Design and Demo of AtlanticWave/SDX

Sean Donovan, Russ Clark
Georgia Tech

Jeronimo Bezerra, Julio Ibarra
Florida International University

NSF International Research
Network Connections (IRNC)
Grant #ACI-1341024

Heidi Morgan

Joaquin Chung, Cas D'Angelo,
Ankita Lamba, John Skandalakis

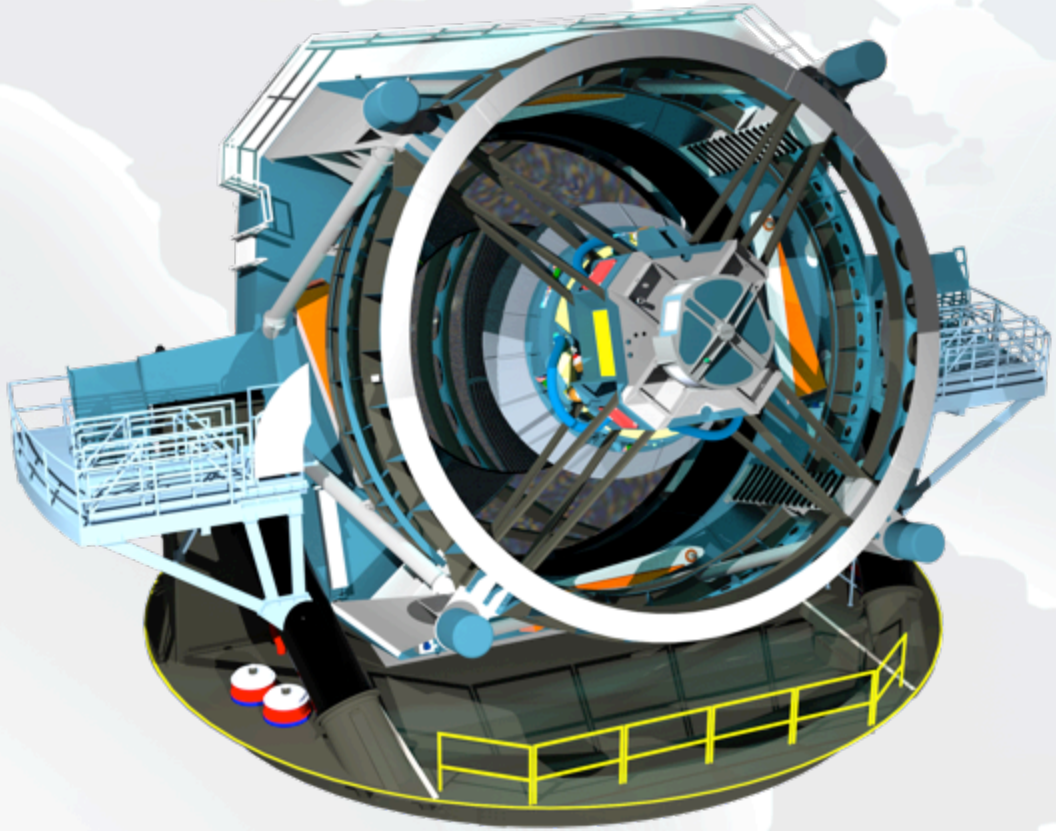
FIU

Georgia Tech

Agenda

- Introduction
- Controller Design Overview
- Controller Features
- Demo 1 – AtlanticWave/SDX
- Demo 2 – RNP Network

Large Synoptic Survey Telescope (LSST)



- High in the mountains in northern Chile
- Engineering First Light in 2019, Science First Light in 2021

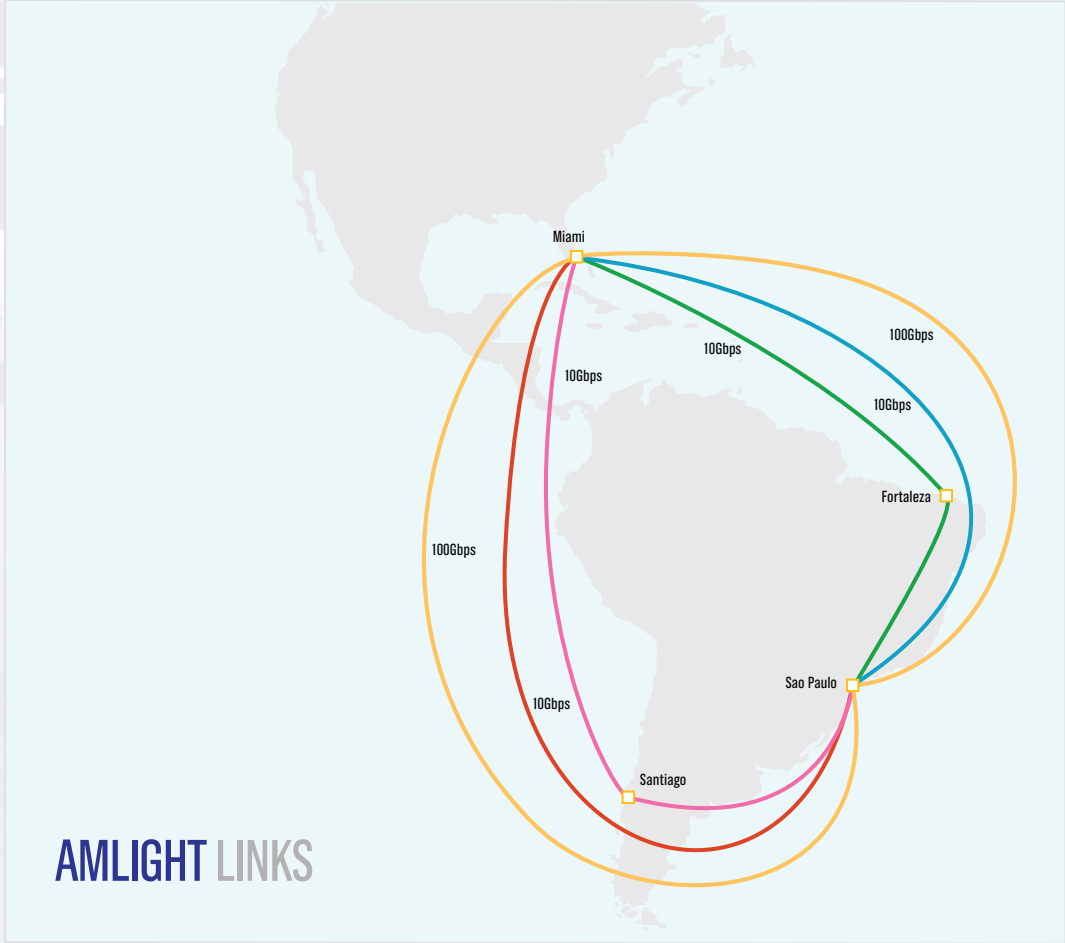
Source: https://www.lsst.org/gallery/telescope-rendering-2013

Huge Bandwidth Requirements

- 8.4 meter primary mirror with 3.2 Gigapixel sensor
- 12.7 GB image taken every 17 seconds
- Needs to be sent from Chile to NCSA/Illinois in 5 seconds
- Peak burst bandwidth of 65 Gbps
- In use all night long

New Connection

- Amlight is commissioning a new 100Gbps network connection between North and South America
- AtlanticWave/SDX sonnects Atlanta, Miami, and São Paulo over the AMLIGHT network
- Opportunity to innovate with the network

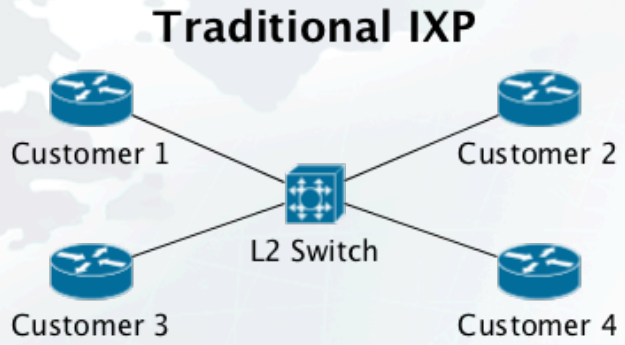


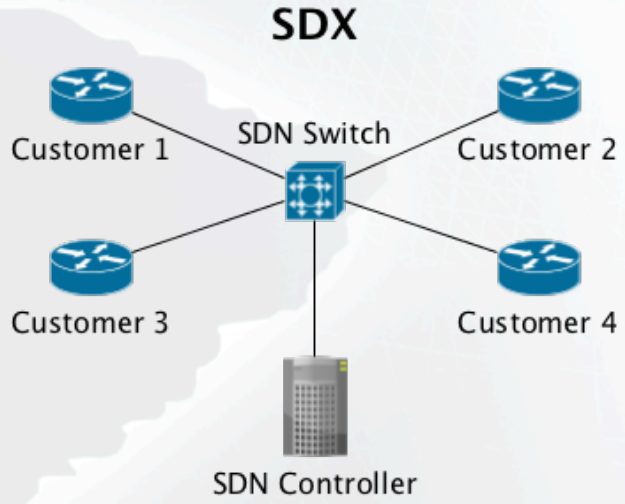
AtlanticWave/SDX

- SDX, but with a twist
  - Multiple, international locations
  - Multiple administrative domains
  - L2 tunneling functionality in addition to SDX functionality
- Lots of telescope data
  - But what about during the day?
  - Have opportunity to do something more interesting

What is an SDX?

- IXP + SDN
  - Not just L2 like an IXP
  - Where participants can write rules
- Multi-site IXP
  - AMS-IX has 10 sites in and around Amsterdam
  - Same administrative domain
- New functionality enabled by SDN at the IXP
  - Not bound by BGP restrictions
  - Application-specific peering

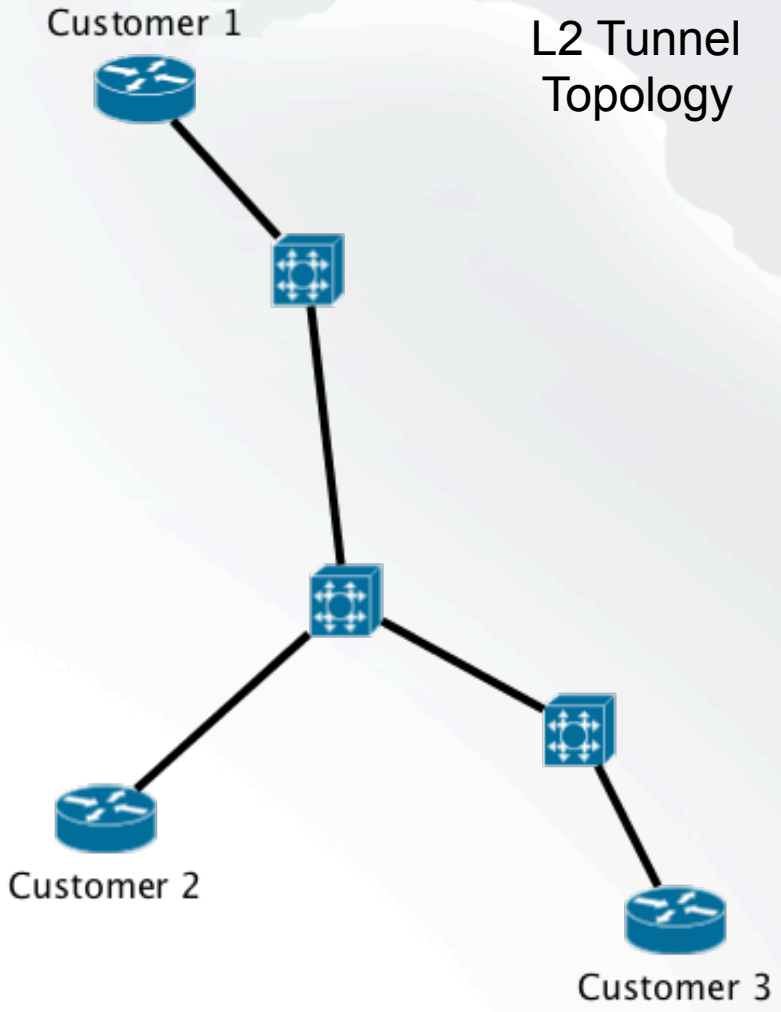


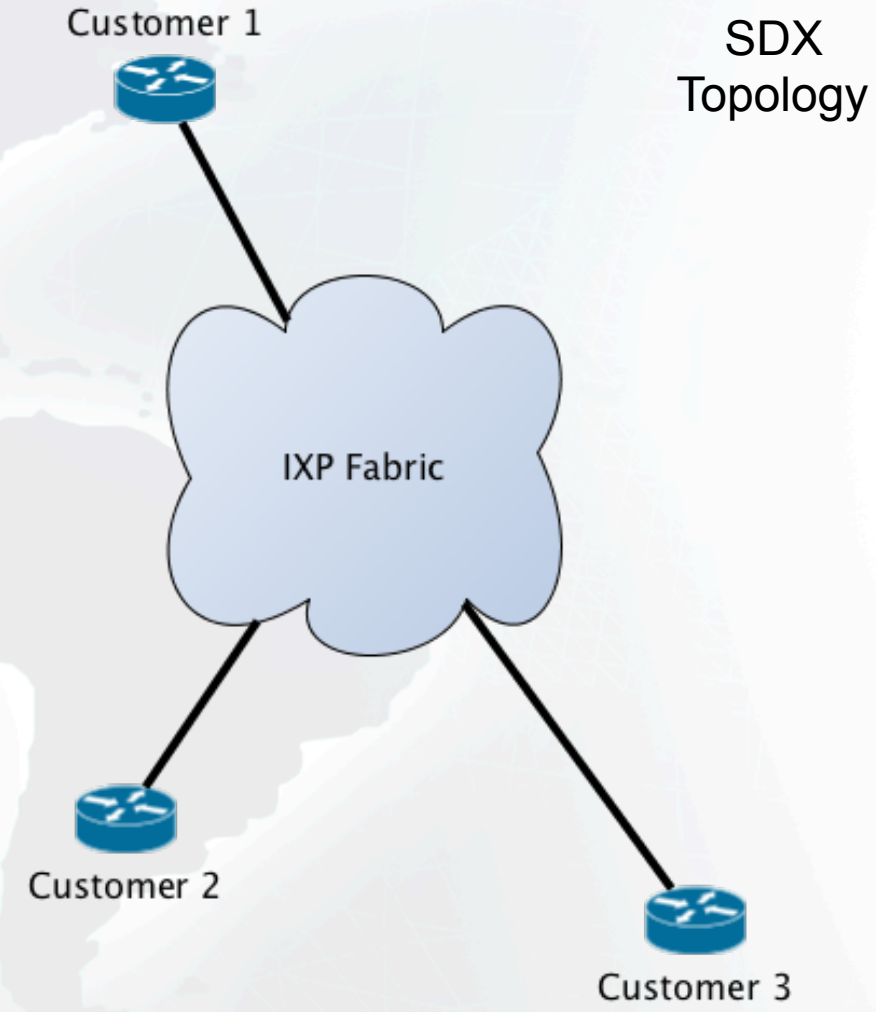


AtlanticWave/SDX

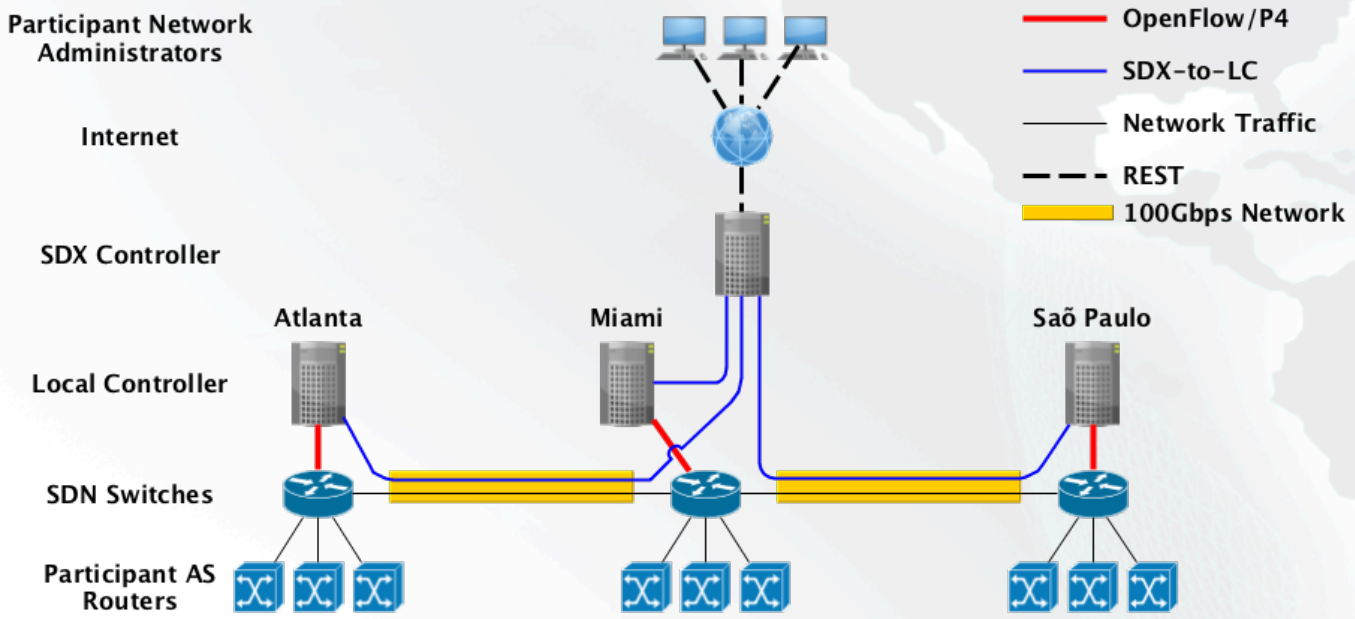
- Two main types of functions we care about
  - L2 Tunnel functionality
    - AL2S, OSCARS, NSI
  - SDX Functionality
    - Useful rules at an IXP, steering traffic
- Why not both?

Different Views For Different Functions



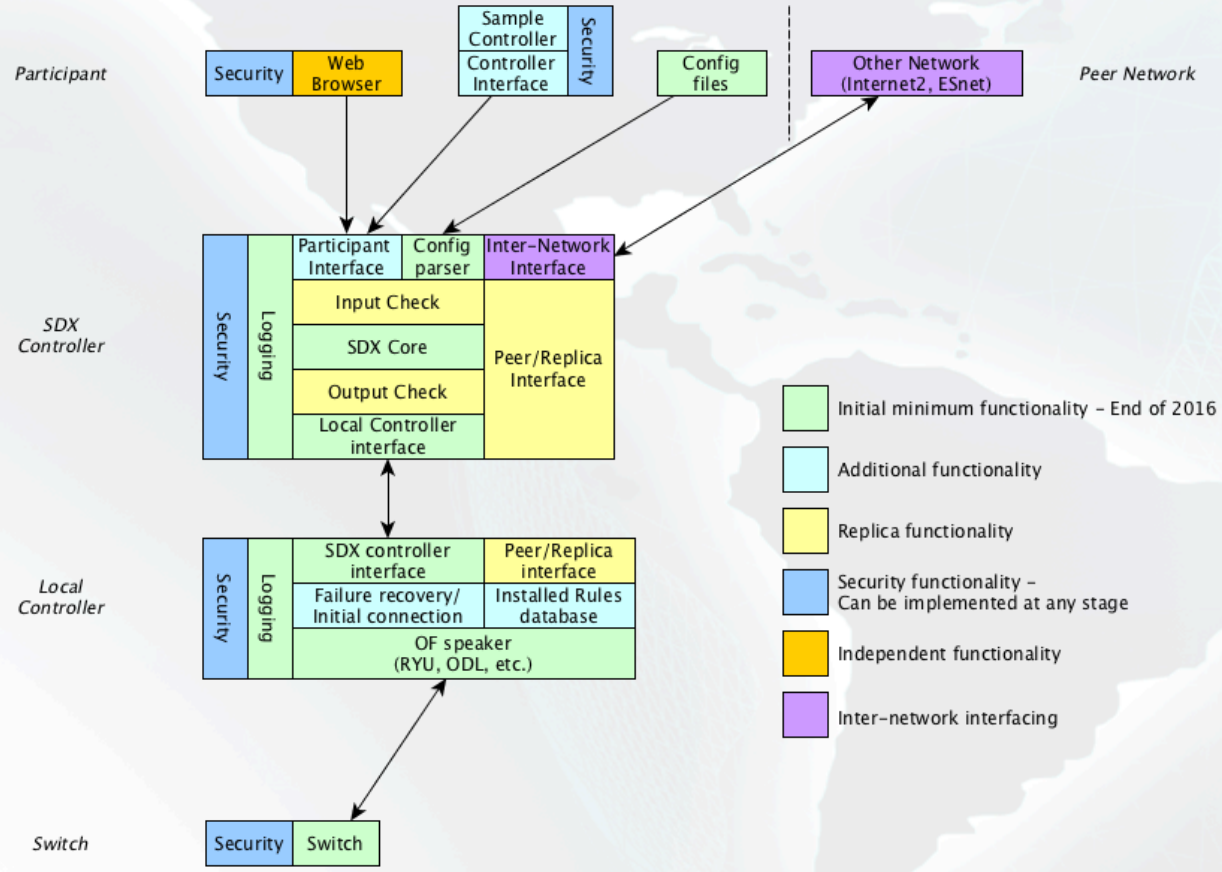


Design Overview



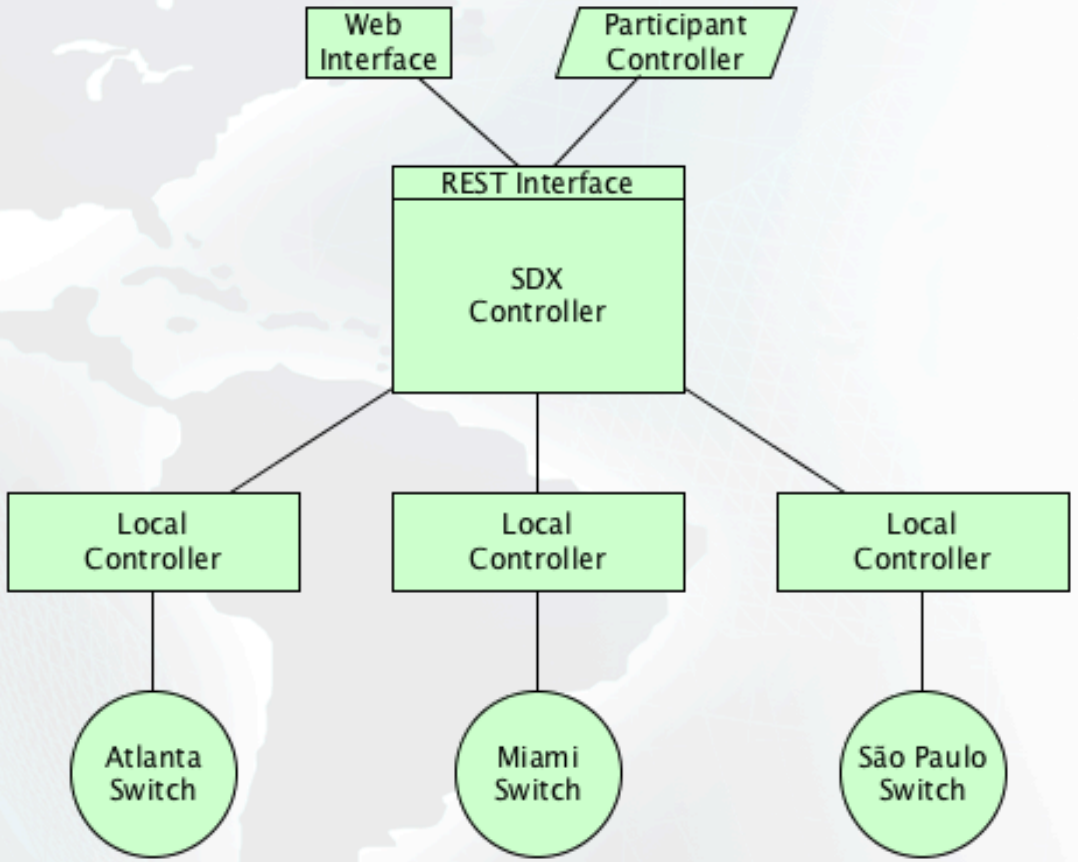
- Initially, three locations to cover
- Thousands of KM of fiber between each location
- Split controller design
  - Central controller for interacting with users
  - Local controllers at each location

Split Controller Design



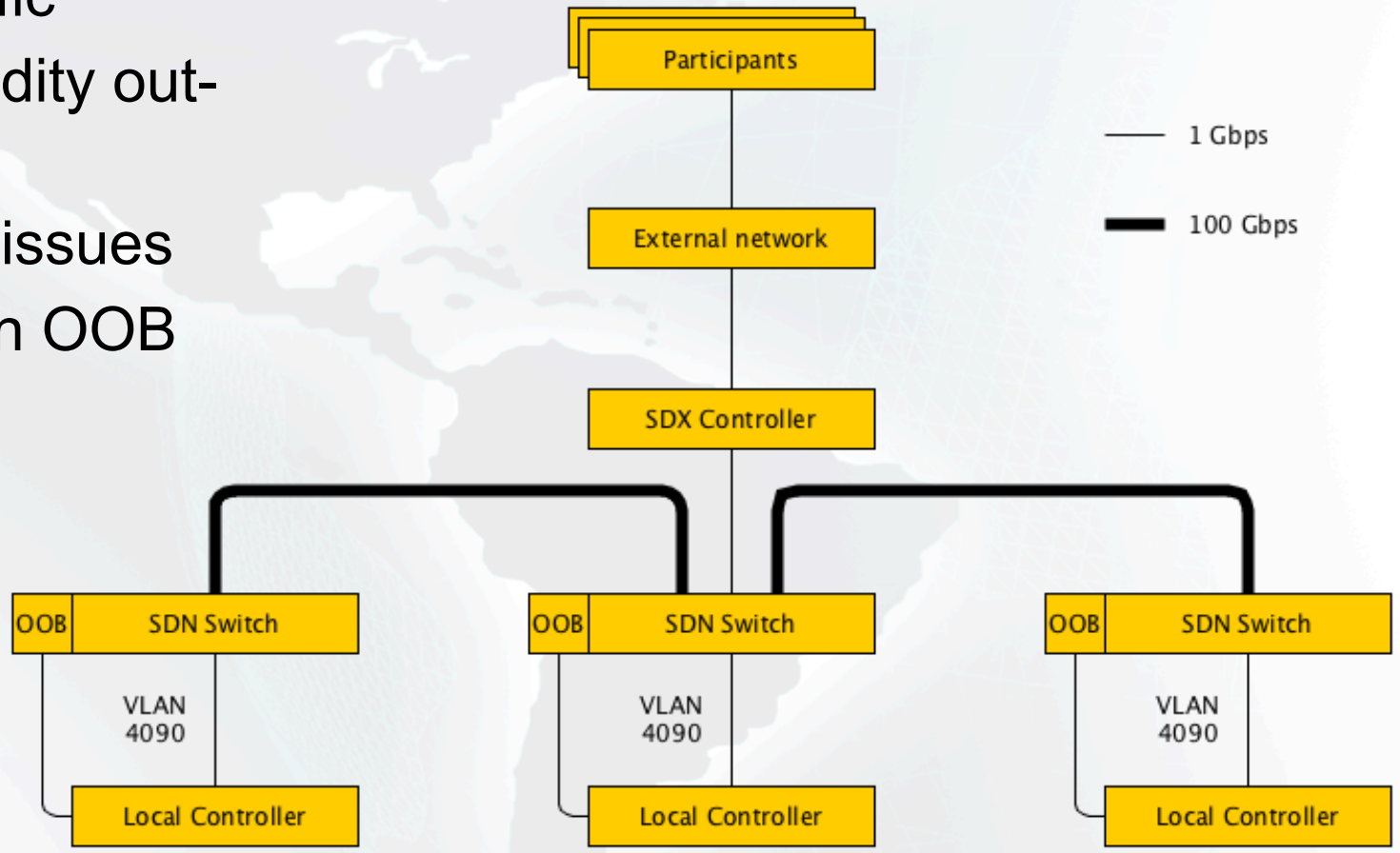
Interfaces

- REST API
- SDX-to-LC
- LC-to-Switch



Management

- In-band management traffic
- Known delays vs. commodity out-of-band connection
- Helps with some security issues
- Switches still controlled on OOB port
- LC bootstraps switches



Multiple Rule Tables

- L2 Tunneling
  - Large amount of traffic will likely be moved through L2 tunnels
- Each participant has two types of SDX rules
  - Inbound – rules for packets coming into the participant's network
    - 0.0.0.0/24 put on VLAN 3, forward to network
    - 128.0.0.0/24 put on VLAN 4, forward to network
  - Outbound – rules for packets leaving participant's network
    - Strip VLAN tag, forward to neighbor
- Learning switch as backup
  - When all else fails…

Functionality

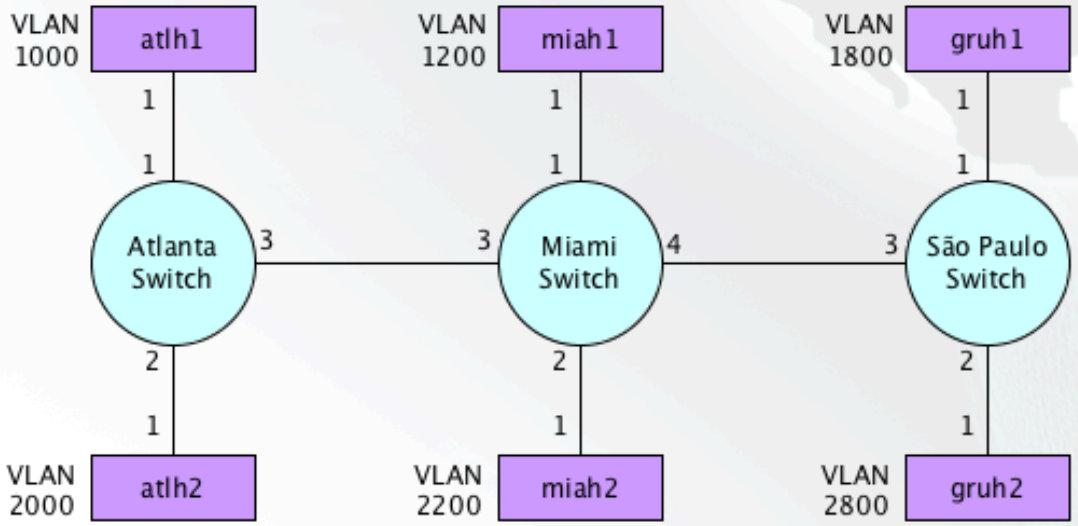
- L2 tunnels
  - Point-to-point
  - Multipoint
- SDX functionality
  - Network-header based rules
  - DNS-based rules
  - Other abstract rules
- In-Band Management
- Local Controller support for:
  - OpenFlow
  - P4
  - Others? Cisco or Juniper?

- NSI Integration for L2 Tunnels
  - Inbound
  - Outbound
- Limited BGP integration for SDX correctness
- User Interfaces
  - Easy to use Web interface
  - REST API
  - Example external controller using REST
- Shibboleth Integration

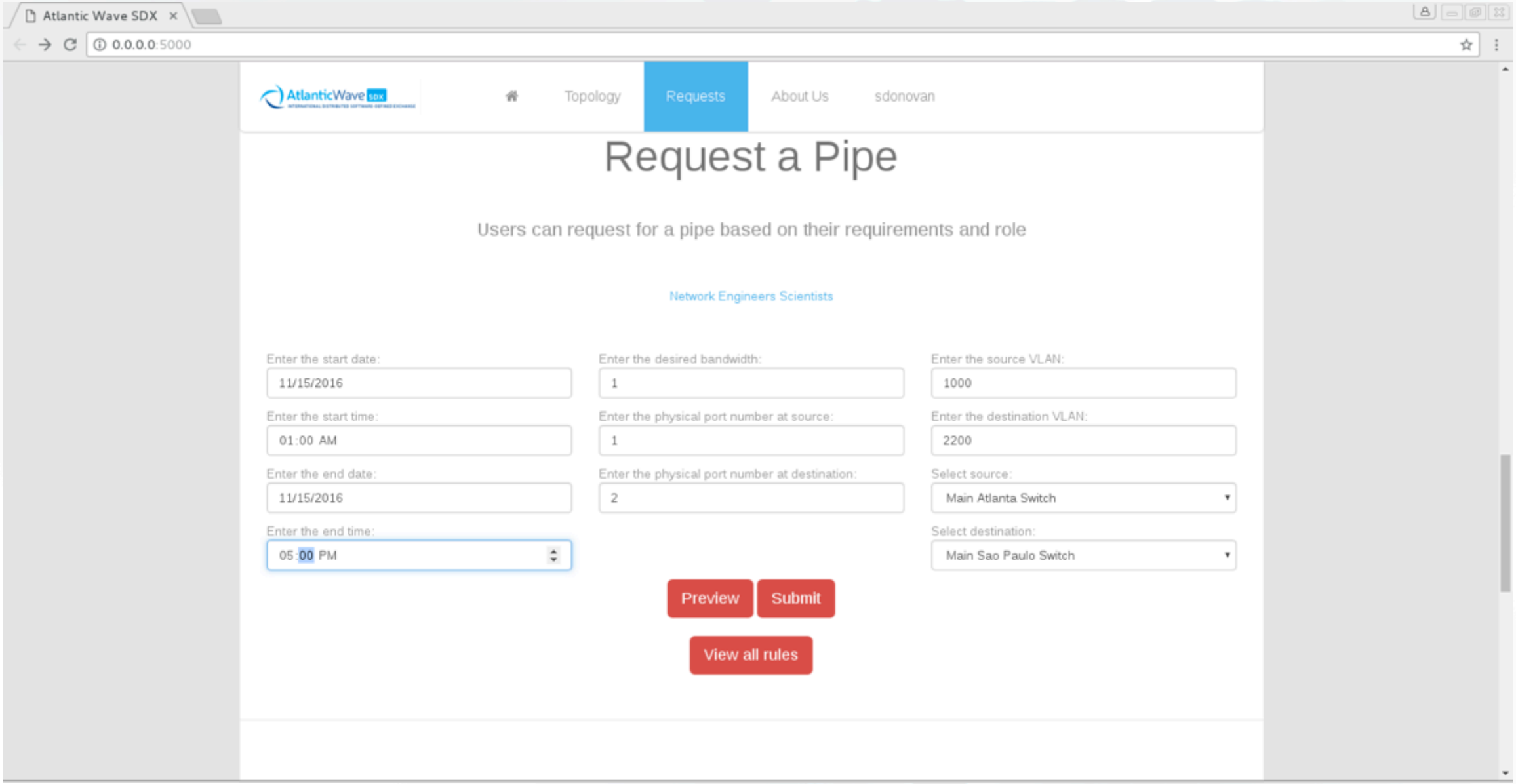
Functionality

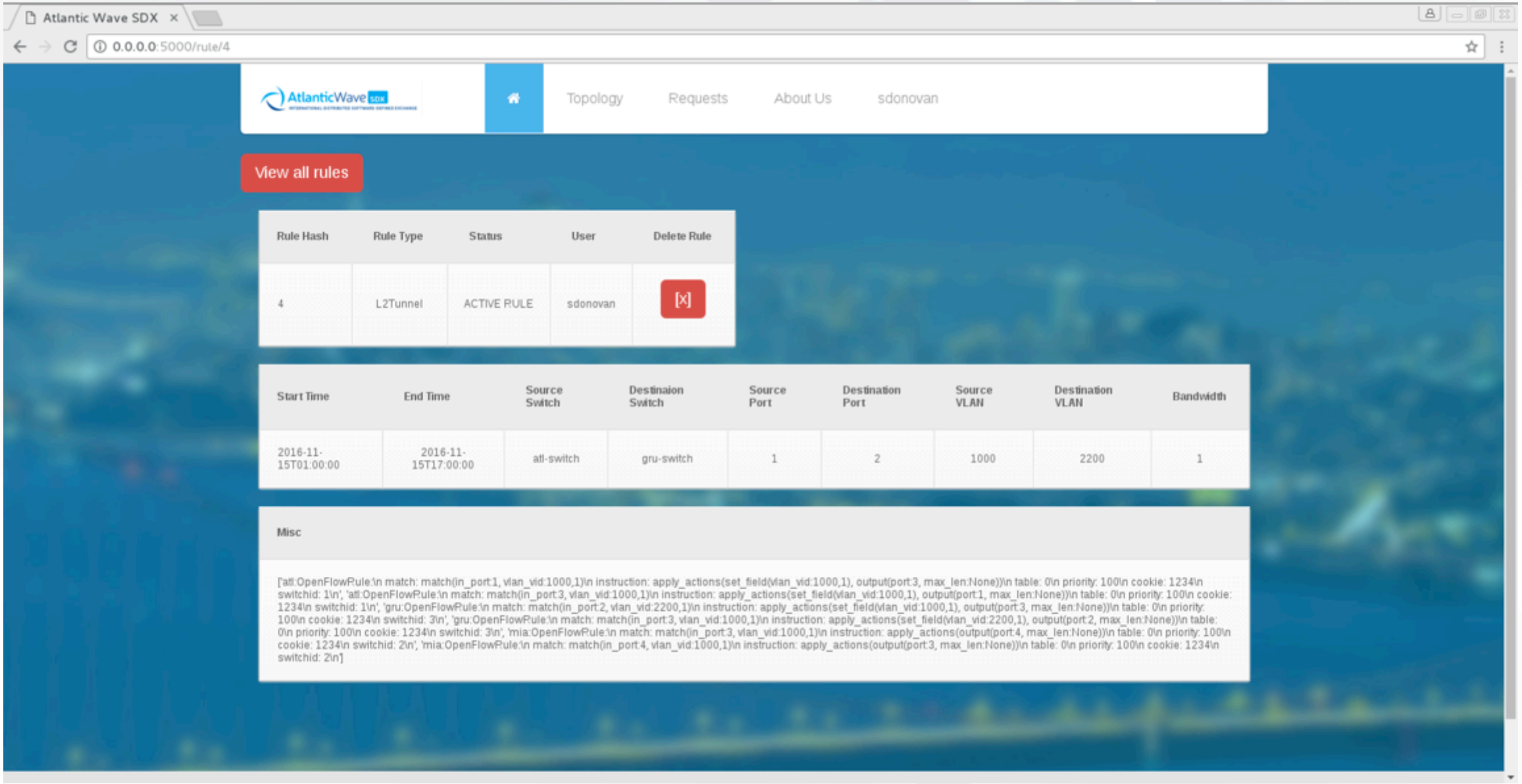
- L2 tunnels
  - Point-to-point
  - Multipoint
- SDX functionality
  - Network-header based rules
  - DNS-based rules
  - Other abstract rules
- In-Band Management
- Local Controller support for:
  - OpenFlow
  - P4
  - Others? Cisco or Juniper?
- NSI Integration for L2 Tunnels
  - Inbound
  - Outbound
- Limited BGP integration for SDX correctness
- User Interfaces
  - Easy to use Web interface
  - REST API
  - Example external controller using REST
- Shibboleth Integration

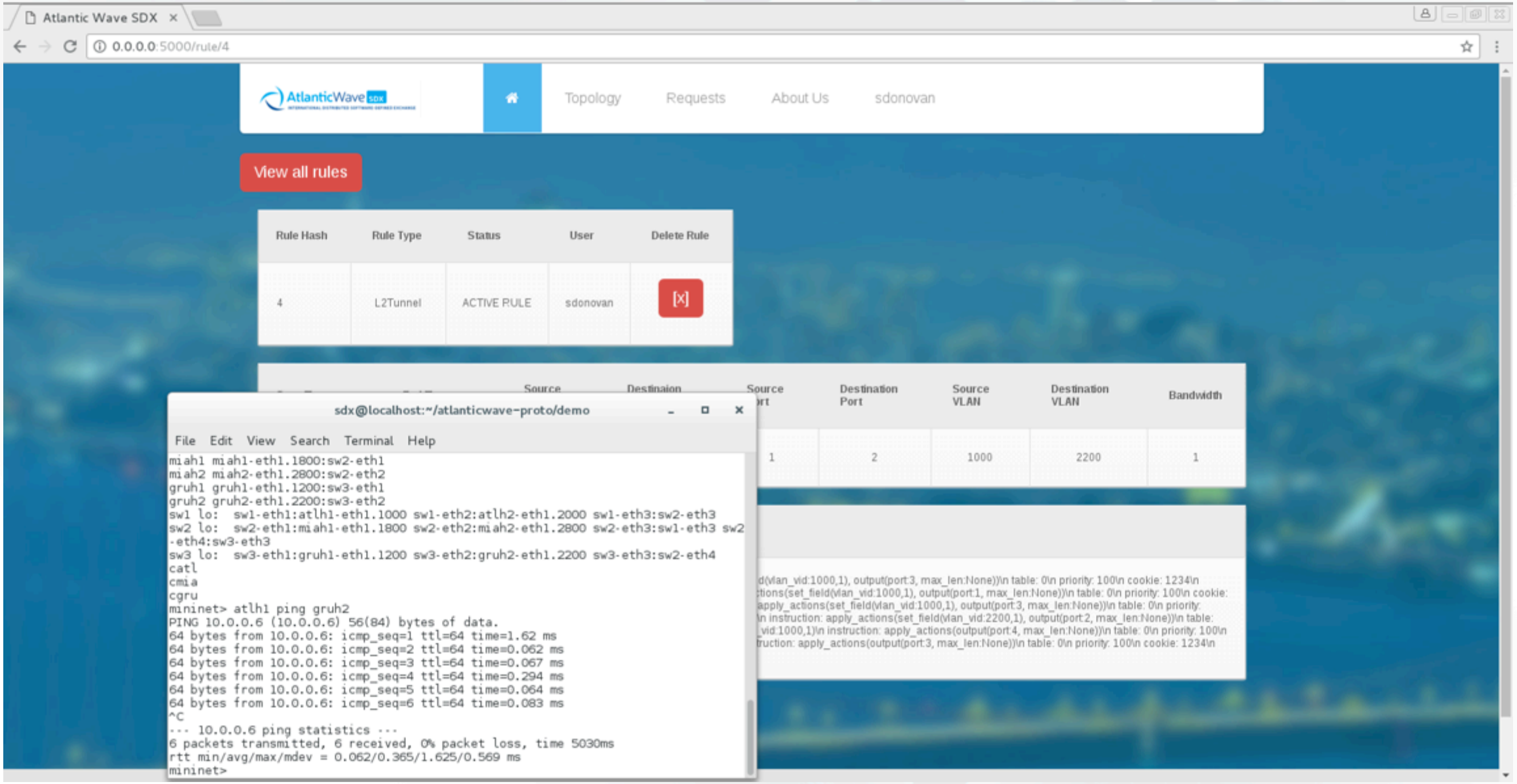
Demo 1



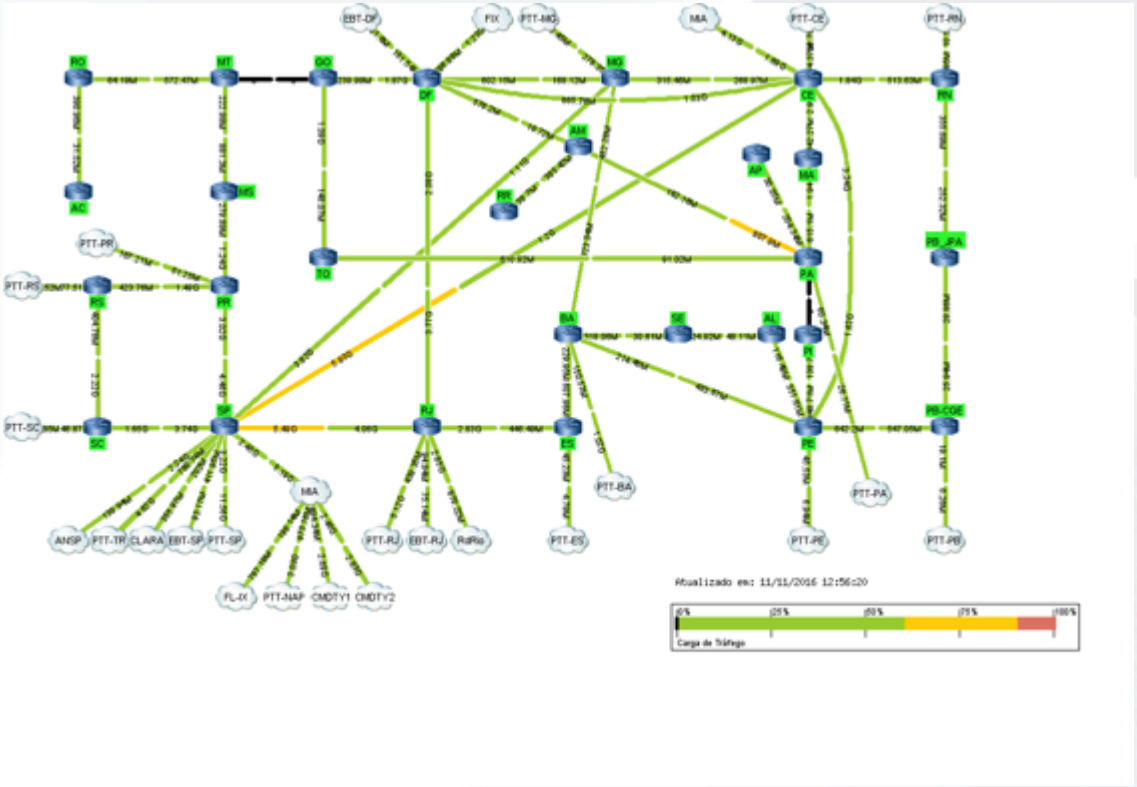
- Representation of Actual topology
- Simulated in Mininet
  - Still waiting on switch hardware
- Each host is on a different VLAN
- Showing real time connection establishment and removal


Atlantic Wave SDX
0.0.0.0:5000
AtlanticWave SDX
Topology
Requests
About Us
sdonovan
Request a Pipe
Users can request for a pipe based on their requirements and role
Network Engineers Scientists
Enter the start date:
11/15/2016
Enter the start time:
01:00 AM
Enter the end date:
11/15/2016
Enter the end time:
05:00 PM
Enter the desired bandwidth:
1
Enter the physical port number at source:
1
Enter the physical port number at destination:
2
Enter the source VLAN:
1000
Enter the destination VLAN:
2200
Select source:
Main Atlanta Switch
Select destination:
Main Sao Paulo Switch
Preview
Submit
View all rules

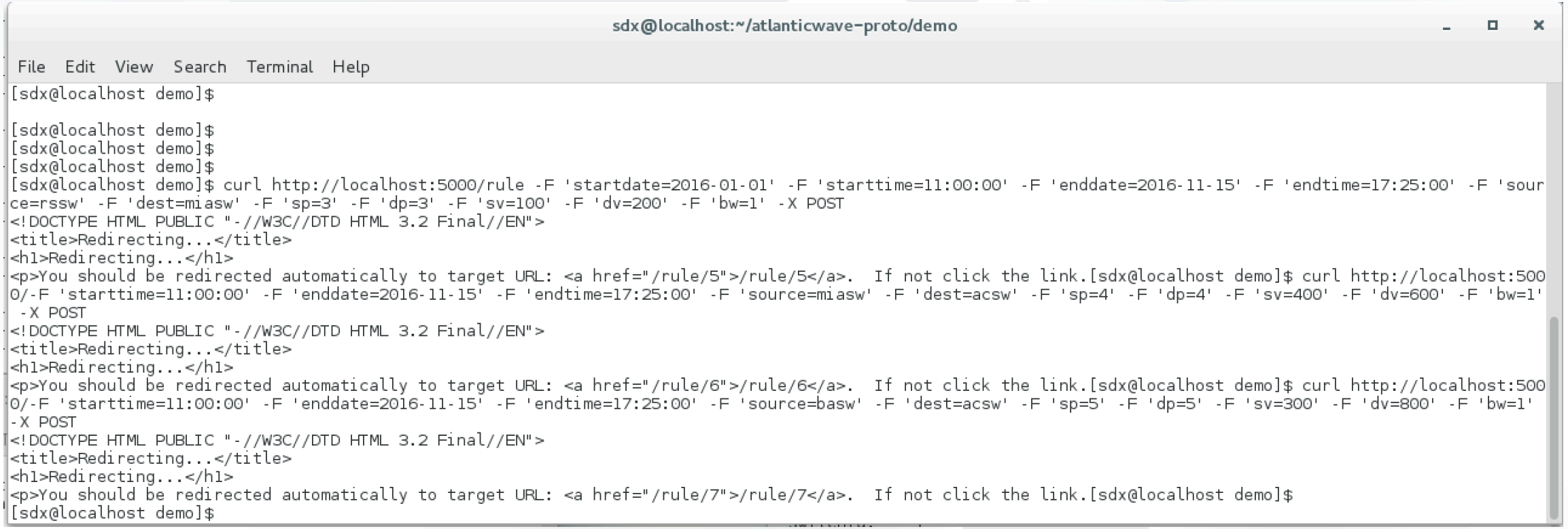

Atlantic Wave SDX
0.0.0.0:5000/rule/4
AtlanticWave SDX
Topology
Requests
About Us
sdonovan
View all rules
Rule Hash
Rule Type
Status
User
Delete Rule
4
L2Tunnel
ACTIVE RULE
sdonovan
[x]
Start Time
End Time
Source Switch
Destinaion Switch
Source Port
Destination Port
Source VLAN
Destination VLAN
Bandwidth
2016-11-15T01:00:00
2016-11-15T17:00:00
atl-switch
gru-switch
1
2
1000
2200
1
Misc
['atl:OpenFlowRule:\n match: match(in_port:1, vlan_vid:1000,1)\n instruction: apply_actions(set_field(vlan_vid:1000,1), output(port:3, max_len:None))\n table: 0\n priority: 100\n cookie: 1234\n switchid: 1\n', 'atl:OpenFlowRule:\n match: match(in_port:3, vlan_vid:1000,1)\n instruction: apply_actions(set_field(vlan_vid:1000,1), output(port:1, max_len:None))\n table: 0\n priority: 100\n cookie: 1234\n switchid: 1\n', 'gru:OpenFlowRule:\n match: match(in_port:2, vlan_vid:2200,1)\n instruction: apply_actions(set_field(vlan_vid:1000,1), output(port:3, max_len:None))\n table: 0\n priority: 100\n cookie: 1234\n switchid: 3\n', 'gru:OpenFlowRule:\n match: match(in_port:3, vlan_vid:1000,1)\n instruction: apply_actions(set_field(vlan_vid:2200,1), output(port:2, max_len:None))\n table: 0\n priority: 100\n cookie: 1234\n switchid: 3\n', 'mia:OpenFlowRule:\n match: match(in_port:3, vlan_vid:1000,1)\n instruction: apply_actions(output(port:4, max_len:None))\n table: 0\n priority: 100\n cookie: 1234\n switchid: 2\n', 'mia:OpenFlowRule:\n match: match(in_port:4, vlan_vid:1000,1)\n instruction: apply_actions(output(port:3, max_len:None))\n table: 0\n priority: 100\n cookie: 1234\n switchid: 2\n']

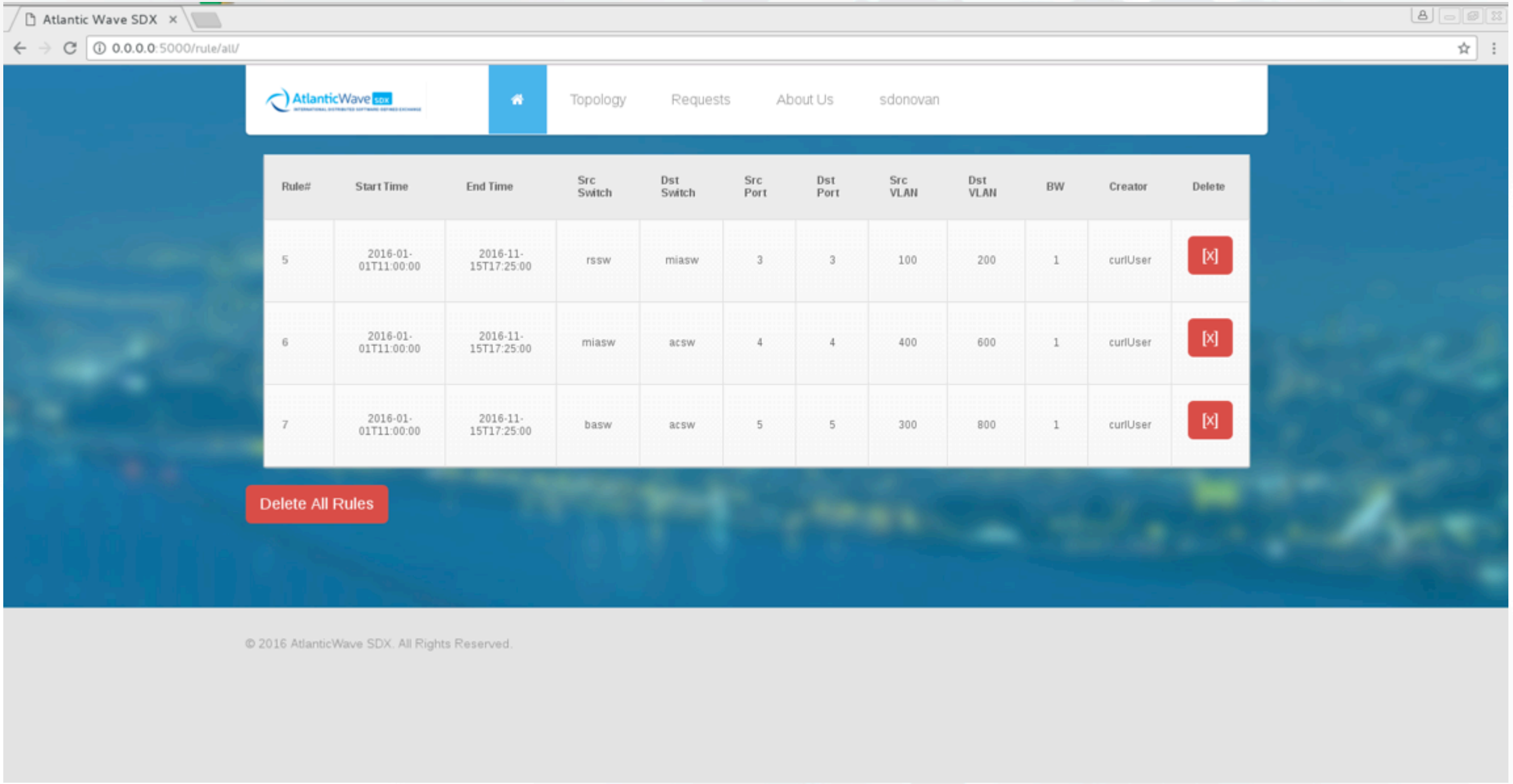

Atlantic Wave SDX
0.0.0.0:5000/rule/4
AtlanticWave SDX
Topology
Requests
About Us
sdonovan
View all rules
Rule Hash
Rule Type
Status
User
Delete Rule
4
L2Tunnel
ACTIVE RULE
sdonovan
[x]
Source Port
Destination Port
Source VLAN
Destination VLAN
Bandwidth
1
2
1000
2200
1
sdx@localhost:~/atlanticwave-proto/demo
File Edit View Search Terminal Help
miah1 miah1-eth1.1800:sw2-eth1
miah2 miah2-eth1.2800:sw2-eth2
gruh1 gruh1-eth1.1200:sw3-eth1
gruh2 gruh2-eth1.2200:sw3-eth2
sw1 lo: sw1-eth1:atlh1-eth1.1000 sw1-eth2:atlh2-eth1.2000 sw1-eth3:sw2-eth3
sw2 lo: sw2-eth1:miah1-eth1.1800 sw2-eth2:miah2-eth1.2800 sw2-eth3:sw1-eth3 sw2-eth4:sw3-eth3
sw3 lo: sw3-eth1:gruh1-eth1.1200 sw3-eth2:gruh2-eth1.2200 sw3-eth3:sw2-eth4
catl
cmia
cgru
mininet> atlh1 ping gruh2
PING 10.0.0.6 (10.0.0.6) 56(84) bytes of data.
64 bytes from 10.0.0.6: icmp_seq=1 ttl=64 time=1.62 ms
64 bytes from 10.0.0.6: icmp_seq=2 ttl=64 time=0.062 ms
64 bytes from 10.0.0.6: icmp_seq=3 ttl=64 time=0.067 ms
64 bytes from 10.0.0.6: icmp_seq=4 ttl=64 time=0.294 ms
64 bytes from 10.0.0.6: icmp_seq=5 ttl=64 time=0.064 ms
64 bytes from 10.0.0.6: icmp_seq=6 ttl=64 time=0.083 ms
^C
--- 10.0.0.6 ping statistics ---
6 packets transmitted, 6 received, 0% packet loss, time 5030ms
rtt min/avg/max/mdev = 0.062/0.365/1.625/0.569 ms
mininet>

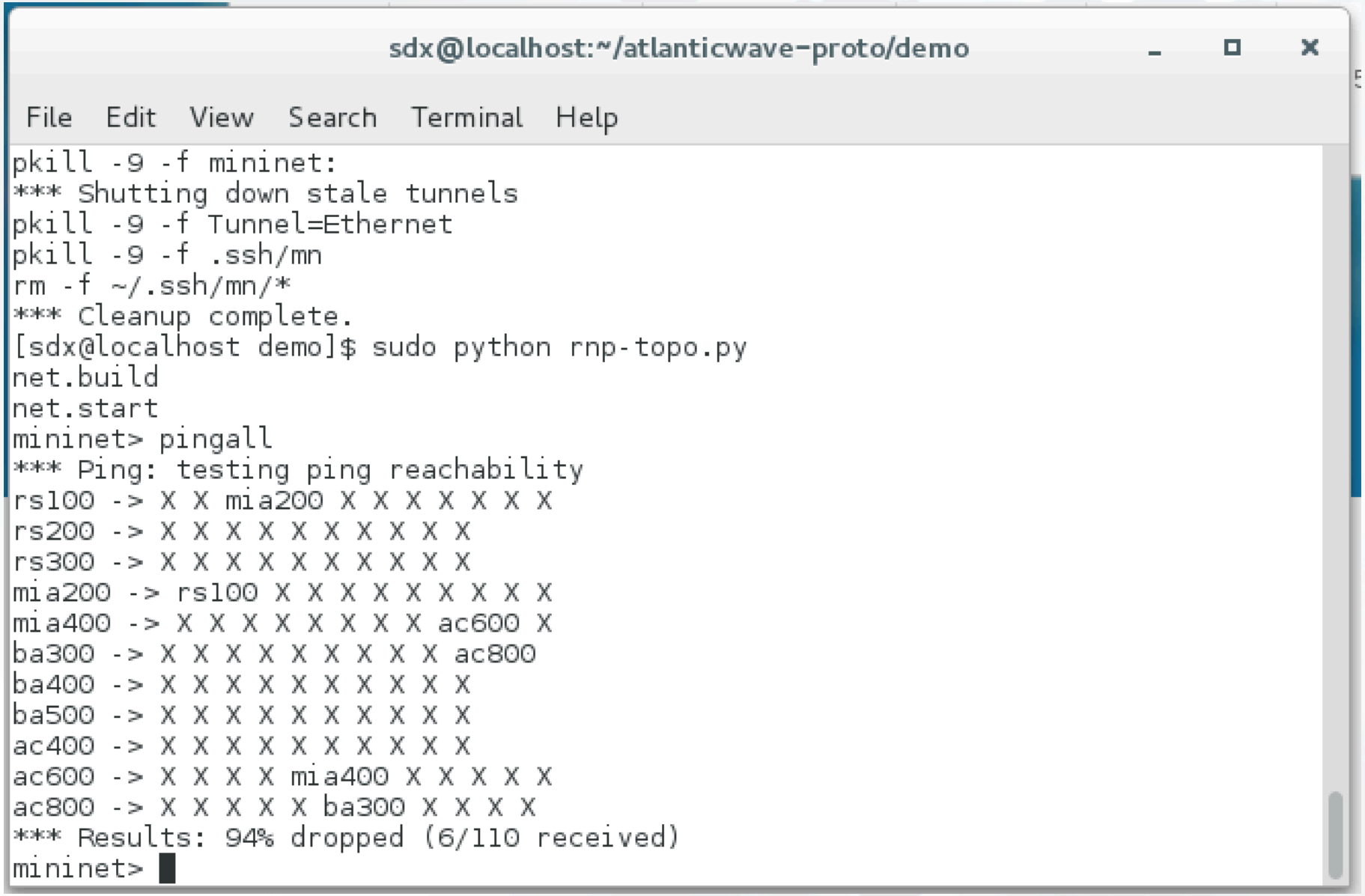
Demo 2

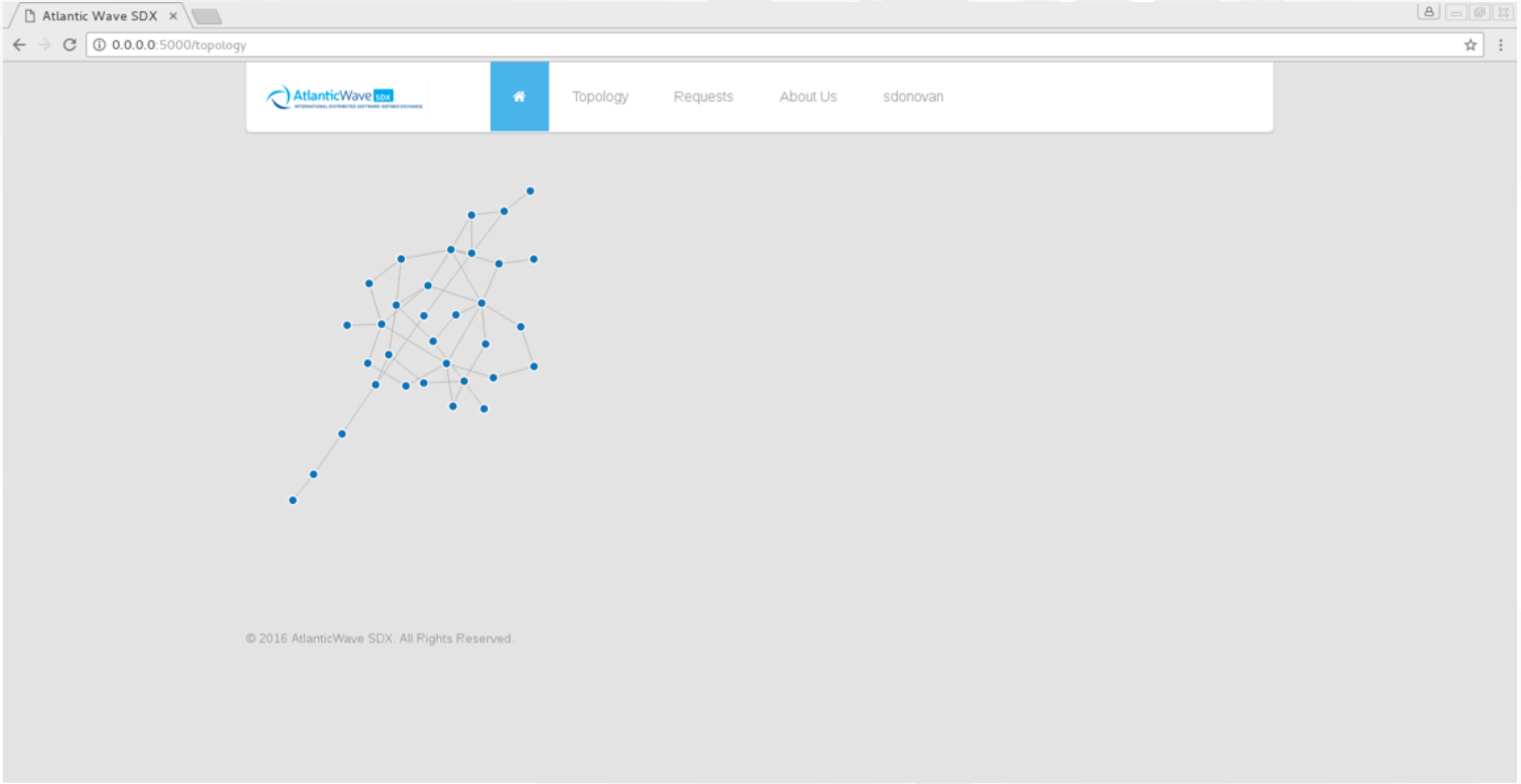


- RNP network topology in Brazil
- Simulated in Mininet
  - Slow, LC doesn't scale well in current form
- Adding a bunch of rules through REST


sdx@localhost:~/atlanticwave-proto/demo
File Edit View Search Terminal Help
[sdx@localhost demo]$
[sdx@localhost demo]$
[sdx@localhost demo]$
[sdx@localhost demo]$
[sdx@localhost demo]$ curl http://localhost:5000/rule -F 'startdate=2016-01-01' -F 'starttime=11:00:00' -F 'enddate=2016-11-15' -F 'endtime=17:25:00' -F 'source=rssw' -F 'dest=miasw' -F 'sp=3' -F 'dp=3' -F 'sv=100' -F 'dv=200' -F 'bw=1' -X POST
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<title>Redirecting...</title>
<h1>Redirecting...</h1>
<p>You should be redirected automatically to target URL: /rule/5. If not click the link.[sdx@localhost demo]$ curl http://localhost:5000/-F 'starttime=11:00:00' -F 'enddate=2016-11-15' -F 'endtime=17:25:00' -F 'source=miasw' -F 'dest=acsw' -F 'sp=4' -F 'dp=4' -F 'sv=400' -F 'dv=600' -F 'bw=1' -X POST
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<title>Redirecting...</title>
<h1>Redirecting...</h1>
<p>You should be redirected automatically to target URL: /rule/6. If not click the link.[sdx@localhost demo]$ curl http://localhost:5000/-F 'starttime=11:00:00' -F 'enddate=2016-11-15' -F 'endtime=17:25:00' -F 'source=basw' -F 'dest=acsw' -F 'sp=5' -F 'dp=5' -F 'sv=300' -F 'dv=800' -F 'bw=1' -X POST
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<title>Redirecting...</title>
<h1>Redirecting...</h1>
<p>You should be redirected automatically to target URL: /rule/7. If not click the link.[sdx@localhost demo]$
[sdx@localhost demo]$


Atlantic Wave SDX
0.0.0.0:5000/rule/all/
AtlanticWave SDX
Topology
Requests
About Us
sdonovan
Rule# Start Time End Time Src Switch Dst Switch Src Port Dst Port Src VLAN Dst VLAN BW Creator Delete
5 2016-01-01T11:00:00 2016-11-15T17:25:00 rssw miasw 3 3 100 200 1 curlUser [x]
6 2016-01-01T11:00:00 2016-11-15T17:25:00 miasw acsw 4 4 400 600 1 curlUser [x]
7 2016-01-01T11:00:00 2016-11-15T17:25:00 basw acsw 5 5 300 800 1 curlUser [x]
Delete All Rules
© 2016 AtlanticWave SDX. All Rights Reserved.


sdx@localhost:~/atlanticwave-proto/demo
File Edit View Search Terminal Help
pkill -9 -f mininet:
*** Shutting down stale tunnels
pkill -9 -f Tunnel=Ethernet
pkill -9 -f .ssh/mn
rm -f ~/.ssh/mn/*
*** Cleanup complete.
[sdx@localhost demo]$ sudo python rnp-topo.py
net.build
net.start
mininet> pingall
*** Ping: testing ping reachability
rs100 -> X X mia200 X X X X X X X
rs200 -> X X X X X X X X X X
rs300 -> X X X X X X X X X X
mia200 -> rs100 X X X X X X X X X
mia400 -> X X X X X X X X ac600 X
ba300 -> X X X X X X X X X ac800
ba400 -> X X X X X X X X X X
ba500 -> X X X X X X X X X X
ac400 -> X X X X X X X X X X
ac600 -> X X X X mia400 X X X X X
ac800 -> X X X X X ba300 X X X X
*** Results: 94% dropped (6/110 received)
mininet>


Atlantic Wave SDX
0.0.0.0:5000/topology
AtlanticWave SDX
Topology
Requests
About Us
sdonovan
© 2016 AtlanticWave SDX. All Rights Reserved.

Thanks!

http://www.atlanticwave-sdx.net/

https://github.com/atlanticwave-sdx/atlanticwave-proto

Sean Donovan

sdonovan@gatech.edu

Russ Clark

russ.clark@gatech.edu

Jeronimo Bezerra

jbezerra@fiu.edu

Challenges

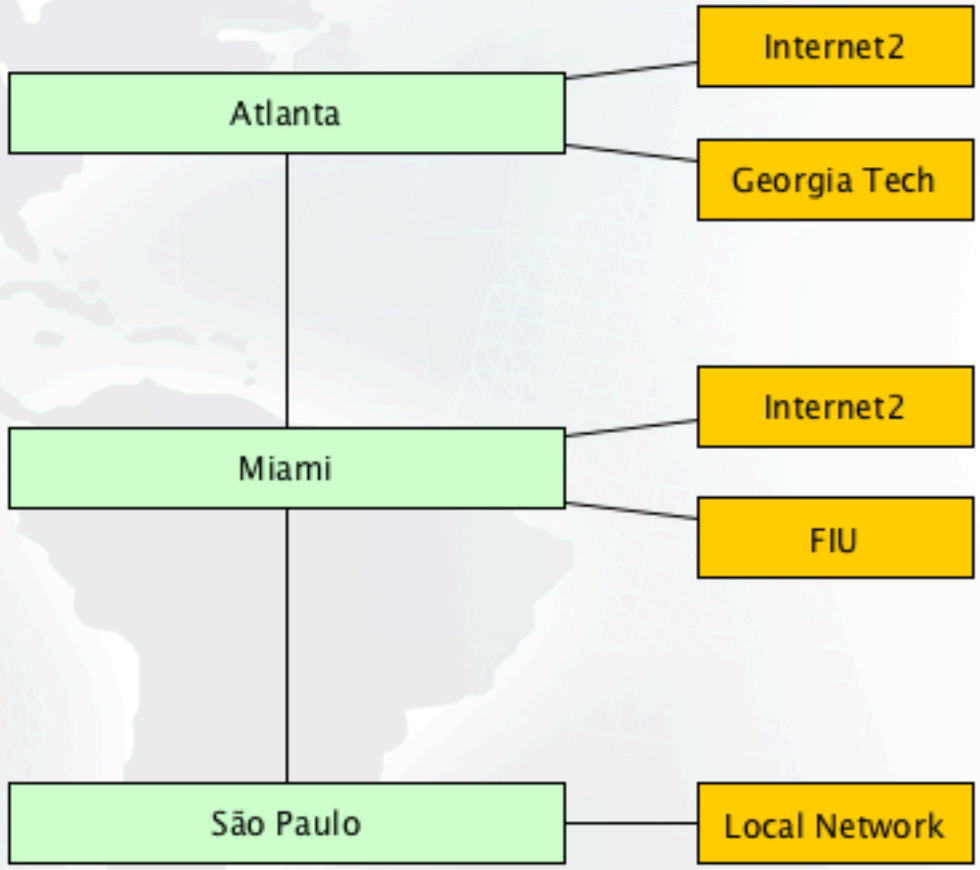
- Like any system, it's complicated
  - But there are some rather unique challenges
- Some solved, but lots of open questions
  - We'd like operator and user help with some of these challenges
- What would *you* want?
  - Network operators, domain scientists

Hardware

- We have some specific requirements
  - Multiple Table support
    - To reduce rule sizes dramatically
    - Cross Multiplication problem
  - 100Gbps
    - Based on the data rates that we expect
  - Support for most, if not all of OpenFlow 1.3
    - Features in OpenFlow 1.3 that are useful
    - OF Groups, for instance

100Gbps OpenFlow Equipment is Hard to Find

- Only a few manufacturers have OF 100Gbps gear and big interface buffers
- A lot only have 1 or 2 ports, need 3 or 4, depending on location



OpenFlow 1.3 (non) Support

- Many vendors claim 1.3 support
  - Often single table
  - Only rules X and Y, but not Z
  - Limited number of rules
    - TCAM limitations
- Study about support being overblown
  - Di Lallo et al., IEEE/IFIP NOMS 2016

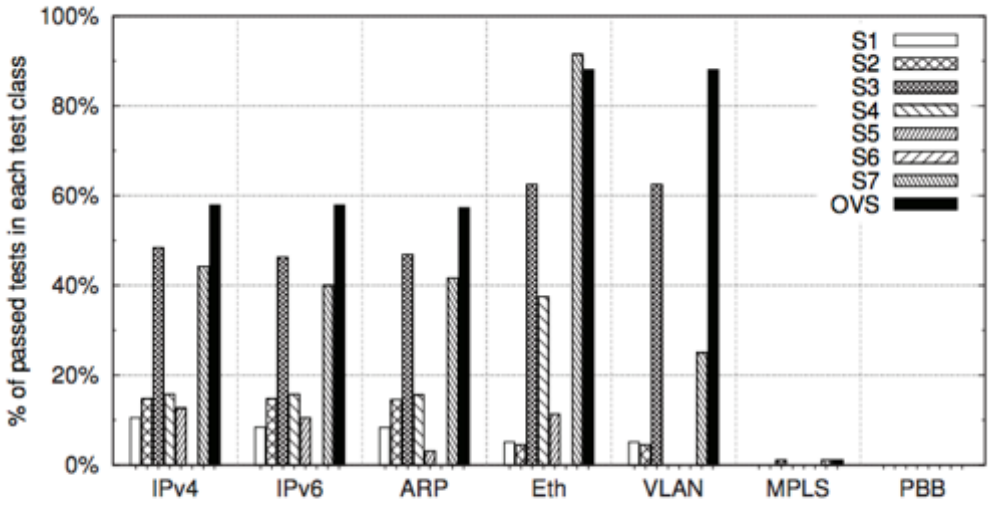


Fig. 5. Percentage of passed OF 1.3 Ryu tests for packets carrying specific protocols.

100Gbps + OpenFlow 1.3 + Multiple Tables

- Rather hard to find!
- Equipment's now trickling out

http://noviflow.com/products/noviswitch/
http://www8.hp.com/us/en/products/networking-switches/product-detail.html?oid=4177453
http://www.corsa.com/products/dp6440/
http://www.brocade.com/en/backend-content/pdf-page.html?/content/dam/common/documents/content-types/datasheet/brocade-mlx-2x100gbe-cfp2-ds.pdf

Abstractions

- What functionality do people need?
  - Point-to-point paths?
  - Point-to-multipoint?
  - Arbitrary routing?
- What should the API look like?
  - REST good enough?
  - Web-based interface?
- Who should it be tailored to?
  - Network admins?
  - Domain scientists?
  - General users?

APIs for Different Audiences

- Administrators

```
{"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}}
```

- Domain scientists

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

What Functionality Would be Useful?

- NSI-like interface planned
  - Partially working now
    - Bandwidth restriction is not implemented.
  - With inter-network NSI integration in the future
- SDX rules based on DNS
  - Based on NetAssay
  - match(domain='example.com')
- Any suggestions?
  - SDX-based rules *and* rules outside of SDX functionality

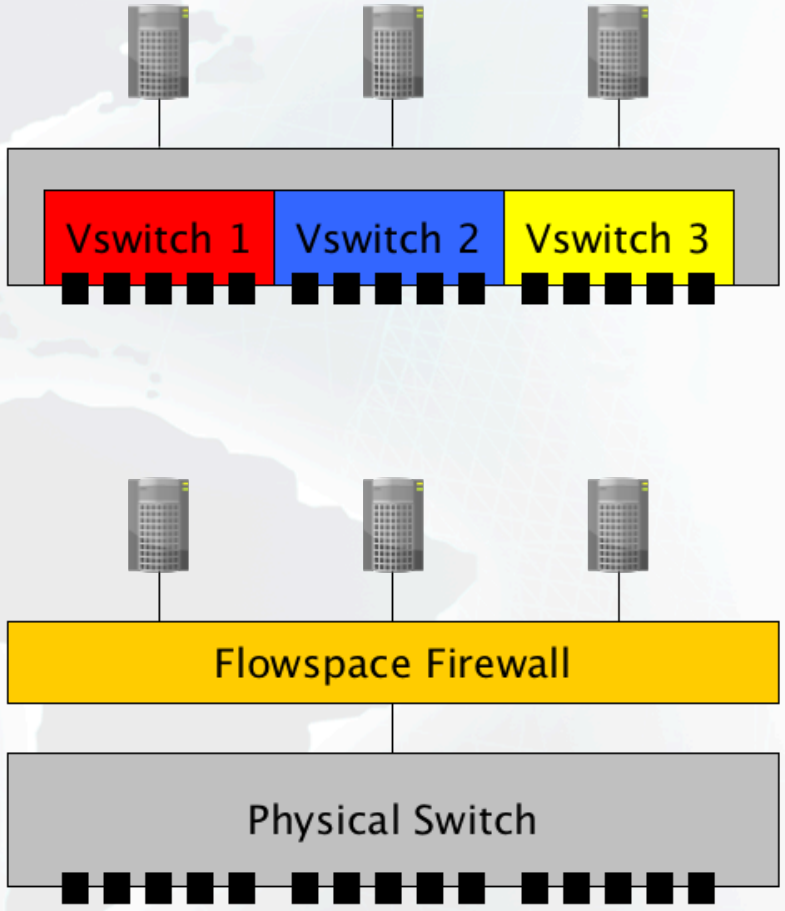
Do Administrators Care about Functionality Beyond BGP?

- Application-based peering
  - YouTube through Level3
  - Netflix through Cogent
  - Everything else through AT&T
  - Impossible with BGP
- Shared services at the SDX
  - Shared IDS for small businesses connection to the SDX
  - Web caching at the SDX

- Would administrators be interested in this type of functionality?

Federation

- Multiple Controllers with a Single Switch
  - Hardware virtualization
    - Per port, typically
    - New switches allow for per VLAN
  - Software Hypervisor
    - Use something like FlowSpace Firewall
      - Below the LC, for AtlanticWave/ SDX
      - FSF does *not* support OF1.3



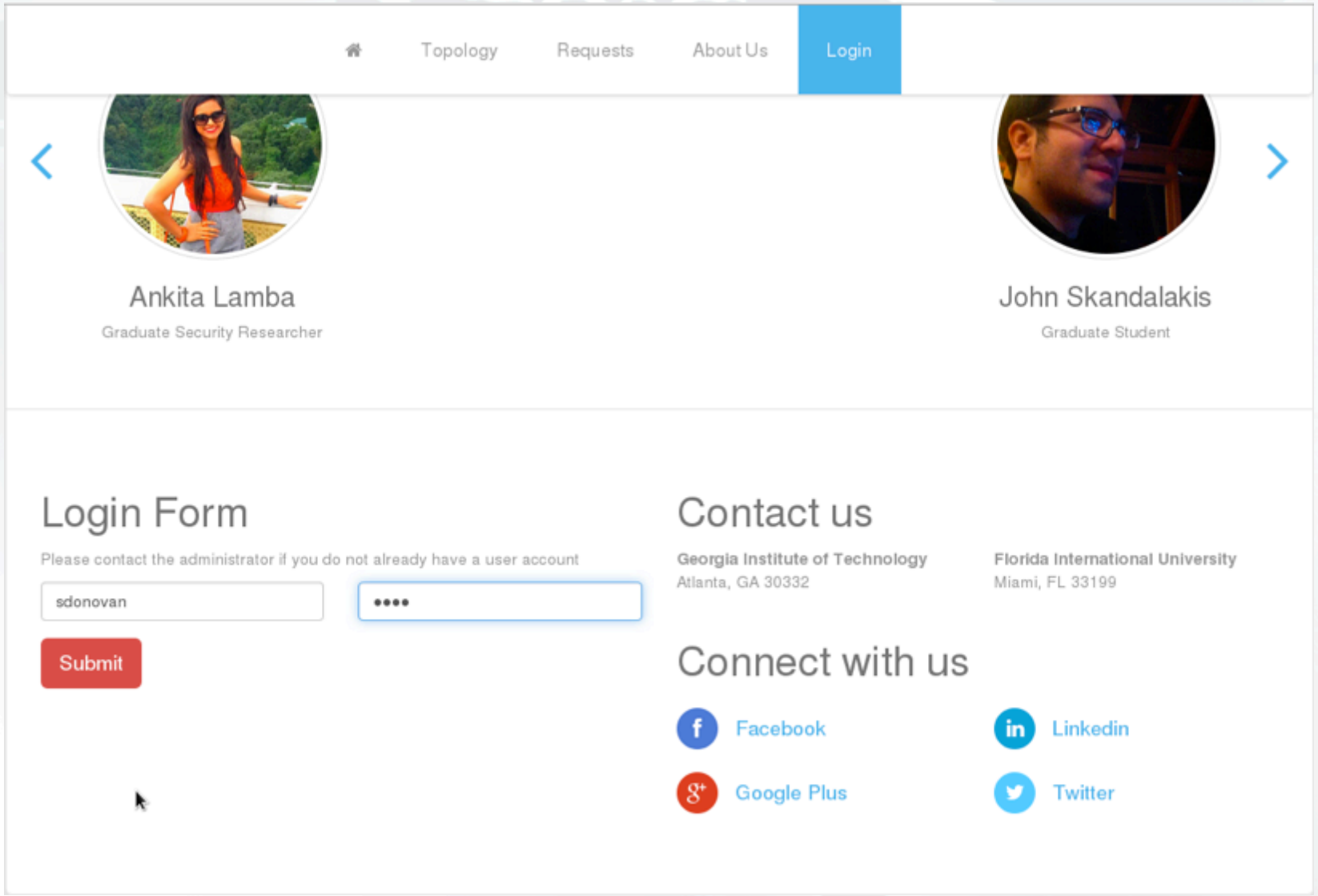
Federation

- Integrating other Networks
  - Integration with NSI
    - There are a number of NSI speakers that could be used to integrate with AtlanticWave/SDX
  - Shibboleth connectivity
    - Users will be academics, primarily
    - MS student actively working on this

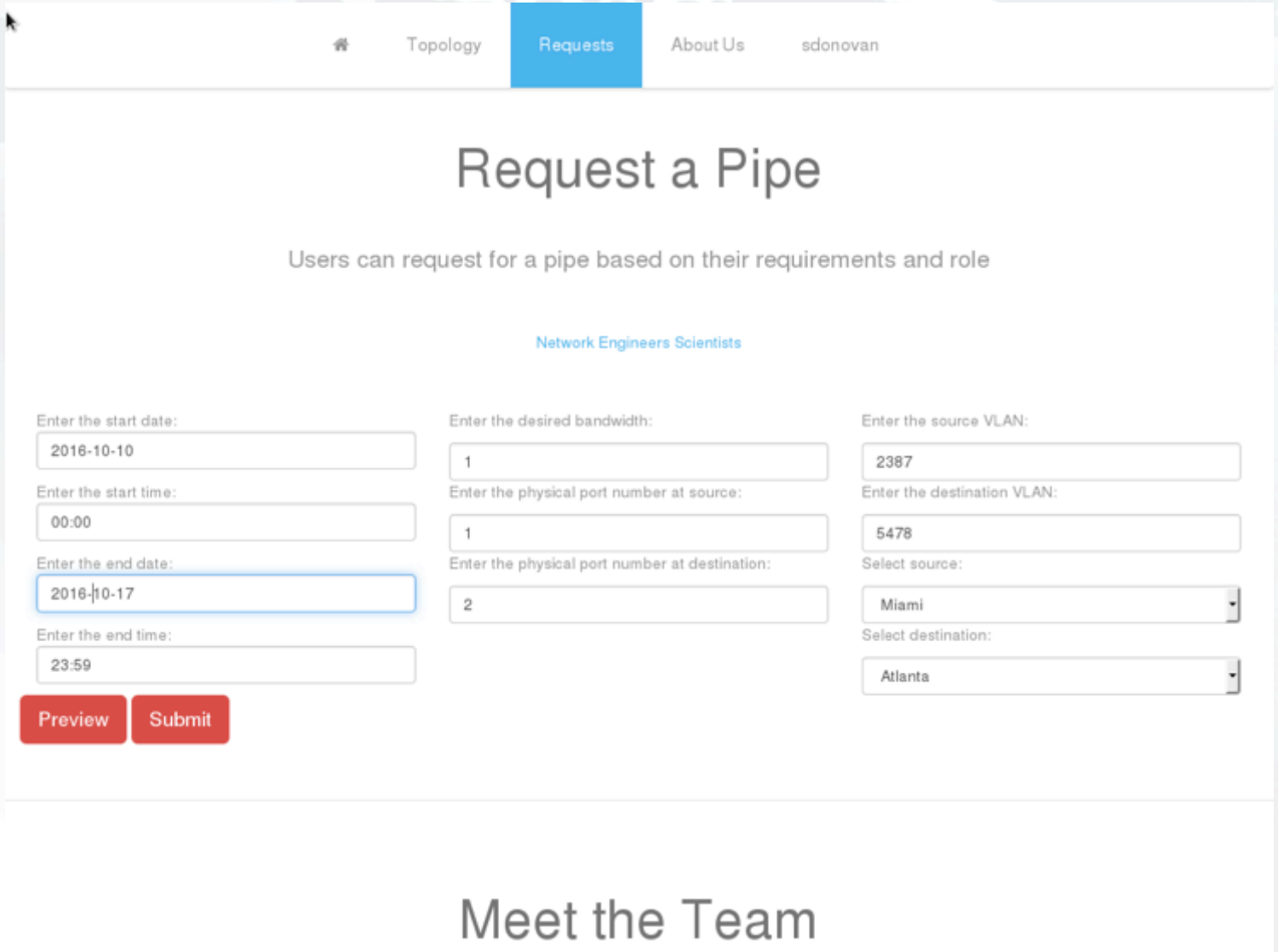
Current Status

- Focusing on NSI-like functionality right now
  - Default IXP behavior will follow
- Initial version of the controller is built
  - Has limitations, but being continuously developed
- Prototype Web Interface
  - Limited to adding rules
- Configuration files for static configurations
  - Users and topology are static at startup

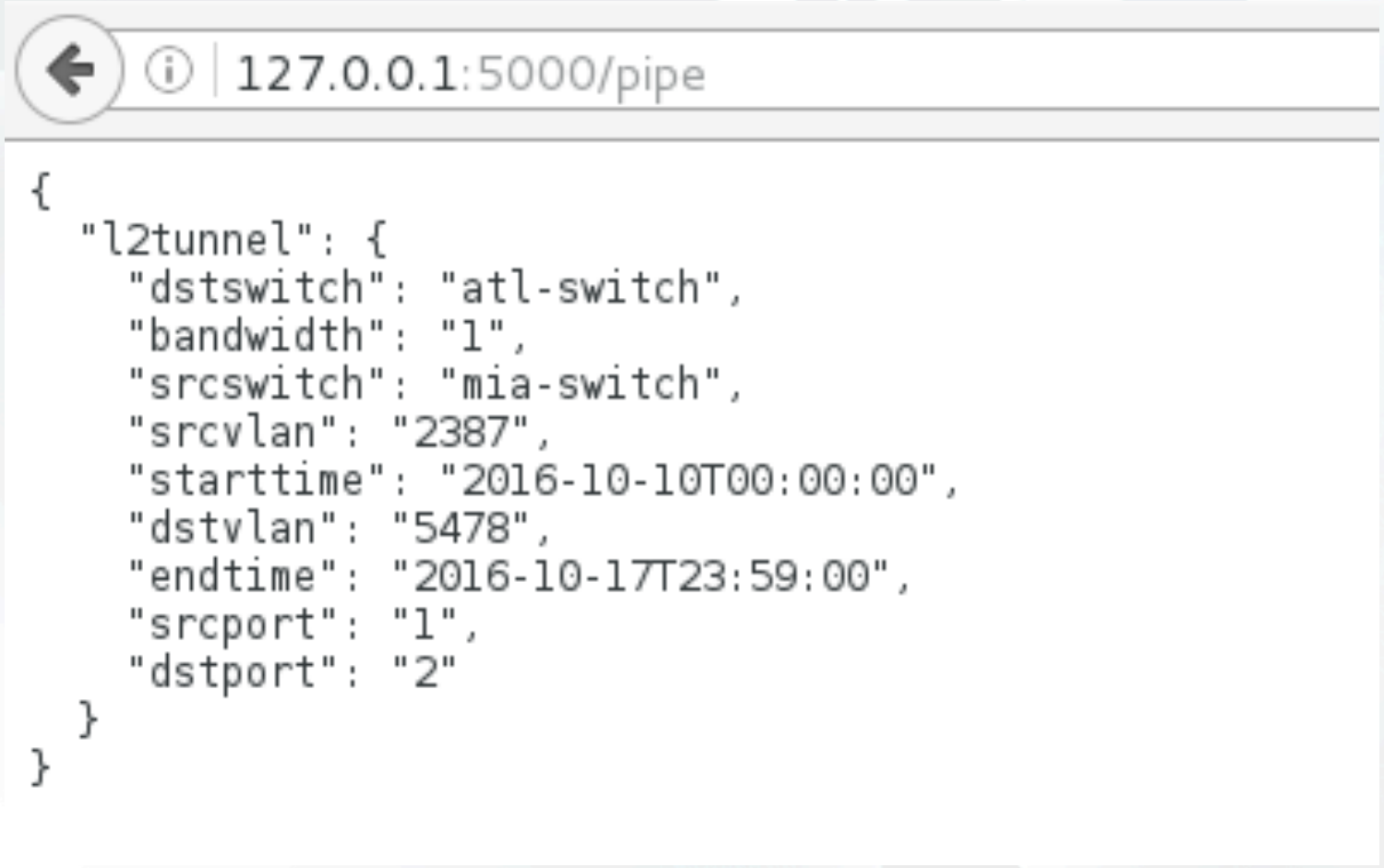
Web Interface



Web Interface



Web Interface



Timeline

- Public Github
  - https://github.com/atlanticwave-sdx/atlanticwave-proto
- October for NSI/AL2S-like functionality completed
  - Missing bandwidth reservation as of today
- Early November for DTN-to-DTN for domain scientists
- November for running on hardware switches
- December for initial SDX functionality

References

- Stringer, Jonathan Philip, et al. "Cardigan: Deploying a distributed routing fabric." *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*. ACM, 2013.
- Stringer, Jonathan, et al. "Cardigan: SDN distributed routing fabric going live at an Internet exchange." *2014 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2014.
- Gupta, Arpit, et al. "SDX: a software defined internet exchange." *ACM SIGCOMM Computer Communication Review* 44.4 (2015): 551-562.
- Gupta, Arpit, et al. "An industrial-scale software defined internet exchange point." *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*. 2016.
- Chung, Joaquin, Henry Owen, and Russell Clark. "SDX architectures: A qualitative analysis." *SoutheastCon 2016*. IEEE, 2016.
- di Lallo, Roberto, et al. "On the practical applicability of SDN research." *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2016.

Current SDX Deployments

- Cardigan – Wellington Internet Exchange and REANNZ
  - Very, very early implementation
  - In early 2014, was deployed for 9 months with only 1134 flows
  - Rather traditional IXP
- Maryland/WIX
  - Controller lives “above” Oscars
  - Adding compute to the mix
- PacificWave-SDX
  - This is the most like AtlanticWave/SDX, distributed on the west coast of the US
  - Also a distributed exchange between Seattle, Sunnyvale, CA, and Los Angeles, CA
  - SDX in parallel with their traditional fabric

Current Examples of SDX Research

- Gupta et al., SIGCOMM 2014 – Initial work, where our definition comes from
- Gupta et al., NSDI 2016 – Optimization work, to allow for scalability
- GENI SDX – Early work at deploying an SDX using GENI project infrastructure, still ongoing
- Work at Starlight – Working on evaluating various SDX design
- SDX taxonomy in Chung et al., SoutheastCon 2016.

Cross Multiplication

	A-in	B-in	C-in
A-out			
B-out			
C-out			

Cross Multiplication

	A-in	B-in	C-in
A-out	A-in*A-out	B-in*A-out	C-in*A-out
B-out	A-in*B-out	B-in*B-out	C-in*B-out
C-out	A-in*C-out	B-in*C-out	C-in*C-out

- $O(N^2)$ sets of rules
- Some optimizations are possible
  - The diagonal can be eliminated
  - Gupta, et. al., 2014 discusses other optimizations

Cross Multiplication

	A-in	B-in	C-in
A-out		B-in*A-out	C-in*A-out
B-out	A-in*B-out		C-in*B-out
C-out	A-in*C-out	B-in*C-out	

- $O(N^2)$ sets of rules
- Some optimizations are possible
  - The diagonal can be eliminated
  - Gupta, et. al., 2014 discusses other optimizations

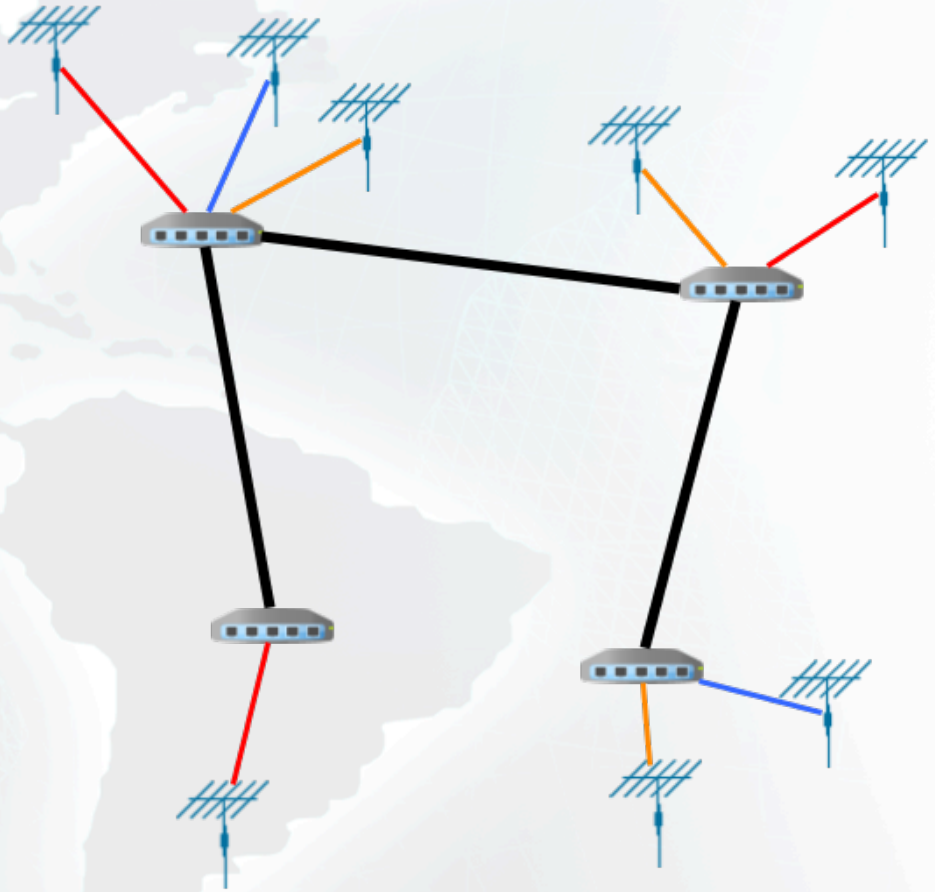
Multiple tables are better

Table 1	Table 2
A-out	A-in
B-out	B-in
C-out	C-in

- With multiple tables, we can pipeline the outbound and inbound rules
- O(2N) sets of rules
  - Much better than $O(N^2)$
- Think of a dozen participants:
  - ~144 sets of rules vs ~24 sets
- Much simpler to implement

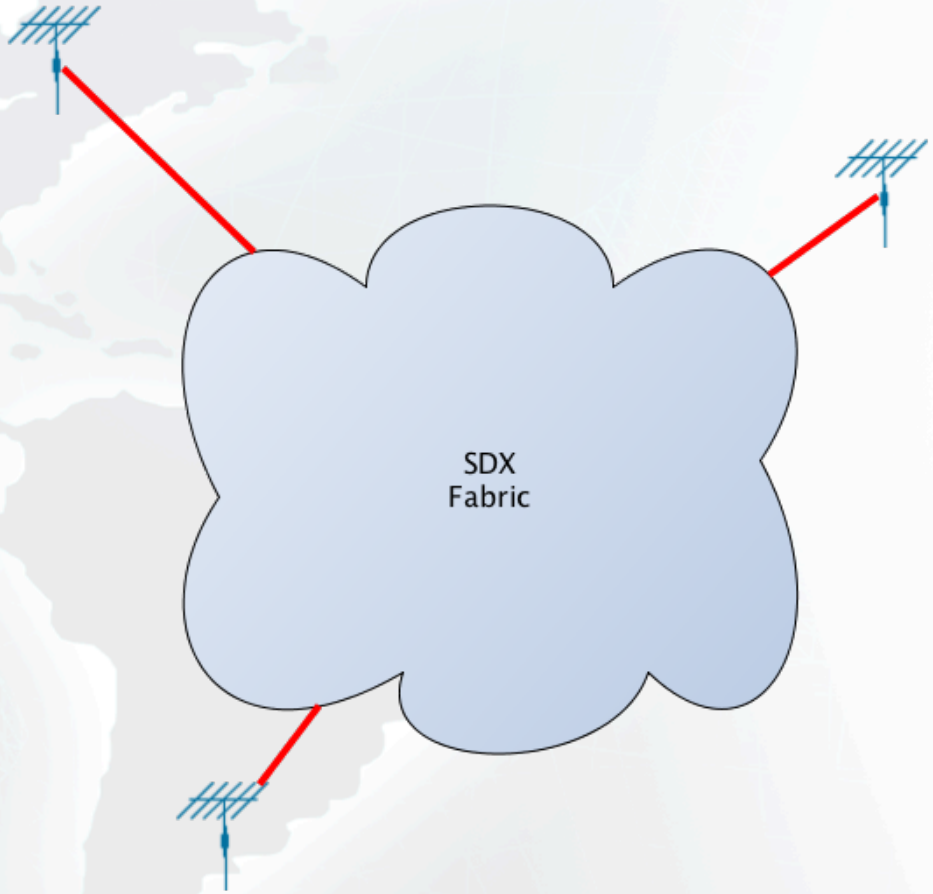
Deployment Outside of AtlanticWave/SDX

- Example deployment
  - In a city with a distributed SDX, like AMS-IX
  - Mobile phone backbone for multiple carriers
- Does this change what sorts of abstractions someone would want?



Deployment Outside of AtlanticWave/SDX

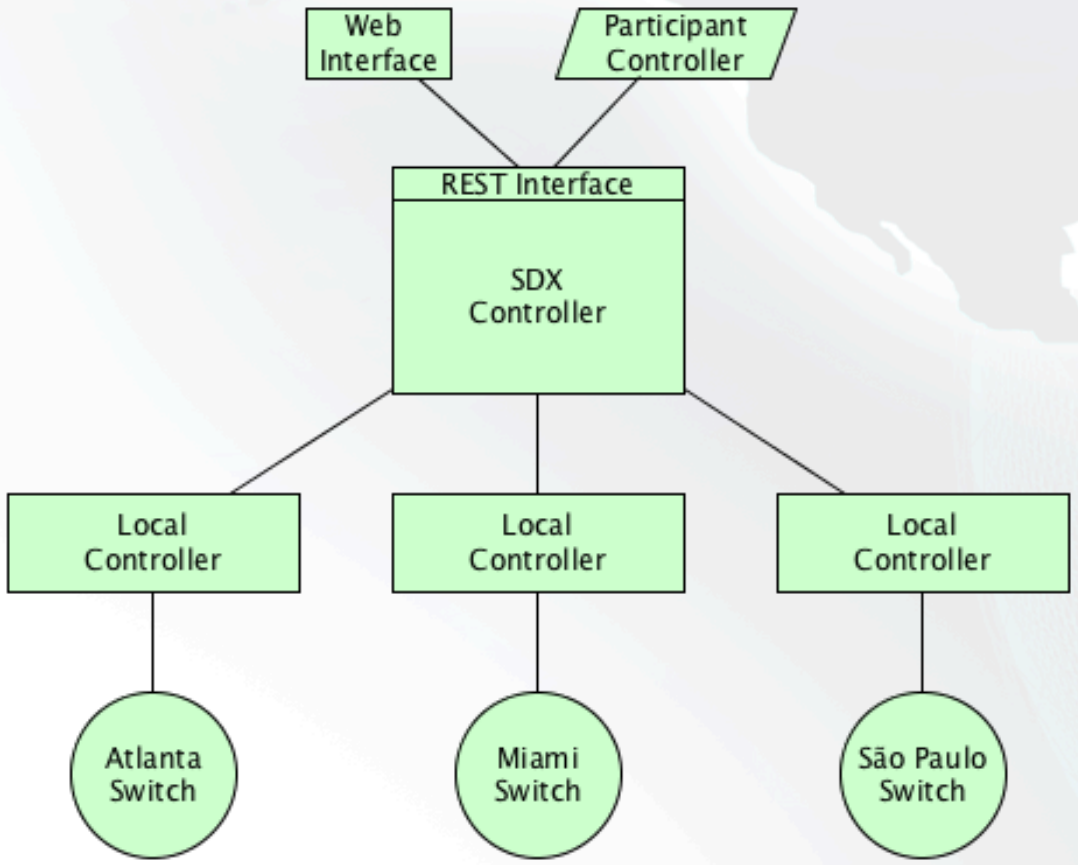
- Example deployment
  - In a city with a distributed SDX, like AMS-IX
  - Mobile phone backbone for multiple carriers
- Does this change what sorts of abstractions someone would want?



Security

- SDN and Security isn't discussed nearly enough
  - Most academic work glosses over security aspects of what they developed
  - New attacks are possible due to the design change over traditional networking
- This is being deployed
  - So we care a lot about security

Security Issues in AtlanticWave/SDX Design



- Information leakage
  - Rules/data leaking to unauthorized users
- DoS attacks
  - REST API is susceptible
  - In-band SDX-to-LC should mitigate
- Policy overlap
  - New user policies must not violate other user's policies

Authentication

- User authentication
  - TLS certificate authentication
  - Would an SSH tunnel with a certificate be enough?
- Local controller and SDX controller
  - Prevent unauthorized rules coming from a fake SDX controller
  - Prevent snooping from a fake local controller
  - Bi-directional TLS authentication with certificates

Authorization

	Admins	Domain Scientists	Data Agent	Research Assistant
GT				
FIU				
NCSA				
UofA				

- What's the correct level of granularity in authorization?
  - Roles
  - Organizations
- What Actions should be authorized?
  - At what granularity should actions be authorized?
  - Positive or negative authorization?
- Future project
  - MS Student

Actions requiring authorization

- Installing rules
  - Per port
  - Per switch
  - Types of rules
- Removing rules
  - Own rules
  - Same org. rules
- Get Statistics
  - To authorize automated collection methods
- View Rules
  - Per user
  - Per organization
  - Per switch

Management

- Failover
  - Distance = Latency
  - Latency = Problems
  - AtlanticWave/SDX is not a physically small network
  - Should there be more autonomy at the LC for failover?

	Atlanta	Miami	São Paulo
Atlanta	-	13ms	119ms
Miami	81 MB	-	106ms
São Paulo	743 MB	662 MB	-

https://wondernetwork.com/pings, FIU/AmLight

Sustainability

- Currently supported by NSF Grant #ACI-1341024 2015-2020
- How to make this self sufficient/sustainable?
- What's a good business model?
- Other research networks are facing the same question (e.g., GENI)