AmLight SDN/SDX: Updates

SAACC 2018

May 18, 2018

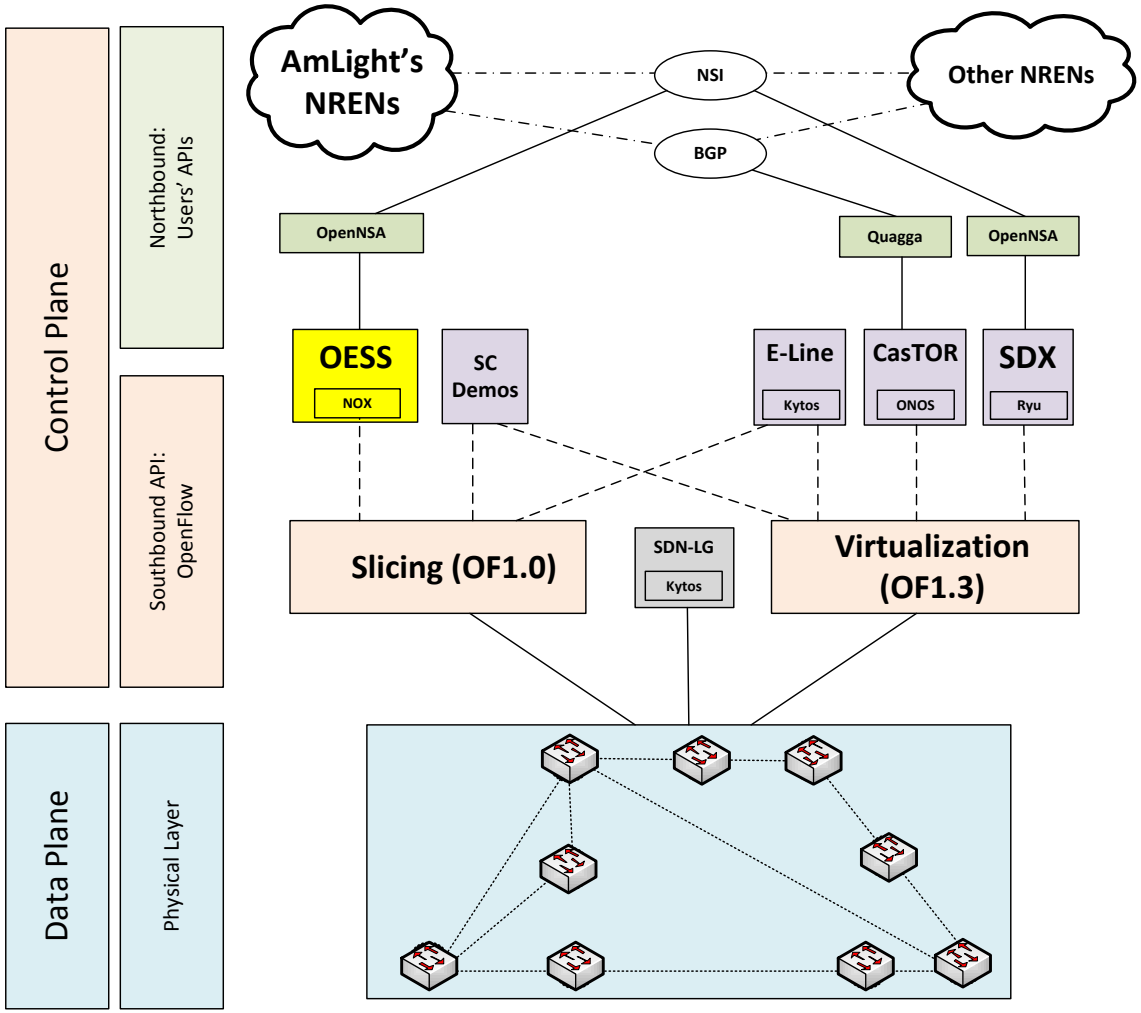
Jeronimo Bezerra and Vinicius Arcanjo

Outline

- AmLight SDN/SDX – Jeronimo
- Cooperative Troubleshooting between ALMA and NERSC - Vinicius

Software-Defined Networking @ AmLight

- **AmLight became an SDN network in 2014:**
  - Optimized provisioning time
  - Network Virtualization
- **Researchers can use virtualization to prototype their own applications**
  - Can implement testbeds with real network devices
  - Can validate their research in a production environment, and at scale
- **Current testbeds:**
  - ONOS/CasTOR, FIBRE, Kytos E-Line, Awave-SDX



Challenging Projects

- New scientific instruments being designed and deployed will increase the need for large, real-time data transfers among scientists throughout the world:
  - LSST will produce 12.7 GB data sets that must be transmitted to the U.S. in 5 seconds and it will be remotely operated;
  - LHC performing massive data transfer to multiple sites
  - ScienceDMZ data transfers, SKA, …
- End-to-end paths must provide high resilience, low delay, multiple paths, high bandwidth and an *efficient control plane* to act in all status changes

Challenging Projects [2]

- Most of the R&E networks can accommodate some of these requirements:
  - Multiple paths with multiple 100G links
  - Dynamic provisioning, bandwidth reservation, network programmability, etc.
- But R&E networks are interconnected through Academic Exchange Points:
  - Almost no support for programmability
- High demanding end-to-end applications require that all networks in the path support QoS and Programmability
  - Including at the Academic Exchange Points
- *Software Defined Exchanges/SDX as a possible solution*

SDX Motivation

- A Software Defined eXchange (SDX) seeks to introduce Software Defined Networking (SDN) technologies into Academic Exchange Points to optimize resource sharing and allocation
  - Inter-domain R&E network programmability
  - End-to-End QoS coordination and enforcement

IRNC: AtlanticWave-SDX

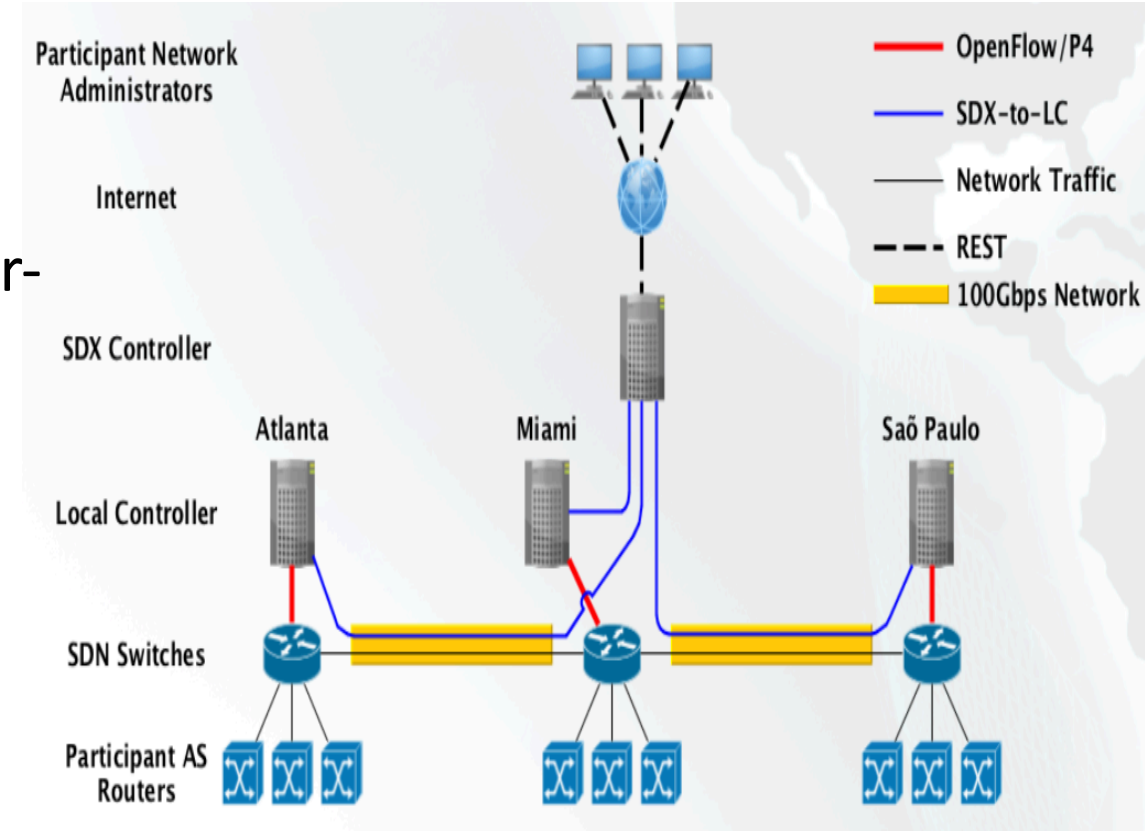
- AtlanticWave-SDX (Awave-SDX) is building a distributed intercontinental experimental SDX in response to a growing demand to:
  - Support end-to-end services capable of
    - Spanning multiple SDN domains
    - Dynamic provisioning of end-to-end circuits
    - Providing network programmability
  - Provide more intelligent network services to
    - Foster innovation
    - Increase network efficiency

- Florida International University (FIU) and Georgia Institute of Technology (GT) are implementing AtlanticWave-SDX, in collaboration with other exchange points supporting SDN

- Four sites will be part of the AtlanticWave-SDX:
  - SOX (Atlanta), AMPATH (Miami), SouthernLight (Sao Paulo), and AndesLight (Chile)

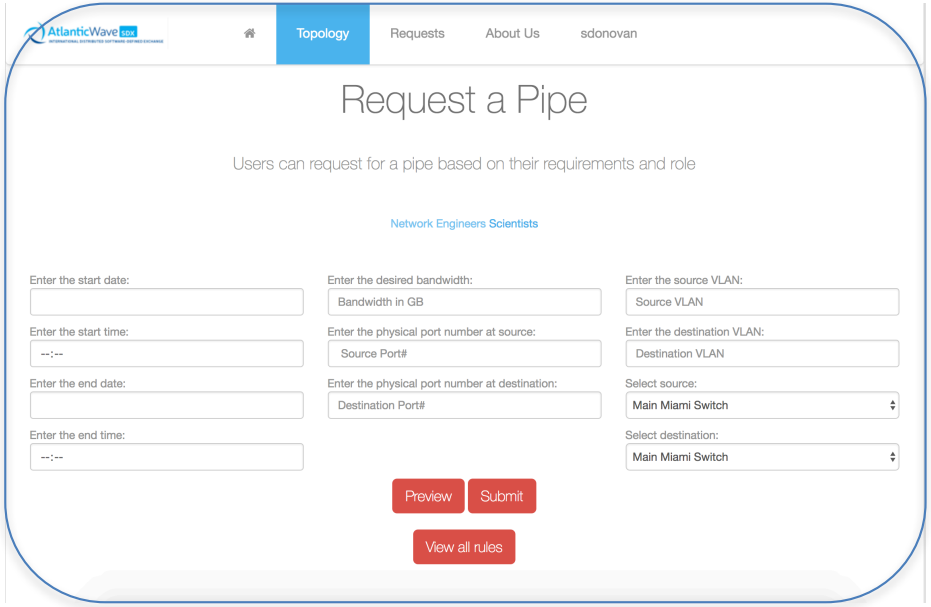
Current Network Design

- The SDX Controller:
  - Interfaces with external requests
  - Coordinates intra and inter-domain provisioning
- Each site will have its Local Controller:
  - Bootstrapping, topology discovery, southbound translation



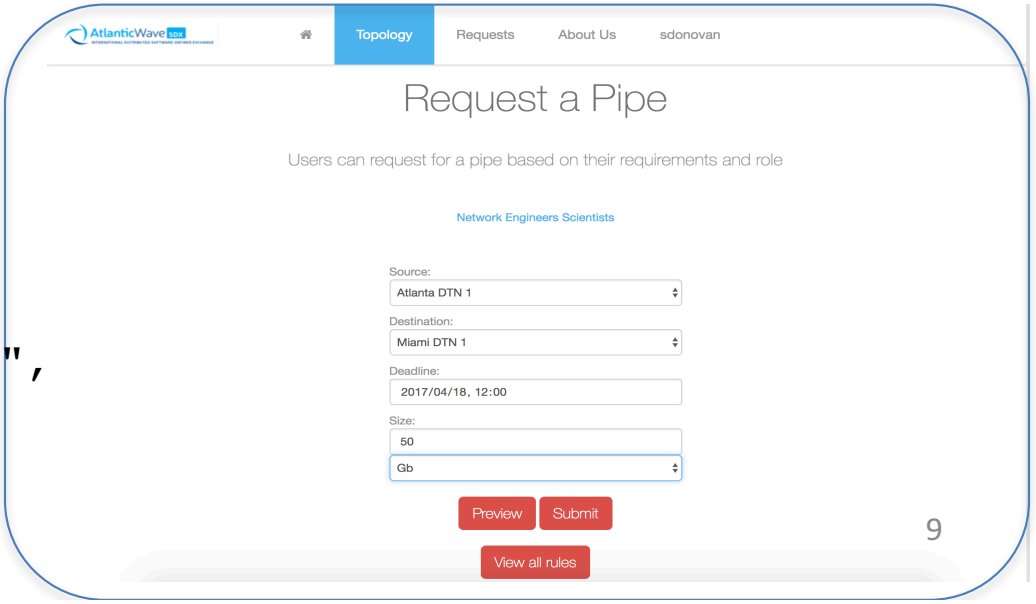
Multiple Interfaces

- User requests via WEB UI or REST calls
- Interface for Network Engineers and Domain Scientists




```
{"l2tunnel":
   {"starttime":"2016-10-12T23:20:50",
    "endtime":"2016-10-13T23:20:50",
    "srcswitch":"atl-switch",
    "dstswitch":"mia-switch",
    "srcport”:5,
    "dstport”:7,
    “srcvlan":1492,
    "dstvlan":1789,
    "bandwidth":1}}
```

```
{"dtntunnel":
   {"quantity":"7TB",
    "deadline":"2016-10-30T23:59:59",
    "srcdtn":"gt-dtn",
    "dstdtn":"fiu-dtn"}}
```



More Policy examples

- On-demand Virtual Circuit provisioning

```
if (current_latency > SLA_latency)
    secondary = findSecondaryPath()

while (current_latency > SLA_latency)
    LoadBalance(primary, secondary)
```

- Bandwidth Calendaring

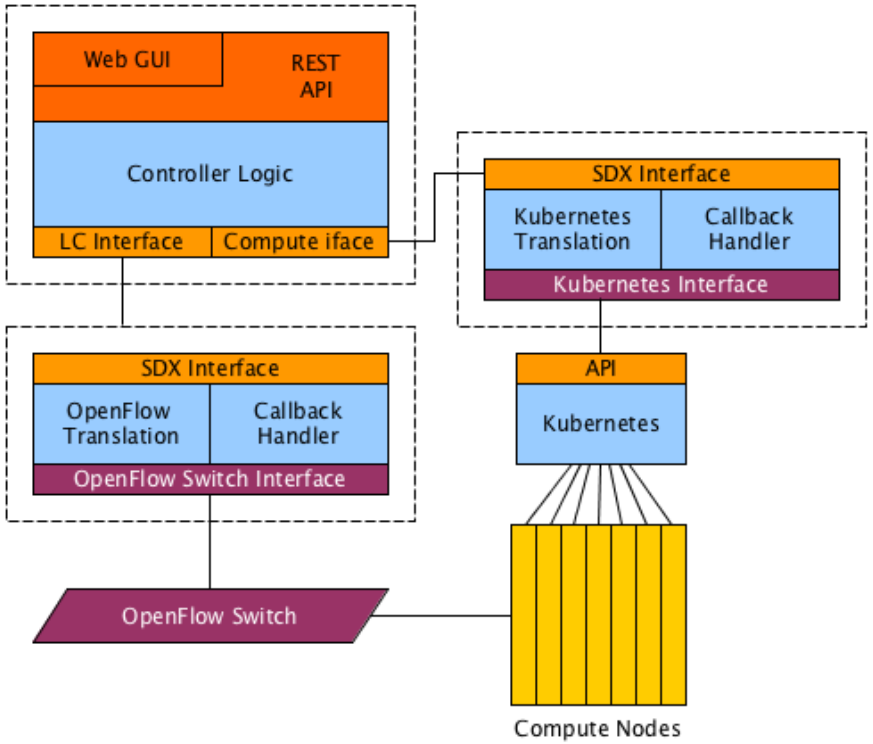
```
scheduled_time = 21:00:00 GMT -5
if (current_time == scheduled_time) {
    BW = 90 // Bandwidth in Mbps
    t = 60 // Reservation time
    OnDemandVC(BW, t)
}
```

AtlanticWave-SDX: Current Features

- **Layer 2 Point-to-Point and *Point-to-Multipoint* circuits with bandwidth reservation and MAC learning**
- Web UI and REST calls customized per user profile
- REST supporting HTML and *JSON* replies
- Support for complex data plane pipelines and arbitrary advanced rules
- Support for Inbound NSI requests for inter-domain provisioning
- Support for Docker and Vagrant images
- Shibboleth

Features Planned for 2018

- Per-User Resource Authorization
- In-Band Controller Communication/Bootstrapping
- *Investigation* of possible integration with compute resources



Deployment Plans for 2018

- Installation of SDX switches in Brazil and Chile
- A L2VPN will be created connecting SOX to AMPATH via Internet2
- L2VPNs will connect AMPATH, AndesLight, and SouthernLight's SDX switches

- An Awave-SDX *prototype* will connect all sites

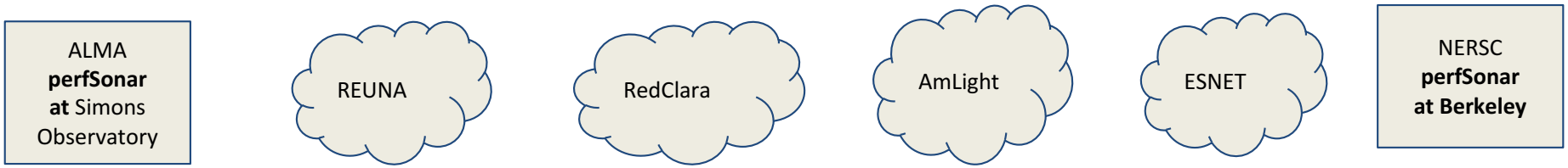
Astronomers/Domain Scientists will be invited to try it out!

Cooperative Troubleshooting perfSonar data transfer between ALMA and NERSC

Vinicius Arcanjo
<vinicius.arcanjo@rnp.br>

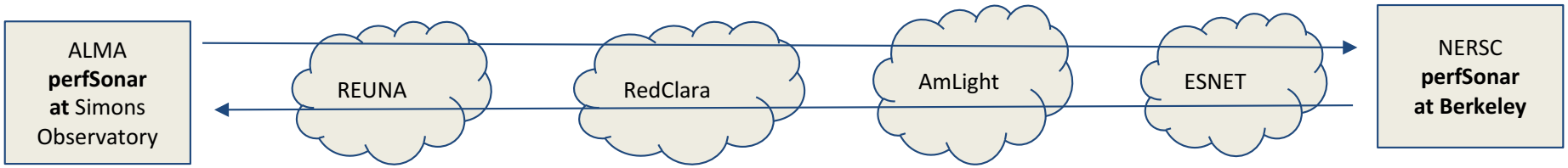
Abstract Topology

- Several NRENs involved:



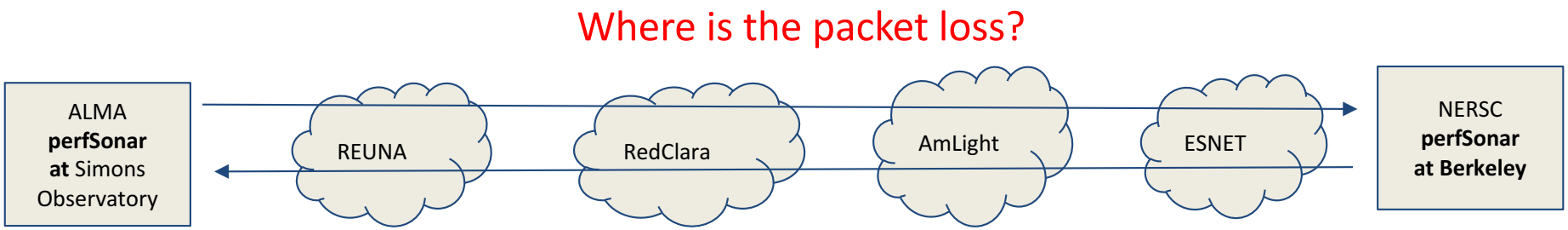
Abstract Topology

- Poor performance for data transfers identified between ALMA Simons and NERSC in Berkeley



Abstract Topology

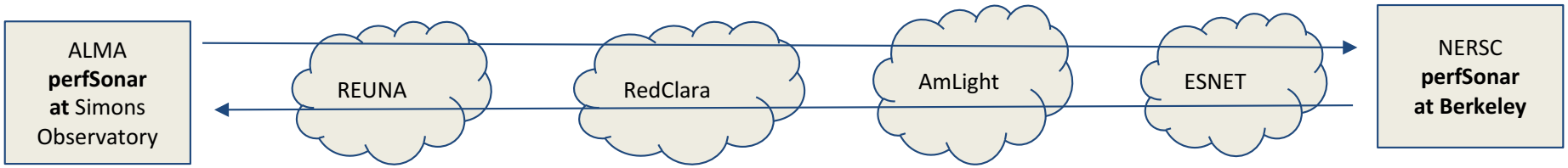
- Data transfer between ALMA Simons and NERSC in Berkeley:

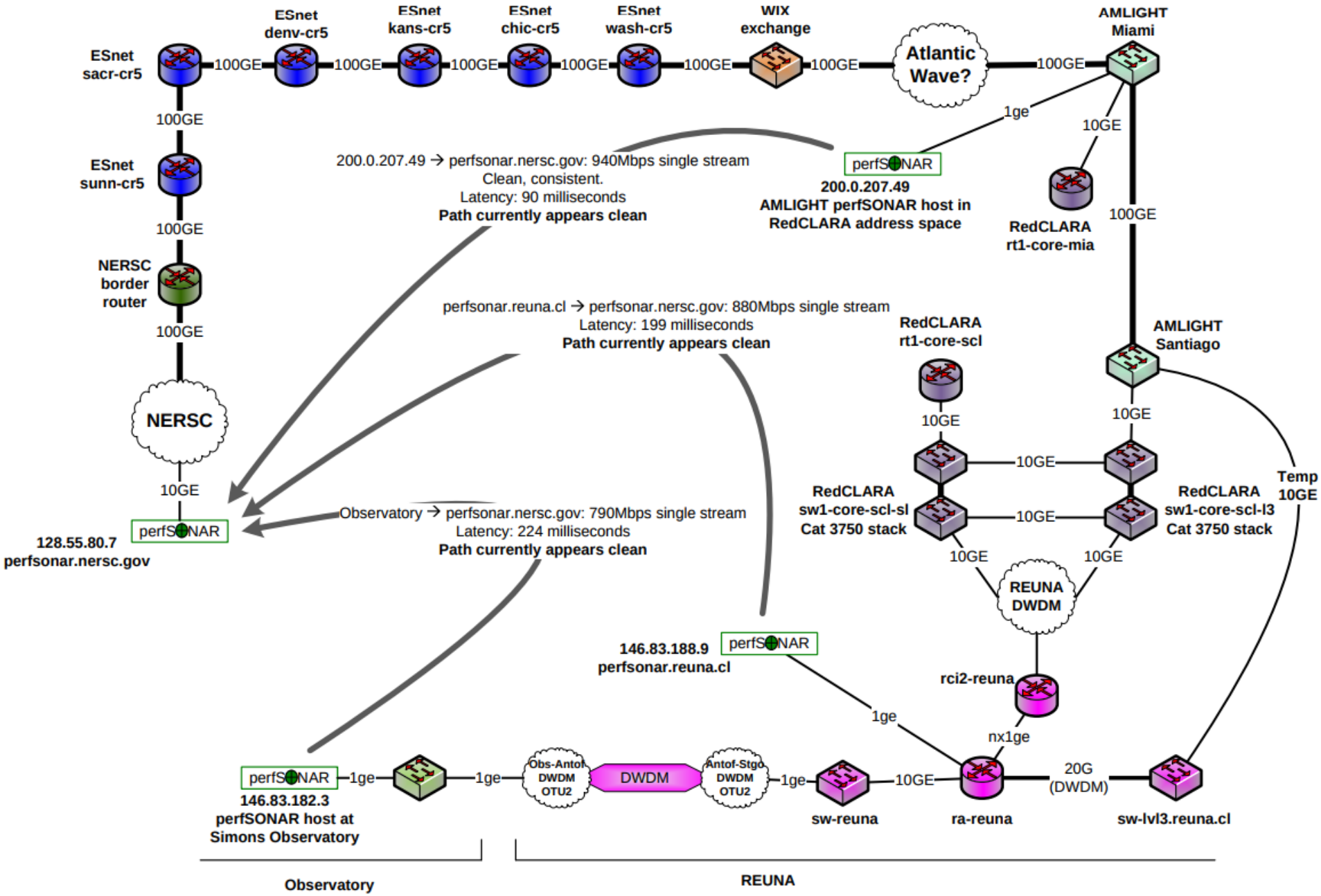


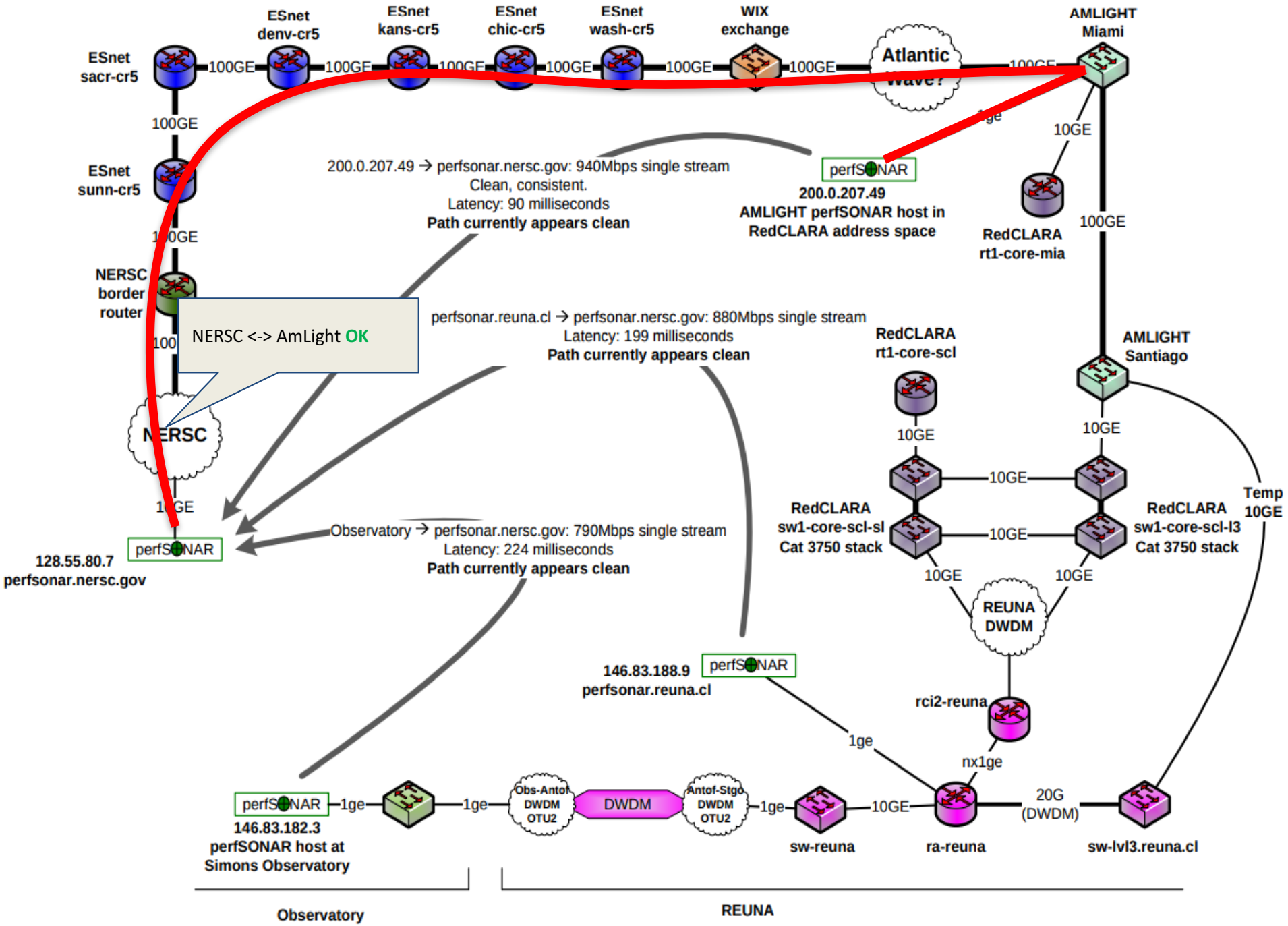
Abstract Topology

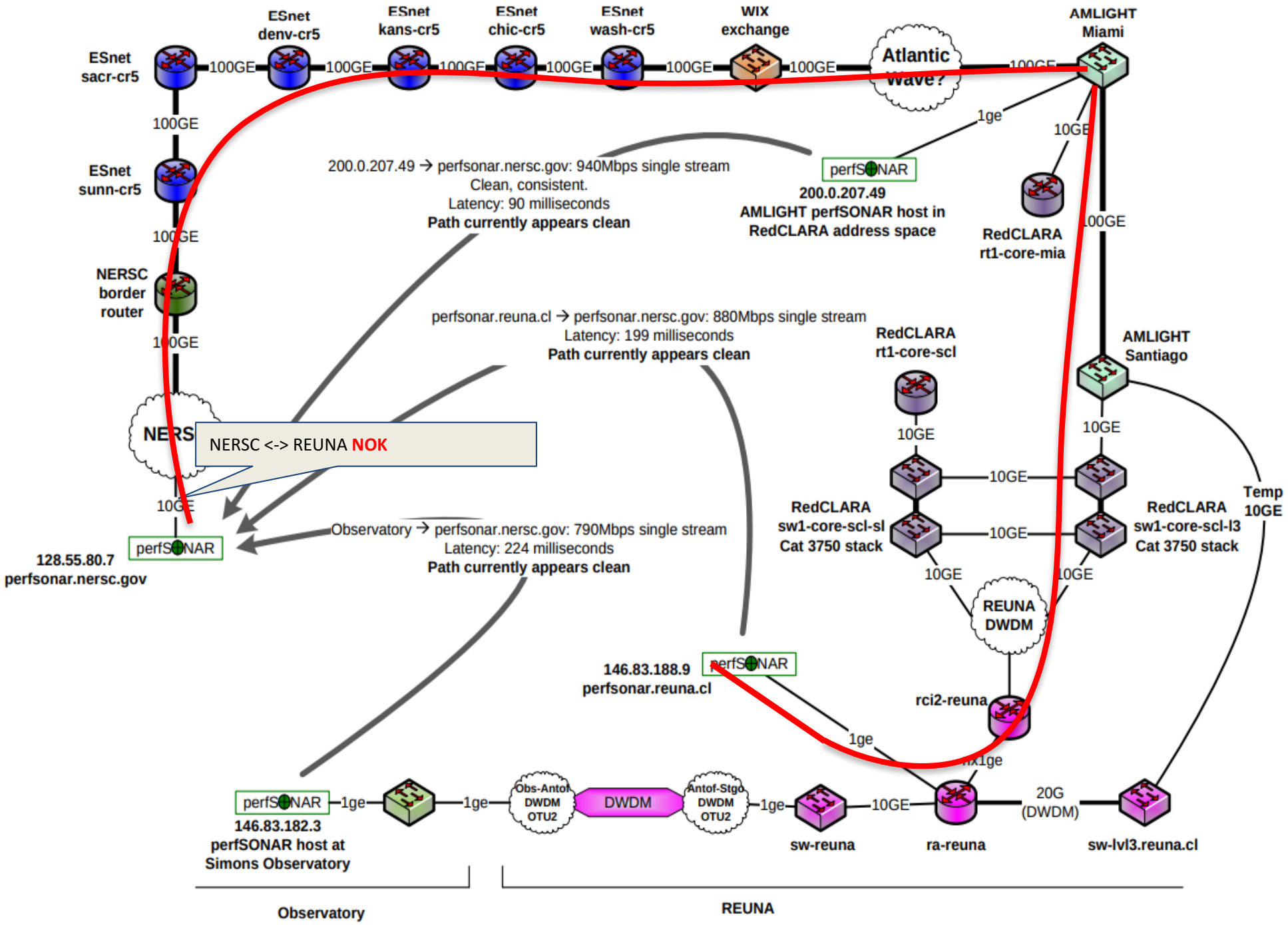
- Isolating link faults in each domain….

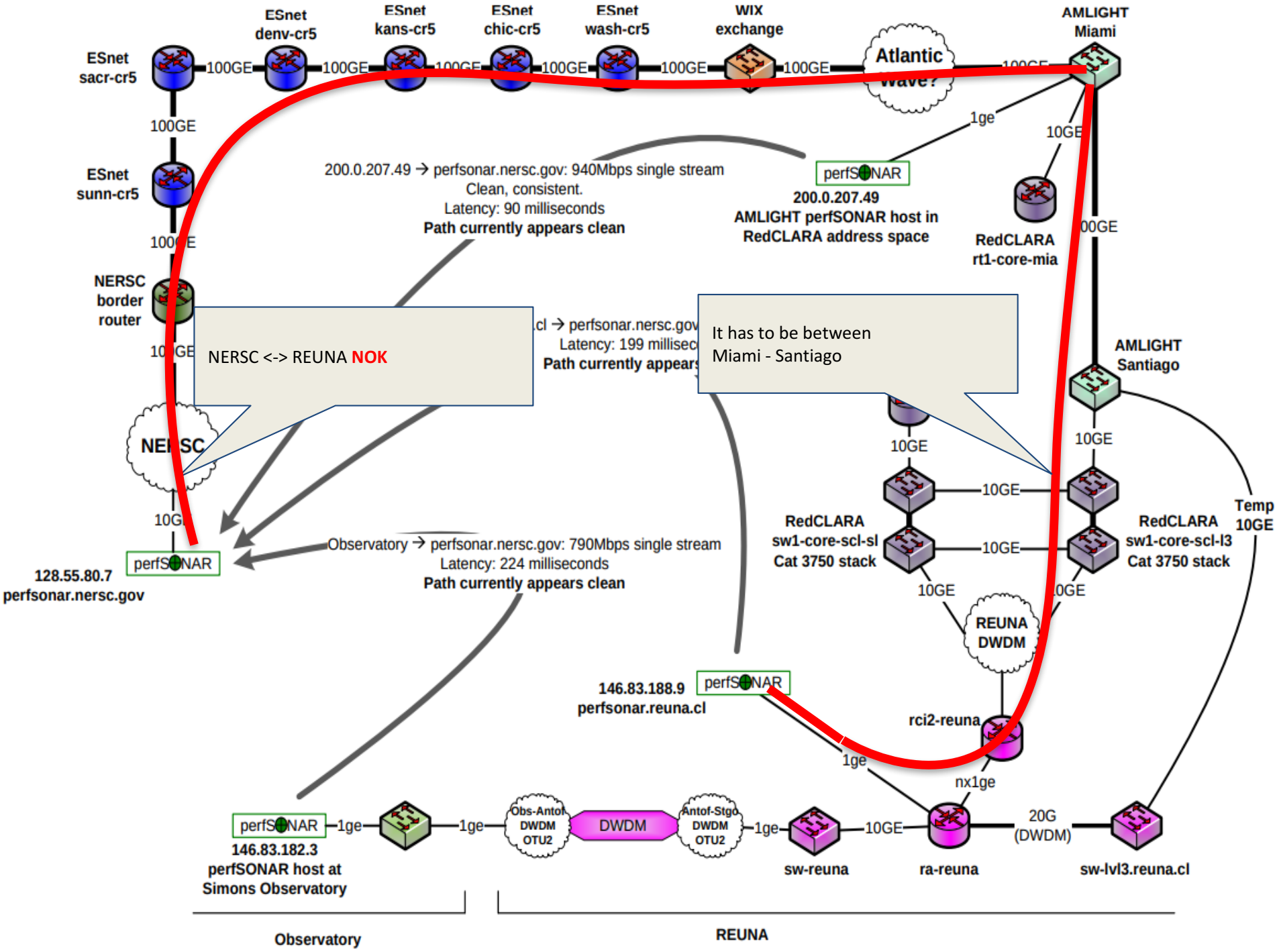
Let’s zoom in on each cloud…..

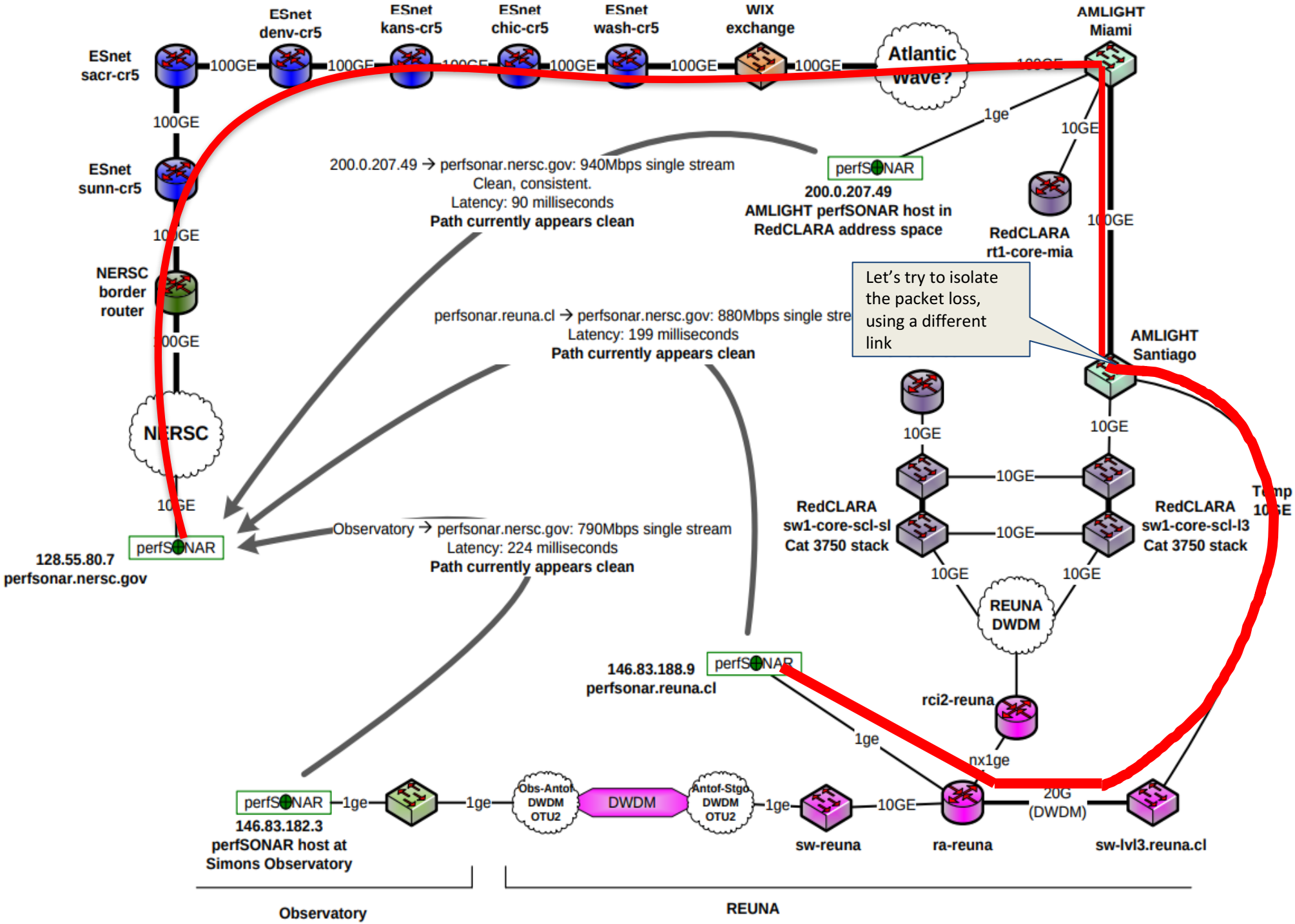


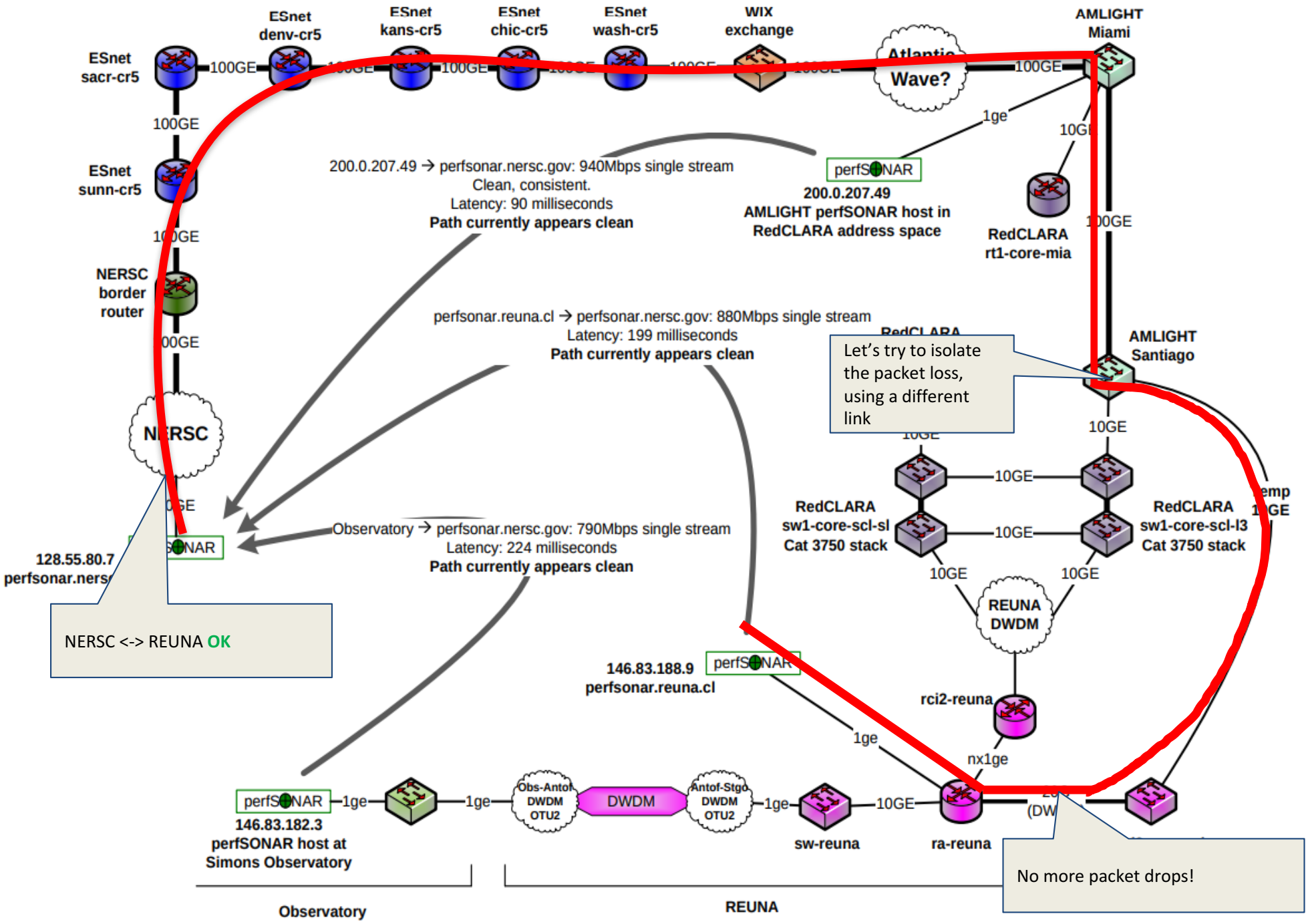

ESnet sacr-cr5
ESnet denv-cr5
ESnet kans-cr5
ESnet chic-cr5
ESnet wash-cr5
WIX exchange
Atlantic Wave?
AMLIGHT Miami
100GE
ESnet sunn-cr5
NERSC border router
NERSC
10GE
perfSONAR
128.55.80.7
perfsonar.nersc.gov
200.0.207.49 → perfsonar.nersc.gov: 940Mbps single stream
Clean, consistent.
Latency: 90 milliseconds
Path currently appears clean
perfsonar.reuna.cl → perfsonar.nersc.gov: 880Mbps single stream
Latency: 199 milliseconds
Path currently appears clean
Observatory → perfsonar.nersc.gov: 790Mbps single stream
Latency: 224 milliseconds
Path currently appears clean
200.0.207.49
AMLIGHT perfSONAR host in RedCLARA address space
1ge
RedCLARA rt1-core-mia
RedCLARA rt1-core-scl
AMLIGHT Santiago
RedCLARA sw1-core-scl-sl Cat 3750 stack
RedCLARA sw1-core-scl-l3 Cat 3750 stack
Temp 10GE
REUNA DWDM
146.83.188.9
perfsonar.reuna.cl
rci2-reuna
nx1ge
20G (DWDM)
sw-lvl3.reuna.cl
146.83.182.3
perfSONAR host at Simons Observatory
Obs-Antof DWDM OTU2
DWDM
Antof-Stgo DWDM OTU2
sw-reuna
ra-reuna
Observatory
REUNA

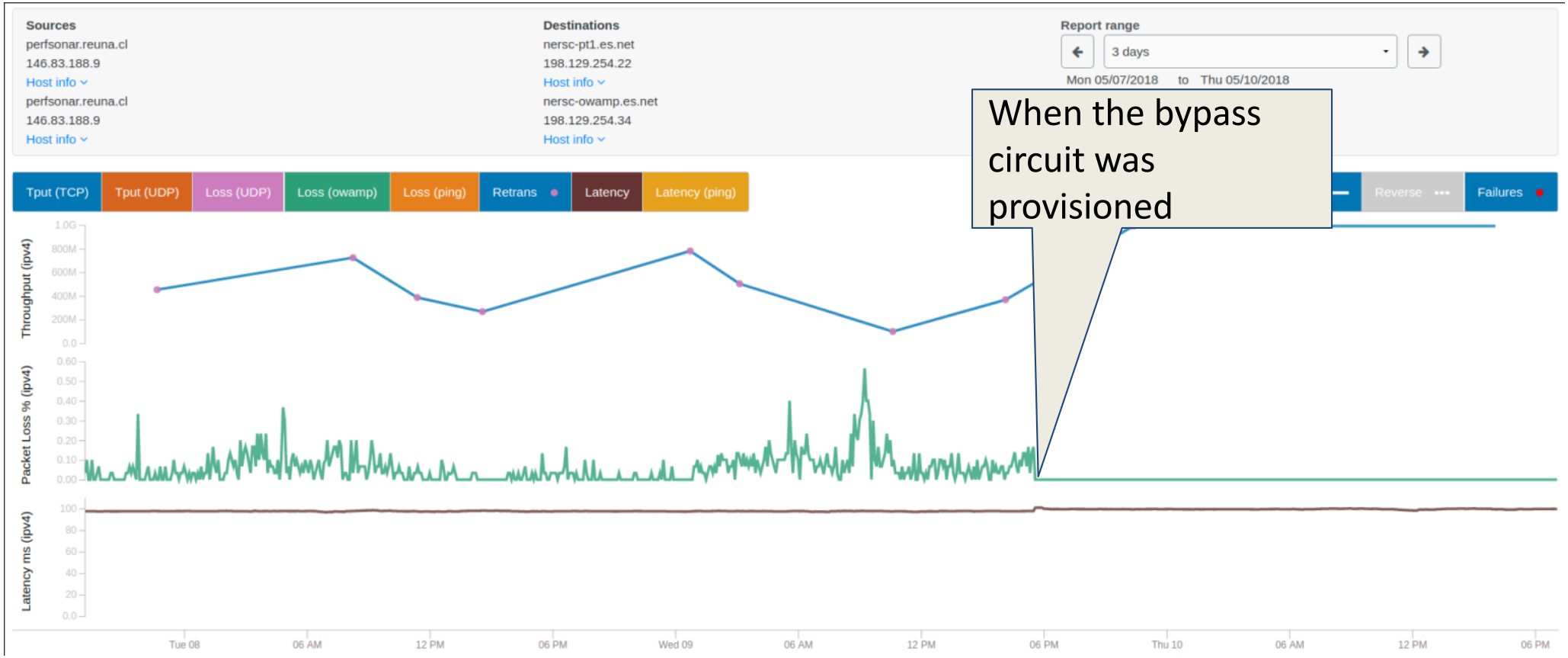

ESnet sacr-cr5
ESnet denv-cr5
ESnet kans-cr5
ESnet chic-cr5
ESnet wash-cr5
WIX exchange
Atlantic Wave?
AMLIGHT Miami
100GE
ESnet sunn-cr5
NERSC border router
NERSC
128.55.80.7
perfsonar.nersc.gov
perfSONAR
200.0.207.49 → perfsonar.nersc.gov: 940Mbps single stream
Clean, consistent.
Latency: 90 milliseconds
Path currently appears clean
200.0.207.49
AMLIGHT perfSONAR host in RedCLARA address space
RedCLARA rt1-core-mia
10GE
NERSC <-> AmLight OK
perfsonar.reuna.cl → perfsonar.nersc.gov: 880Mbps single stream
Latency: 199 milliseconds
Path currently appears clean
RedCLARA rt1-core-scl
AMLIGHT Santiago
RedCLARA sw1-core-scl-sl Cat 3750 stack
RedCLARA sw1-core-scl-l3 Cat 3750 stack
Temp 10GE
REUNA DWDM
Observatory → perfsonar.nersc.gov: 790Mbps single stream
Latency: 224 milliseconds
Path currently appears clean
146.83.188.9
perfsonar.reuna.cl
rci2-reuna
1ge
nx1ge
146.83.182.3
perfSONAR host at Simons Observatory
Obs-Antof DWDM OTU2
DWDM
Antof-Stgo DWDM OTU2
sw-reuna
ra-reuna
20G (DWDM)
sw-lvl3.reuna.cl
Observatory
REUNA


ESnet sacr-cr5
ESnet denv-cr5
ESnet kans-cr5
ESnet chic-cr5
ESnet wash-cr5
WIX exchange
Atlantic Wave?
AMLIGHT Miami
100GE
ESnet sunn-cr5
NERSC border router
200.0.207.49 → perfsonar.nersc.gov: 940Mbps single stream
Clean, consistent.
Latency: 90 milliseconds
Path currently appears clean
perfSONAR
200.0.207.49
AMLIGHT perfSONAR host in RedCLARA address space
RedCLARA rt1-core-mia
1ge
10GE
perfsonar.reuna.cl → perfsonar.nersc.gov: 880Mbps single stream
Latency: 199 milliseconds
Path currently appears clean
RedCLARA rt1-core-scl
AMLIGHT Santiago
NERS
NERSC <-> REUNA NOK
RedCLARA sw1-core-scl-sl Cat 3750 stack
RedCLARA sw1-core-scl-l3 Cat 3750 stack
Temp 10GE
Observatory → perfsonar.nersc.gov: 790Mbps single stream
Latency: 224 milliseconds
Path currently appears clean
128.55.80.7
perfsonar.nersc.gov
REUNA DWDM
146.83.188.9
perfsonar.reuna.cl
rci2-reuna
1ge
nx1ge
perfSONAR
1ge
Obs-Antof DWDM OTU2
DWDM
Antof-Stgo DWDM OTU2
10GE
20G (DWDM)
146.83.182.3
perfSONAR host at Simons Observatory
sw-reuna
ra-reuna
sw-lvl3.reuna.cl
Observatory
REUNA


ESnet sacr-cr5
ESnet denv-cr5
ESnet kans-cr5
ESnet chic-cr5
ESnet wash-cr5
WIX exchange
Atlantic Wave?
AMLIGHT Miami
100GE
ESnet sunn-cr5
NERSC border router
200.0.207.49 → perfsonar.nersc.gov: 940Mbps single stream
Clean, consistent.
Latency: 90 milliseconds
Path currently appears clean
perfSONAR
200.0.207.49
AMLIGHT perfSONAR host in RedCLARA address space
RedCLARA rt1-core-mia
1ge
10GE
NERSC <-> REUNA NOK
.cl → perfsonar.nersc.gov
Latency: 199 millisec
Path currently appears
It has to be between Miami - Santiago
AMLIGHT Santiago
NERSC
128.55.80.7
perfsonar.nersc.gov
Observatory → perfsonar.nersc.gov: 790Mbps single stream
Latency: 224 milliseconds
Path currently appears clean
RedCLARA sw1-core-scl-sl Cat 3750 stack
RedCLARA sw1-core-scl-l3 Cat 3750 stack
Temp 10GE
REUNA DWDM
146.83.188.9
perfsonar.reuna.cl
rci2-reuna
nx1ge
perfSONAR
1ge
Obs-Antof DWDM OTU2
DWDM
Antof-Stgo DWDM OTU2
146.83.182.3
perfSONAR host at Simons Observatory
sw-reuna
ra-reuna
20G (DWDM)
sw-lvl3.reuna.cl
Observatory
REUNA


ESnet sacr-cr5
ESnet denv-cr5
ESnet kans-cr5
ESnet chic-cr5
ESnet wash-cr5
WIX exchange
Atlantic Wave?
AMLIGHT Miami
100GE
ESnet sunn-cr5
NERSC border router
NERSC
perfSONAR
128.55.80.7
perfsonar.nersc.gov
200.0.207.49 → perfsonar.nersc.gov: 940Mbps single stream
Clean, consistent.
Latency: 90 milliseconds
Path currently appears clean
perfsonar.reuna.cl → perfsonar.nersc.gov: 880Mbps single stre
Latency: 199 milliseconds
Path currently appears clean
Observatory → perfsonar.nersc.gov: 790Mbps single stream
Latency: 224 milliseconds
Path currently appears clean
perfSONAR
200.0.207.49
AMLIGHT perfSONAR host in RedCLARA address space
1ge
10GE
RedCLARA rt1-core-mia
Let's try to isolate the packet loss, using a different link
AMLIGHT Santiago
RedCLARA sw1-core-scl-sl Cat 3750 stack
RedCLARA sw1-core-scl-l3 Cat 3750 stack
REUNA DWDM
146.83.188.9
perfsonar.reuna.cl
rci2-reuna
nx1ge
20G (DWDM)
sw-reuna
ra-reuna
sw-lvl3.reuna.cl
Obs-Antof DWDM OTU2
DWDM
Antof-Stgo DWDM OTU2
146.83.182.3
perfSONAR host at Simons Observatory
Observatory
REUNA


ESnet sacr-cr5
ESnet denv-cr5
ESnet kans-cr5
ESnet chic-cr5
ESnet wash-cr5
WIX exchange
Atlantic Wave?
AMLIGHT Miami
100GE
ESnet sunn-cr5
NERSC border router
200.0.207.49 → perfsonar.nersc.gov: 940Mbps single stream
Clean, consistent.
Latency: 90 milliseconds
Path currently appears clean
perfSONAR
200.0.207.49
AMLIGHT perfSONAR host in RedCLARA address space
RedCLARA rt1-core-mia
1ge
10GE
perfsonar.reuna.cl → perfsonar.nersc.gov: 880Mbps single stream
Latency: 199 milliseconds
Path currently appears clean
Let's try to isolate the packet loss, using a different link
AMLIGHT Santiago
NERSC
128.55.80.7
perfsonar.nersc
NERSC <-> REUNA OK
Observatory → perfsonar.nersc.gov: 790Mbps single stream
Latency: 224 milliseconds
Path currently appears clean
RedCLARA sw1-core-scl-sl Cat 3750 stack
RedCLARA sw1-core-scl-l3 Cat 3750 stack
REUNA DWDM
146.83.188.9
perfsonar.reuna.cl
rci2-reuna
nx1ge
Obs-Antof DWDM OTU2
DWDM
Antof-Stgo DWDM OTU2
146.83.182.3
perfSONAR host at Simons Observatory
sw-reuna
ra-reuna
No more packet drops!
Observatory
REUNA


Sources
perfsonar.reuna.cl
146.83.188.9
Host info
perfsonar.reuna.cl
146.83.188.9
Host info
Destinations
nersc-pt1.es.net
198.129.254.22
Host info
nersc-owamp.es.net
198.129.254.34
Host info
Report range
3 days
Mon 05/07/2018 to Thu 05/10/2018
When the bypass circuit was provisioned
Tput (TCP)
Tput (UDP)
Loss (UDP)
Loss (owamp)
Loss (ping)
Retrans
Latency
Latency (ping)
Reverse
Failures
Throughput (ipv4)
1.0G
800M
600M
400M
200M
0.0
Packet Loss % (ipv4)
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Latency ms (ipv4)
100
80
60
40
20
0.0
Tue 08
06 AM
12 PM
06 PM
Wed 09
06 AM
12 PM
06 PM
Thu 10
06 AM
12 PM
06 PM

- Thank you all Engineers involved:
  - Eli Dart, Alejandro, Albert, Sandra, Marco Teixeira, Gustavo, and many others.

- Thank you all Engineers involved!
- There’s a long road of troubleshooting ahead!

- Thank you all Engineers involved
- There’s a long road of troubleshooting ahead!
- Fortunately, everyone is engaged!

- Thank you all Engineers involved
- There’s a long road of troubleshooting ahead!
- Fortunately, everyone is engaged!
- The collaboration happened in a short time frame

- Thank you all Engineers involved
- There’s a long road of troubleshooting ahead!
- Fortunately, everyone is engaged!
- The collaboration happened in a short time frame.
- SDN/SDX can bring even more agility to support data science!