ESnet
ENERGY SCIENCES NETWORK

ESnet High-Touch Telemetry Platform

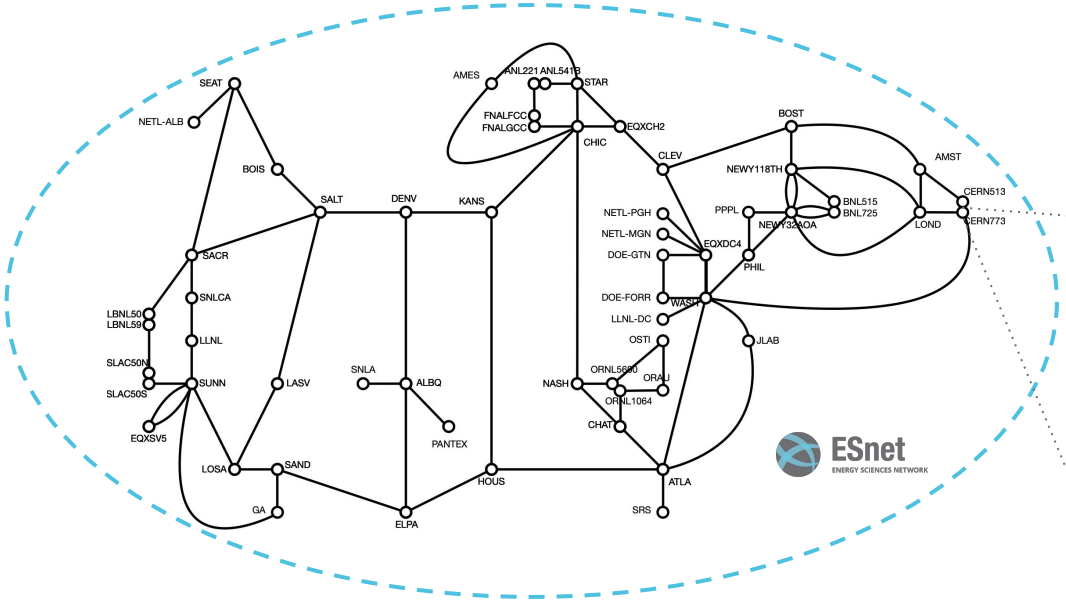
Chin Guok
Chief Technology Officer
Scientific Networking Division
Lawrence Berkeley National Laboratory

4GRP Workshop
Limassol, Cyprus
Oct 10, 2023

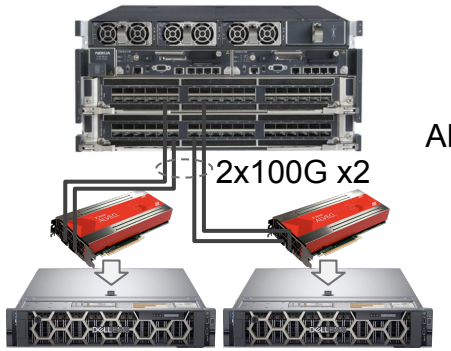
U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

ESnet6 High-Touch Platform Field Deployment



High-Touch Server Hardware Deployment



ESnet6 Core Router

AMD Xilinx Alveo U280 FPGA (1 per server)

FPGA Servers (2 per hub site)

- 42 deployment locations
- Near 100% perimeter coverage*
- 100% packet inspection

ESnet

*NB: Certain ports are omitted due to security sensitivities.

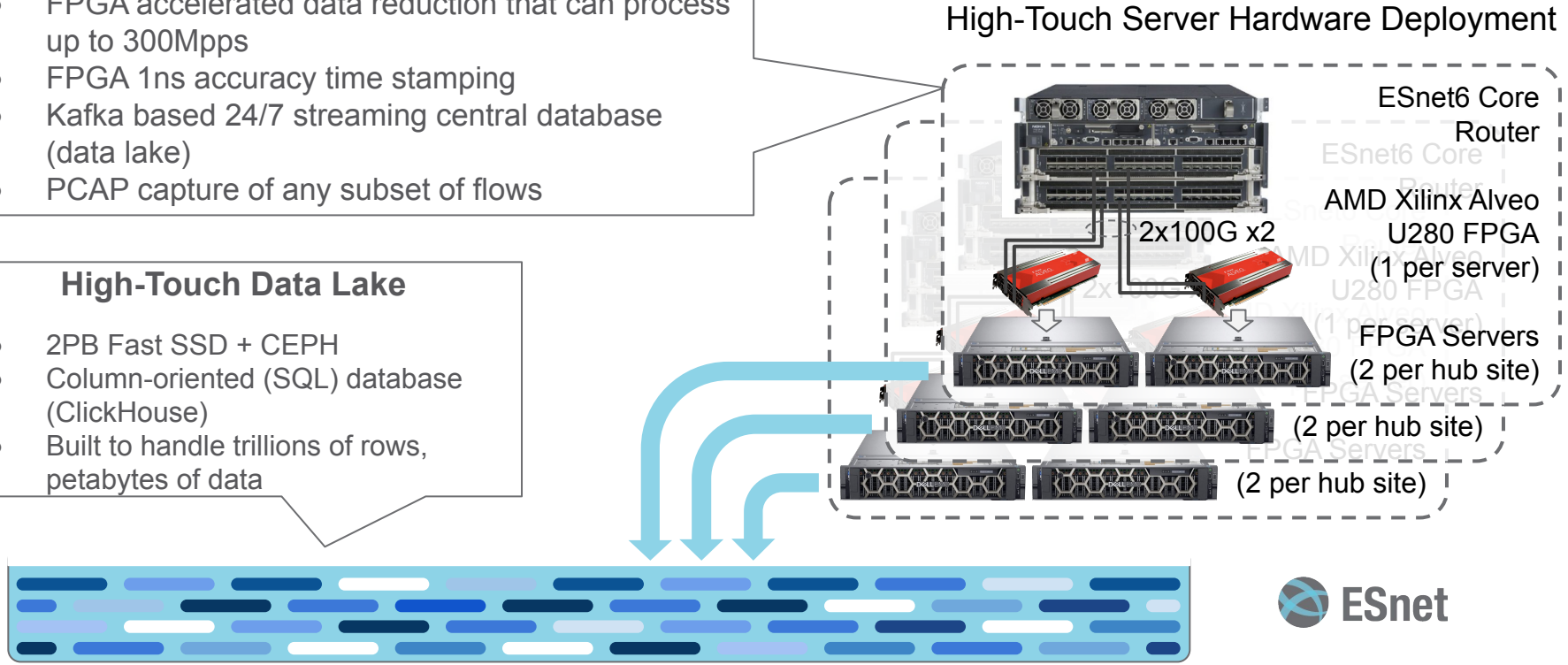
ESnet6 High-Touch System Deployment

High-Touch Platform

- Router packet mirroring with (128 byte) packet truncation at line rate
- FPGA accelerated data reduction that can process up to 300Mpps
- FPGA 1ns accuracy time stamping
- Kafka based 24/7 streaming central database (data lake)
- PCAP capture of any subset of flows

High-Touch Data Lake

- 2PB Fast SSD + CEPH
- Column-oriented (SQL) database (ClickHouse)
- Built to handle trillions of rows, petabytes of data



ESnet

Flexibility in Tool Selection

SQL CLI

```
SELECT
    exporting_node,
    count(*) AS total_records
FROM ht.all_flows
GROUP BY exporting_node
FORMAT PrettyCompactMonoBlock

Query id: 3101a6b0-9ae6-4e94-b78c-d
```

exporting_node	total_records
bnl515-ht1	282710771
bost-ht2	1104520
eqxch2-ht2	1809320
slac50s-ht1	8545951
salt-ht2	568492
lbnl59-ht2	882130013
newy1118th-ht2	1071826
elpa-ht2	480642
ornl5600-ht1	45841556
atla-ht2	772929
eqxsv5-ht1	2835203590
anl541b-ht2	924414
slac50n-ht1	2221787835
nash-ht2	696555
atla-ht1	108835837
bois-ht2	432433
anl541b-ht1	65303026
eqxsv5-ht2	1189111
ornl5600-ht2	721476
newy1118th-ht1	2991159948
sand-ht2	387820
bnl515b-ht2	829636

Jupyter Lab / Pandas

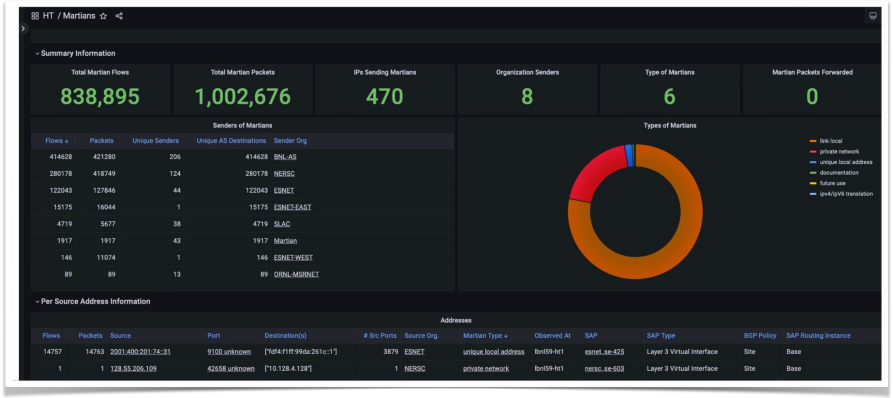
```
[9]: # Quick preview to see what the data frame looks like
     df.drop_duplicates(subset='flow_id')
```

[9]:

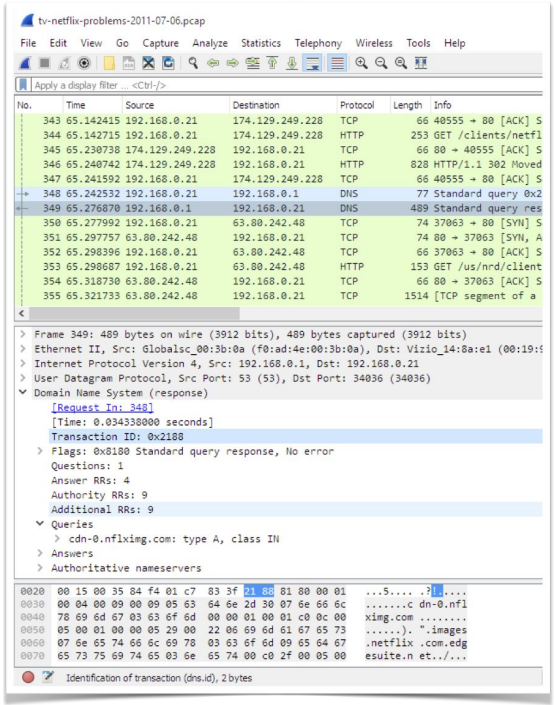
	dst	dport	src	sport	proto	time	dst_org	src_org	src_network	dst_network	flow_id
0	104.196.237.25	42948	198.129.224.35	80	6	2022-02-15T11:09:59.994-0800	UNDEF	ESNET	198.129.224.0/24	104.196.237.0/24	1
1	134.79.146.247	34432	45.137.21.208	48138	6	2022-02-15T11:09:59.984-0800	SLAC	UNDEF	45.137.21.0/24	134.79.146.0/24	2
2	134.79.25.243	33566	92.63.196.25	47014	6	2022-02-15T11:09:59.974-0800	SLAC	UNDEF	92.63.196.0/24	134.79.25.0/24	3
3	198.128.14.236	48690	198.124.155.24	22	6	2022-02-15T11:09:59.974-0800	ESNET	ESNET	198.124.155.0/24	198.128.14.0/24	4
4	128.3.18.26	53422	128.55.136.54	27017	6	2022-02-15T11:09:59.964-0800	LBNL	NERSC	128.55.136.0/24	128.3.18.0/24	5
...	...	...	...	...	...	...	...	...	...	...	...
166483	128.55.244.94	50264	89.248.168.172	56292	6	2022-02-15T10:20:00.504-0800	NERSC	UNDEF	89.248.168.0/24	128.55.244.0/24	86741
166484	131.225.205.56	34749	128.55.224.115	57714	6	2022-02-15T10:20:00.474-0800	FNAL	NERSC	128.55.224.0/24	131.225.205.0/24	86742
166487	84.220.141.235	57150	128.55.206.106	443	6	2022-02-15T10:20:00.404-0800	UNDEF	NERSC	128.55.206.0/24	84.220.141.0/24	86743
166493	128.55.109.9	15529	112.31.169.97	4183	6	2022-02-15T10:20:00.093-0800	NERSC	UNDEF	112.31.169.0/24	128.55.109.0/24	86744
166494	198.129.217.96	57599	173.194.152.170	443	17	2022-02-15T10:20:00.083-0800	SLAC	UNDEF	173.194.152.0/24	198.129.217.0/24	86745

86745 rows × 11 columns

Grafana / Stardust



Wireshark



ESnet

Internet Background Radiation

All the dst ports that get a SYN and no further packets

All the ports that accept a TCP-SYN and return a SYNACK

Organization name

# of IPs used to probe

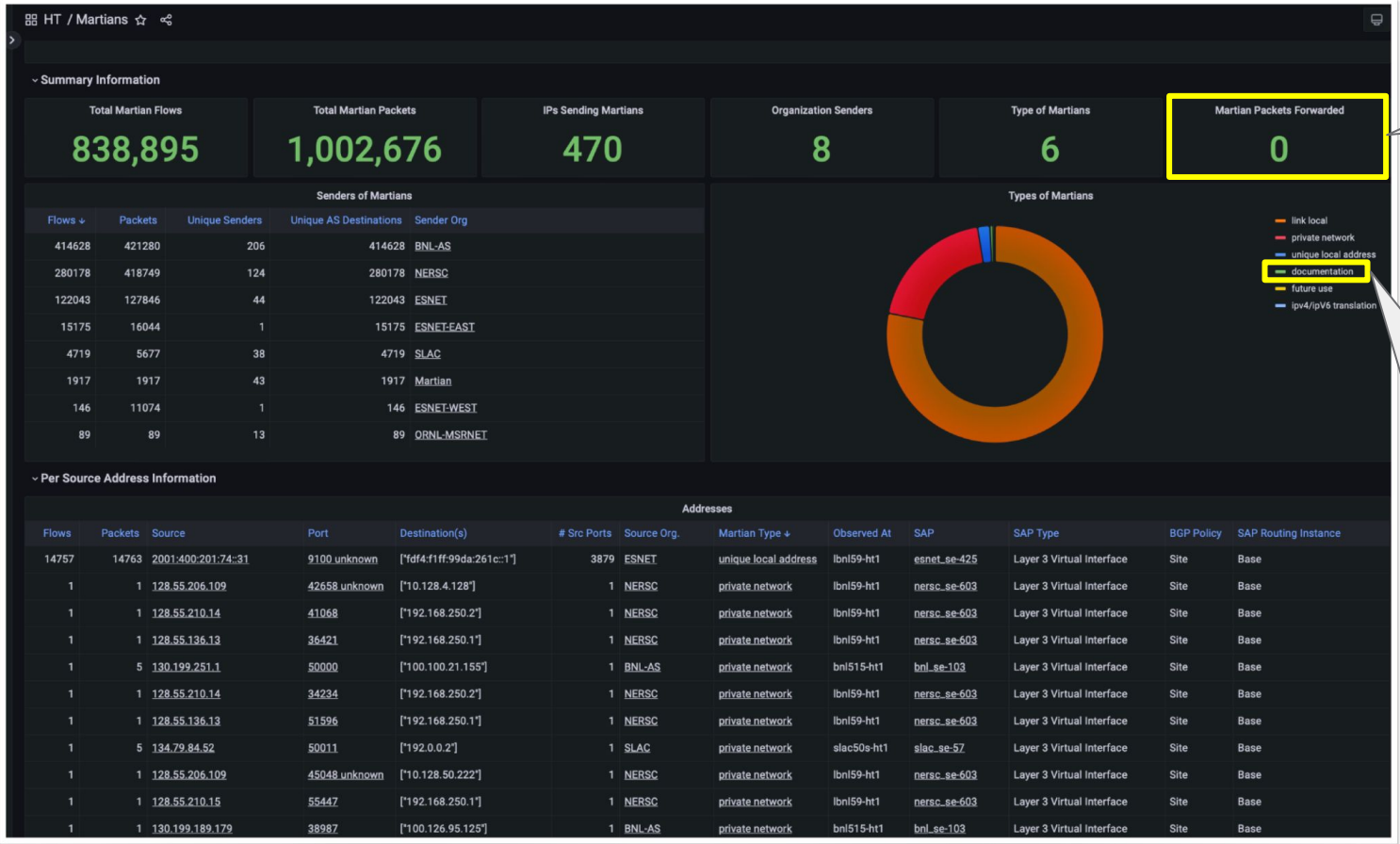
Comparison of probes that generated a bi-directional syn_ack when sent a syn probe

The effectiveness of syn_only probes can be checked against the number of syn_ack responses they generate. In this diagram, we plot the total number of syn … ne below, we plot only the probes that got a syn_ack message in response. This provides a measure of hosts, and ports that are successfully enumerated by the scanner.

- By looking at the Number of IP's probed bar chart, we can see that only a small number of hosts being scanned, are returning any respons…
- By looking at the number of source IPs we can see that some of the scanners source IPs's are completely unsuccessful at getting a…

[7] … di , df] , 'top' , 25)

Network Audits - Martians



Bogon filtering at work!

Why do we have “documentation” (e.g., TEST-NET) addresses?

- **SONIC** - Open Source Switches
- **Docker** - Data Center Auto Config

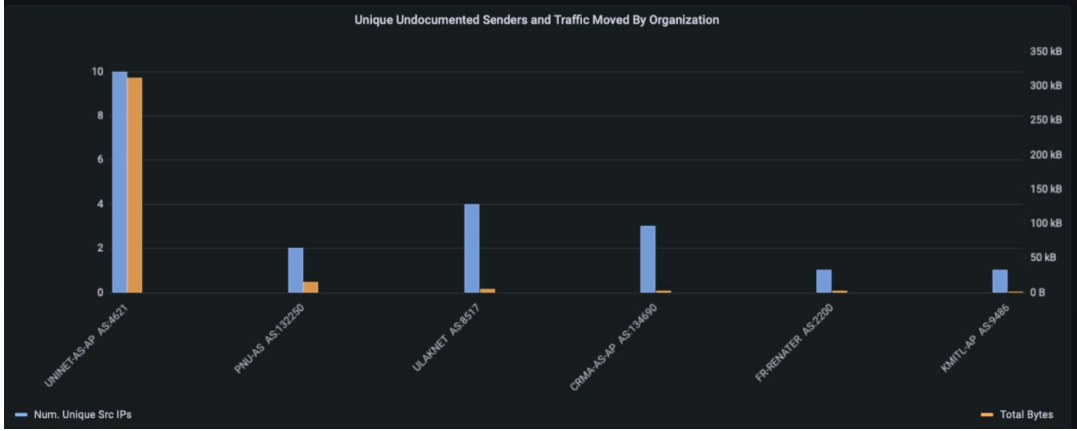
ESnet

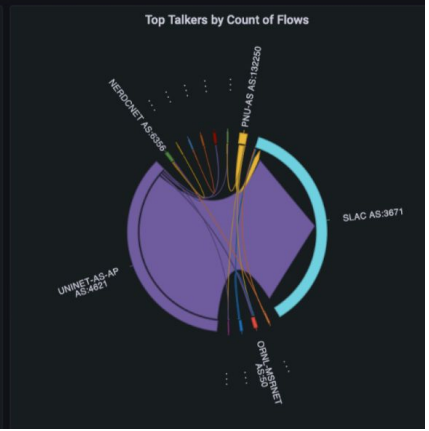
LHCONE Traffic - CRIC Audit

Undocumented Sources to Documented Destinations

LHC One Undocumented Sources

Num. Flows	Total Bytes	Router	Interface	Src ASN	Src Org.	Dst IP	Dst ASN	Dst Org.
39	1.56 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.24.197	3671	SLAC
38	1.52 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.110.7	3671	SLAC
38	1.52 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.60.133	3671	SLAC
38	1.52 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.244.2	3671	SLAC
38	1.56 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.188.23	3671	SLAC
37	1.48 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.22.48	3671	SLAC
36	1.44 kB	atla-cr6	1/1/c3/1	4621	UNINET-AS-AP	128.227.246.204	6356	NERDCNET
36	1.44 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.140.185	3671	SLAC
36	1.48 kB	slac50s-cr6	1/1/c2/1	4621	UNINET-AS-AP	134.79.14.233	3671	SLAC





WLCG CRIC
Database for LHCONE prefixes (wlcg-cric.cern.ch)

JSON Query

ESnet High-Touch Data lake

SQL Query

Select * NOT in JSON

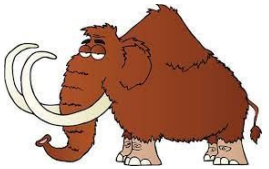
[]

Should be empty set

ESnet

Y. Kumar at el

Network Planning - Flow Volumes



SELECT * WHERE peak_rate > 10gbps AND duration > 10secs ORDER BY rate

	caida_org_name_src	caida_org_name_dst	ip_src	ip_dst	Gbps	hostname_src	hostname_dst
0	U-CHICAGO-AS	ARGONNE-AS	192.170.224.134	140.221.68.2	30.037561	scidmz-ps4.scidmz.uchicago.net.	typhoon.pub.alcf.anl.gov
1	ARGONNE-AS	U-CHICAGO-AS	140.221.68.2	192.170.224.134	27.532194	typhoon.pub.alcf.anl.gov.	scidmz-ps4.scidmz.uchicago.net
2	ESNET	ESNET	2001:400:f010:200::1	2001:400:f010:240::1	26.215328	eqxch2-ps-tp.lhcone.es.net.	fnalfcc-ps-tp.lhcone.es.net
3	ESNET	ESNET	2001:400:ee00:20::1	2001:400:ee00:21::1	26.209250	lbnl59-ps-tp.es.net.	lbnl50-ps-tp.es.net
4	ESNET	ESNET	2001:400:f010:640::1	2001:400:f010:641::1	26.208939	bnl515-ps-tp.lhcone.es.net.	bnl515b-ps-tp.lhcone.es.net
5	ESNET	ESNET	2001:400:ee00:880::1	2001:400:ee00:881::1	26.208344	ornl1064-ps-tp.es.net.	ornl5600-ps-tp.es.net
6	ESNET	ESNET	2001:400:ee00:221::1	2001:400:ee00:220::1	26.208284	anl541b-ps-tp.es.net.	anl221-ps-tp.es.net
7	ESNET	ESNET	2001:400:ee00:881::1	2001:400:ee00:880::1	26.207954	ornl5600-ps-tp.es.net.	ornl1064-ps-tp.es.net
8	ESNET	ESNET	2001:400:ee00:601::1	2001:400:ee00:600::1	26.207889	newy1118th-ps-tp.es.net.	newy32aoa-ps-tp.es.net
9	ESNET	ESNET	2001:400:ee00:881::1	2001:400:ee00:882::1	26.207831	ornl5600-ps-tp.es.net.	orau-ps-tp.es.net
10	ESNET	ESNET	2001:400:ee00:200::1	2001:400:ee00:201::1	26.206976	eqxch2-ps-tp.es.net.	chic-ps-tp.es.net
11	ESNET	ESNET	2001:400:f010:200::1	2001:400:f010:221::1	26.206912	eqxch2-ps-tp.lhcone.es.net.	anl541b-ps-tp.lhcone.es.net
12	ESNET	ESNET	2001:400:ee00:200::1	2001:400:ee00:202::1	26.206903	eqxch2-ps-tp.es.net.	star-ps-tp.es.net
13	ESNET	ESNET	2001:400:ee00:882::1	2001:400:ee00:881::1	26.206468	orau-ps-tp.es.net.	ornl5600-ps-tp.es.net
14	ESNET	ESNET	2001:400:f010:240::1	2001:400:f010:221::1	26.206126	fnalfcc-ps-tp.lhcone.es.net.	anl541b-ps-tp.lhcone.es.net
15	ESNET	ESNET	2001:400:ee00:200::1	2001:400:ee00:220::1	26.205755	eqxch2-ps-tp.es.net.	anl221-ps-tp.es.net
16	ESNET	ESNET	2001:400:ee00:240::1	2001:400:ee00:221::1	26.205489	fnalfcc-ps-tp.es.net.	anl541b-ps-tp.es.net
17	ESNET	ESNET	2001:400:f010:221::1	2001:400:f010:220::1	26.204826	anl541b-ps-tp.lhcone.es.net.	anl221-ps-tp.lhcone.es.net
18	ESNET	ESNET	2001:400:f010:200::1	2001:400:f010:220::1	26.204172	eqxch2-ps-tp.lhcone.es.net.	anl221-ps-tp.lhcone.es.net
19	ESNET	ESNET	2001:400:ee00:220::1	2001:400:ee00:200::1	26.203990	anl221-ps-tp.es.net.	eqxch2-ps-tp.es.net
20	ESNET	ESNET	2001:400:f010:241::1	2001:400:f010:200::1	26.203445	fnalgcc-ps-tp.lhcone.es.net.	eqxch2-ps-tp.lhcone.es.net

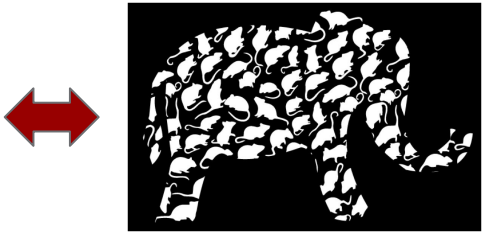
Primarily perfSONAR transfers

Observation

- perf5ONAR traffic appears to be the primary contributor of “large” flows in this dataset.

Supposition*

- Large data movement tools (e.g., Globus, etc) utilize massively parallel small flows to reduce the impact of packet loss.

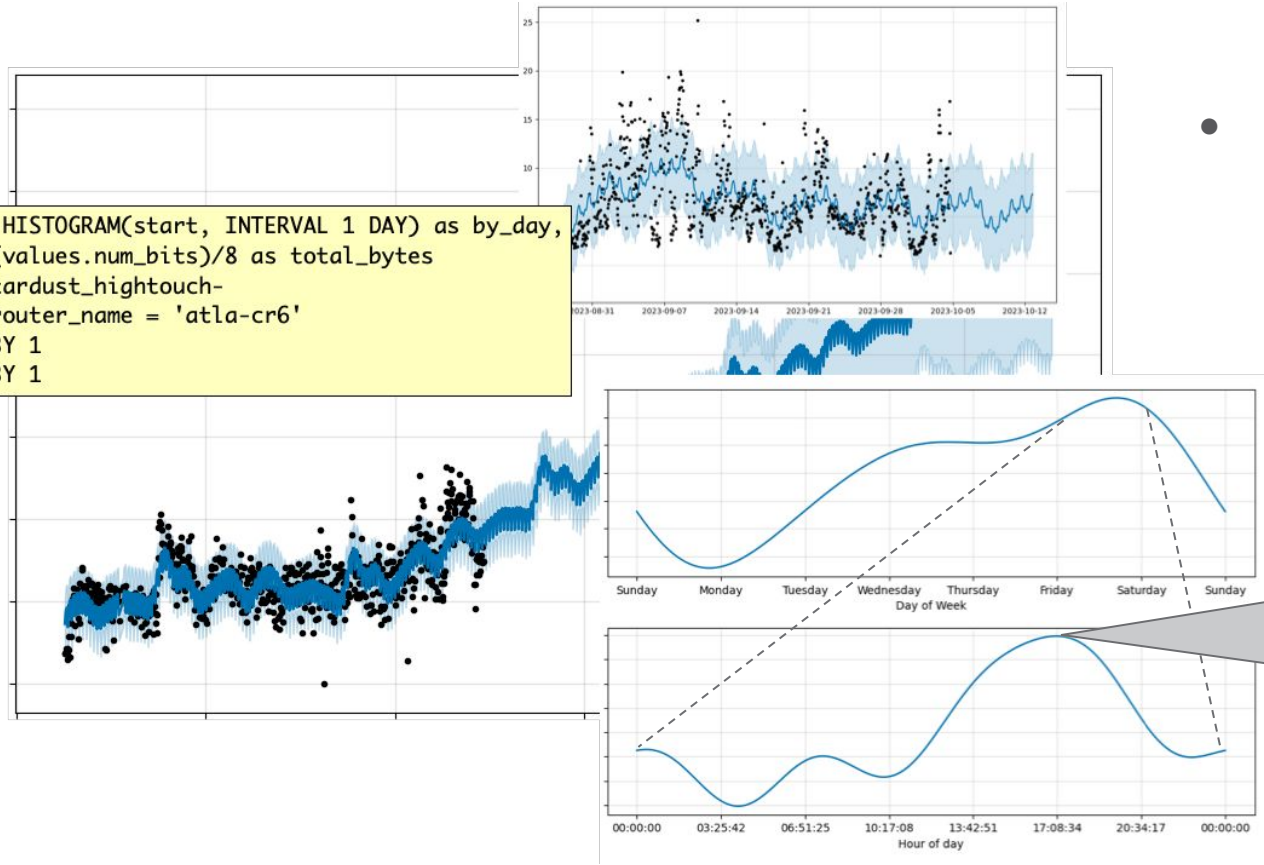


ESnet

*NB: Findings based on flow analysis on a single 24hr-period dataset, further investigation is needed to quantify the finding.

Trend Analysis / (Peak) Capacity Planning


```
SELECT HISTOGRAM(start, INTERVAL 1 DAY) as by_day,
   sum(values.num_bits)/8 as total_bytes
FROM stardust_hightouch-
WHERE router_name = 'atla-cr6'
GROUP BY 1
ORDER BY 1
```

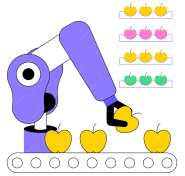


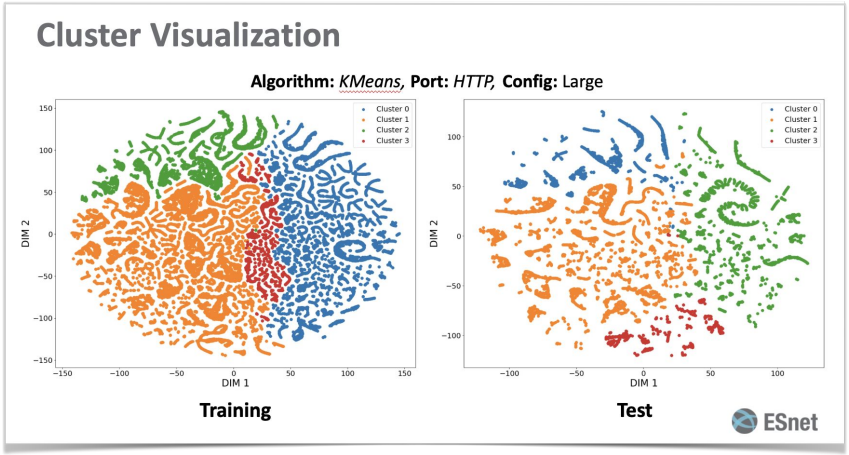
- High-Touch data is exported into existing measurement systems (i.e., Stardust), allow us to leverage existing analysis, but with much higher fidelity

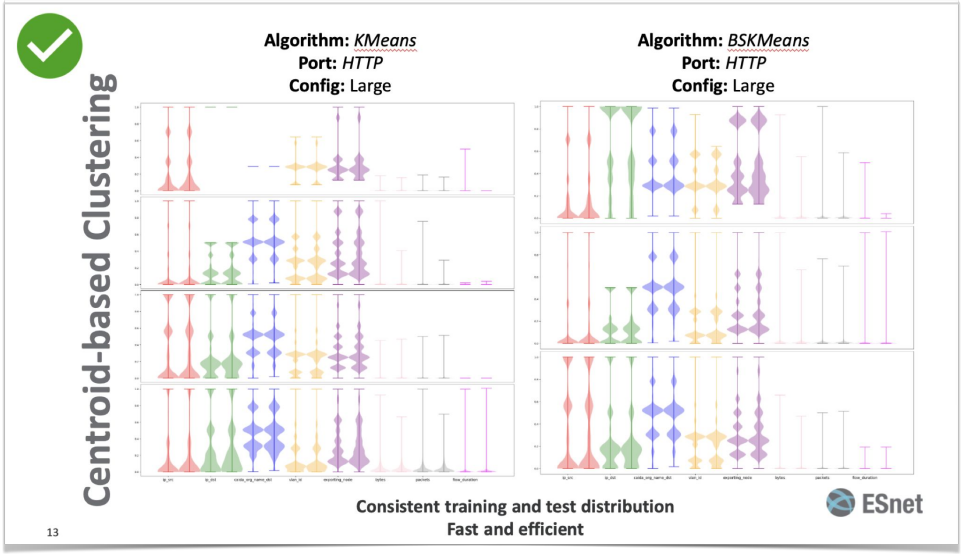
Analysis (for a single router, **atla-cr6**) using hourly aggregates, shows that 17:00 UTC and Fridays are peak demand times.

ESnet

ML - Clustering / Self Similarity / Prediction







Observation

- We are seeing consistent auto clustering over the 24-hr (12 billion) flow data set.

Implications

- We can do capacity planning and prediction without worrying about wild fluctuations.
- We can automate exception/anomaly detection.

UC San Diego ESnet

O. Gungor, T. Rosing *Y. Kumar at el*

Summary

1. What problem are we solving?

   Gaining a complete understanding of high-speed network traffic, leading to improved performance and reliability for science workflows.

2. How is what we are doing different from what is done today?

   A combination of programmable hardware and software enables custom collection and analysis.

3. How does this benefit the user community?

   It helps network and security engineers troubleshoot issues and investigate anomalies, resulting in improved service.

4. How can the user community get involved?

   Send email to: hightouch@es.net

ESnet

Questions...

ESnet