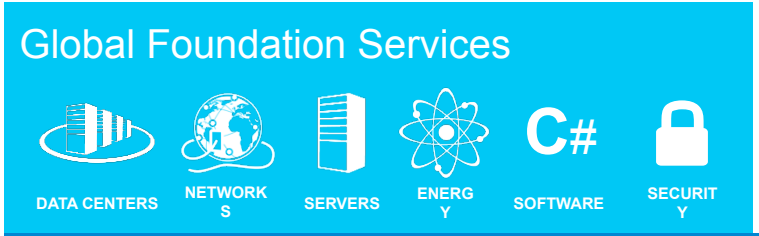


Microsoft

Large-Scale Passive Monitoring using SDN

Mohan Nanduri
mnanduri@microsoft.com

Justin Scott
juscott@microsoft.com

Onur Karaagaoglu
onurka@microsoft.com

What is this about?

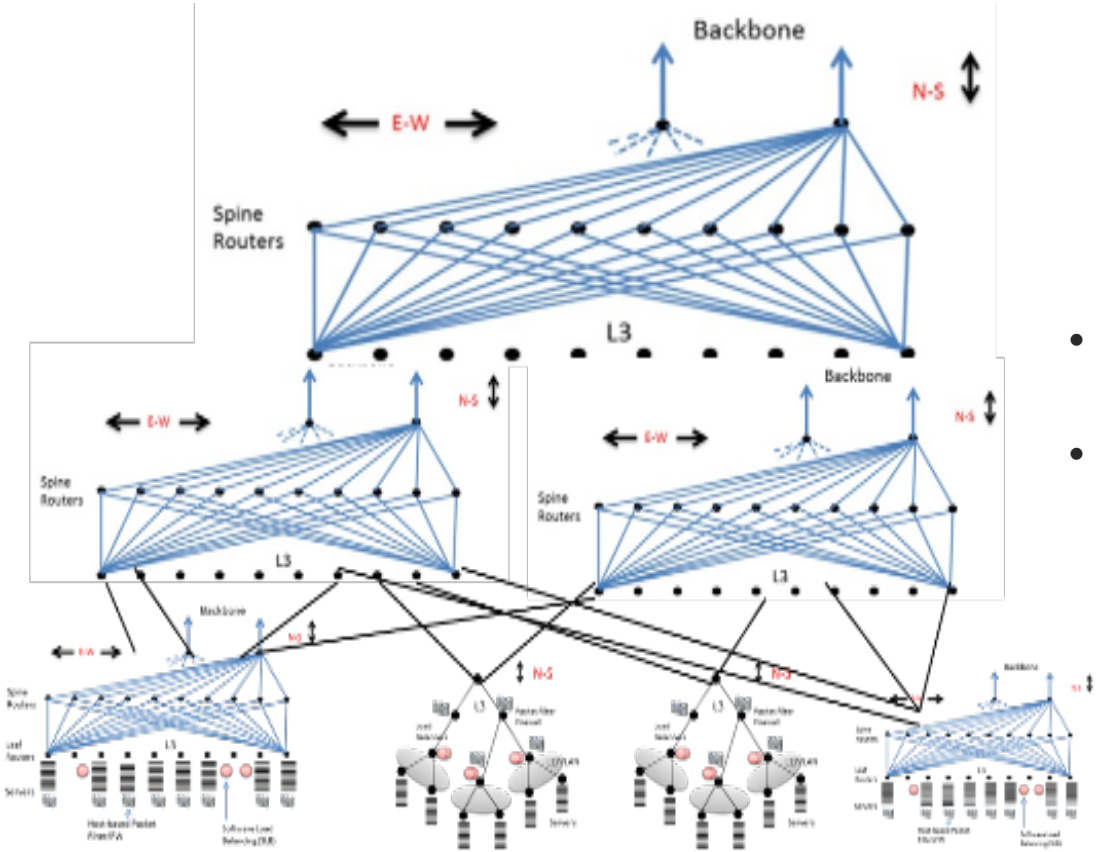
A different way to aggregate data from an optical TAP/SPAN

Architecture based on Openflow and commodity merchant silicon

Why?

Exonerate network quickly, reduce MTTR and MTTM
Lack of ability to data mine telemetry data at the network layer
Reduce operations engagement to troubleshoot service
Multi-tenant offering
Solution that’s on-demand and always available
Provides large-scale tap aggregation leveraging commodity hardware

Hyper Scale



- Thousands of 10G links per Data Center
- Cost makes it a non-starter with commercial solutions

Prior Attempts

Capture-Net

- Off the shelf aggregation gear, was too expensive at scale
- Resulted in lots of gear gathering dust
- Operations not mature enough to back such a solution

PMA/PUMA –”Passive Measurement Architecture”

- Lower cost than Capture-net
- Designed for a specific environment and not intended to scale
- Extremely feature rich

Stuck with shuffling sniffers around

NOT

THE

END

Just took a step back

What Features Make Up a Packet Broker?

Terminates taps
Match on a 5-tuple
Duplication
Packets unaltered
Low Latency
Statistics

80%

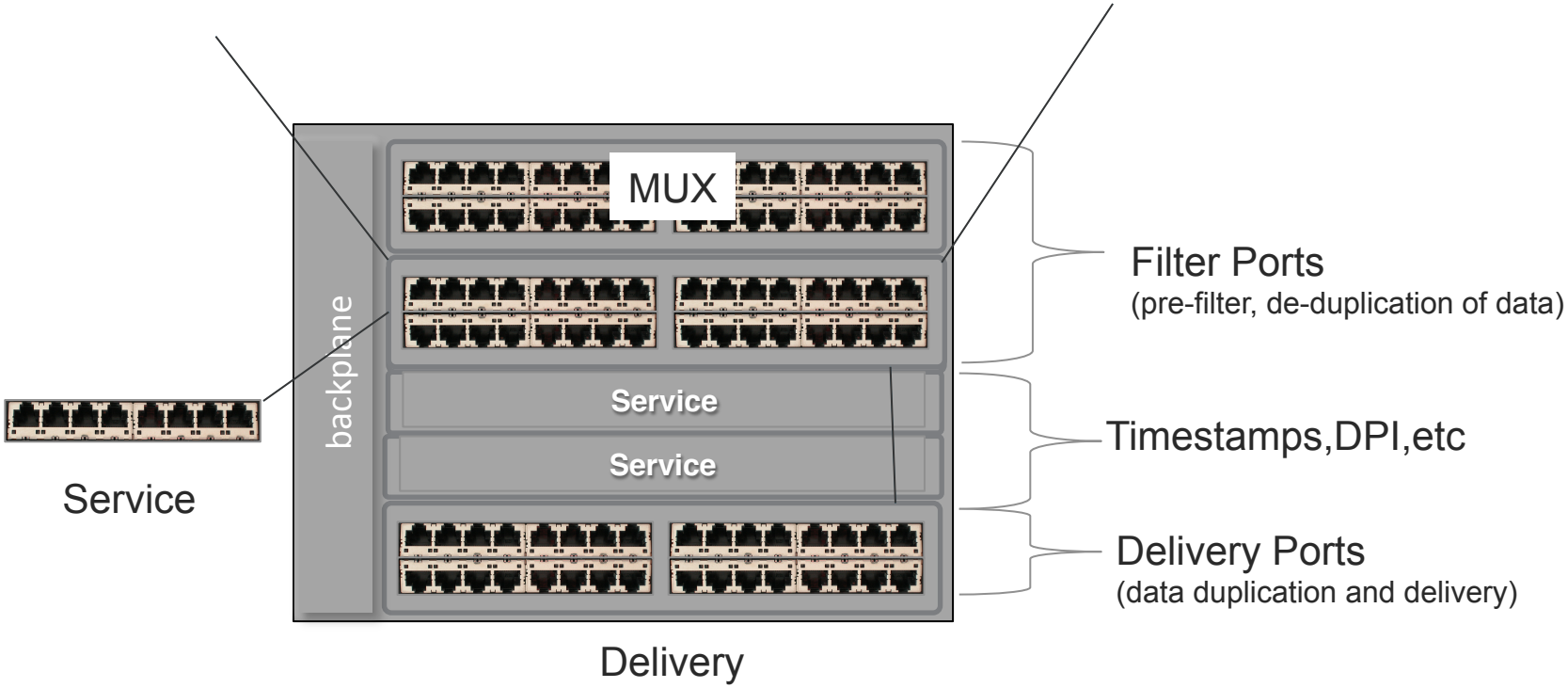
→ Merchant Silicon

Layer 7 packet inspection
Time stamps
Frame Slicing
Microburst detection

20%

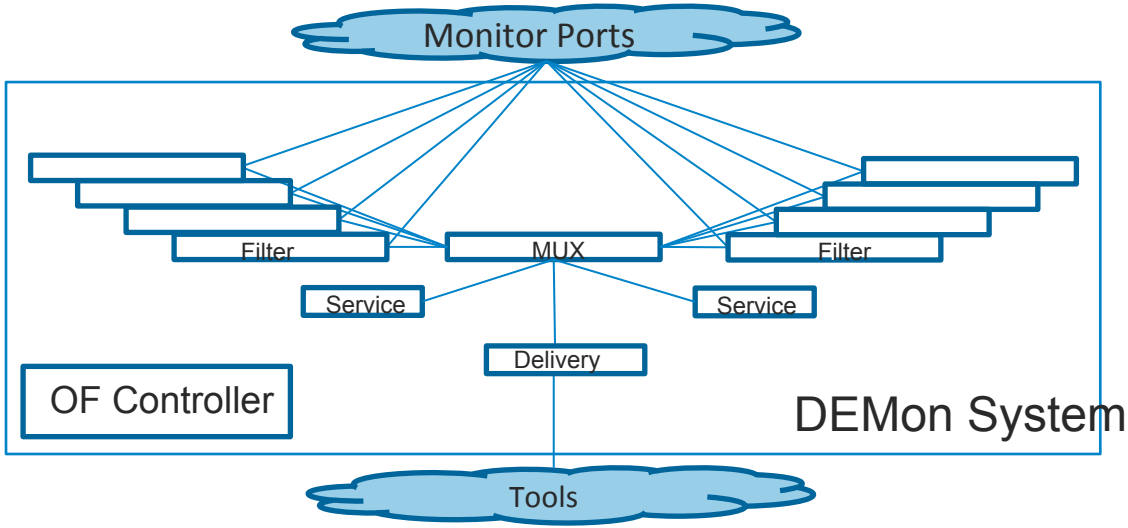
→ Requires specialized Hardware

Reverse Engineering Packet Broker

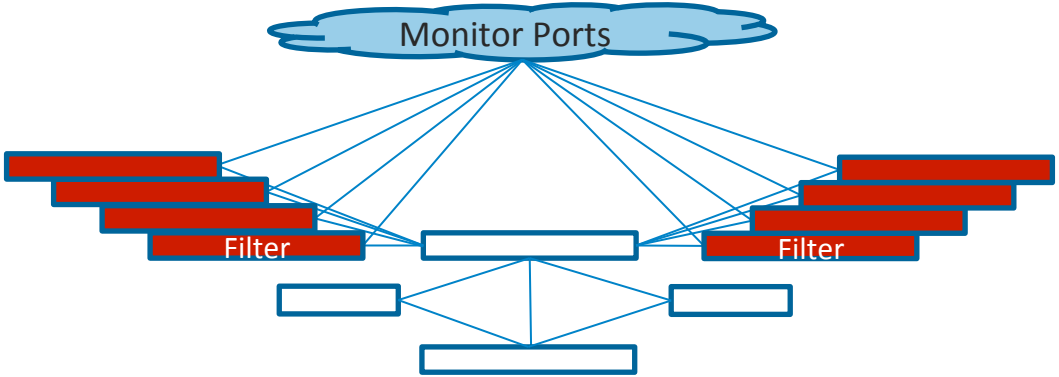


Architecture

Architecture Overview

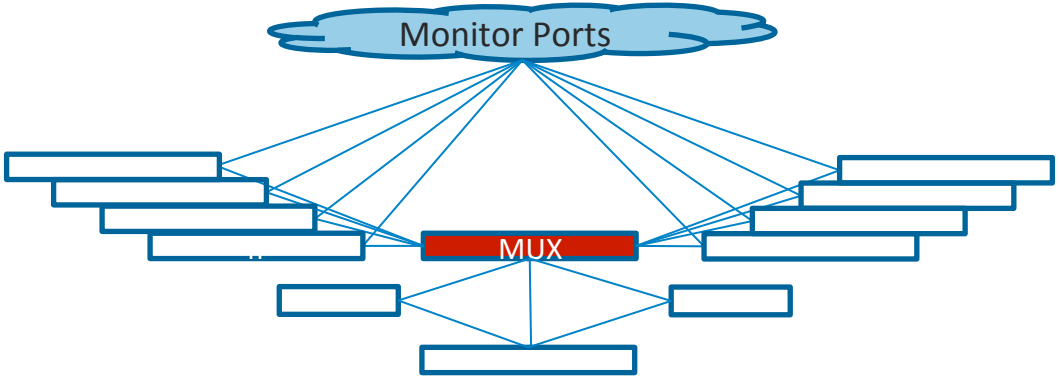


Filter Layer



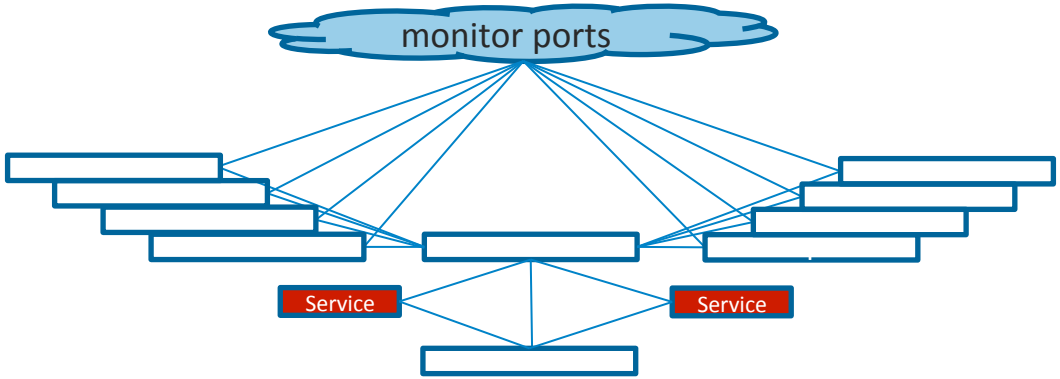
Terminates all monitor ports
Drops all traffic by default
De-duplication of data if needed
Aggressive sFlow exports

MUX Layer



Aggregates all filter switches in a data center
Directs traffic to either service nodes or delivery interfaces
Enables service chaining per policy

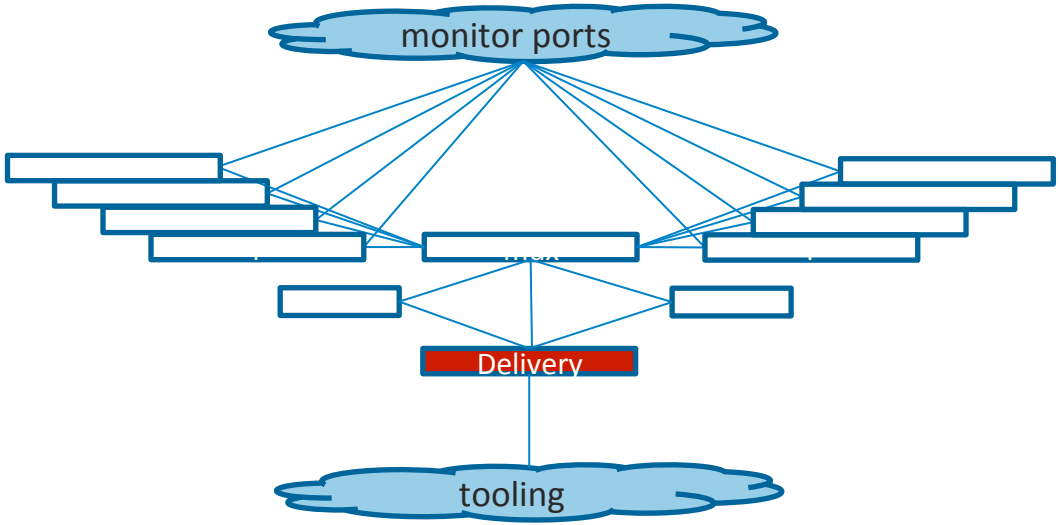
Service Layer



Aggregated by MUX layer
Flows are sent through services nodes to perform extended functions
Resulting flows are sent back to the delivery switch and then to tools

Some Applications:
Deeper (layer 7) filtering
Time stamping
Microburst detection
Traffic Ratio's (SYN/SYNC ACK)
Frame slicing (64, 128 and 256 byte)
Payload removal for compliance
Rate limiting

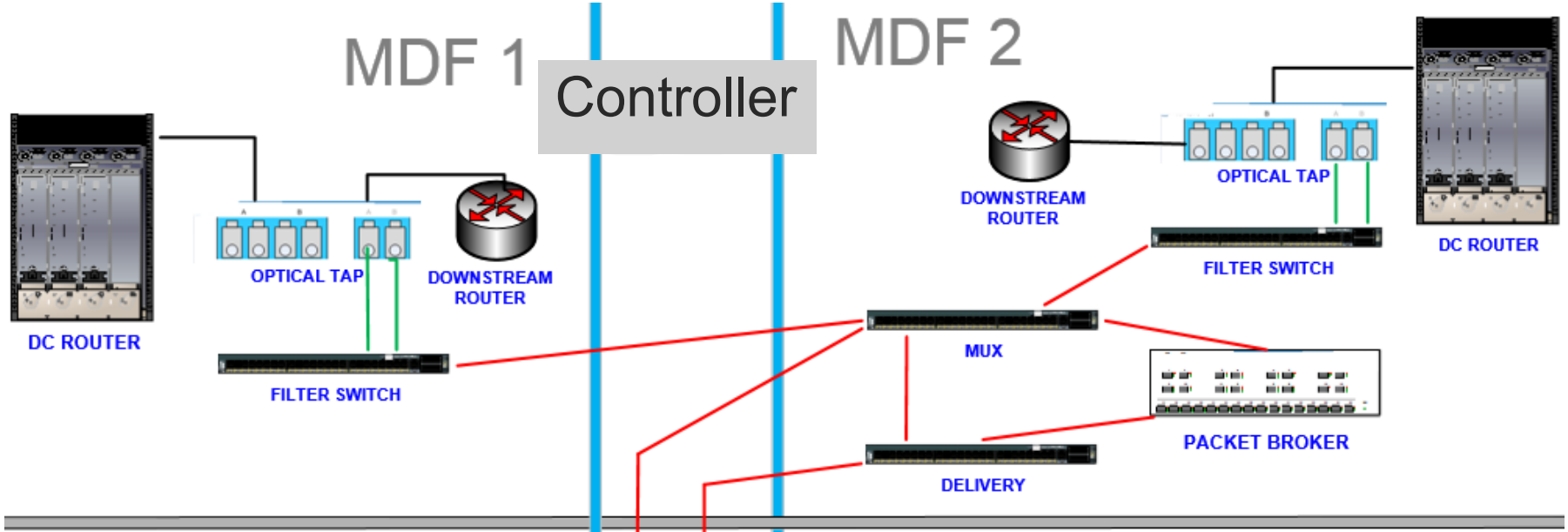
Delivery Layer

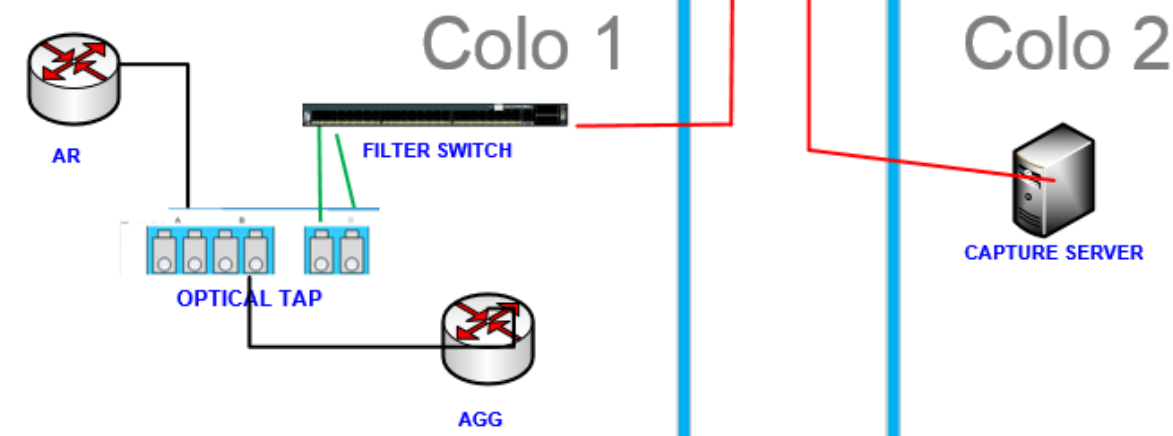


1:N and N:1 delivery to duplication of data
Delivery to local or tunnel traffic to remote tools

SDN Controller

OpenFlow 1.0 based
Discovers topology via LLDP
Roles of each layer are assigned and discovered automatically



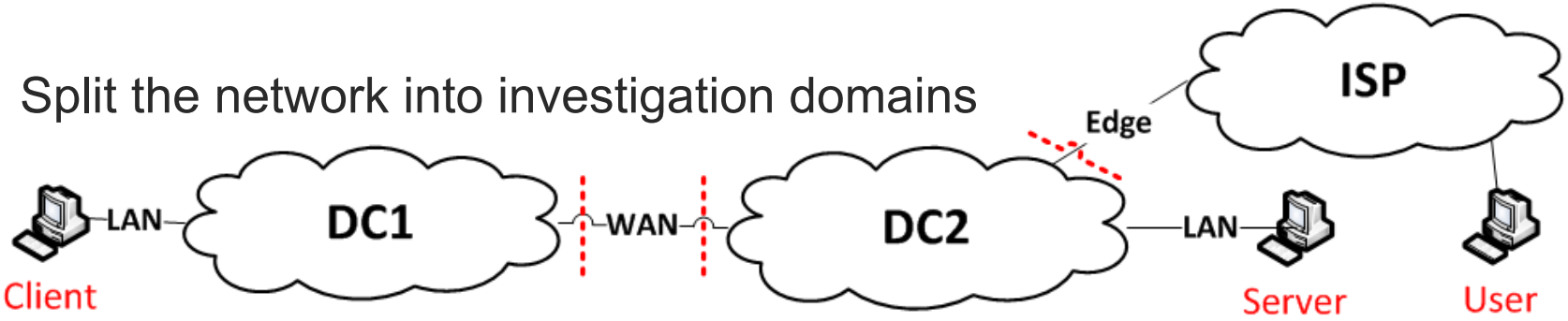


policy demo
description 'Ticket 12345689'
1 match tcp dst-port 80
2 match tcp src-port 80
filter-interface 1_2
filter-interface 1_1
filter-interface 2_1
filter-interface 2_2
delivery-interface Capture_server_NIC_1
use-service tcp_trunk_1 sequence 1
action forward

Use Cases and Examples

Reactive

Split the network into investigation domains

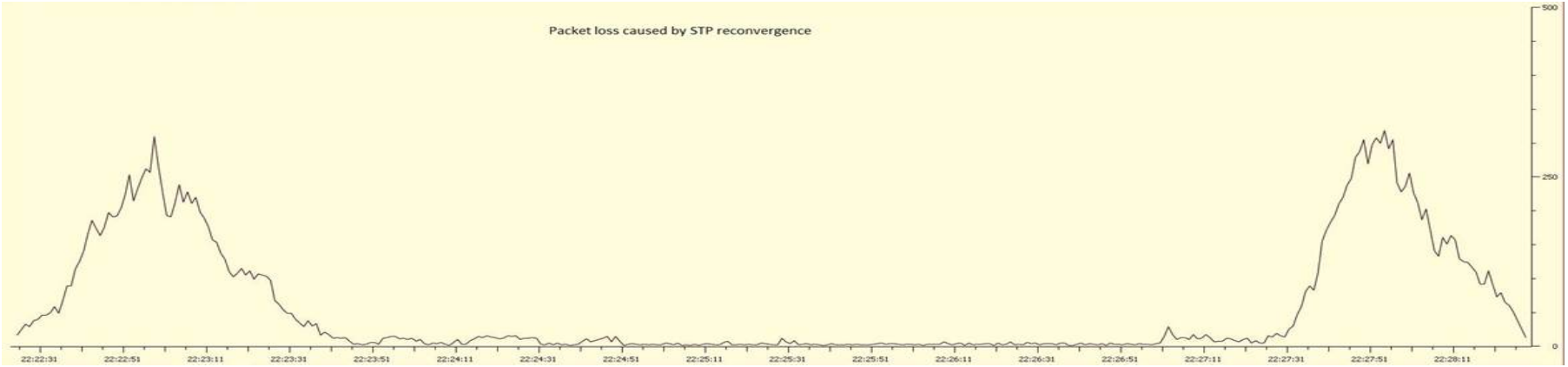


Quickly exonerate or implicate a network segment

Verify TCP intelligent network appliance are operating as expected

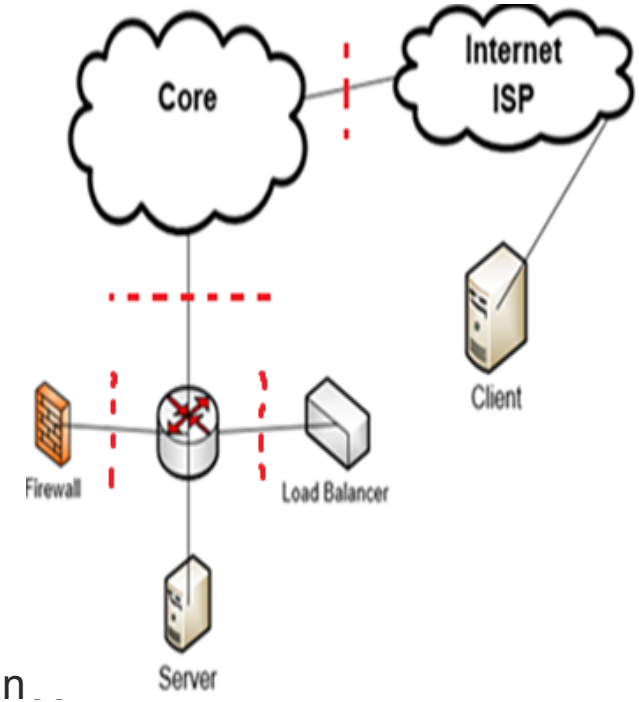
Proactive Monitoring

Relying sole on SNMP polling and syslog's gives you false confidence
Performance data can be gleaned from exposing TCP telemetry data
Ability to detect re-transmissions (TCP-SACK)



IPv6

Users were unable to connect to the service intermittently via IPv6



Repro facts:

3-way TCP connection setup's up
9-way SSL handshake fails
Ack for client hello was not making it back to load balancer

Solution:

Implicates or exonerates Layer7 devices that are commonly finger pointed

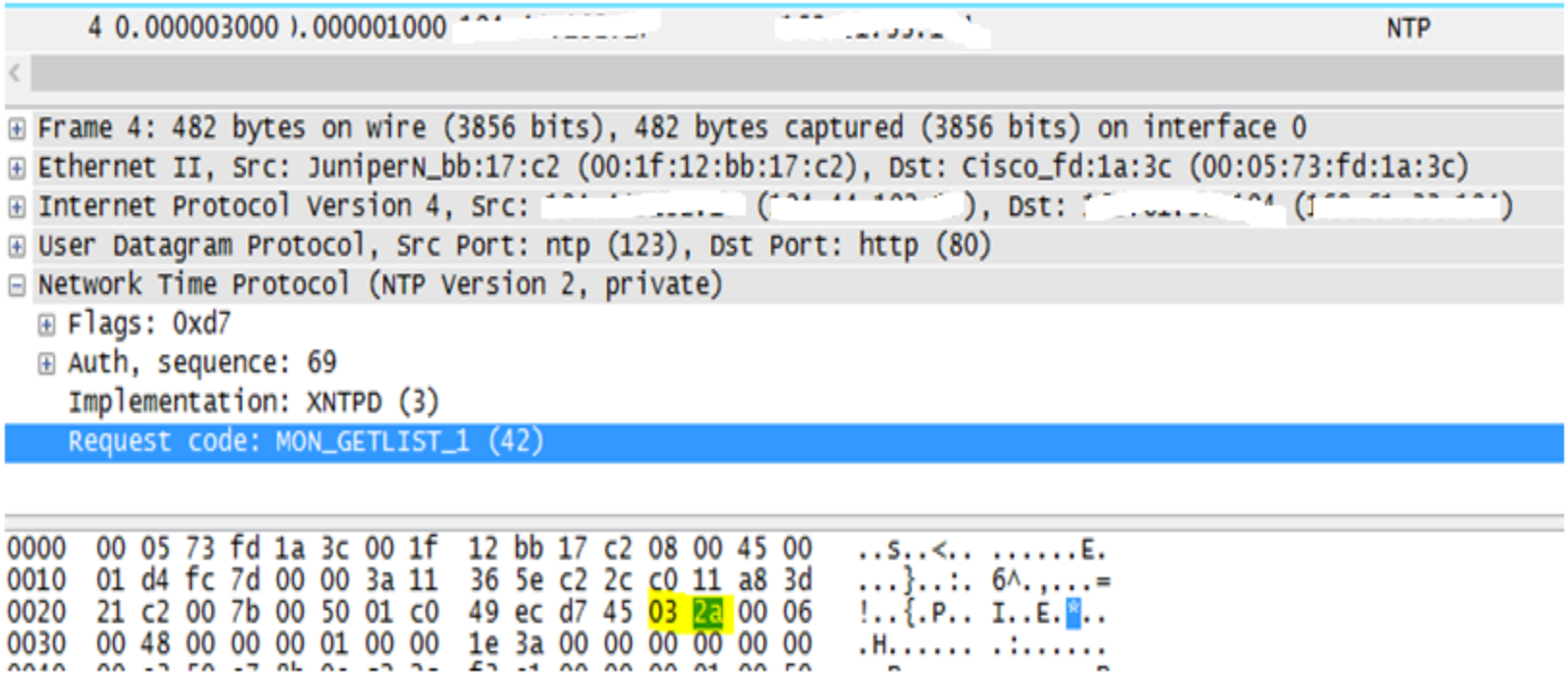
Root cause:

Race condition - If the client hello was received on the load balan
before the backend connection was made it would trigger the bug

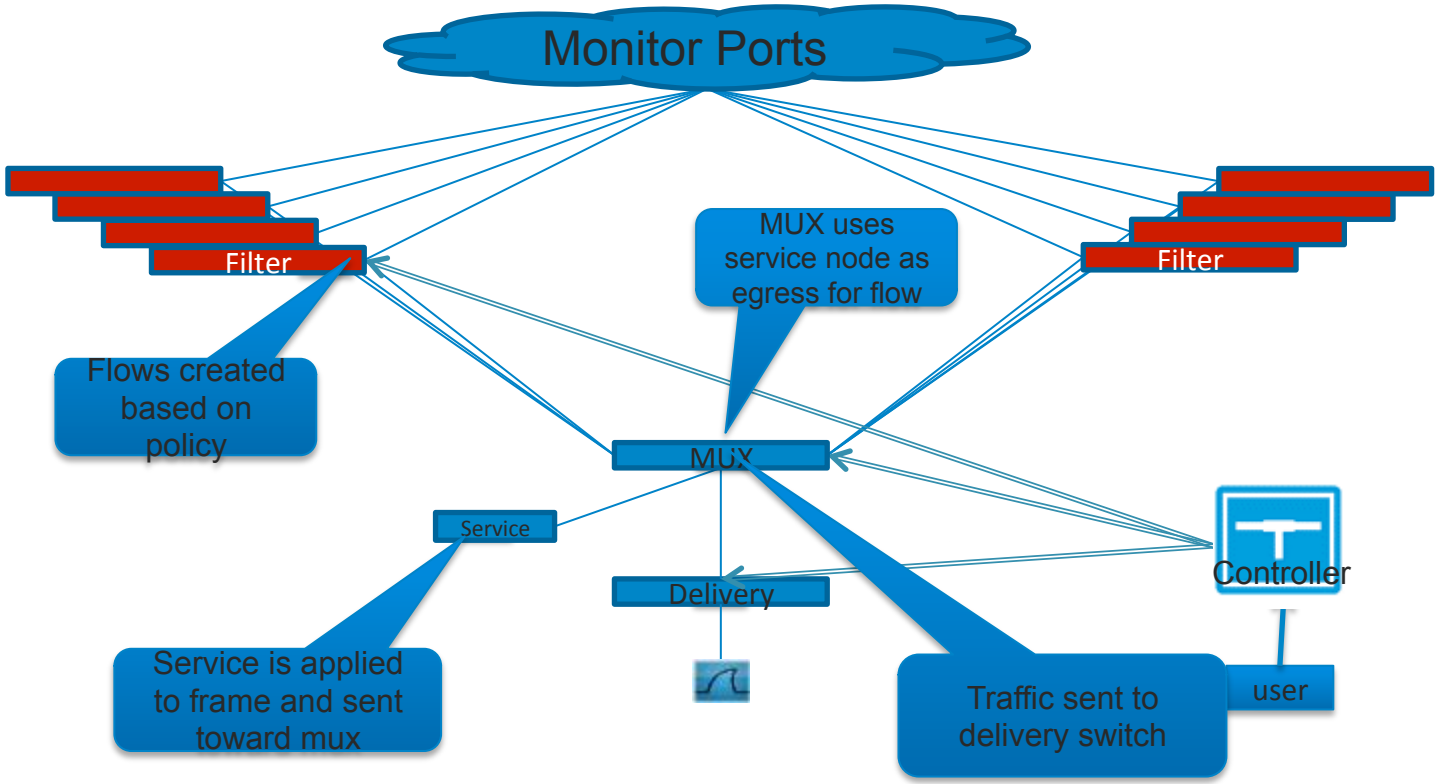
DDOS

5 Minutes later

DDOS: Packet Capture



sFLOW

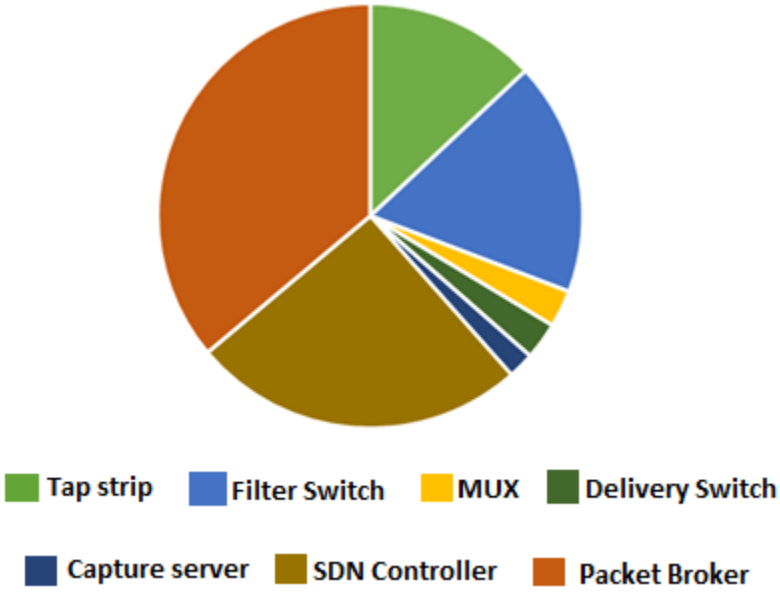

Monitor Ports
Filter
Filter
MUX uses service node as egress for flow
Flows created based on policy
MUX
Service
Delivery
Controller
user
Service is applied to frame and sent toward mux
Traffic sent to delivery switch

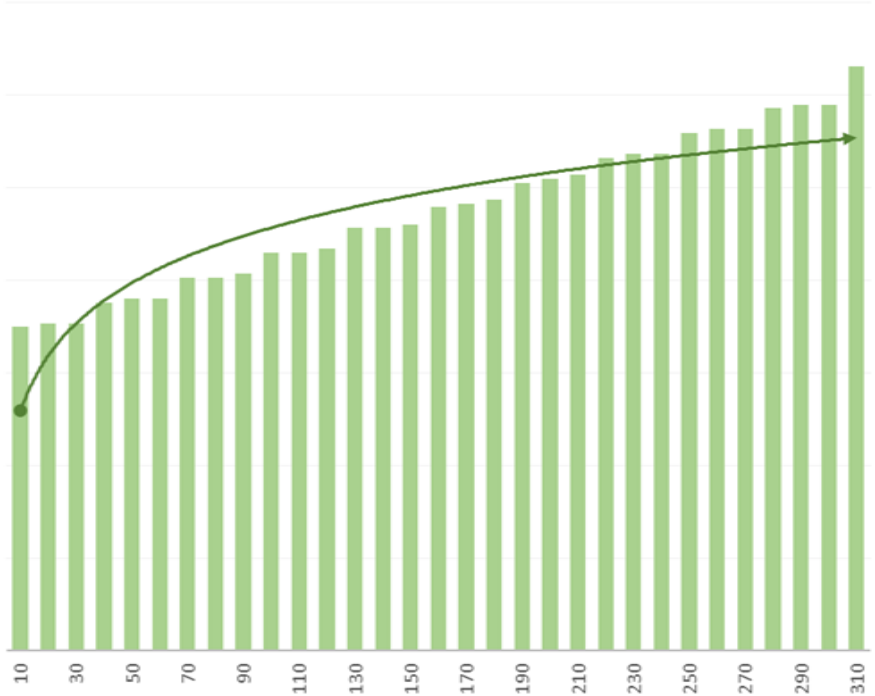
Caveats and Cost

Caveats

TCP/IP fields with MPLS encapsulated packets cannot be matched
Lack of IPv6 source and destination matching in OF 1.0
Limited number of flow rules due to TCAM limitation
Policy will not load balance traffic amongst ECMP links
Not all switch vendors OF implementation is the same
Commercial controller support is splintering

Cost breakdown



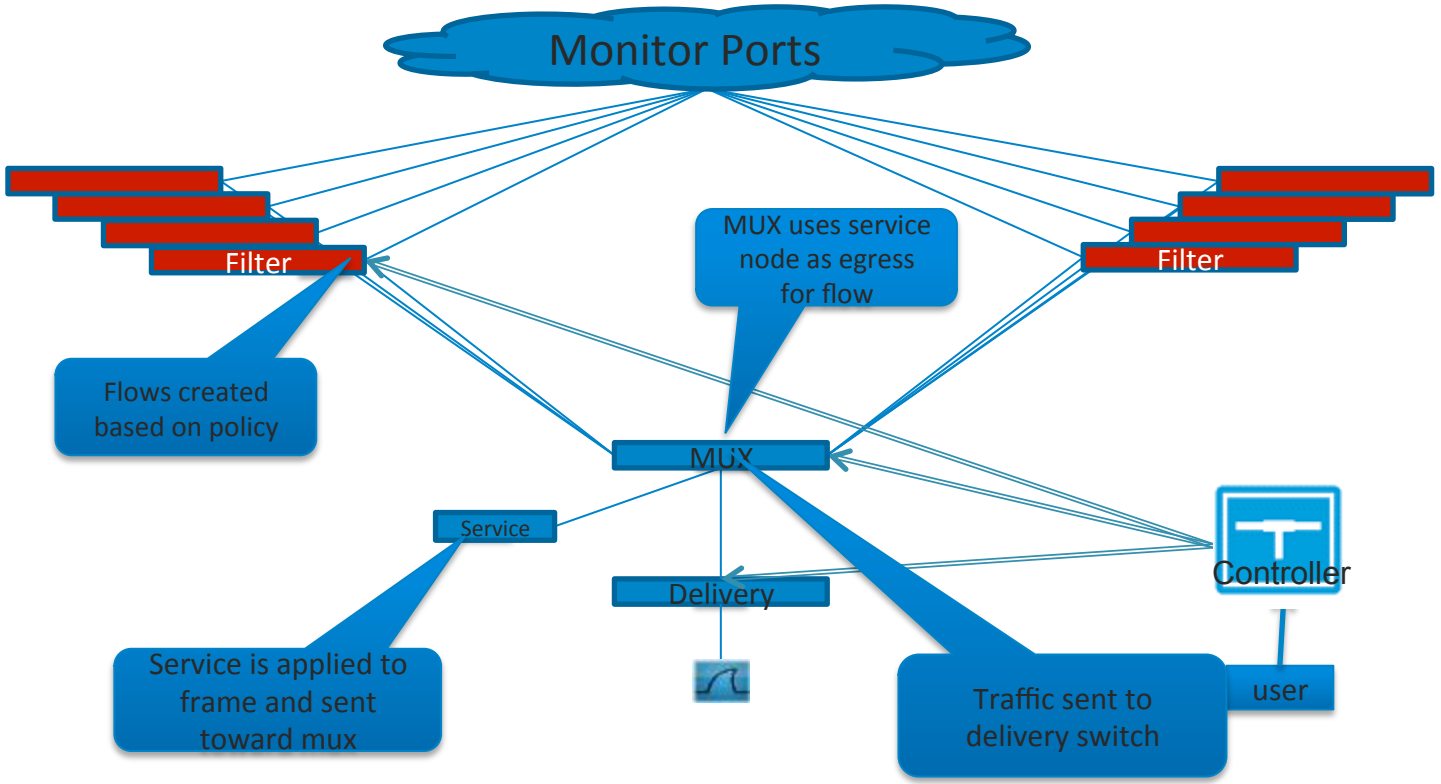


Number of links

Questions?

Backup

Creating a Service Chain



Sniffer Features

Terminate taps

Match on 5-tuple

Duplication of packets

Low latency

Layer 7 packet inspection

Time stamping

Frame slicing