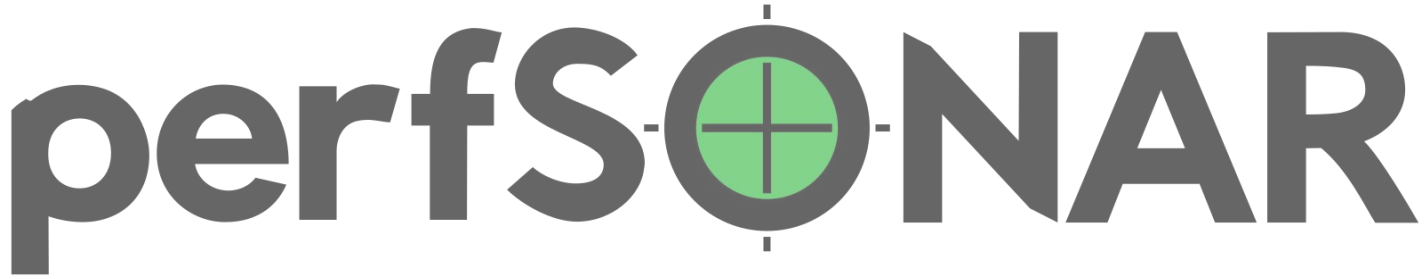
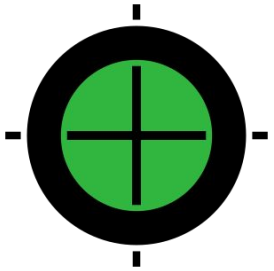


CLI Testing and Troubleshooting

Doug Southworth · Texas Advanced Computing Center · dsouthworth@tacc.utexas.edu

perfSONAR is developed by a partnership of



If you're wondering why I've brought you here…

Motivations and Methods

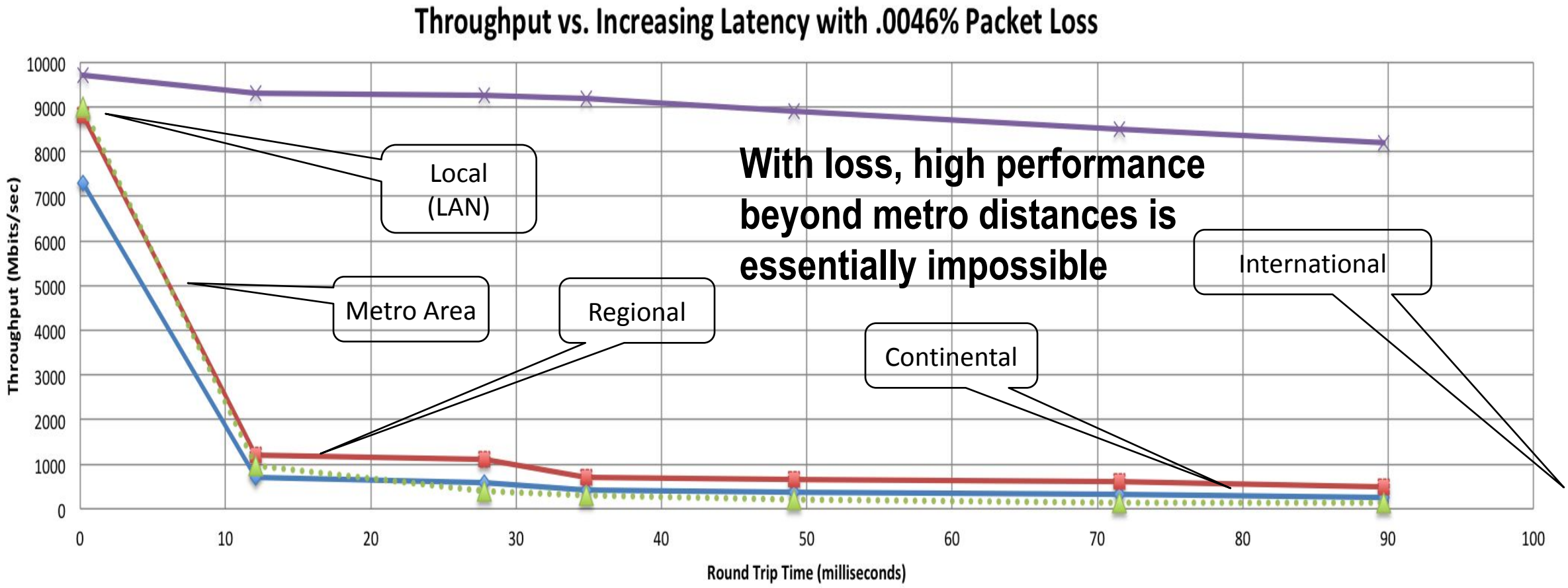
What is perfSONAR?

- perfSONAR is a tool to:
  - Set network performance expectations
  - Find network problems (“soft failures”)
  - Help fix these problems
  - All in multi-domain environments

- These problems are all harder when multiple networks are involved

- perfSONAR is provides a standard way to publish active and passive monitoring data
  - This data is interesting to network researchers as well as network operators

Soft Failures Cause Packet Loss and Degraded TCP Performance



perfSONAR Dashboard: Improving network visibility

Status at-a-glance

- Packet loss
- Throughput
- Correctness

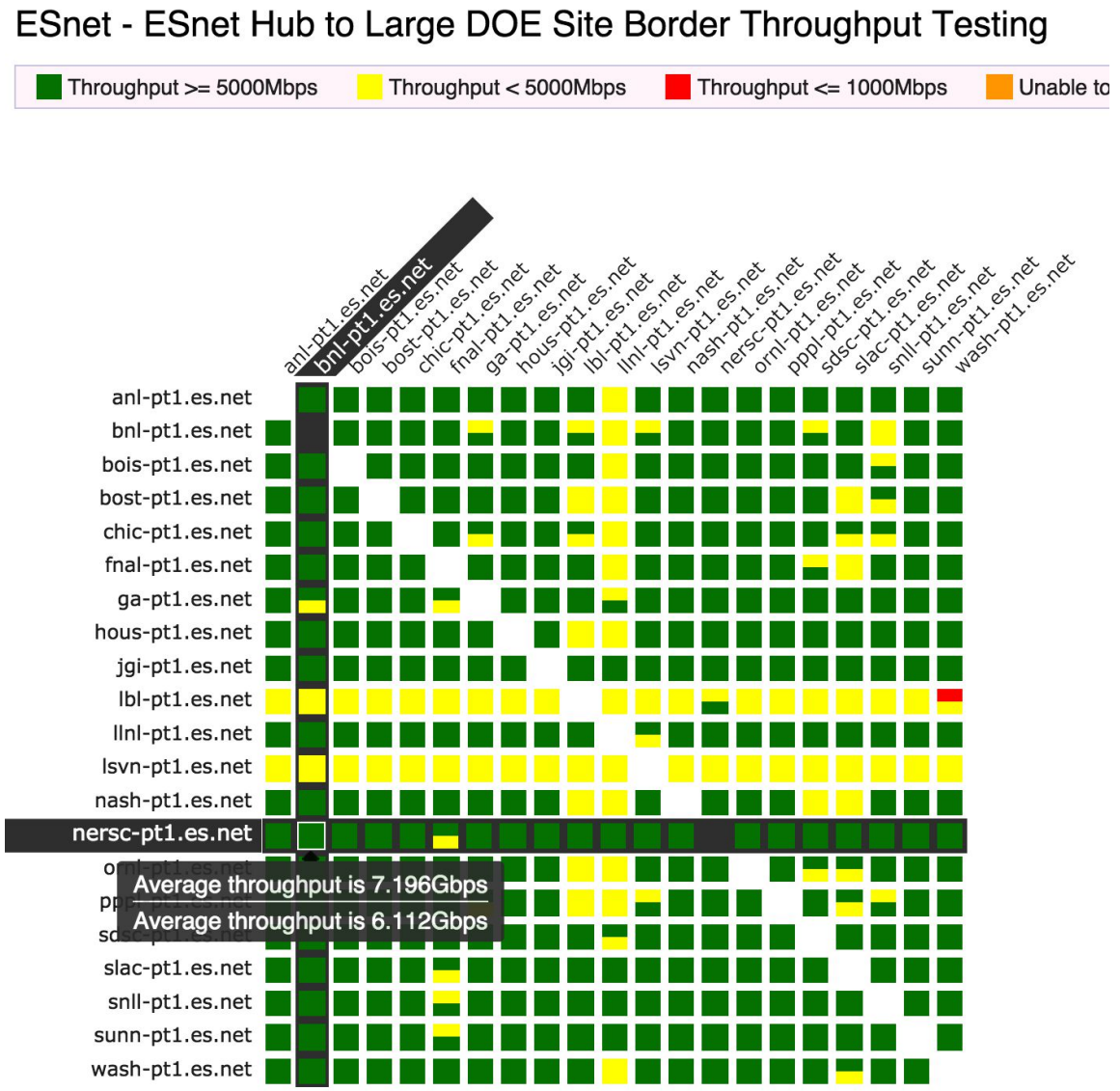
Current live instances at

http://pas.net.internet2.edu/

http://ps-dashboard.es.net/

Drill-down capabilities:

- Test history between hosts
- Ability to correlate with other events
- Very valuable for fault localization and isolation



Stop letting that pS node gather dust

perfSONAR has a lot of value as a day-to-day tool; it's not just a software package that puts a green square on a mesh

- If you use it more often, you'll tend to keep it updated and operational
- Having an outdated, insecure node is worse than not having one at all

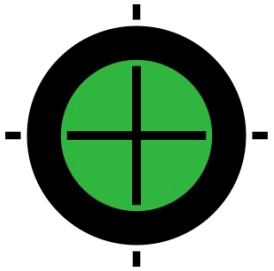
The goal is proactive analysis, troubleshooting, and resolution

- R&E networks, like old British cars, need us: they aren't self sufficient
- Researchers rely on our networks to do their work, and they won't necessarily speak up when there's a problem
- pS gives you eyes to see what's going on beyond your borders

WE NEED YOU!

perfSONAR is a community effort, and not just for development.

Your node gives us all eyes into the global R&E infrastructure



Commands and Syntax

Simulating Performance

- It's infeasible to perform at-scale data movement all the time – as we see in other forms of science, we need to rely on simulations
- Network performance comes down to a couple of key metrics:
  - Throughput (e.g. "how much can I get out of the network")
  - Latency (time it takes to get to/from a destination)
  - Packet loss/duplication/ordering (for some sampling of packets, do they all make it to the other side without serious abnormalities occurring?)
  - Network utilization (the opposite of "throughput" for a moment in time)
- We can get many of these from a selection of active and passive measurement tools – enter the perfSONAR Toolkit

pscheduler – The Secret Sauce

- pscheduler is the engine that drives perfSONAR
  - Coordinates timeslots and schedules tests between nodes
  - Creates a common syntax that all tools use
  - Handles the storage of results
- This approach to test management ensures that:
  - Tests that could impact each other's performance, like throughput, are never run simultaneously
  - Simplified coordination, where you don't need an engineer at the other end to start a daemon, open a port, etc.
  - Access control is maintained and test limits are enforced

Basic syntax

pscheduler task [options] test-type [test-options]

- *task* is what you want pscheduler to do
- *test-type* is how you want pscheduler to do it

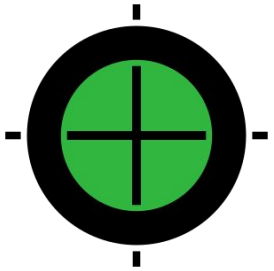
There's more than one tool to run many of these tests, and pscheduler gives you the option to choose that tool:

- *iperf2/iperf3/nuttcp* for bandwidth
- *traceroute/tracepath/paris-traceroute* for routing
- *--help is your friend!*

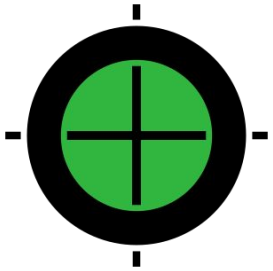
Remote Commands, AKA, Your Best Friends

- --source and --dest flags do what you expect they would, but neither end has to be your node
  - You can run most tests between two remote nodes
  - This includes perfSONAR's built-in self-diagnostic tools

- This core piece of functionality is what makes perfSONAR useful in day-to-day troubleshooting activities and makes it more than just a simple performance data collector



```
[ps-iniu@thrpt10ge-1 ~]$ pscheduler task throughput --source perf.newy32aoa.neaar.net --dest test.seat.transpac.org
Submitting task...
Task URL:
https://perf.newy32aoa.neaar.net/pscheduler/tasks/666b16fc-c32d-4960-a8ac-ef66b2eca183
Running with tool 'iperf3'
Fetching first run...

Next scheduled run:
https://perf.newy32aoa.neaar.net/pscheduler/tasks/666b16fc-c32d-4960-a8ac-ef66b2eca183/runs/31b57ed0-c39c-4533-8654-60e8
a2137d64
Starts 2021-05-26T14:18:53Z (~6 seconds)
Ends   2021-05-26T14:19:12Z (~18 seconds)
Waiting for result...

* Stream ID 5
Interval       Throughput     Retransmits    Current Window
0.0 - 1.0      4.67 Gbps      1              148.25 MBytes
1.0 - 2.0      9.89 Gbps      0              148.39 MBytes
2.0 - 3.0      9.90 Gbps      0              148.39 MBytes
3.0 - 4.0      9.90 Gbps      2              152.96 MBytes
4.0 - 5.0      9.90 Gbps      0              152.96 MBytes
5.0 - 6.0      9.89 Gbps      0              152.96 MBytes
6.0 - 7.0      9.90 Gbps      0              152.96 MBytes
7.0 - 8.0      9.90 Gbps      0              152.96 MBytes
8.0 - 9.0      9.90 Gbps      0              152.96 MBytes
9.0 - 10.0     9.89 Gbps      0              152.96 MBytes

Summary
Interval       Throughput     Retransmits    Receiver Throughput
0.0 - 10.0     9.37 Gbps      3              9.20 Gbps

No further runs scheduled.
```



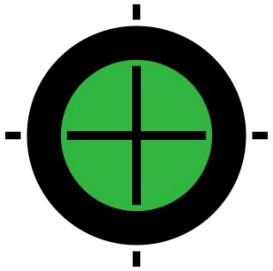
```
[ps-iniu@thrpt10ge-1 ~]$ pscheduler task --tool tracepath trace --source perf.newy32aoa.neaar.net --dest test.seat.transpac.org
Submitting task...
Task URL:
https://perf.newy32aoa.neaar.net/pscheduler/tasks/a447400a-125a-46cb-9e78-020ab537cca1
Running with tool 'tracepath'
Fetching first run...

Next scheduled run:
https://perf.newy32aoa.neaar.net/pscheduler/tasks/a447400a-125a-46cb-9e78-020ab537cca1/runs/19416479-28b0-40a3-9844-2bd9c4f93a70
Starts 2021-05-26T14:30:25Z (~1 seconds)
Ends   2021-05-26T14:32:06Z (~100 seconds)
Waiting for result...

1       vlan-150.rtr.newy32aoa.neaar.net (192.203.116.32) AS396390 0.142 ms mtu 9000 bytes
          INDIANA-UNIVERSITY-NEAAR, US
2       et-2-1-5.127.rtsw.newy32aoa.net.internet2.edu (198.71.45.192) AS11537 0.789 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
3       ae-3.4079.rtsw.wash.net.internet2.edu (162.252.70.138) AS11537 6.192 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
4       ae-0.4079.rtsw2.ashb.net.internet2.edu (162.252.70.137) AS11537 6.62 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
5       ae-2.4079.rtsw.ashb.net.internet2.edu (162.252.70.74) AS11537 6.54 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
6       ae-20.4079.rtsw.clev.net.internet2.edu (162.252.70.129) AS11537 13.544 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
7       ae-3.4079.rtsw3.eqch.net.internet2.edu (162.252.70.131) AS11537 20.07 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
8       ae-5.4079.rtsw.eqch.net.internet2.edu (162.252.70.162) AS11537 26.237 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
9       ae-0.4079.rtsw.minn.net.internet2.edu (162.252.70.107) AS11537 27.898 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
10      ae-1.4079.rtsw.seat.net.internet2.edu (162.252.70.172) AS11537 59.964 ms mtu 9000 bytes
          INTERNET2-RESEARCH-EDU, US
11      207.231.240.24 AS53965 59.659 ms mtu 9000 bytes
          CCSEBGP, US
12      test.seat.transpac.org (192.203.115.2) AS22388 59.635 ms mtu 9000 bytes
          TRANSPAC, US
```

Wait, did you say built-in diagnostics??

- *pscheduler troubleshoot*
  - See if your node has the basic, necessary services running, if PMTUD is working, if the clock is synched to an NTP source, etc.

- *pscheduler troubleshoot $remote_node1*
  - Same as above, but adds in the same tests on a remote node to ensure your two nodes can successfully complete a test

- *pscheduler troubleshoot --host=$remote_node1 $remote_node2*
  - *Same as the first two, but checking two remote nodes against each other*



```
[ps-iniu@thrpt10ge-1 ~]$ pscheduler troubleshoot --host perf.newy32aoa.neaar.net test.seat.transpac.org
Performing basic troubleshooting of perf.newy32aoa.neaar.net and test.seat.transpac.org.

perf.newy32aoa.neaar.net:

  Measuring MTU... 9000+
  Looking for pScheduler... OK.
  Fetching API level... 4
  Checking clock... OK.
  Exercising API... Status... Tests... Tools... OK.
  Fetching service status... OK.
  Checking services... ticker... scheduler... runner... archiver... OK.
  Idle test.... 4 seconds.... Checking archiving... OK.

test.seat.transpac.org:

  Measuring MTU... 9000+
  Looking for pScheduler... OK.
  Fetching API level... 4
  Checking clock... OK.
  Exercising API... Status... Tests... Tools... OK.
  Fetching service status... OK.
  Checking services... ticker... scheduler... runner... archiver... OK.
  Idle test.... 4 seconds.... Checking archiving... OK.

perf.newy32aoa.neaar.net and test.seat.transpac.org:

  Checking IP addresses... IPv4
  Measuring MTU... 9000+
  Checking timekeeping... OK.
  Simple stream test.... 11 seconds.... OK.

pScheduler on both hosts appears to be functioning normally.
```

Yeah, but what about pscheduler monitor?

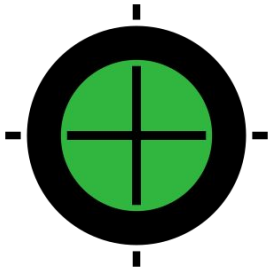
Yep, that works too.

pscheduler monitor --host=$remote_host

You can even plot a host’s schedule:

pscheduler plot-schedule --host=$remote_host

```
2021-05-27T14:38:19-04:00                    pScheduler Monitor                              test.seat.transpac.org

2021-05-27T18:37:29Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:37:31Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:37:31Z  +0 Finished    latencybg --source test.seat.transpac.org --dest 192.35.145.5 --packet-count 600+
2021-05-27T18:37:39Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:37:47Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:37:50Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:38:00Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:38:06Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:38:06Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:38:11Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:15Z     Running     latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T03:53:16Z     Running     latencybg --source test.seat.transpac.org --dest 128.171.64.53 --packet-count 60+
2021-05-27T03:53:16Z     Running     latencybg --source test.seat.transpac.org --dest 192.35.145.5 --packet-count 600+
2021-05-27T18:38:16Z  +0 Finished    latencybg --source test.seat.transpac.org --source-node test.seat.transpac.org -+
2021-05-27T18:48:21Z     Pending     throughput --source 192.35.145.5 --dest test.seat.transpac.org --duration PT10S
2021-05-27T18:56:13Z     Pending     throughput --source 128.171.64.53 --dest test.seat.transpac.org --duration PT10S
2021-05-27T18:56:52Z     Pending     throughput --source test.seat.transpac.org --dest 192.35.145.5 --duration PT10S
2021-05-27T18:57:52Z     Pending     throughput --source 2404:a8:19::e --source-node [2404:a8:19::e] --dest test.seat+
2021-05-27T19:00:59Z     Pending     throughput --source test.seat.transpac.org --source-node test.seat.transpac.org +
2021-05-27T19:13:48Z     Pending     throughput --source test.seat.transpac.org --dest 128.171.64.53 --duration PT10S
2021-05-27T19:15:31Z     Pending     throughput --source 128.171.64.53 --dest test.seat.transpac.org --duration PT10S
2021-05-27T19:44:55Z     Pending     throughput --source test.seat.transpac.org --dest 192.35.145.5 --duration PT10S
2021-05-27T19:45:21Z     Pending     throughput --source 150.129.185.14 --source-node 150.129.185.14 --dest test.seat
```



In The Wild

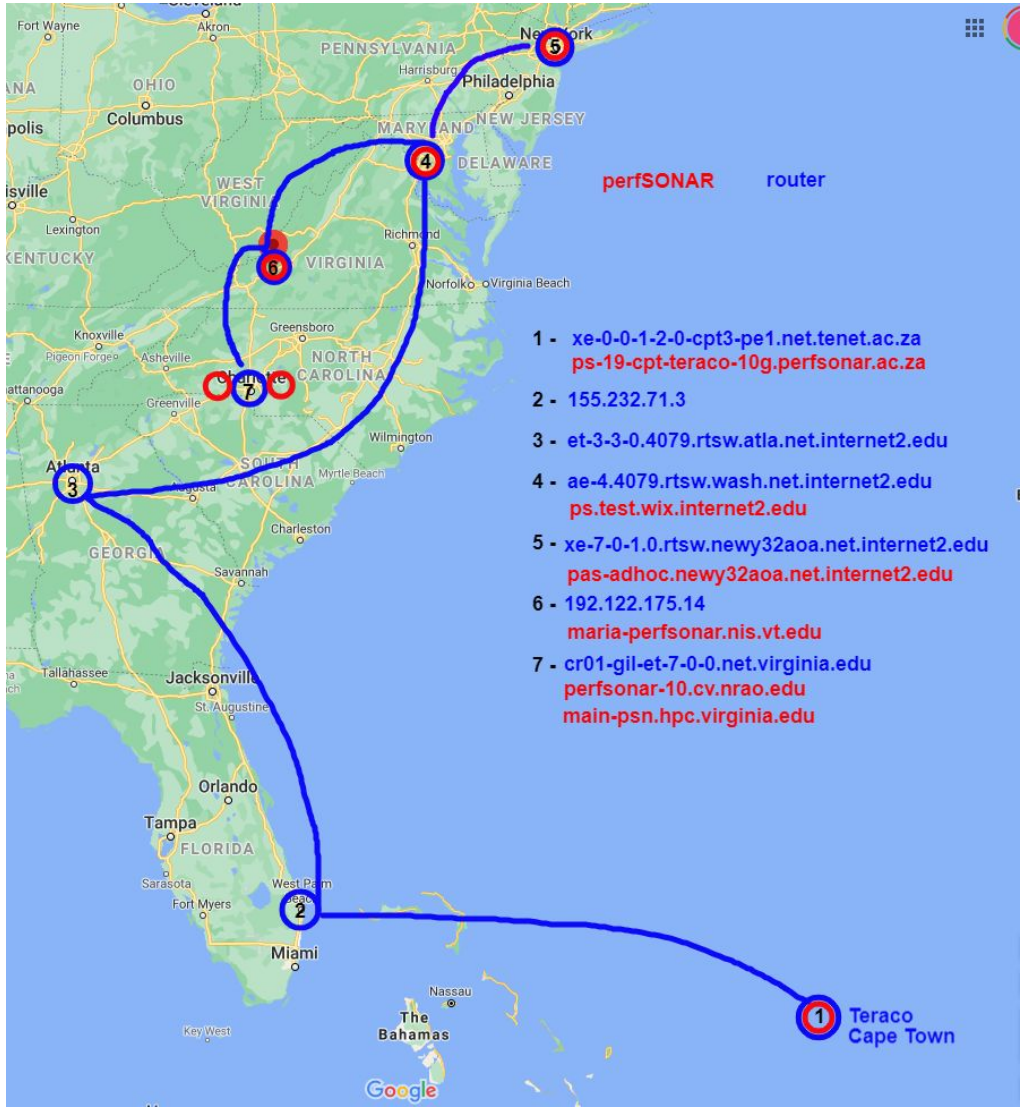
Examples from real world scenarios

NRAO/UVA <> SARAO Performance Problem

- Data sharing from the National Radio Astronomy Observatory, located on the University of Virginia campus, to the South African Radio Astronomy Observatory
  - Low performance – 4.8Mbps

- Initial testing from the South African side revealed a few potential problems, such as asymmetric routing and paths with unnecessarily circuitous routes.
  - These were identified using normal traceroutes and quickly corrected
  - No appreciable change in performance

Identification of possible pS toolkits



1 te0-1-0-1-cpt2-pe1.net.tenet.ac.za (155.232.40.9) AS2018 0.703 ms

…

5 155.232.71.3 AS2018 166.6 ms
TENET-1, ZA

6 et-3-3-0.4079.rtsw.atla.net.internet2.edu (162.252.70.42) AS11537 172.712 ms
INTERNET2-RESEARCH-EDU, US

7 ae-4.4079.rtsw.wash.net.internet2.edu (198.71.45.7) AS11537 185.775 ms
INTERNET2-RESEARCH-EDU, US

8 ae-0.4079.rtsw2.ashb.net.internet2.edu (162.252.70.137) AS11537 186.419 ms
INTERNET2-RESEARCH-EDU, US

9 ae-2.4079.rtsw.ashb.net.internet2.edu (162.252.70.74) AS11537 185.845 ms
INTERNET2-RESEARCH-EDU, US

10 192.122.175.14 AS40220 186.368 ms
MARIA, US

11 br01-udc-et-1-0-0-20.net.virginia.edu (192.35.48.33) AS225 188.065 ms
VIRGINIA-AS, US

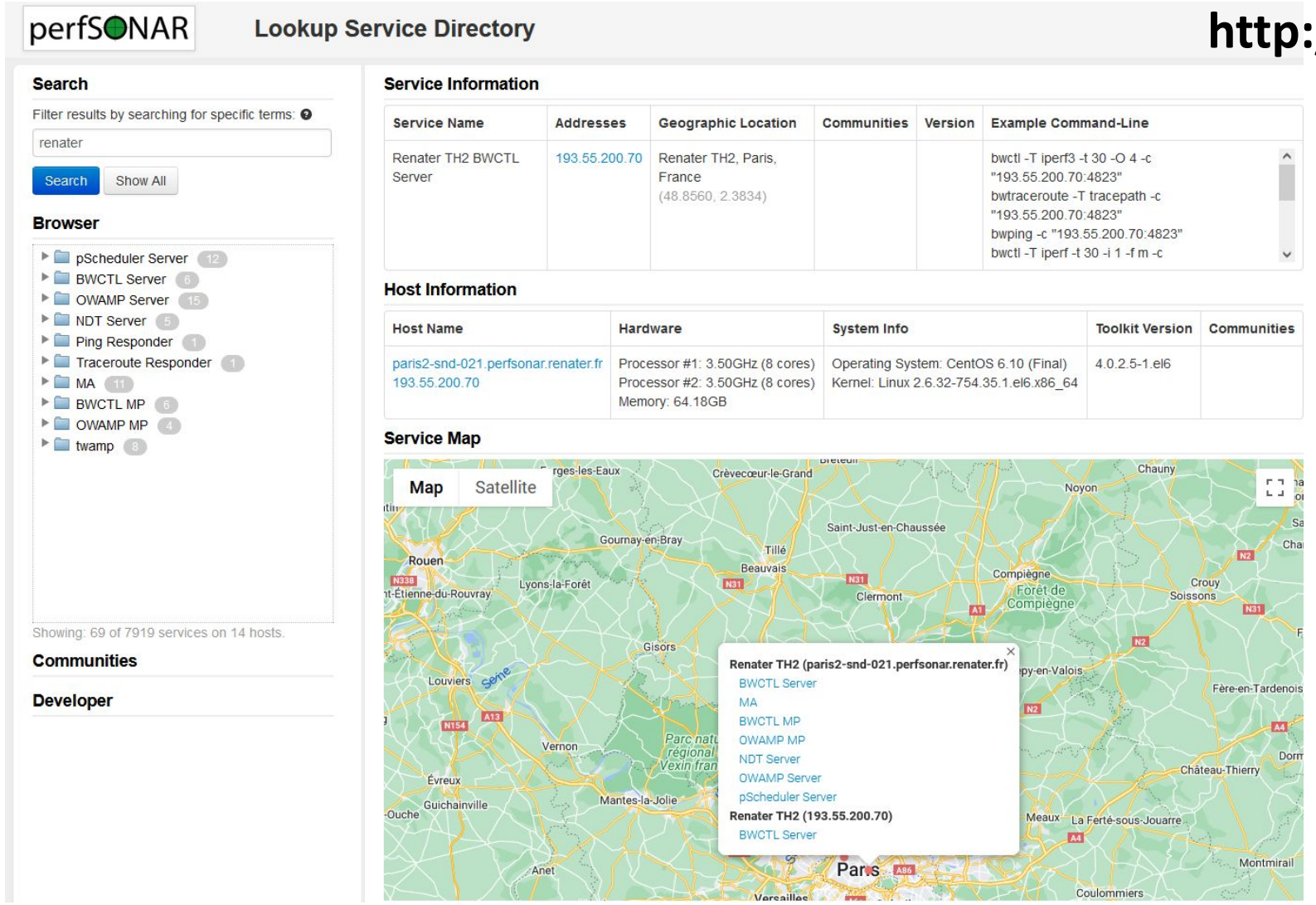
12 cr01-udc-et-4-2-0.net.virginia.edu (128.143.236.6) AS225 188.448 ms
VIRGINIA-AS, US

13 cr01-gil-et-7-0-0.net.virginia.edu (128.143.236.89) AS225 203.281 ms
VIRGINIA-AS, US

14 perfsonar-10.cv.nrao.edu (198.51.208.55) AS225 188.179 ms
VIRGINIA-AS, US

perfSONAR Lookup Service Directory

http://stats.es.net/ServicesDirectory/

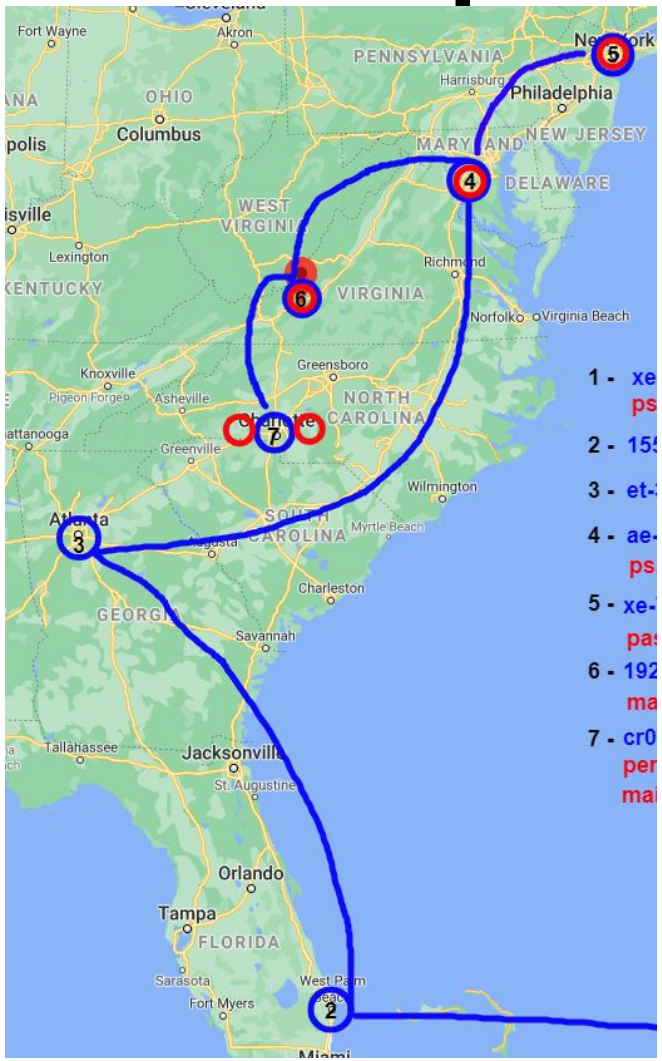


Lookup listed testpoints and toolkits by almost any criteria:

- Hostname
- IP address
- Institution
- City
- Country
- REN

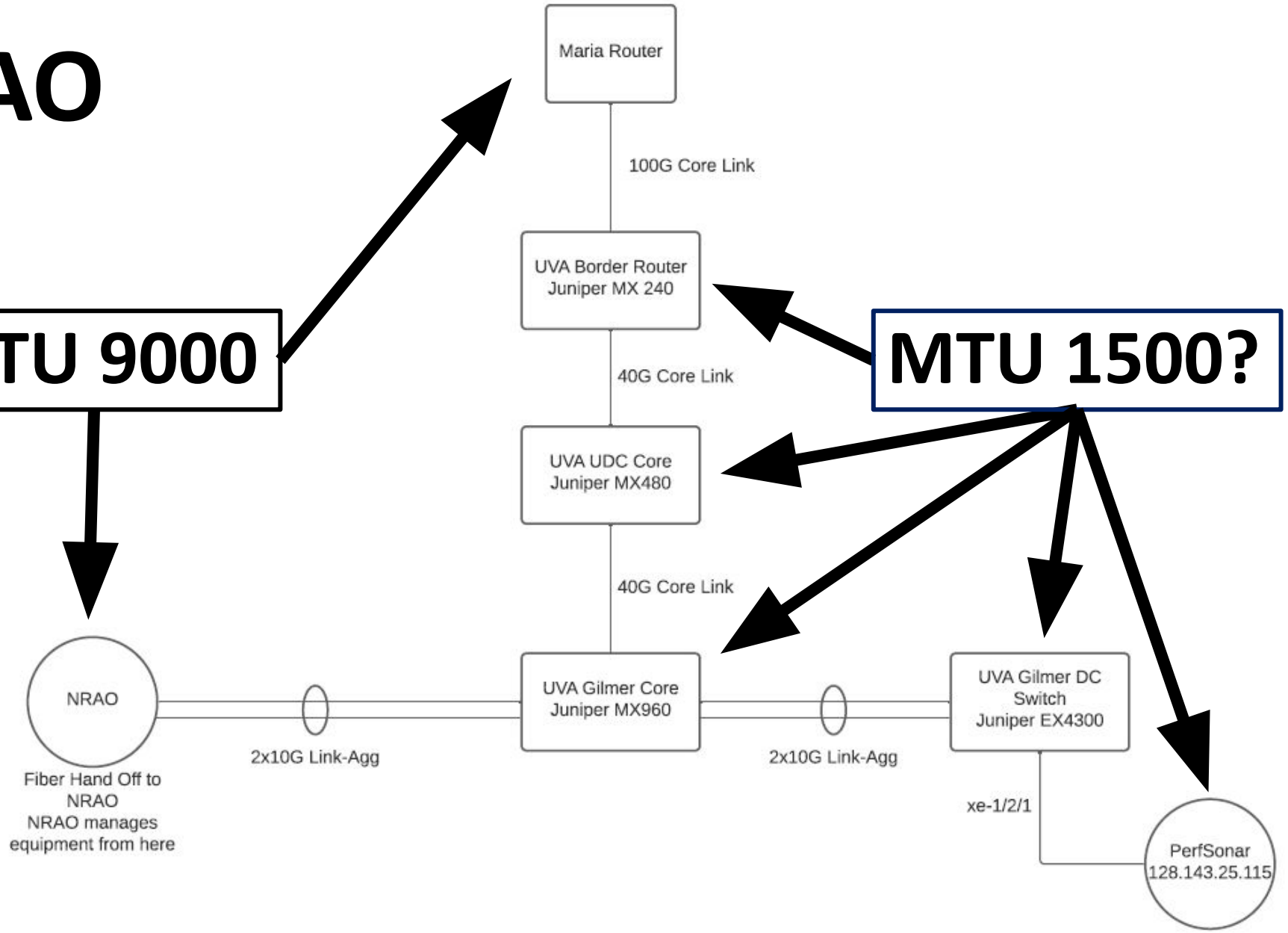
pS instance must have commodity internet access to be listed.

Initial problem isolation



- Tests from various domestic and international perfSONAR nodes to UVAs campus were telling:
  - CHPC South Africa -> Internet2 Washington - 6.67 Gbps
  - Internet2 Albany -> Internet2 Washington - 9.893 Gbps
  - Internet2 Washington -> NRAO - 3.31 Gbps (lots of retries)
  - Internet2 Washington -> HPC University Virginia - 2.21 Gbps (lots of retries)
  - NRAO -> HPC University Virginia - 6.64 Gbps (lots of retries)

UVA/NRAO Network



Path MTU Discovery (PMTUD)

- Is a layer 3 construct
- Requires UDP and ICMP to function
  - UDP packets larger than the MTU setting of the receiving router interface will trigger an ICMP “unreachable” message back to the sending router, which in turn causes a renegotiation to a lower MTU
- All is not lost if PMTUD doesn’t work
  - Smart transfer tools can figure out a common MTU, at the cost of time
  - Packets sent at 9K can be fragmented to adhere to a smaller MTU, at the cost of performance…unless the no-fragment flag is set
  - Neither of these scenarios is good for high performance. PMTUD should be made to work and common MTUs enforced wherever possible

Further isolation

Working inward from a known good ESnet perfSONAR node to UVA:

Interval	Throughput	Retransmits	Current Window
0.0 - 1.0	9.13 Gbps	22	33.17 MBytes
1.0 - 2.0	9.35 Gbps	0	33.58 MBytes
2.0 - 3.0	9.38 Gbps	0	33.58 MBytes
3.0 - 4.0	9.38 Gbps	0	33.58 MBytes
4.0 - 5.0	9.38 Gbps	0	33.58 MBytes
5.0 - 6.0	9.35 Gbps	0	33.58 MBytes
6.0 - 7.0	9.36 Gbps	0	33.58 MBytes
7.0 - 8.0	9.38 Gbps	0	33.58 MBytes
8.0 - 9.0	9.37 Gbps	0	33.58 MBytes
9.0 - 10.0	9.37 Gbps	0	33.58 MBytes

Summary

Interval	Throughput	Retransmits	Receiver Throughput
0.0 - 10.0	9.35 Gbps	22	9.25 Gbps

This test looks good, because the hosts successfully negotiate 1500 MTU

Negotiations break down

Working inward from a known good ESnet perfSONAR node to NRAO: (Keep in mind, we know MTU 9000 on both ends, but with a step down to 1500 in the middle of the UVA campus)

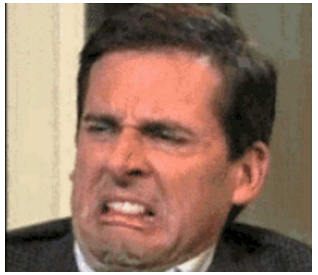
Interval	Throughput	Retransmits	Current Window
0.0 - 1.0	2.54 Mbps	2	8.95 KBytes
1.0 - 2.0	0.00bps	1	8.95 KBytes
2.0 - 3.0	0.00bps	0	8.95 KBytes
3.0 - 4.0	0.00bps	31	3.07 KBytes
4.0 - 5.0	0.00bps	67	5.12 KBytes
5.0 - 6.0	4.45 Mbps	2	17.41 KBytes
6.0 - 7.0	8.26 Mbps	0	33.79 KBytes
7.0 - 8.0	23.28 Mbps	0	94.21 KBytes
8.0 - 9.0	51.75 Mbps	0	218.11 KBytes
9.0 - 10.0	83.88 Mbps	0	392.19 KBytes

Summary

Interval	Throughput	Retransmits	Receiver Throughput
0.0 - 10.0	17.42 Mbps	103	10.29 Mbps

9000B packets failing

1500B packets after re-negotiation



Traceroute: ESnet to NRAO

```
traceroute to perfsonar-10.cv.nrao.edu (198.51.208.55), 30 hops max, 60 byte packets
 1  esneteastrt1-eastdcpt1.es.net (198.124.238.37)  0.549 ms  0.544 ms  0.547 ms
 2  newycr5-ip-a-esneteastrt1.es.net (198.124.218.17)  1.969 ms  1.963 ms  1.953 ms
 3  aofacr5-ip-a-newycr5.es.net (134.55.37.77)  2.330 ms 2.304 ms  2.313 ms
 4  et-2-1-5.197.rtsw.newy32aoa.net.internet2.edu (64.57.28.14)  2.323 ms  2.324 ms  2.327
ms
 5  ae-3.4079.rtsw.wash.net.internet2.edu (162.252.70.138)  7.571 ms  7.672 ms  7.528 ms
 6  ae-0.4079.rtsw2.ashb.net.internet2.edu (162.252.70.137)  8.095 ms  8.077 ms  8.061 ms
 7  ae-2.4079.rtsw.ashb.net.internet2.edu (162.252.70.74)  28.089 ms  18.414 ms  18.454 ms
 8  192.122.175.14 (192.122.175.14)  8.221 ms  8.179 ms  8.205 ms
 9  br01-udc-et-1-0-0-20.net.virginia.edu (192.35.48.33)  10.310 ms  10.310 ms  10.383 ms
10  cr01-udc-et-4-2-0.net.virginia.edu (128.143.236.6)  12.609 ms  12.603 ms  12.638 ms
11  cr01-gil-et-7-0-0.net.virginia.edu (128.143.236.89)  12.407 ms  12.403 ms  12.393 ms
12  perfsonar-10.cv.nrao.edu (198.51.208.55)  10.058 ms  10.032 ms  10.022 ms
```

Well, that looks good. Let's try tracepath and see where the MTU changes

Tracepath: ESnet to NRAO

```
1?: [LOCALHOST]                                      pmtu 9000
 1:  esneteastrt1-eastdcpt1.es.net                   0.788ms
 1:  bnlmr2-bnlpt1.es.net                            0.728ms
 2:  no reply
 3:  aofacr5-ip-b-newycr5.es.net                     2.411ms asymm  2
 4:  et-2-1-5.197.rtsw.newy32aoa.net.internet2.edu  2.468ms asymm  3
 5:  ae-3.4079.rtsw.wash.net.internet2.edu           8.176ms asymm  4
 6:  ae-0.4079.rtsw2.ashb.net.internet2.edu          8.889ms asymm  5
 7:  ae-2.4079.rtsw.ashb.net.internet2.edu           8.242ms asymm  6
 8:  192.122.175.14                                  8.522ms asymm  7
 9:  no reply
10:  no reply
11:  no reply
12:  no reply
```

Traceroute works, but tracepath doesn't??

Different Tools, Different Packets

- Traceroute uses small 60B UDP packets
- Tracepath uses larger 64KB UDP packets

So, somewhere we have a roadblock. Small packets can make it through, but larger ones are dropped (not fragmented).

How do we figure out the max size? Trial and error. Start at 9K and cut the size in half until you get a response, then sneak back up until the packets disappear again.

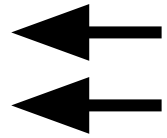
Tracepath: ESnet to NRAO, 1509 bytes

1: esneteastrt1-eastdcpt1.es.net 0.340ms
2: no reply
3: aofacr5-ip-a-newycr5.es.net 2.279ms asymm 2
4: et-2-1-5.197.rtsw.newy32aoa.net.internet2.edu 2.310ms asymm 3
5: ae-3.4079.rtsw.wash.net.internet2.edu 7.574ms asymm 4
6: ae-0.4079.rtsw2.ashb.net.internet2.edu 9.422ms asymm 5
7: ae-2.4079.rtsw.ashb.net.internet2.edu 7.986ms asymm 6
8: 192.122.175.14 8.123ms asymm 7
9: no reply

← MARIA
← UVA

Tracepath: ESnet to NRAO, 1508 bytes

```
1:  bnlmr2-bnlpt1.es.net                                0.327ms
2:  no reply
3:  aofacr5-ip-b-newycr5.es.net                          2.332ms asymm  2
4:  et-2-1-5.197.rtsw.newy32aoa.net.internet2.edu        2.338ms asymm  3
5:  ae-3.4079.rtsw.wash.net.internet2.edu                7.668ms asymm  4
6:  ae-0.4079.rtsw2.ashb.net.internet2.edu               9.833ms asymm  5
7:  ae-2.4079.rtsw.ashb.net.internet2.edu                7.872ms asymm  6
8:  192.122.175.14                                     8.166ms asymm  7
9:  br01-udc-et-1-0-0-20.net.virginia.edu                9.998ms asymm  7
9?: br01-udc-et-1-0-0-20.net.virginia.edu                asymm  7
10:  cr01-udc-et-4-2-0.net.virginia.edu                  10.470ms asymm  8
11:  cr01-gil-et-7-0-0.net.virginia.edu                  10.208ms asymm  9
12:  cr01-gil-et-7-0-0.net.virginia.edu                  10.253ms pmtu 1500
12:  perfsonar-10.cv.nrao.edu                            10.154ms !H
    Resume: pmtu 1500
```



MARIA
UVA

Problem located

- The issue was between the MARIA router and the UVA router
  - The MARIA interface was configured for MTU 9192
  - The UVA interface was configured for MTU 1518
- With PMTUD broken there was no hope for external MTU 9000 equipment to negotiate an appropriate MTU with the NRAO node
- UVA changed the MTU on their router interface to match that of MARIA, while keeping their downstream equipment at their campus standard MTU 1500

Ping verification

ping -s 8972 -M do -c 4 perfsonar-10.cv.nrao.edu (don't fragment)

PING perfsonar-10.cv.nrao.edu (198.51.208.55) 8972(9000) bytes of data.
From cr01-gil-et-7-0-0.net.virginia.edu (128.143.236.89) icmp_seq=1 Frag needed and DF set (mtu = 1500)
ping: local error: Message too long, mtu=1500
ping: local error: Message too long, mtu=1500
ping: local error: Message too long, mtu=1500

ping -s 8972 -M dont -c 4 perfsonar-10.cv.nrao.edu (do fragment)

PING perfsonar-10.cv.nrao.edu (198.51.208.55) 8972(9000) bytes of data.
8980 bytes from perfsonar-10.cv.nrao.edu (198.51.208.55): icmp_seq=1 ttl=55 time=10.3 ms
8980 bytes from perfsonar-10.cv.nrao.edu (198.51.208.55): icmp_seq=2 ttl=55 time=10.2 ms
8980 bytes from perfsonar-10.cv.nrao.edu (198.51.208.55): icmp_seq=3 ttl=55 time=10.2 ms
8980 bytes from perfsonar-10.cv.nrao.edu (198.51.208.55): icmp_seq=4 ttl=55 time=10.2 ms

Yeah, yeah, but what about performance??

Before:

pscheduler task throughput --source cpt-chpc-10g.perfsonar.ac.za --dest perfsonar-10.cv.nrao.edu

Summary

Interval	Throughput	Retransmits	Receiver Throughput
0.0 - 10.0	380.37 Kbps	58	108.18 Kbps

After:

pscheduler task throughput -t 30 --source cpt-chpc-10g.perfsonar.ac.za --dest perfsonar-10.cv.nrao.edu

Summary

Interval	Throughput	Retransmits	Receiver Throughput
0.0 - 30.0	2.67 Gbps	0	2.62 Gbps

You think we need one more?
You think we need one more.

Alright. We'll do one more.

City College of New York (CCNY) to Kyutech Institute (JGN)

Reported asymmetric, poor performance across GRE tunnel

- JGN □ CCNY (TCP)
  - No packet loss
  - 79Mbps throughput
- CCNY□ JGN (TCP)
  - 0.082% packet loss
  - 8Mbps throughput

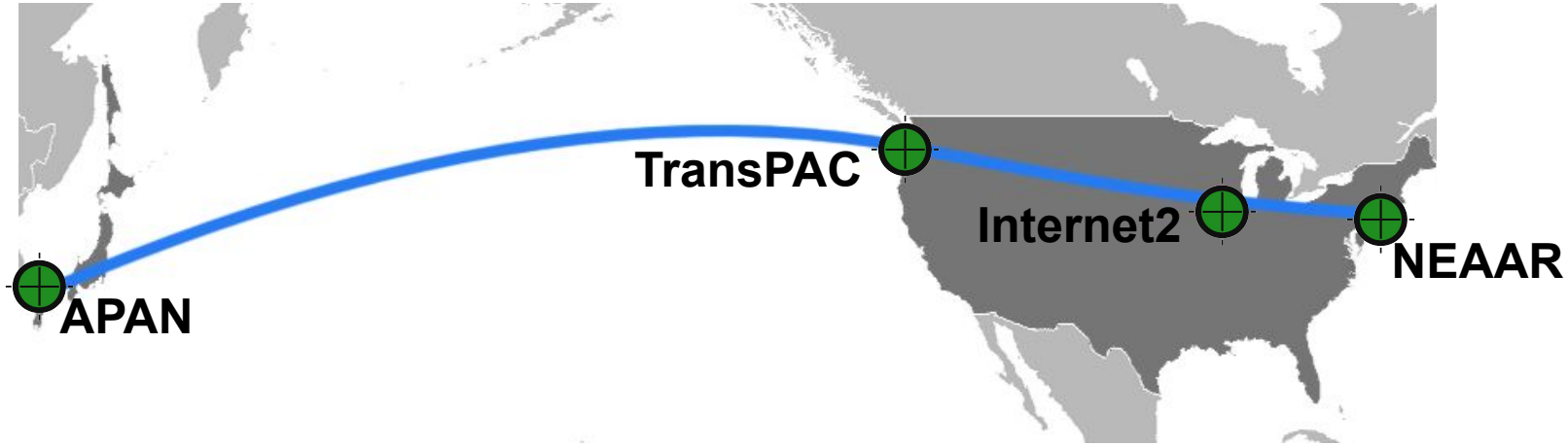
Tested UDP performance, however, was symmetric at 90Mbps either direction

Initial troubleshooting

Used perfSONAR nodes along the path to test to closest open node available-at ManLan.

Nodes located at

- APAN/Tokyo
- TransPAC/Seattle
- Internet2/Chicago
- NEAAR/ManLan



Testing to NYC showed good performance and no packet loss- indicating problem was likely within CCNY

Internal troubleshooting

- CCNY and EPOC engineers installed perfSONAR node in researcher's lab
- Tests from prior locations to lab showed same packet loss as original problem
- Verified issue within campus

Regional troubleshooting

- NYSERNet
  - Regional network for NY
  - Provides R&E connectivity for CCNY
  - Engineers installed a new CCNY pS node at campus edge
- Testing edge to lab
  - Packet fragmentation and MTU issues on the ingress path to CCNY
  - Packet loss isolated to specific segment of the CCNY campus network

Problem located

With this data

- CCNY engineers did additional local troubleshooting
- Cause identified as outdated network security device
- Replacement had been scheduled, expedited due to results

After replacement

- pS tests verified performance was greatly improved

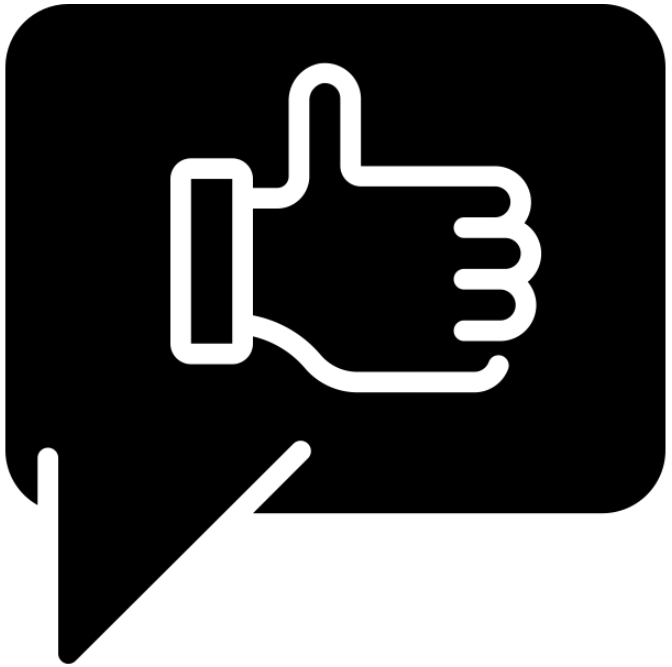
Final Results

- CCNY/JGN GRE tunnel shows consistent, symmetric performance
- JGN □ CCNY (TCP)
  - No packet loss
  - 80Mbps throughput
- CCNY □ JGN (TCP)
  - No packet loss
  - 85Mbps throughput
  - **10-fold improvement**

Question and answer icon by iconosphere from The Noun Project

Questions and Answers

perfSONAR

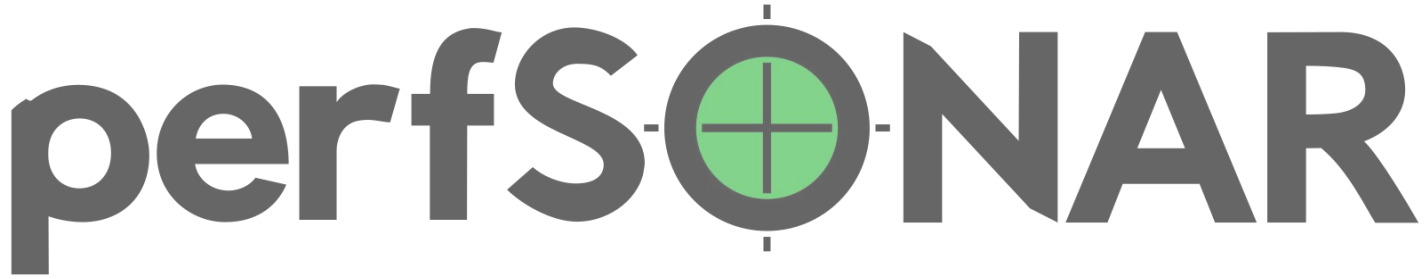


Thanks!

For more information,
please visit our web site:
https://www.perfsonar.net

Thanks icon by priyanka from The Noun Project

perfSONAR is developed by a partnership of ESnet GÉANT INDIANA UNIVERSITY INTERNET2 RNP ORGANIZAÇÃO SOCIAL DO MCTI UNIVERSITY OF MICHIGAN



CLI Testing and Troubleshooting

Doug Southworth · Texas Advanced Computing Center · dsouthworth@tacc.utexas.edu

perfSONAR is developed by a partnership of

INDIANA UNIVERSITY