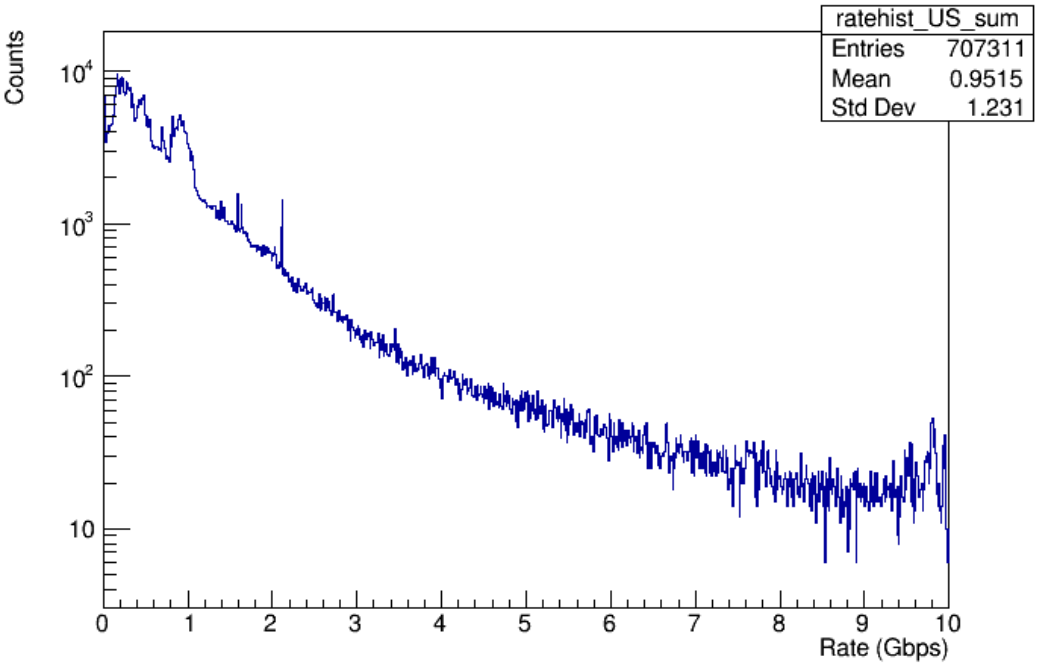
Third busiest site, a Tier-2, as opposed to FNAL and BNL

Not sure why there are gaps in the dataset



Bandwidth by Endpoint





US Only

This shows aggregate bandwidth for endpoints: if a server is simultaneously performing two transfers at 100 Mbps, have one entry at 200 Mbps, instead of two at 100 Mbps

Only have complete information for US to US transfers

Slow Endpoints

- Looking only at US sites greatly reduces the number of very slow endpoints, but there are still some
- Remainder seems to be mainly a combination of CVMFS, worker nodes, and storage nodes at very busy sites, e.g. FNAL, BNL, UCSD

Conclusions

- Individual flows have low bandwidth, but aggregate is large; expected given computing model
- Difficult to isolate actual file transfers between storage elements
- Just scratching the surface of this dataset; feedback on further studies very welcome!