ESnet High Touch

I2 TechEx 2023

Yatish Kumar

ESnet

Structure of the talk

- What is High Touch Telemetry
  - Hardware
  - Grafana / Python / SQL Tools
- Use Cases
  - Network visibility
  - Operations debug / auditing
  - Network Planning
  - AI / ML Research

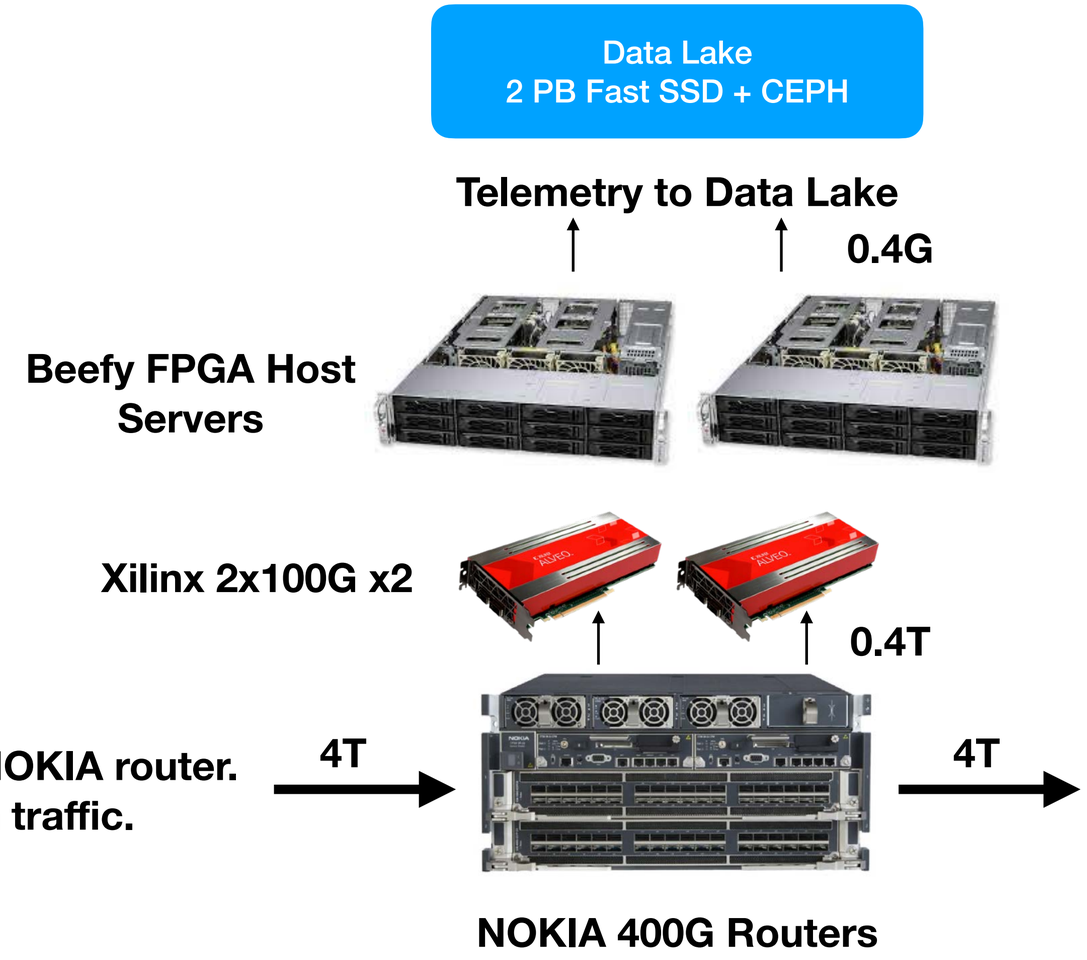
ESnet

High Touch in a nutshell

1:1 Monitoring on up to 4 Terabits of traffic
Hardware (FPGA) Accelerated Data Reduction
Hardware (FPGA) 1ns accurate time stamping

Kafka based 24/7 Streaming to a data lake
PCAP capture of any subset of flows

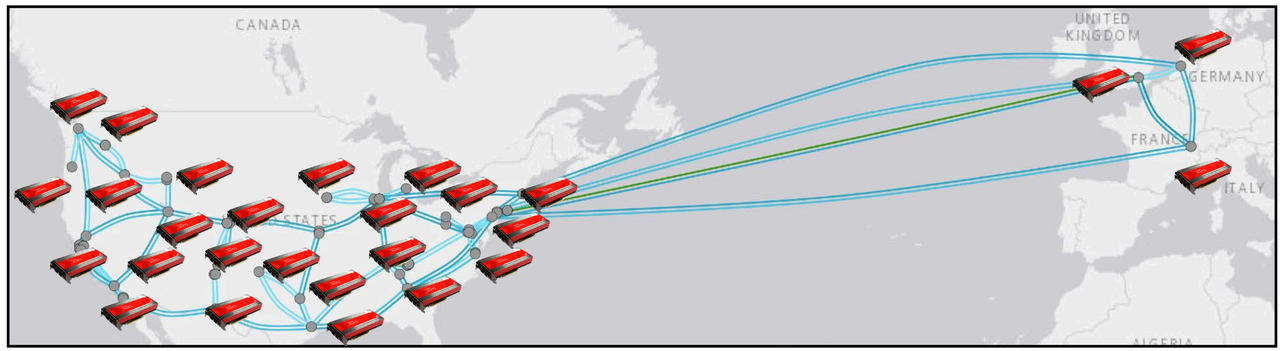
SQL Backend running on fast servers

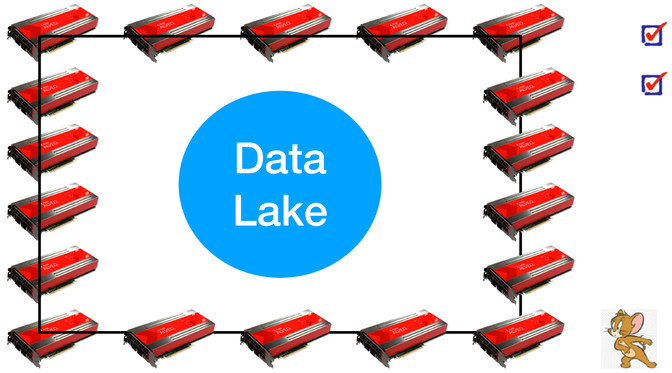


ESnet

High Touch in a nutshell

High Touch Nodes at all ESnet Edge Routers





- 100 % Packet Inspection
- 100 % Perimeter Coverage

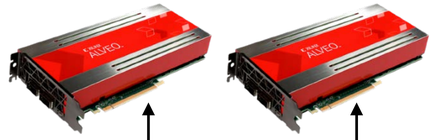
Complete Perimeter for all traffic

Data Lake
2 PB Fast SSD + CEPH

Telemetry to Data Lake

Beefy FPGA Host Servers

Xilinx 2x100G x2



NOKIA 400G Routers

ESnet

Tools

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 │
```

```
SELECT
    count(*) AS total_flows,
    sum(bytes) AS total_bytes,
    sum(packets) AS total_packets
FROM ht.all_flows

Query id: 6b3c7f73-8dda-4950-b819-67f112102944

┌─total_flows─┬─total_bytes─┬─total_packets─┐
│ 13612804466 │ 1475967324390170 │ 774102047235 │
└─────────────┴─────────────┴──────────────┘

1 row in set. Elapsed: 37.291 sec. Processed 13.61 billion rows,
```

Jupyter Lab / Pandas

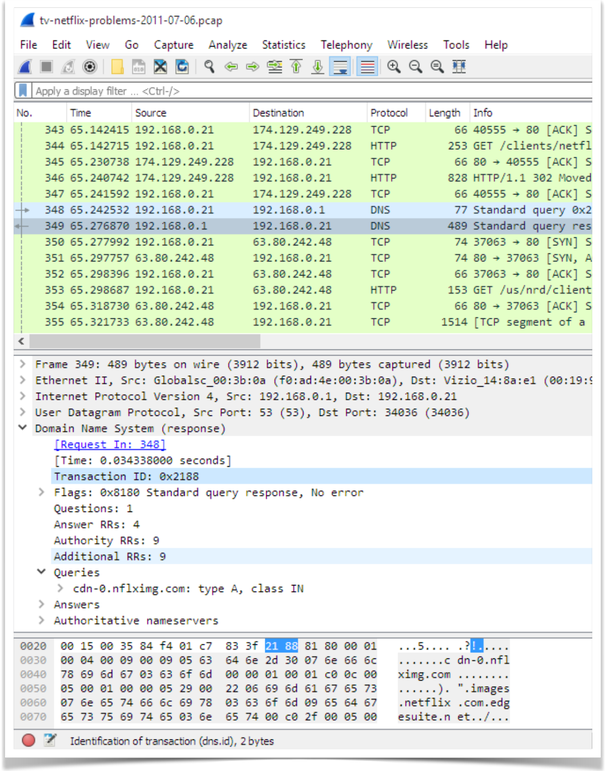
```
[9]: # Quick preview to see what the data frame looks like
     df.drop_duplicates(subset='flow_id')
```

[9]:

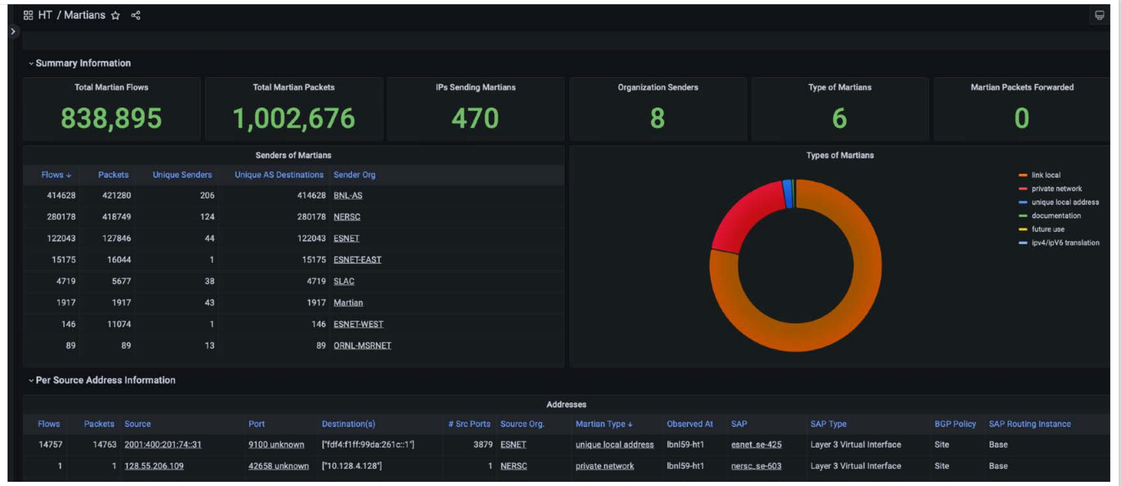
	dst	dport	src	sport	proto	time	dst_org	src_org	src_network	dst_network	flow_id
0	104.196.237.25	42948	198.129.224.35	80	6	2022-02-15T11:09:59.994-0800	UNDEF	ESNET	198.129.224.0/24	104.196.237.0/24	1
1	134.79.146.247	34432	45.137.21.208	48138	6	2022-02-15T11:09:59.984-0800	SLAC	UNDEF	45.137.21.0/24	134.79.146.0/24	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

Wireshark



Grafana / Stardust



ESnet

“Use Cases”

ESnet

Internet Background Radiation

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_only probes, as well in the line 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.

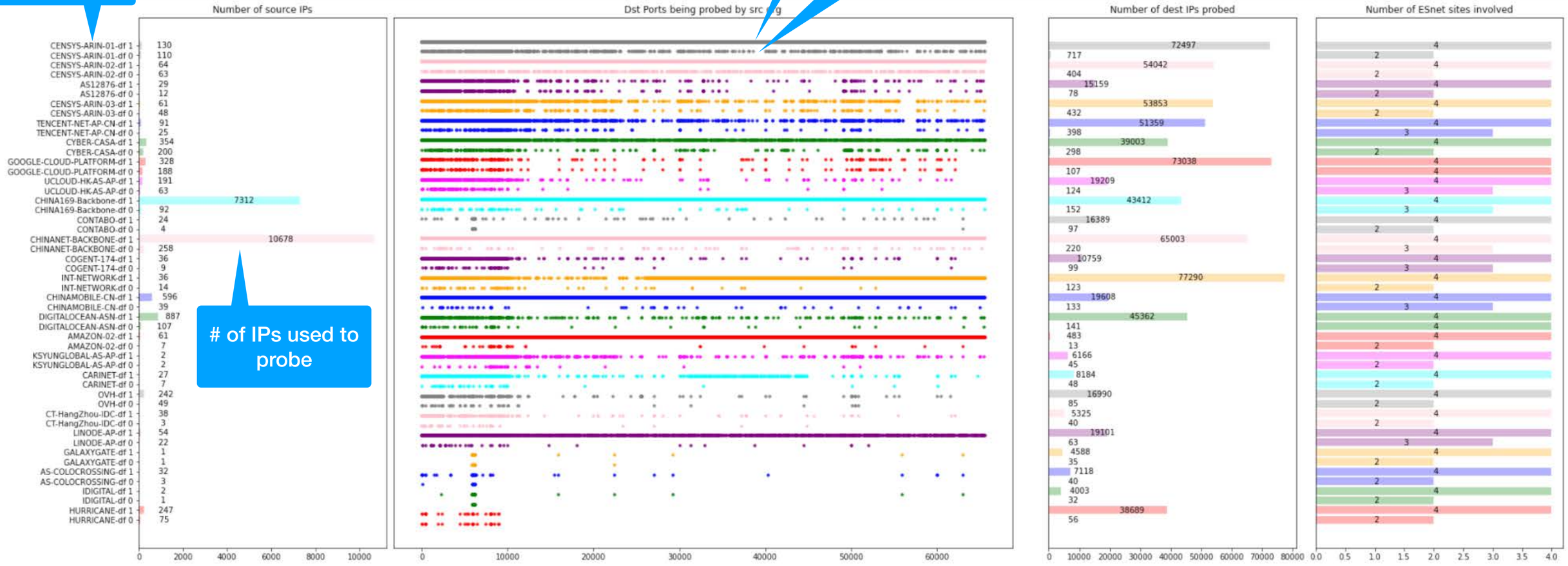
- By looking at the Number of IP's probed bar chart, we can see that only a small number of hosts being scanned, are returning any responses.
- By looking at the number of source IPs we can see that some of the scanners source IPs's are completely unsuccessful at getting any responses.

All the dst ports that GET A SYN and no further packets

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

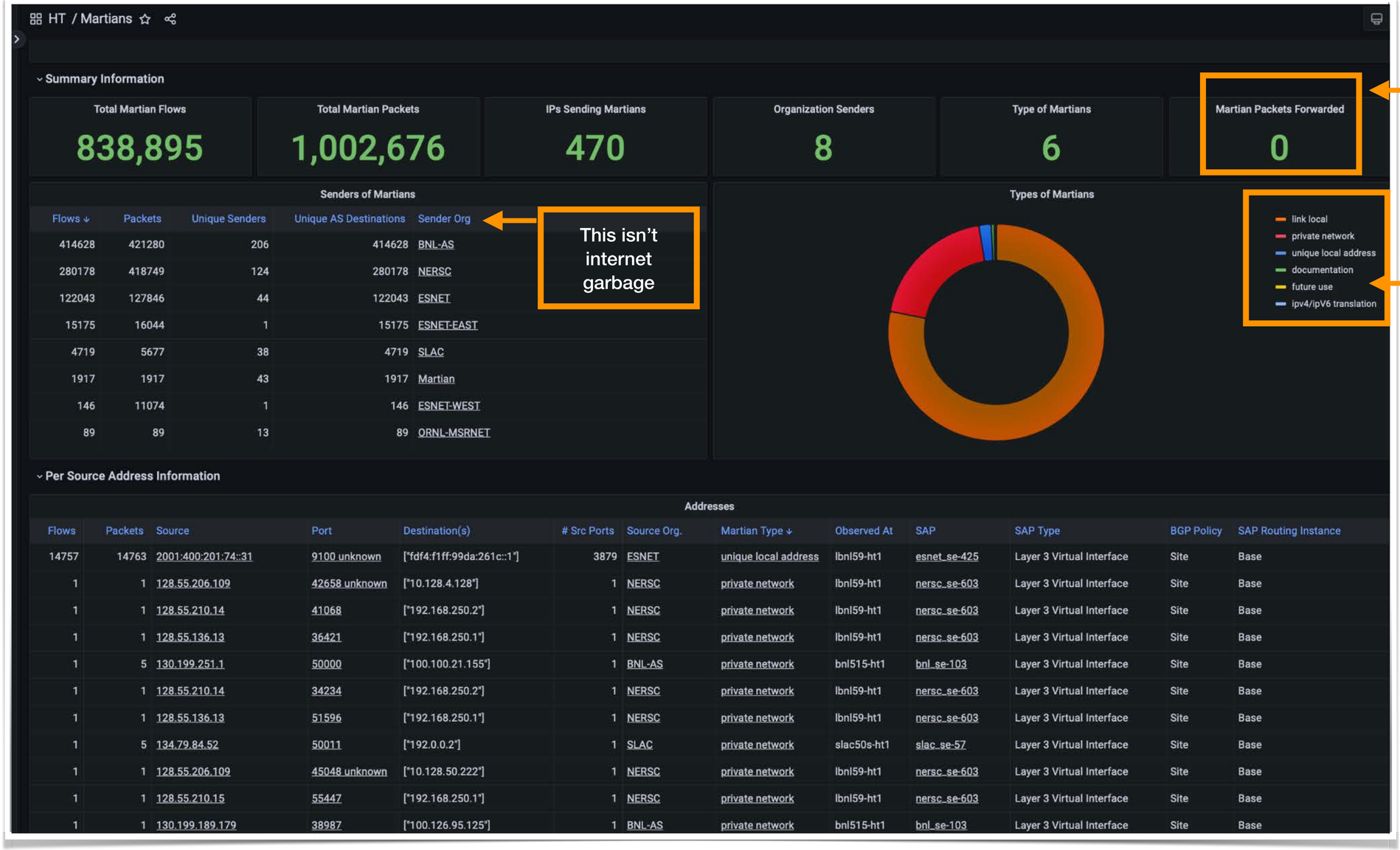
Org Name

```
_bidi , df] , 'top' , 25 )
```



# of IPs used to probe

Network Audits - Martians



NICE ! Kudos ESnet Router Configs

Why do we have documentation addresses in our network ?

SONIC - Open Source Switches
Docker - Data Center Auto Config

Beware when you download and get instant gratification from the internet.

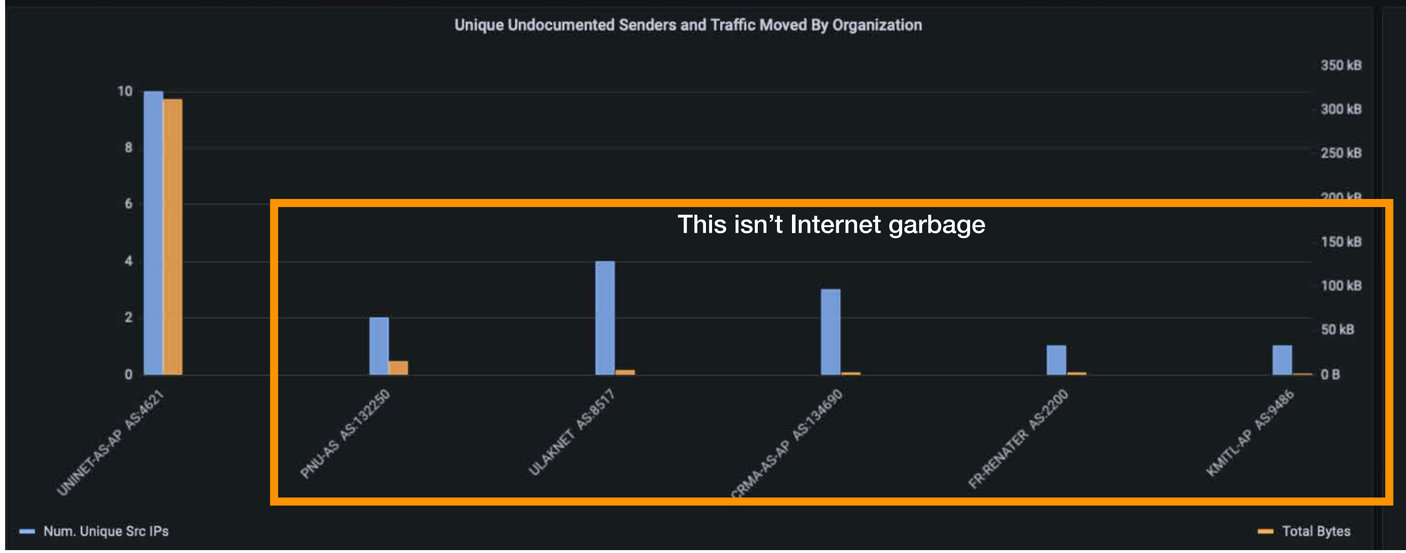
ESnet

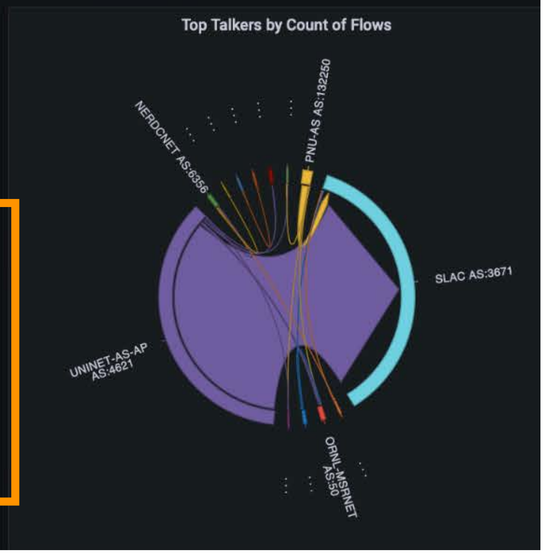
LHCONE - CRIC Audit

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





CERN / CRIC
Source of Truth for what prefixes are on LHCONE

JSON Query

HT SQL Database

Select *
NOT in JSON

[ ]
Should be empty set

ESnet

Network Planning - Elephants or Big Foot

	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.
21	ESNET	ESNET	2001:400:f010:221::1	2001:400:f010:241::1	26.203144	anl541b-ps-tp.lhcone.es.net.	fnalgcc-ps-tp.lhcone.es.net.
22	ESNET	ESNET	2001:400:ee00:b03::1	2001:400:ee00:10::1	26.203090	slac50s-ps-tp.es.net.	eqxsv5-ps-tp.es.net.
23	ESNET	ESNET	2001:400:ee00:b02::1	2001:400:ee00:10::1	26.203027	slac50n-ps-tp.es.net.	eqxsv5-ps-tp.es.net.
24	ESNET	ESNET	2001:400:ee00:20::1	2001:400:ee00:b03::1	26.202994	lbnl59-ps-tp.es.net.	slac50s-ps-tp.es.net.
25	ESNET	ESNET	2001:400:ee00:221::1	2001:400:ee00:240::1	26.202628	anl541b-ps-tp.es.net.	fnalfcc-ps-tp.es.net.
26	ESNET	ESNET	2001:400:ee00:10::1	2001:400:ee00:b03::1	26.202129	eqxsv5-ps-tp.es.net.	slac50s-ps-tp.es.net.
27	ESNET	ESNET	2001:400:ee00:200::1	2001:400:ee00:240::1	26.201956	eqxch2-ps-tp.es.net.	fnalfcc-ps-tp.es.net.
28	ESNET	ESNET	2001:400:ee00:241::1	2001:400:ee00:221::1	26.201614	fnalgcc-ps-tp.es.net.	anl541b-ps-tp.es.net.
29	ESNET	ESNET	2001:400:ee00:240::1	2001:400:ee00:200::1	26.201460	fnalfcc-ps-tp.es.net.	eqxch2-ps-tp.es.net.
30	ESNET	ESNET	2001:400:ee00:200::1	2001:400:ee00:241::1	26.201034	eqxch2-ps-tp.es.net.	fnalgcc-ps-tp.es.net.
31	ESNET	ESNET	2001:400:f010:240::1	2001:400:f010:200::1	26.201015	fnalfcc-ps-tp.lhcone.es.net.	eqxch2-ps-tp.lhcone.es.net.
32	ESNET	ESNET	2001:400:ee00:10::1	2001:400:ee00:b02::1	26.200805	eqxsv5-ps-tp.es.net.	slac50n-ps-tp.es.net.
33	ESNET	ESNET	2001:400:ee00:20::1	2001:400:ee00:b02::1	26.200350	lbnl59-ps-tp.es.net.	slac50n-ps-tp.es.net.
34	ESNET	ESNET	2001:400:ee00:221::1	2001:400:ee00:241::1	26.200129	anl541b-ps-tp.es.net.	fnalgcc-ps-tp.es.net.
35	ESNET	ESNET	2001:400:ee00:10::1	2001:400:ee00:b01::1	26.200096	eqxsv5-ps-tp.es.net.	llnl-ps-tp.es.net.
36	ESNET	ESNET	2001:400:f010:200::1	2001:400:f010:241::1	26.198824	eqxch2-ps-tp.lhcone.es.net.	fnalgcc-ps-tp.lhcone.es.net.
37	NCSA-AS	ESNET	2620:0:c80:300::2	2001:400:ee00:221::1	26.198818	Timeout	anl541b-ps-tp.es.net.
38	ESNET	ESNET	2001:400:f010:241::1	2001:400:f010:221::1	26.198072	fnalgcc-ps-tp.lhcone.es.net.	anl541b-ps-tp.lhcone.es.net.
39	ESNET	ESNET	2001:400:ee00:20::1	2001:400:ee00:10::1	26.197927	lbnl59-ps-tp.es.net.	eqxsv5-ps-tp.es.net.
40	ESNET	ESNET	2001:400:ee00:10::1	2001:400:ee00:b04::1	26.197207	eqxsv5-ps-tp.es.net.	NXDOMAIN
41	ESNET	ESNET	2001:400:ee00:10::1	2001:400:ee00:20::1	26.196922	eqxsv5-ps-tp.es.net.	lbnl59-ps-tp.es.net.
42	ESNET	ESNET	2001:400:ee00:820::1	2001:400:ee00:821::1	26.196839	nash-ps-tp.es.net.	chat-ps-tp.es.net.
43	ESNET	ESNET	2001:400:ee00:b01::1	2001:400:ee00:10::1	26.196467	llnl-ps-tp.es.net.	eqxsv5-ps-tp.es.net.
44	ESNET	ESNET	2001:400:ee00:115::1	2001:400:ee00:110::1	26.192698	losa-ps-tp.es.net.	sand-ps-tp.es.net.
45	ESNET	ESNET	2001:400:ee00:821::1	2001:400:ee00:820::1	26.192622	chat-ps-tp.es.net.	nash-ps-tp.es.net.

Select
(*)
Where
Peak Rate > 10 Gbps
For at least 10 seconds
Order by Rate

What is .ps-tp ?

Because it generates our largest elephant flows.

ESnet

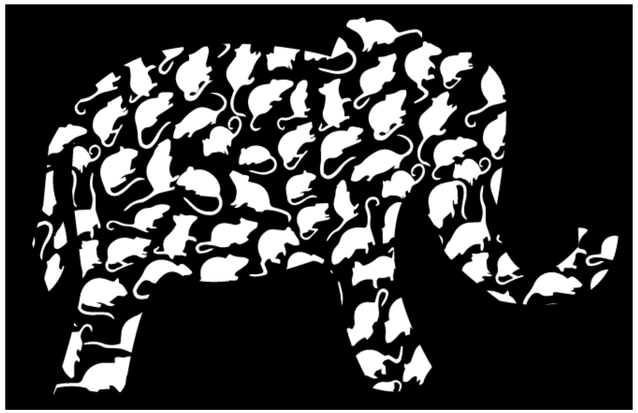
Network Planning - Slightly Bigger Picture

Query time .. 52.55099439620972

	caida_org_name_src	caida_org_name_dst	Gbps
0	U-CHICAGO-AS	ARGONNE-AS	30.037561
1	ARGONNE-AS	U-CHICAGO-AS	27.532194
2	ESNET	ESNET	26.215328
3	NCSA-AS	ESNET	26.198818
4	ESNET	NCSA-AS	26.189637
5	ESNET-WEST	ESNET-WEST	26.151662
6	ESNET-EAST	ESNET-EAST	26.150740
7	ESNET-EAST	ESNET-WEST	26.149878
8	ESNET-WEST	ESNET-EAST	26.145060
9	NCSA-AS	ESNET-WEST	26.136680
10	BNL-AS	ESNET	24.860384
11	ESNET-WEST	NCSA-AS	24.237054
12	ESNET-WEST	ARGONNE-AS	23.859718
13	NCSA-AS	ESNET-EAST	23.723869
14	BNL-AS	ESNET-EAST	22.466708
15	BNL-AS	NU-AS	22.372516
16	BNL-AS	ESNET-WEST	21.165468
17	ESNET-WEST	MISU-231	20.912281
18	MISU-231	ESNET-WEST	19.870178
19	TACCNET	ESNET-WEST	15.653456
20	STANFORD	ESNET-EAST	12.623604
21	TACCNET	ESNET-EAST	12.543568
22	MERIT-AS-6	ESNET	12.453257
23	ESNET-WEST	LBL	10.111514
24	SLAC	LBL	9.996838
25	ESNET	LBL	9.969307
26	ESNET	BNL-AS	9.965247
27	LBL	ESNET	9.956527
28	LBL	ESNET-WEST	9.954399
29	ULTRALIGHT	VANDERBILT	9.939852
30	ESNET	VANDERBILT	9.908270
31	CWRU-AS-1	ESNET-EAST	9.847033
32	OARNET-AS-2	ESNET-EAST	9.843675
33	OARNET-AS-2	ESNET-WEST	9.842789
34	U-CHICAGO-AS	ESNET	9.740794
35	BNL-AS	AMNH	9.710251
36	NCAR-AS	ESNET-EAST	9.660255
37	NCAR-AS	ESNET-WEST	9.625336
38	LBL	ESNET-EAST	9.459778
39	ARGONNE-AS	CSM-AS	9.362607
40	UCLA	ESNET-EAST	9.349517
41	WASH-NSF-AS	ESNET-EAST	9.282061
42	ESNET-EAST	JANET	9.184101
43	JANET	ESNET-EAST	9.056038
44	ULTRALIGHT	ESNET	9.036059
45	UTARLINGTON	ESNET	8.758342
46	TENET-1	ESNET-EAST	8.341057
47	FNAL-AS	SLAC	8.287741
48	CSM-AS	ARGONNE-AS	8.128646
49	ESNET-WEST	WN-AZ-AS	7.991214

These are largely, if not entirely PerfSonar

99 % of our traffic averages 0.5 Gbps per flow.
LHC / Globus / everything…



This is how we do large tranfers.
NO EXCEPTIONS !

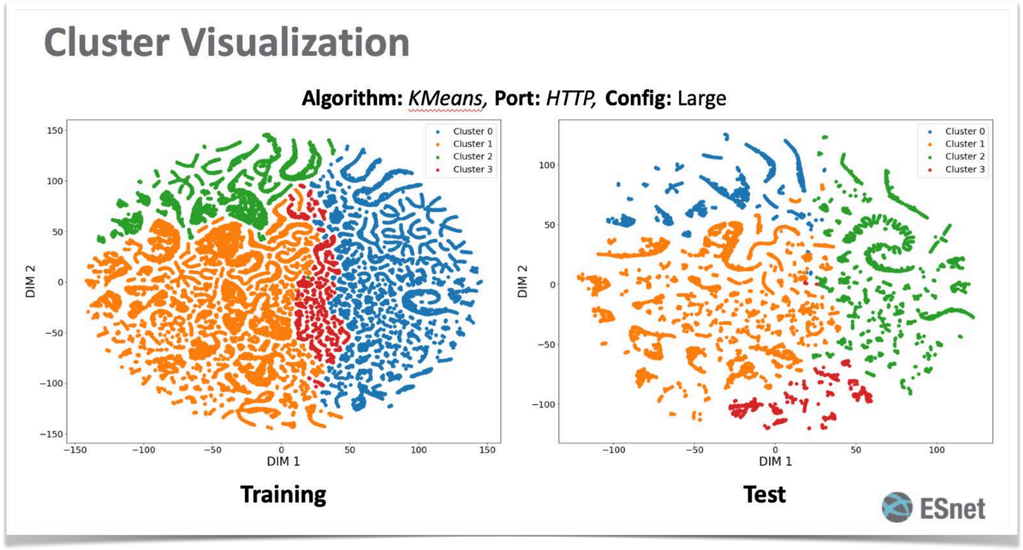
ECMP / LAG
Campus 100G Optimization
FlexE
9K MTUs

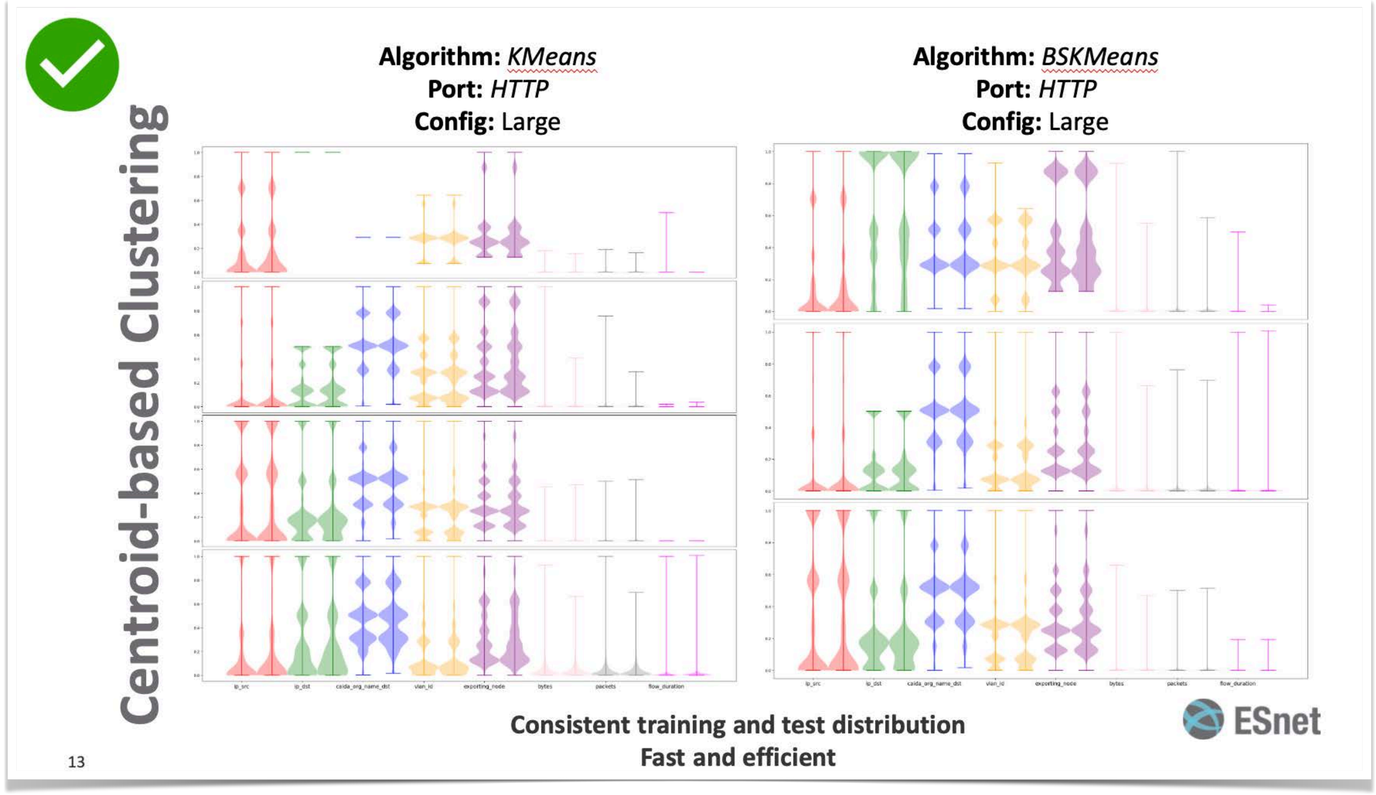
Everything else “peaks” at 10 Gbps
“average” is < 1 Gbps

This provides a context between
“Theory” and “Practice”

ESnet

ML - Clustering / Self Similarity / Prediction etc..





Joint work with UCSD (Onat Gungor / Tajana Rosing)

Reliable auto clustering of our 12 billion flows / day ?
How repeatable are the statistics ?
Ans: 99% of the time the hundreds of thousands of nodes on ESnet do more or less the same thing everyday.

Good ! Then we can do capacity planning and prediction, without worrying about daily chaos.
Good ! Then we can look for exceptions automatically.

ESnet

“Thank You”

Yatish Kumar

ESnet