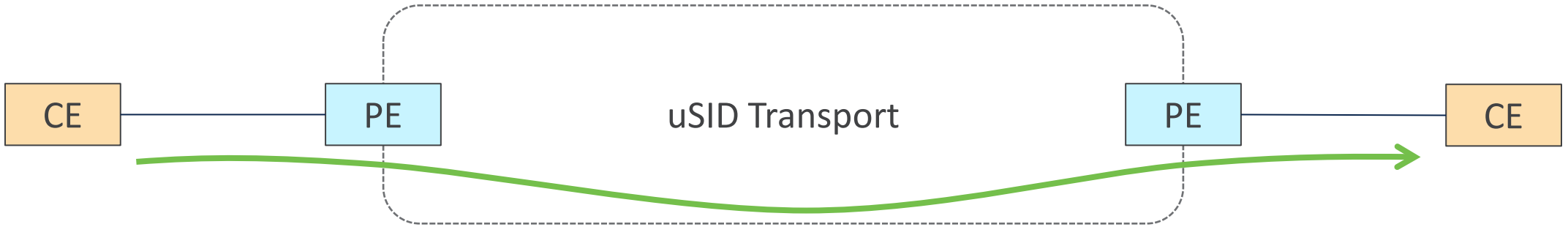
SRv6 uSID: Measurement Analytics & Path Tracing

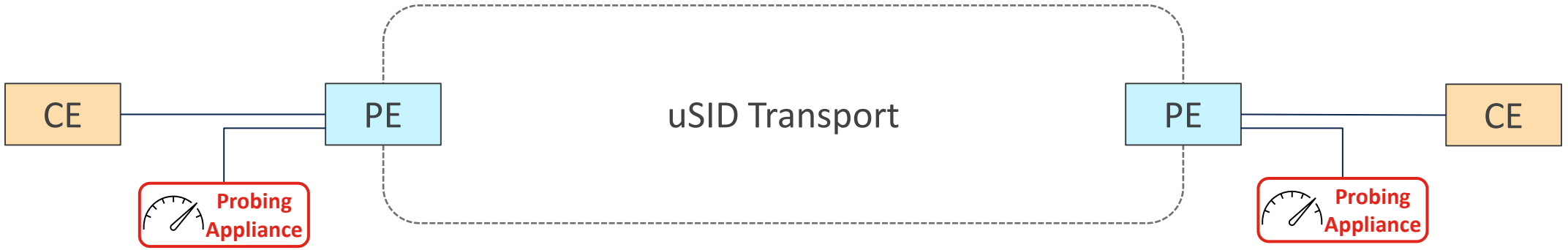
Clarence Filsfils, **Pablo Camarillo**, Ahmed Abdelsalam, Rakesh Gandhi, Sonia Ben Ayed

Service Creation

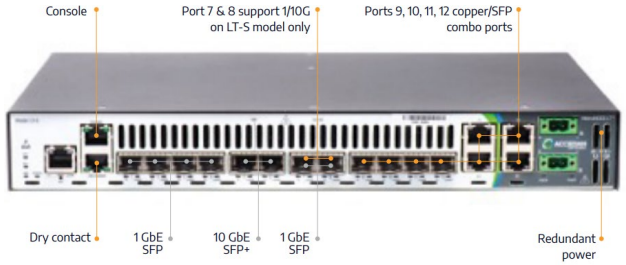


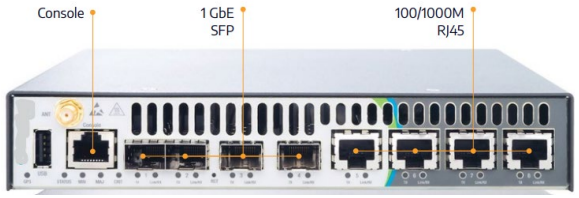
- End-to-End Policy
  from host to Internet through DC, Access, Metro, Core, Cloud
- Solves Any Requirement
  VPN, FRR, TE, NFV
- **We want to understand the SLA provided to customers: How do we measure the service?**

How do we measure the service?

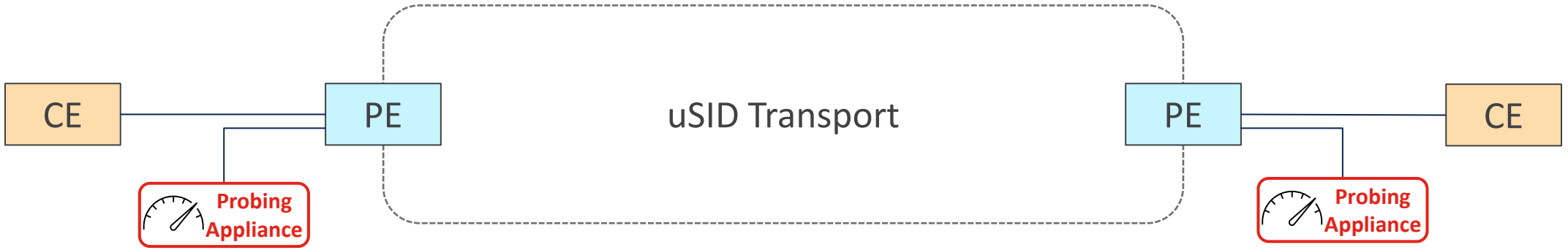


- **External probing appliances**
- x86 hardware not optimized for probing
- Commodity testing: L3 TWAMP Light generation and reflection; L2 Service OAM



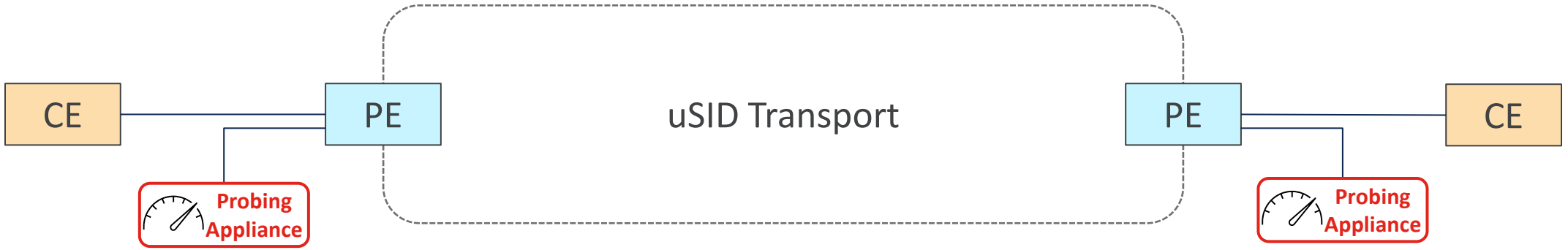


How do we measure the service?

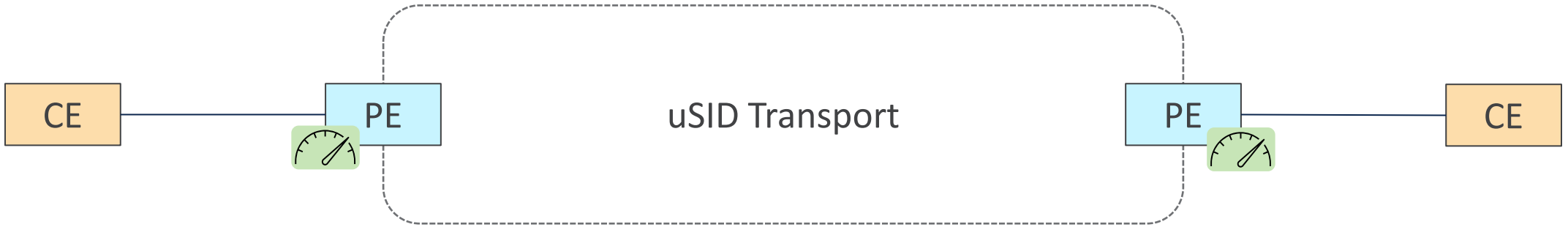


- Complex to deploy (requires physical installation)
- Complex to operate (often require proprietary orchestrator for provisioning)
- Port consumption on PE devices to connect monitoring appliances
- Extra power consumption ~90W (+ SFP)
- Limited monitoring (blind end-to-end; no Routing/CP knowledge; no visibility into underlay fabric)
- Limited analytics (providing min, max, avg doesn't help analyzing issues)
- Extremely expensive:
  "For each 1$ spent on routers, we spend an additional 1$ on probing appliances."

How about integrating it in the Hardware?

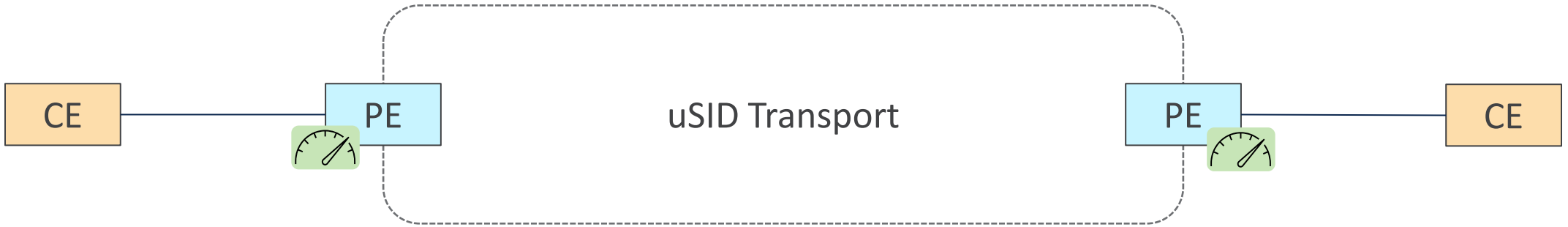


How about integrating it in the Hardware?



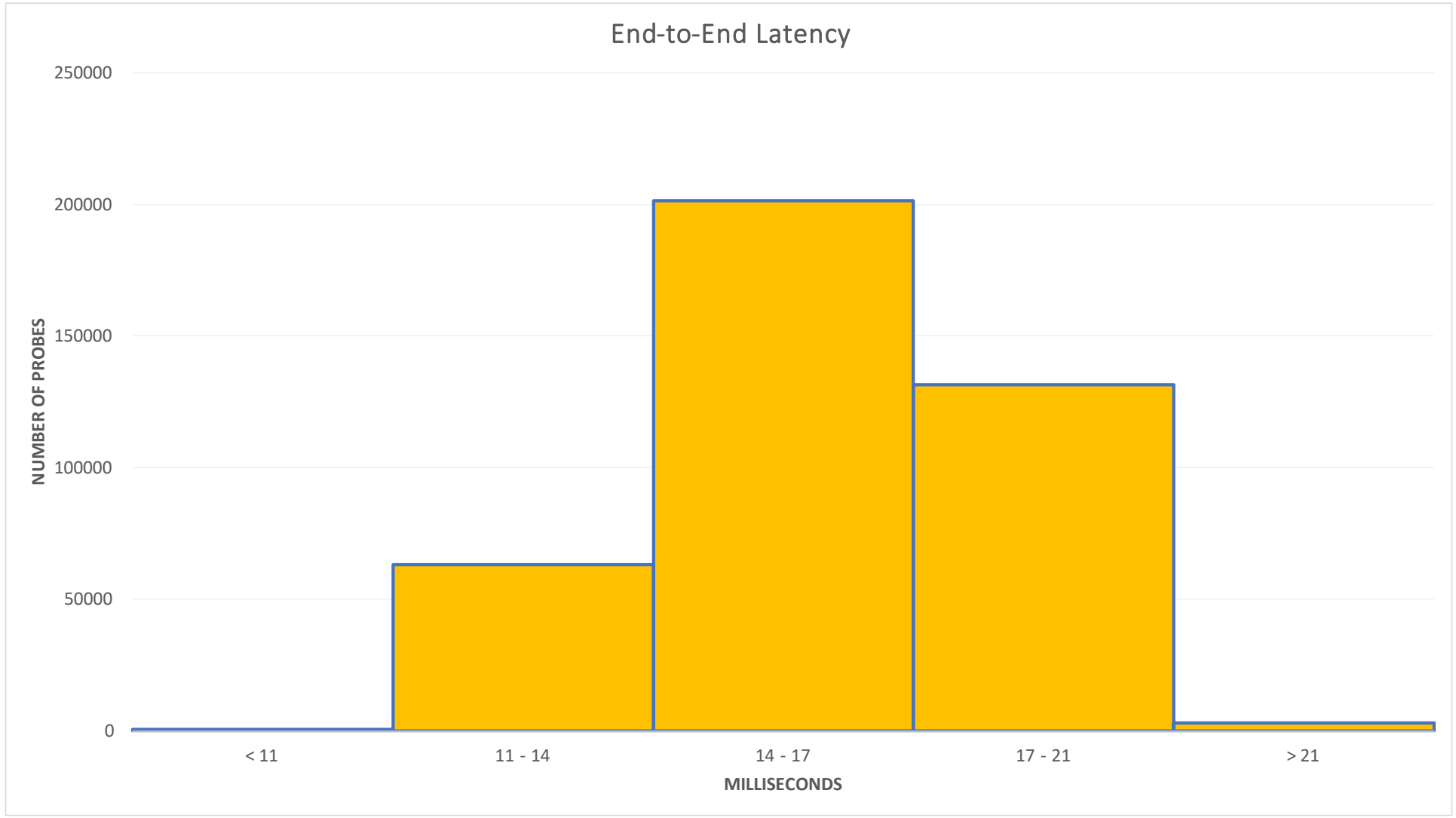
- Leverage router HW capabilities for probe generation/ingestion **natively**
- CAPEX savings:
  - No need to spend $ on external boxes
  - No PE port consumption
- OPEX savings:
  - Already integrated in your router. No physical deployment required; no cost involved in maintenance; same orchestration
  - No additional power consumption
- Richer functionality with routing awareness
  - FlexAlgo, DSCP, ECMP, …

Are we performant?

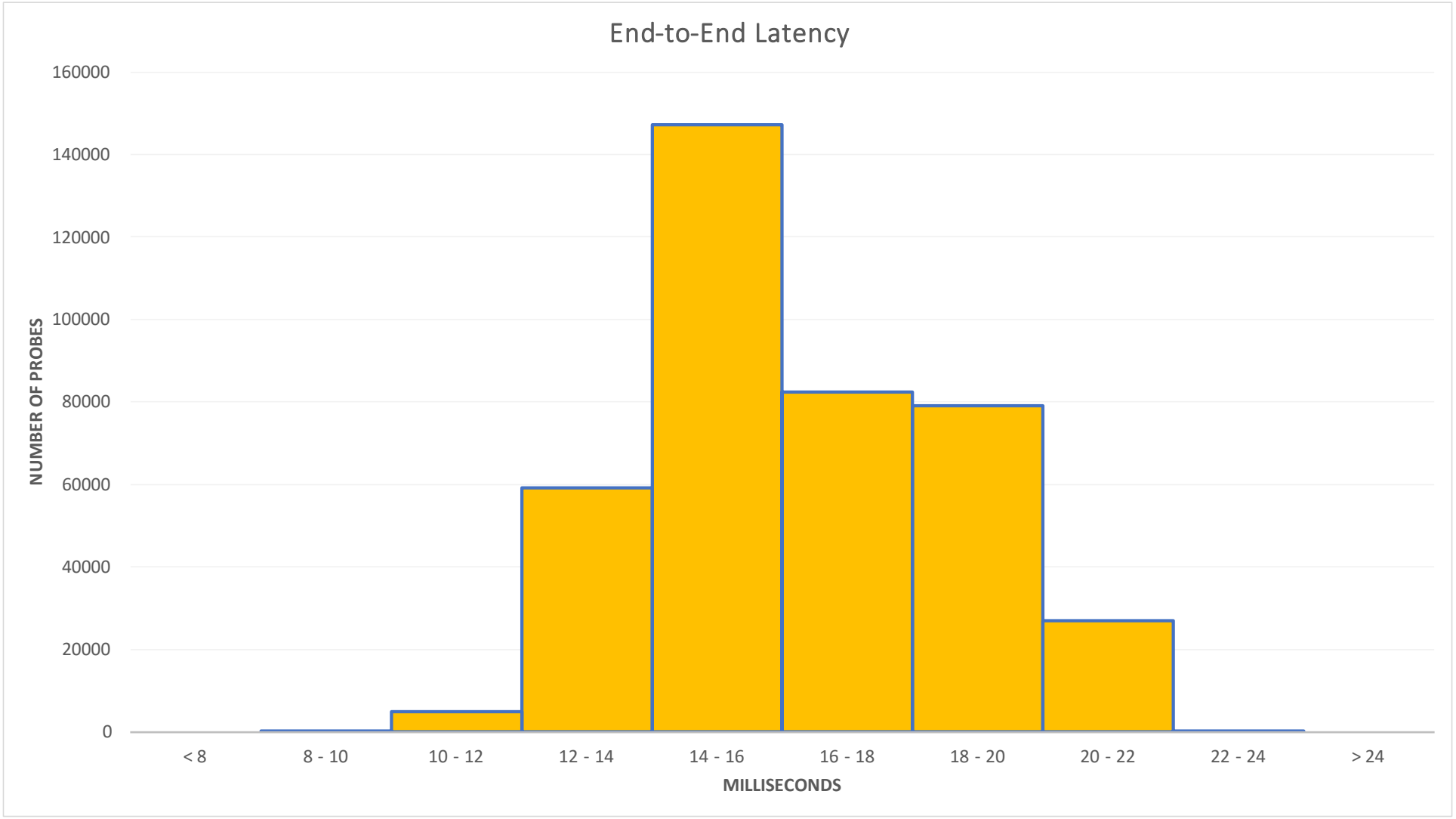


- Dedicated HW engine generates probes at high rate 🚀
  - Silicon One (Q200) delivers > **14 Million Probes Per Second**
  - For free: no impact on service creation
- Dedicated HW engine ingest probes, measures and aggregates data:
  - Granular statistics: No more {min, avg, max}
  - **Latency histogram!**

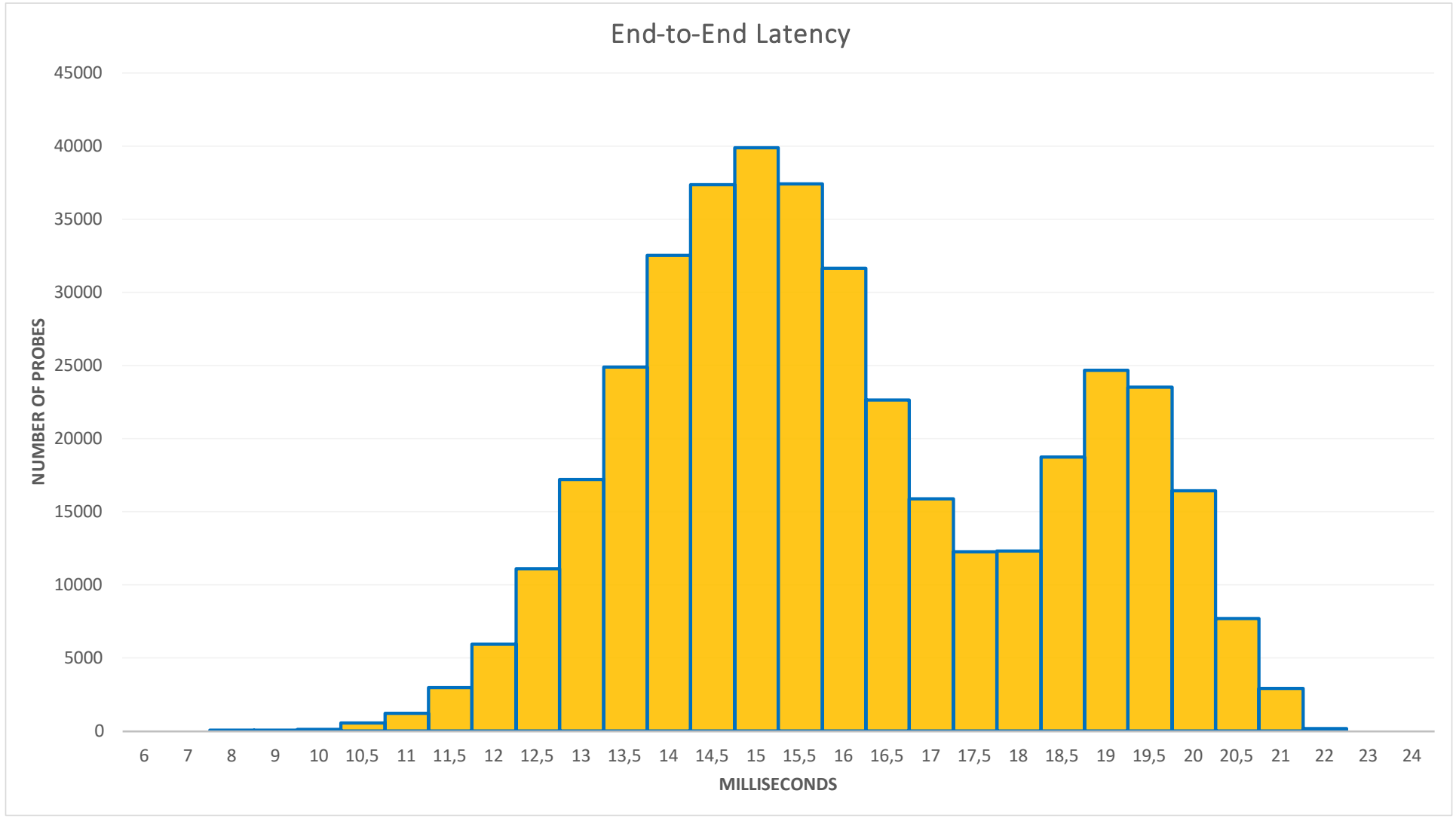
The importance of granular statistics – 5 bins



The importance of granular statistics – 10 bins



The importance of granular statistics – 30 bins



Service Measurement: 3L Probing

- End-to-end measurement of Latency, Loss and Liveness
  - From iPE to ePE
  - Single probe for everything
- Synthetic probing
  - Simulates end-user experience
  - Allows to detect problems before users do
- We monitor the shared transport **AND** the service forwarding path on PEs
- We monitor all ECMP paths from iPE to ePE
- Digging billions of Performance Measurements
  - Pushed through telemetry to network-wide analytics engine
  - Analytics allow to Correlating with Current and Past Routing Data

Network-wide Analytics

Network-wide analytics with SR Services App

- Single pane of glass visibility for network status

Network-wide analytics with SR Services App

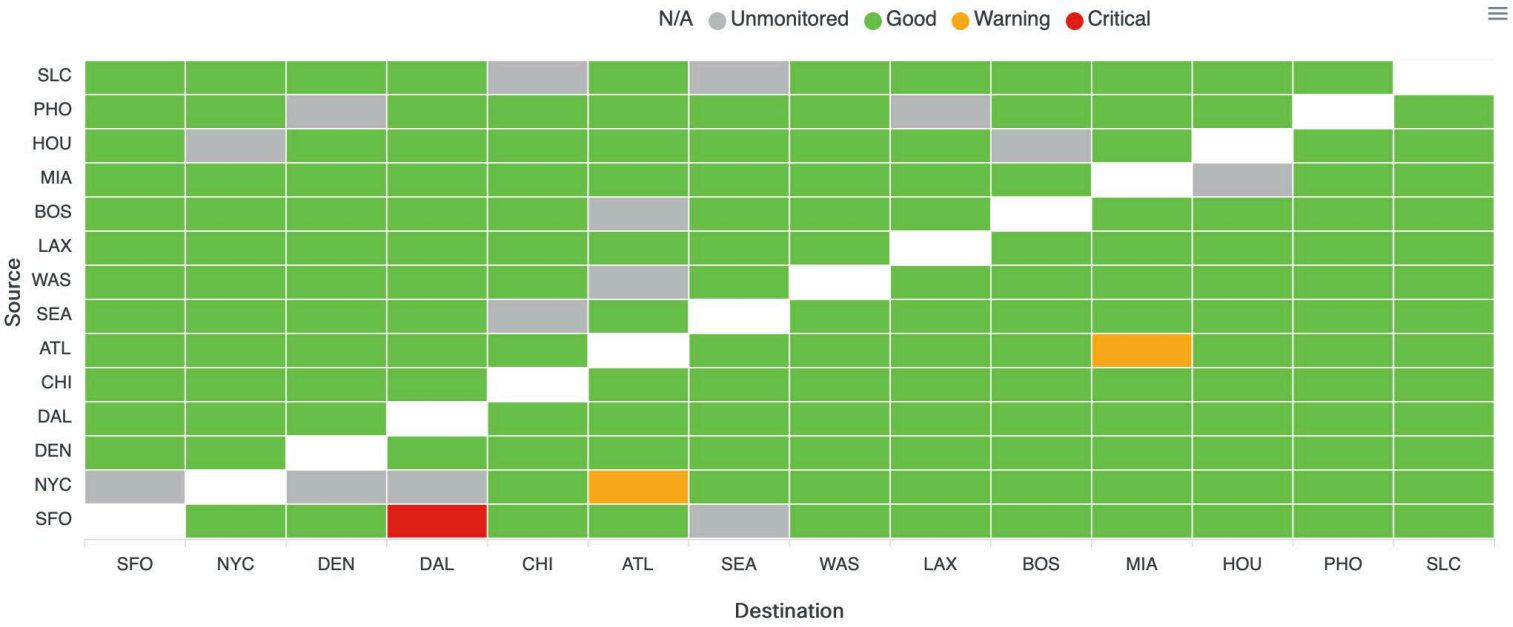
- Single pane of glass visibility for network status
- Intelligent data
  - No point in plotting 40k graphs with latency
  - We process raw/brute data and correlate it with routing to obtain **intelligent** data
  - Automatically drawing your attention to what matters
- Analytics allows to identify suspected *troublemakers* and trigger further troubleshooting

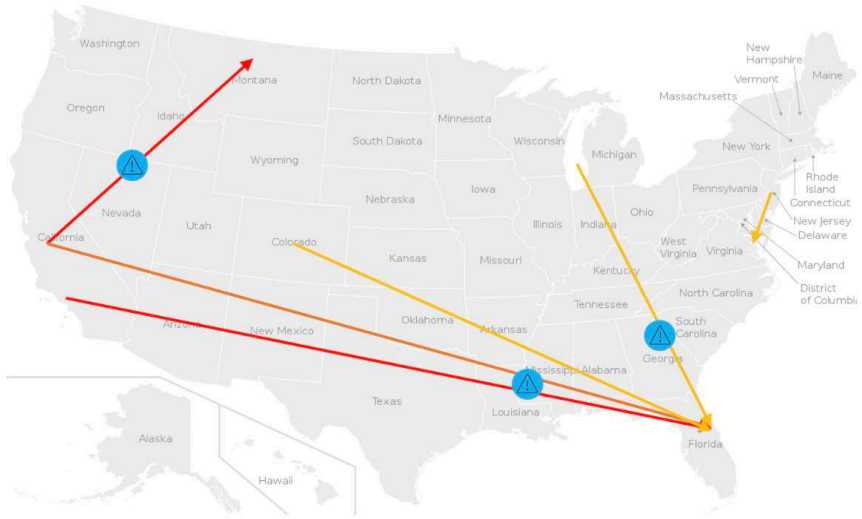
Cisco

3L Loss

Last period | Last stable topology

Critical Threshold: 0.25% Warning Threshold: 0.1% Flex-Algo 0 DSCP 0





Suspected Links

Last 15 minutes (10:00 to 10:15)

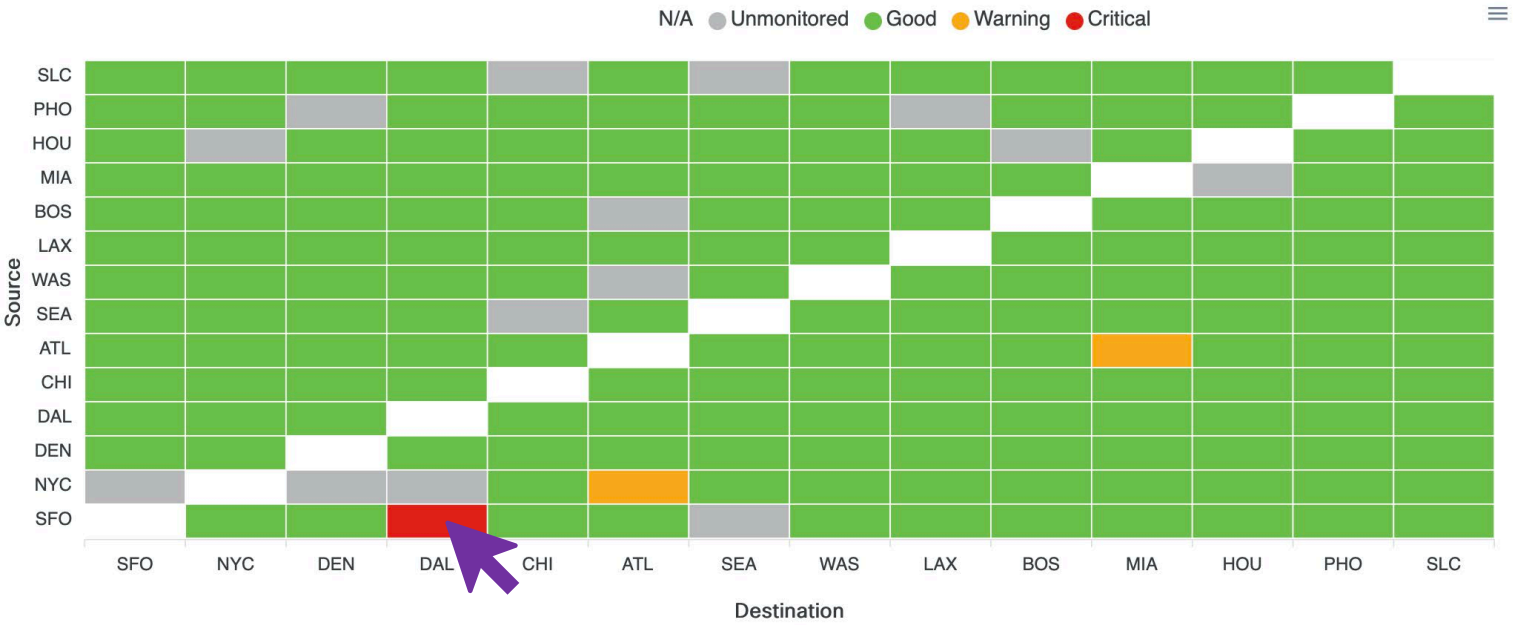
1% (1)	1% (2)	91% (165)	7%(14)
Critical	Warning	Good	Unmonitored

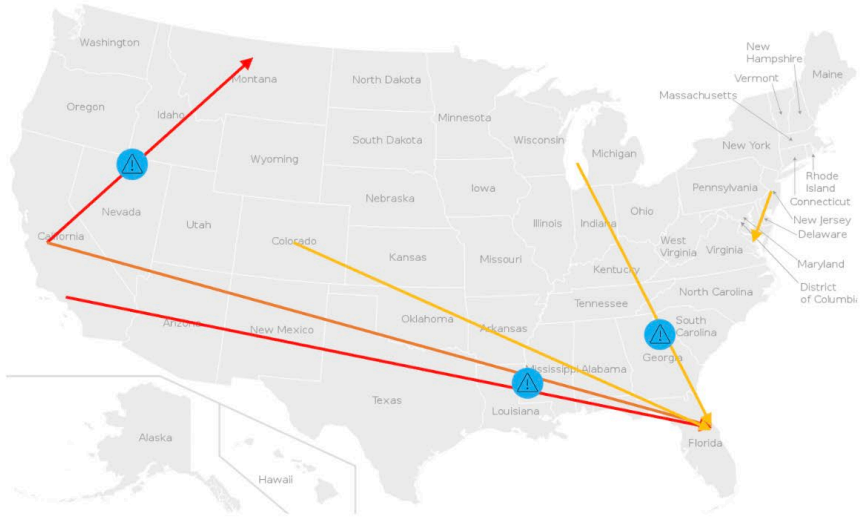
Cisco

3L Loss

Last period | Last stable topology

Critical Threshold: 0.25% Warning Threshold: 0.1% Flex-Algo 0 DSCP 0





Suspected Links

Last 15 minutes (10:00 to 10:15)

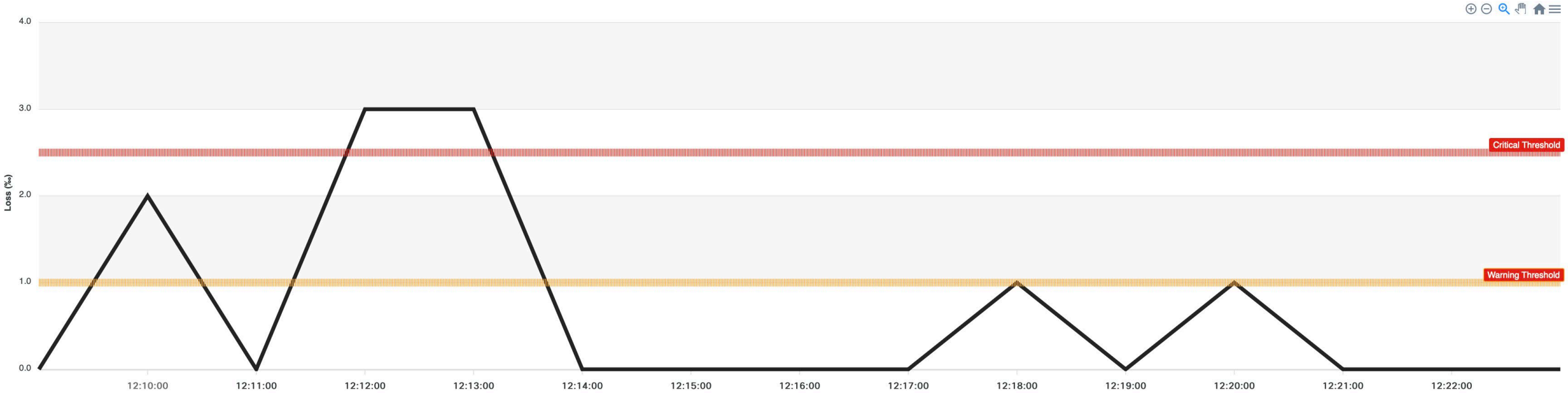
1% (1)	1% (2)	91% (165)	7%(14)
Critical	Warning	Good	Unmonitored

Cisco

3L Loss

Src: SFO | Dst: DAL | Color/QoS: IGP/0 | Timescale: Last 1h | Submit

Critical Threshold: 0.25% Warning Threshold: 0.1% Flex-Algo: 0 DSCP: 0

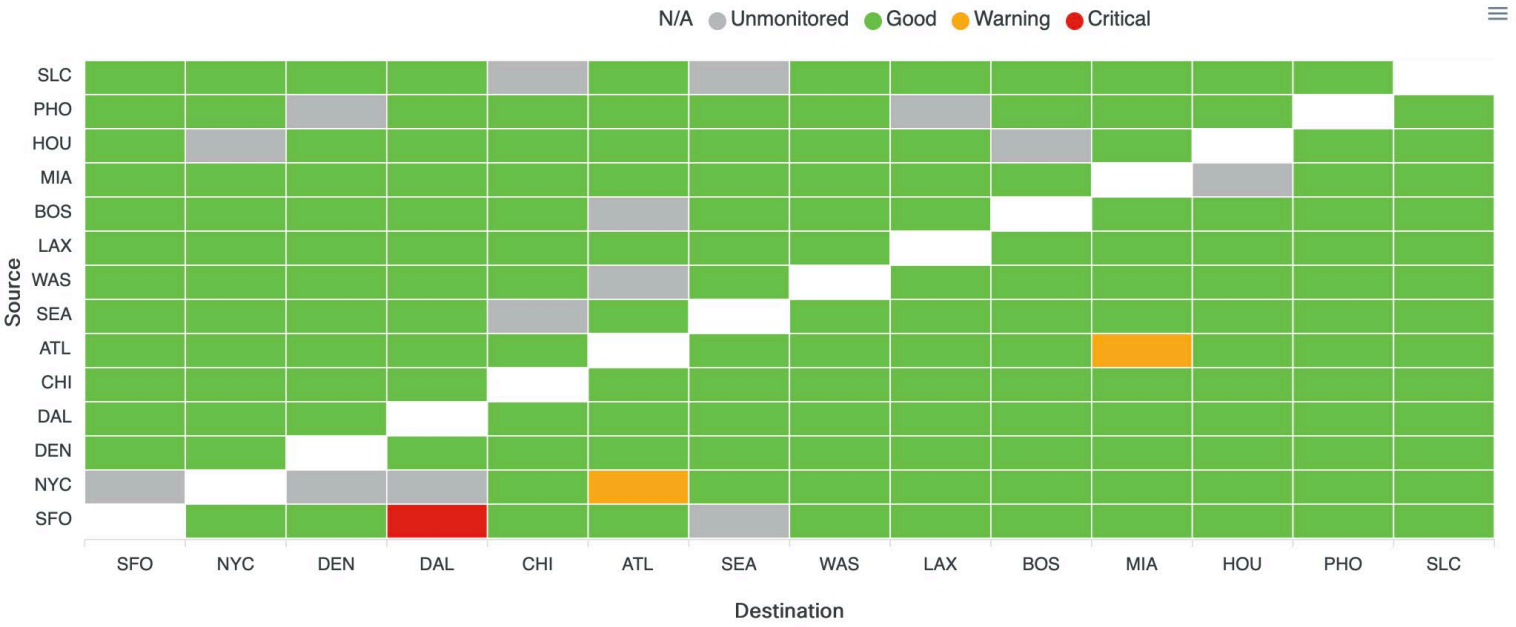


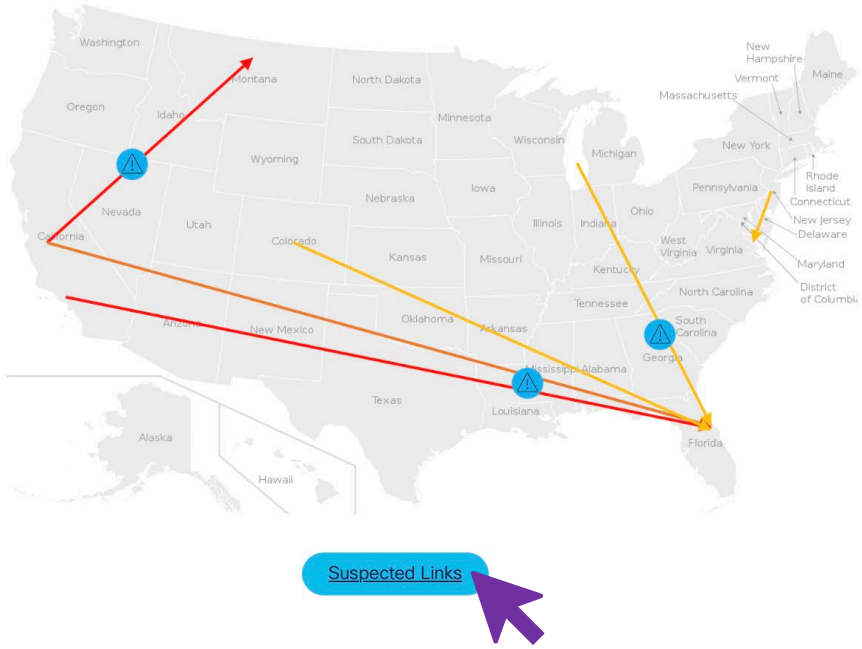
Cisco

3L Loss

Last period | Last stable topology

Critical Threshold: 0.25% Warning Threshold: 0.1% Flex-Algo 0 DSCP 0





Suspected Links

Last 15 minutes (10:00 to 10:15)

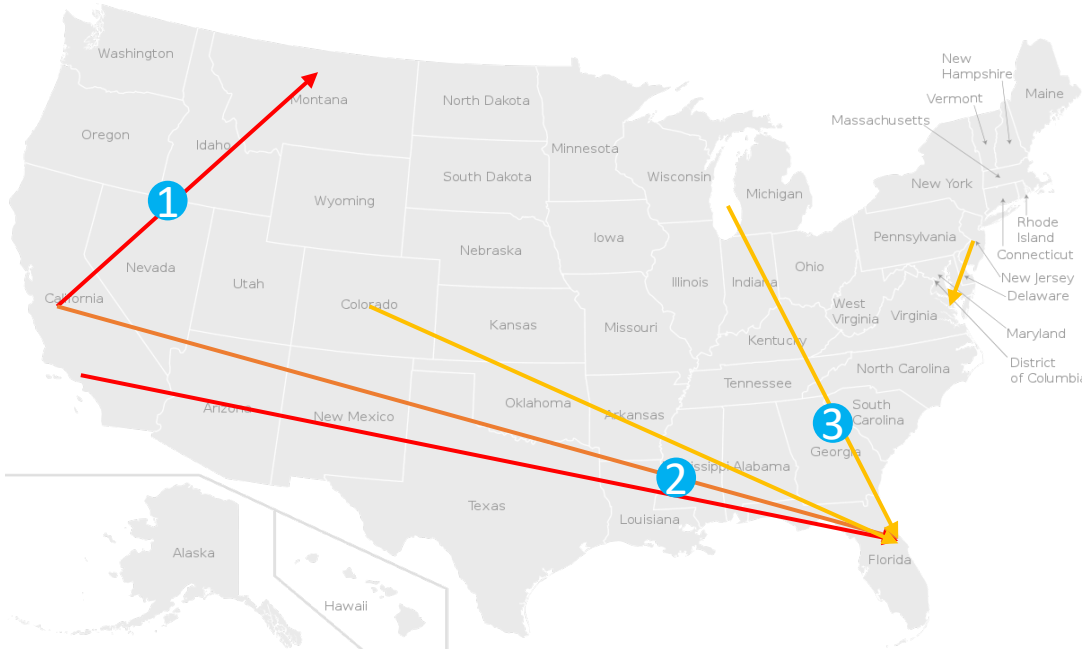
1% (1)	1% (2)	91% (165)	7%(14)
Critical	Warning	Good	Unmonitored

Suspected links for ongoing issues:

1.- Sacramento-HundredGigE0/0/0/1: Load(97%) Test
2.- NewOrleans-HundredGigE0/0/0/2: Correlation Test
3.- Atlanta-HundredGigE0/0/3: Load 92% Test

Historical Troublemakers:

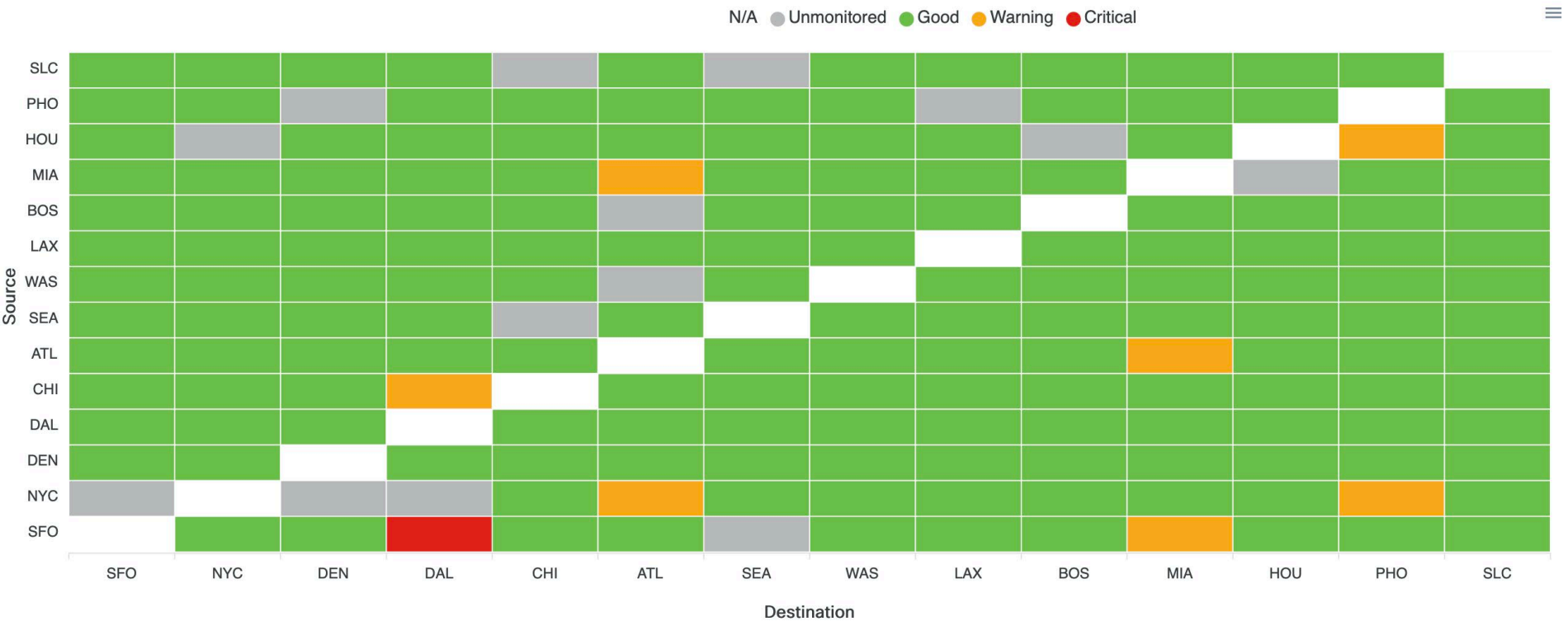
- NewOrleans-HundredGigE0/0/0/2: Correlation Test



Cisco

3L Latency

Last period | Last stable topology



Flex-Algo 0

DSCP 0

Warning Threshold +10% of expected latency

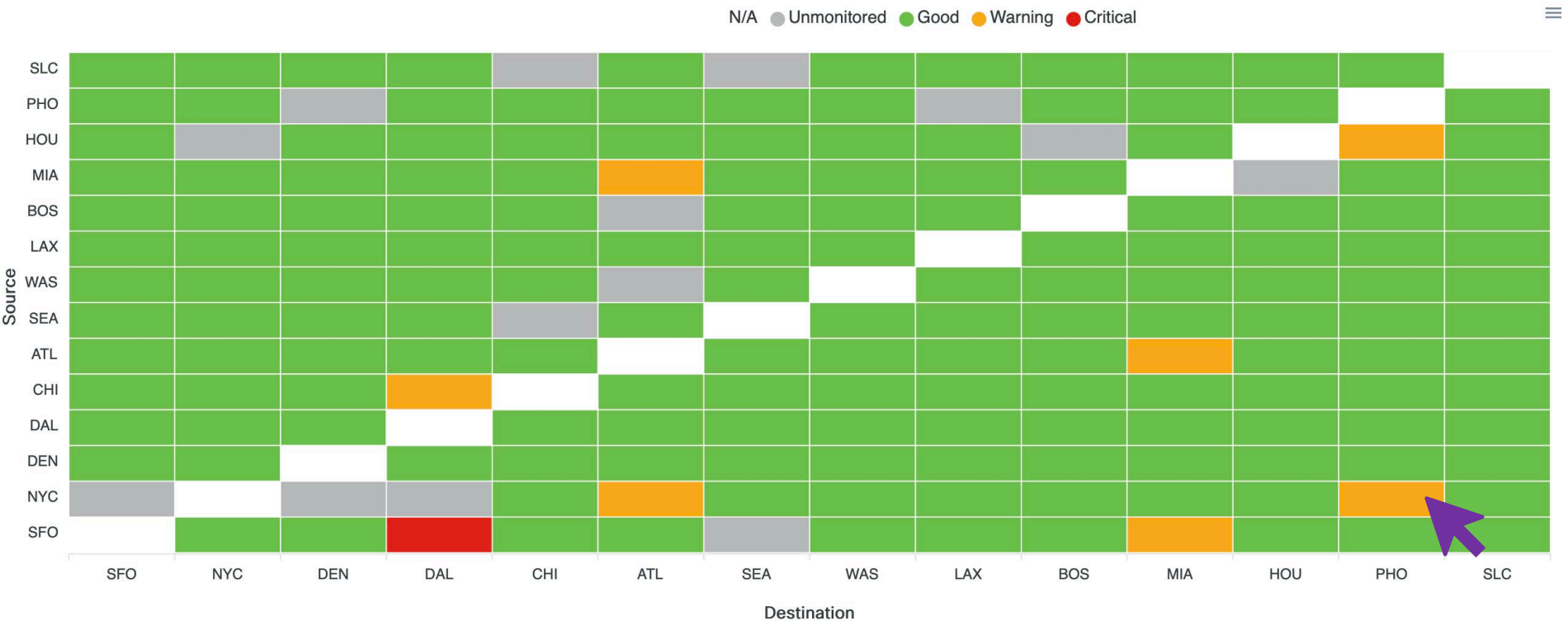
Critical Threshold: +20% of expected latency

1% (1)	4% (7)	88% (160)	7%(14)
Critical	Warning	Good	Unmonitored

Cisco

3L Latency

Last period | Last stable topology



Flex-Algo 0

DSCP 0

Warning Threshold +10% of expected latency

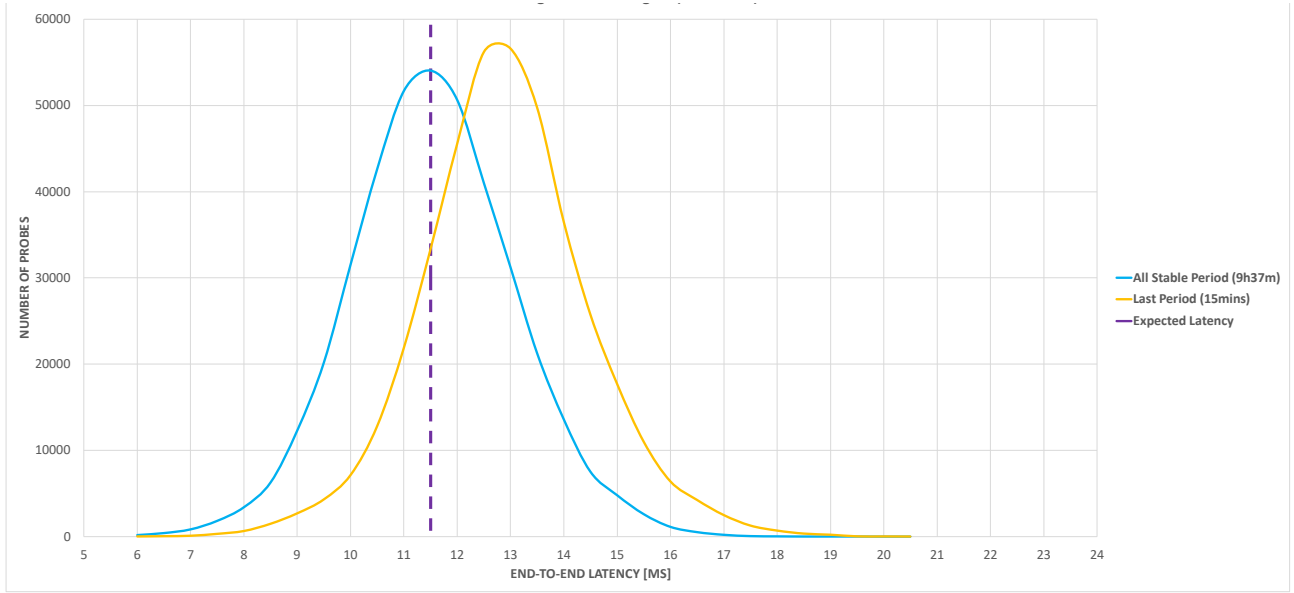
Critical Threshold: +20% of expected latency

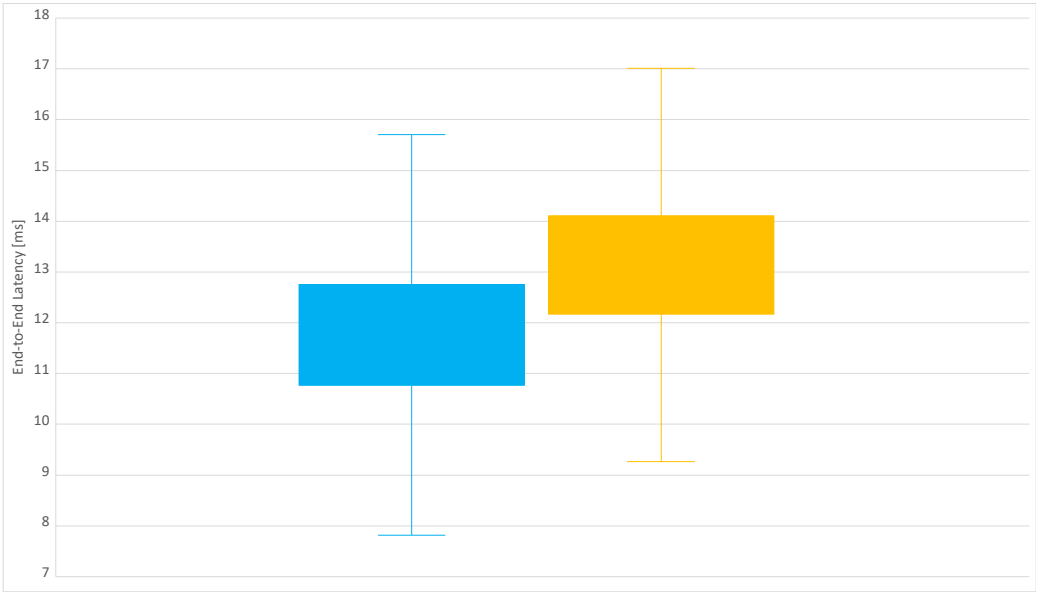
1% (1)	4% (7)	88% (160)	7%(14)
Critical	Warning	Good	Unmonitored

Cisco

3L Latency

Src: NYC | Dst: PHO | Flex-Algo: All | Timescale: Last period | Submit

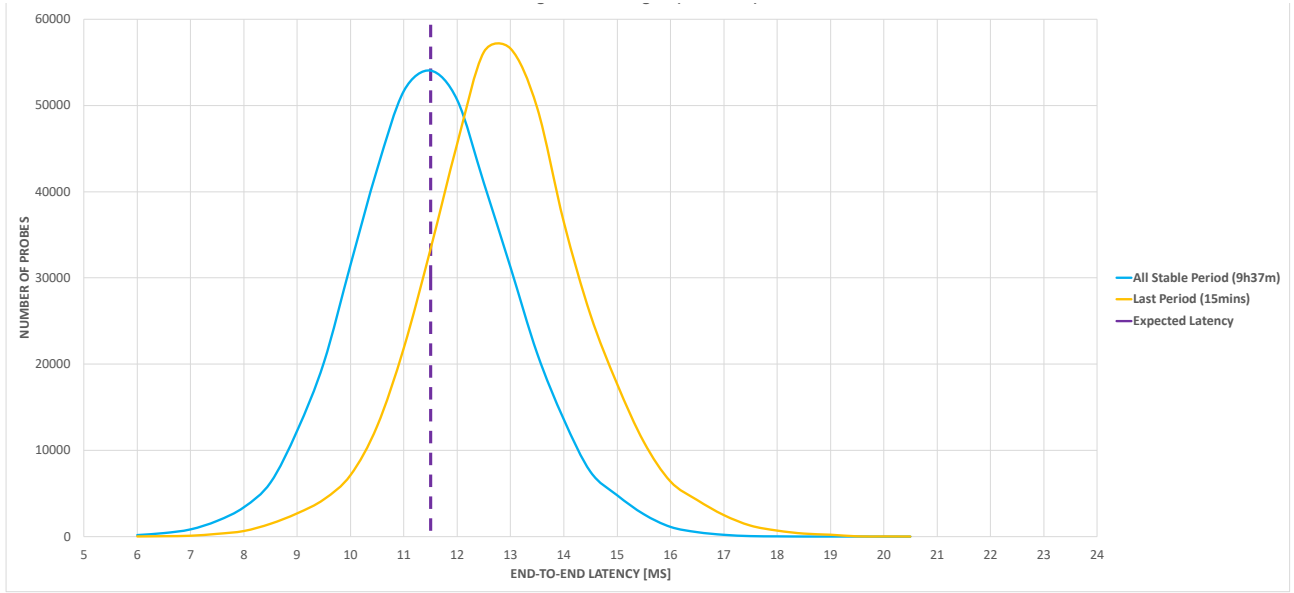


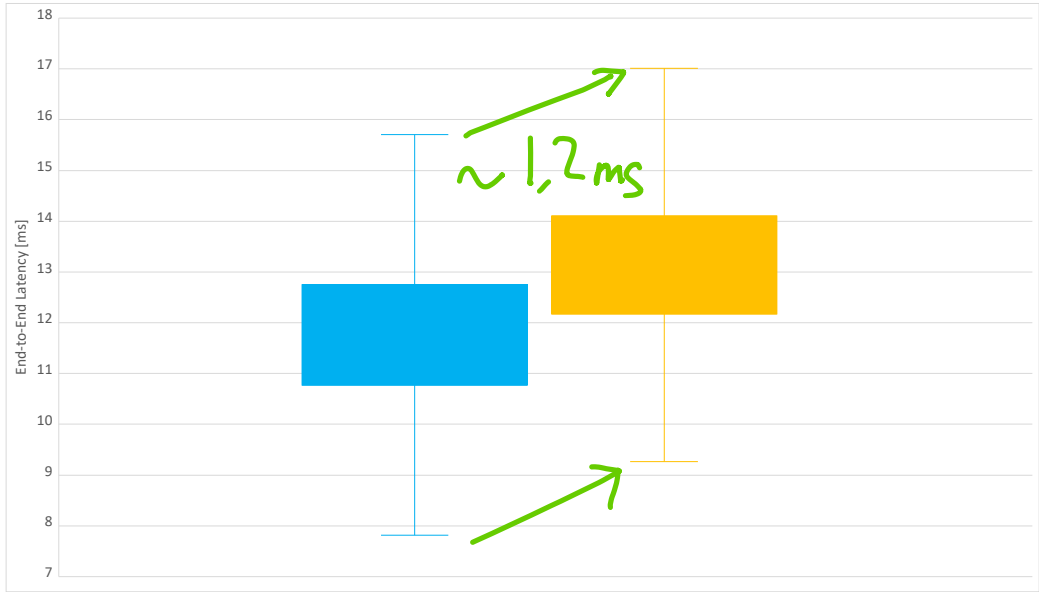


Cisco

3L Latency

Src: NYC | Dst: PHO | Flex-Algo: All | Timescale: Last period | Submit

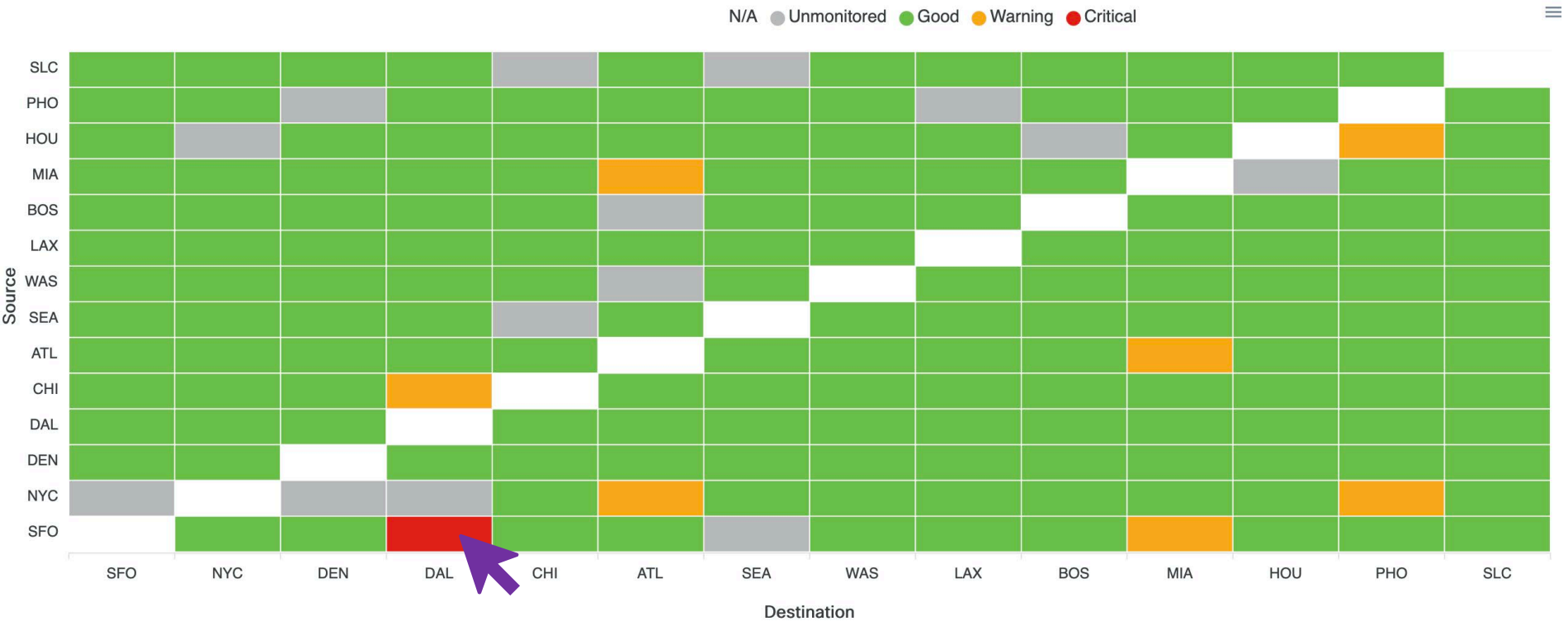




Cisco

3L Latency

Last period | Last stable topology



Flex-Algo 0

DSCP 0

Warning Threshold +10% of expected latency

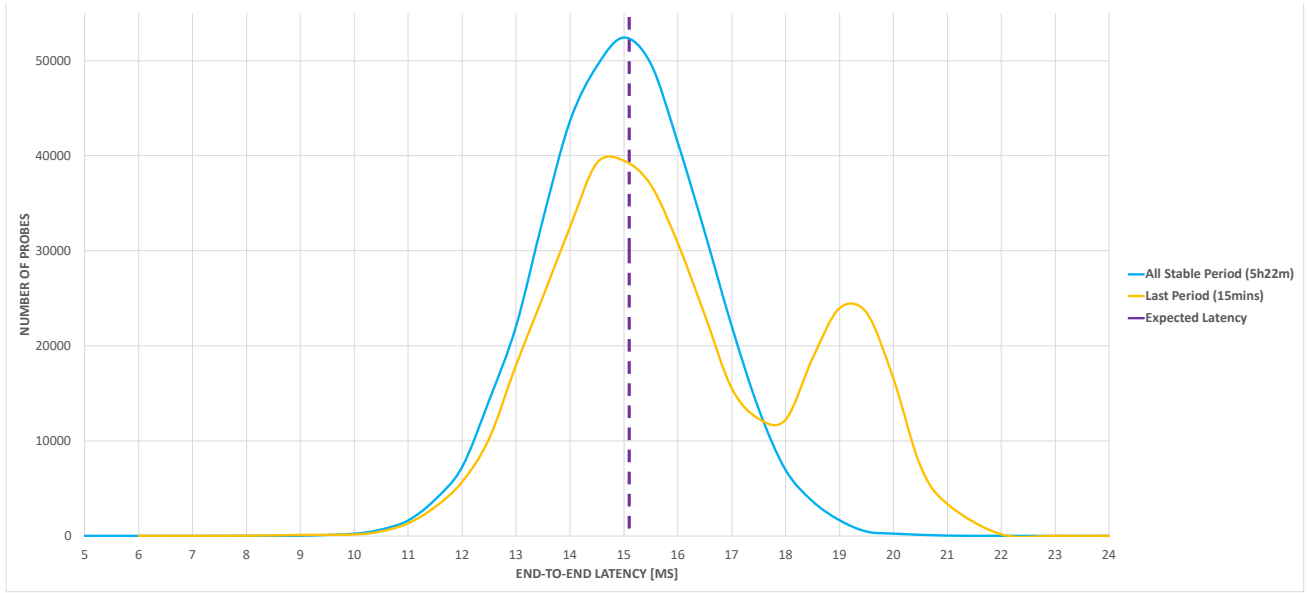
Critical Threshold: +20% of expected latency

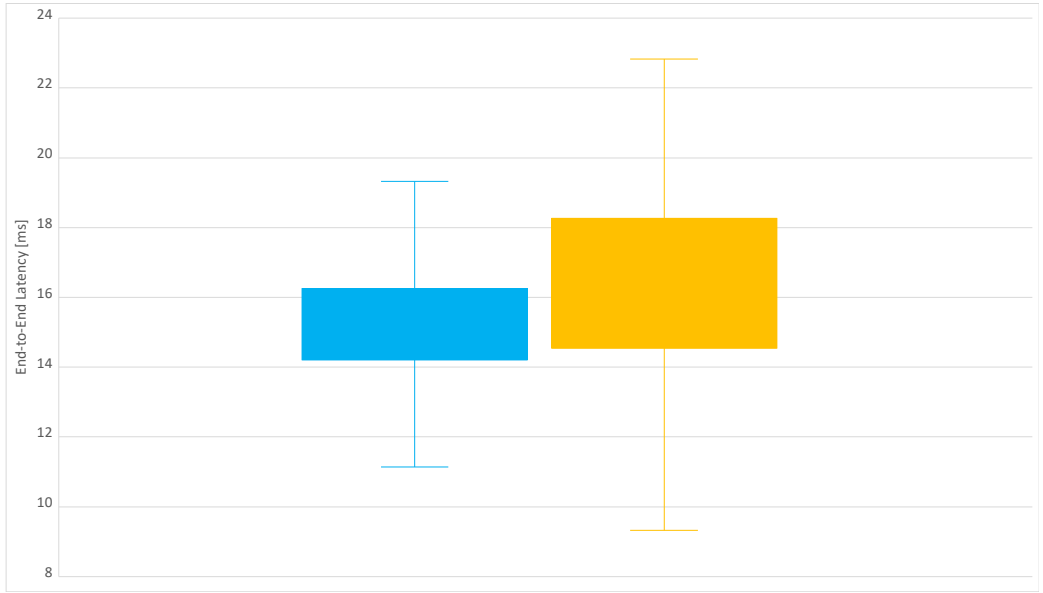
1% (1)	4% (7)	88% (160)	7%(14)
Critical	Warning	Good	Unmonitored

Cisco

3L Latency

Src: SFO | Dst: DAL | Flex-Algo: All | Timescale: Last period | Submit

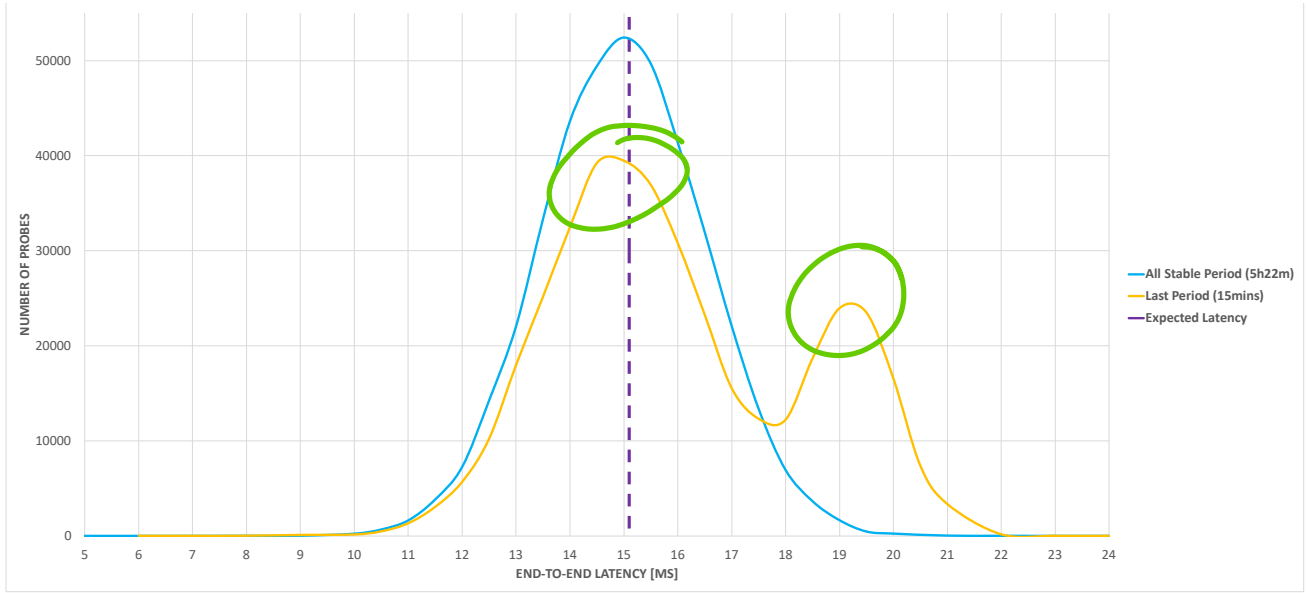


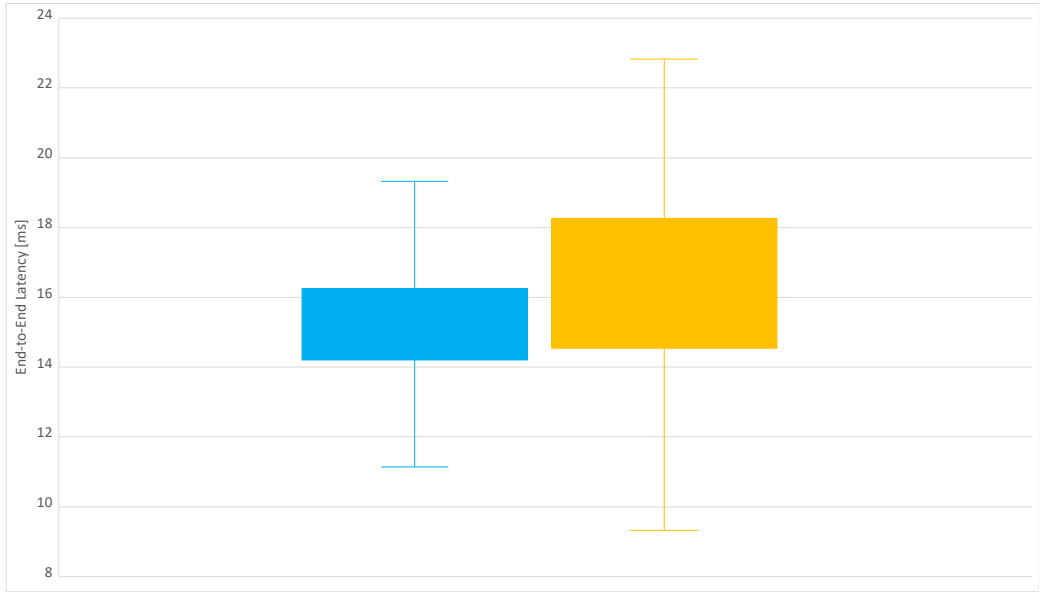


Cisco

3L Latency

Src: SFO | Dst: DAL | Flex-Algo: All | Timescale: Last period | Submit

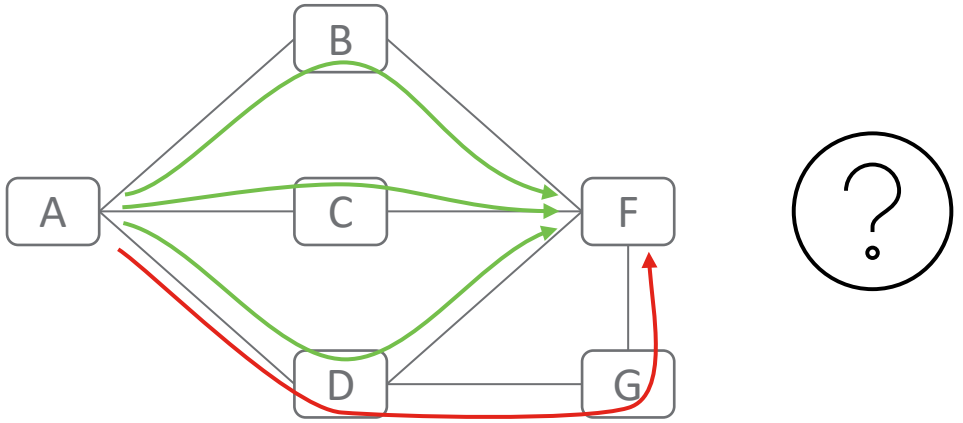




Path Tracing:
unleashing underlay visibility

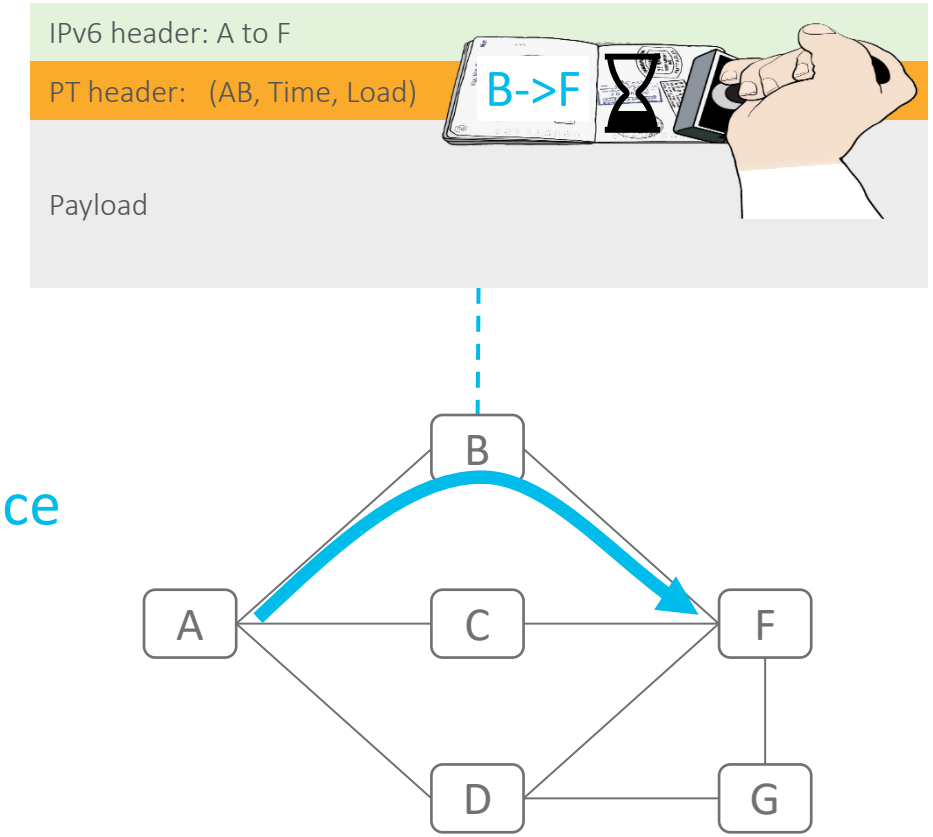
How did the packet arrive from A to F?

- 3 possible “valid” ECMP paths
  - Any drop?
  - End-to-End Latency homogeneity?
- An invalid path is possible
  - Routing or FIB corruptions
- 40-year-old unsolved IP problem



Stamping Trajectory in PT Header

- Each transit router records in PT header:
  - Outgoing interface ID
  - Timestamp (with 60µs accuracy)
  - Egress Queue Load
- Highly compressed for low MTU overhead
  - Only 3 bytes per hop!
- Implemented at linerate: Reports true packet experience
- Native interworking with legacy nodes
  - Seamless deployment
- Hardware/XR feature with analytics app



Mature Eco-System

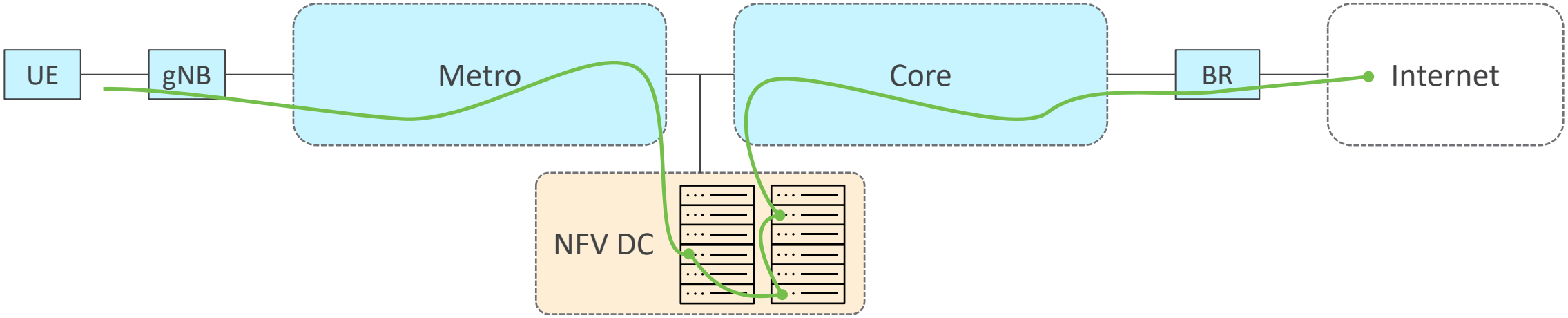
- PT Midpoint Shipping - IOS XR 7.8.1
  - Cisco 8000 (Silicon One Q200; native SDK)
  - NCS5700 (DNX2 - J2; native SDK)
  - ASR9000 (LS)
- Rich Eco-system
  - Cisco, Broadcom, Marvell, +others
  - Linux, FD.io VPP, P4, Wireshark, TCPDUMP
  - SAI/SONiC in progress
- Ongoing standardization
  - draft-filsfils-spring-path-tracing

ECMP Analytics

- Detects
  - Blackholing paths
  - An ECMP path that is not expected (routing/dataplane corruption)
  - Incoherent latency between ECMP paths
- EDM measures
  - End-to-end latency of each path (60µsec in WAN, 200ns in DC)
- Current technique of sending probes from anywhere to anywhere without any PT data requires AI processing of huge data sets

Demo available

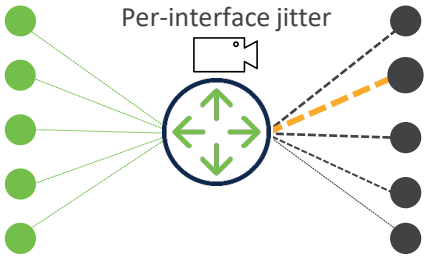
NFV: Latency Analytics and Proof of Transit

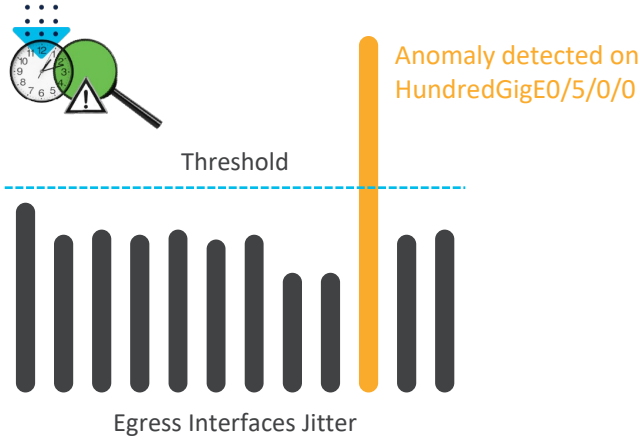


- Deterministic confirmation of NFV processing
- Deterministic latency measurement of the NFV processing
  - And/or the detour to the DC to get the NFV processing (e.g., MUP use-case)

Jitter Analytics

- ECMP Analytics probing created an extensive dataset
  - Dataplane Timestamps at each hop
  - 60µs accuracy in the WAN (200nsec in DC)
- Jitter Analytics studies this dataset on a per-node/per-intf
  - Jitter introduced by that node and egress interface
  - Min, Avg, Per50, Per80, Per90…
  - Across different queues
  - AI-based Alerts
- Per-Interface Jitter at 60µsec in live network has never been done before



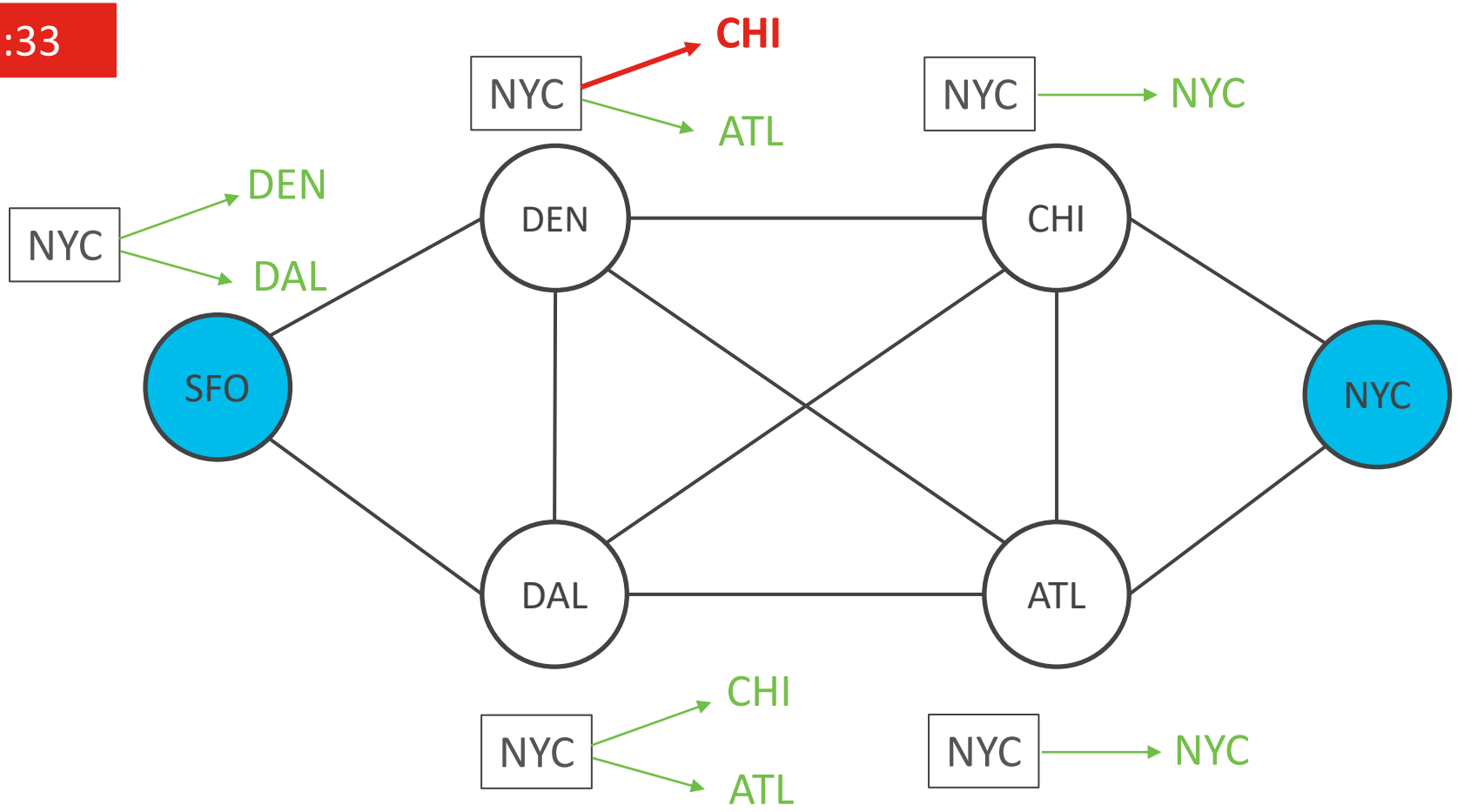


PT ECMP Analytics Demo

PT Demo use case 1 :
Blackholing paths

Corruption causing blackholes

- Hardware corruption on DEN
  - Traffic to NYC via CHI is dropped.

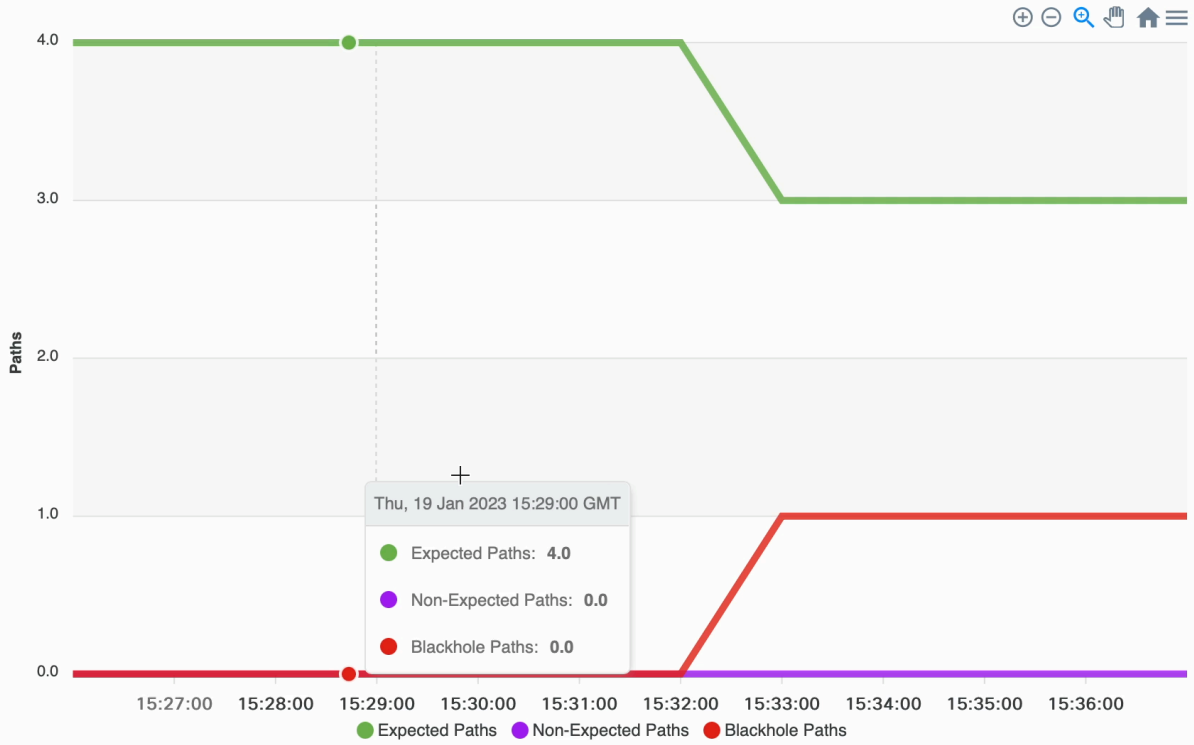


ECMP Analytics SFO -> NYC

25% of ECMP paths are blackholing.

Cisco Path Tracing

ECMP Paths monitoring



ECMP events

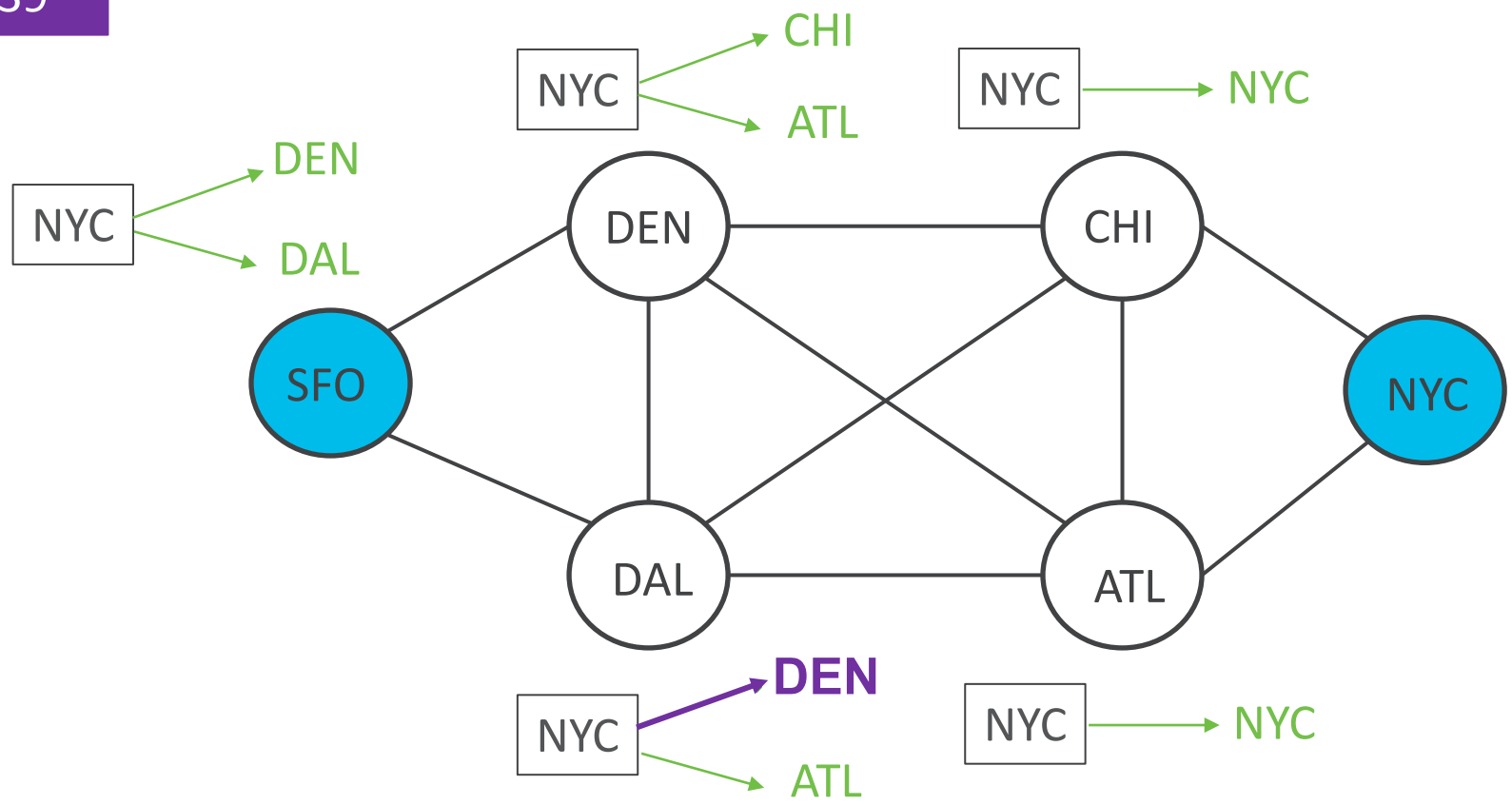
Thu, 19 Jan 2023 15:33:00 GMT — 1 Blackhole paths.

PT Demo use case 2 :
Non-expected / Wrong path

Corruption causing non-expected path

- Hardware/FIB corruption on DAL.
  - Traffic to NYC is taking a wrong path.

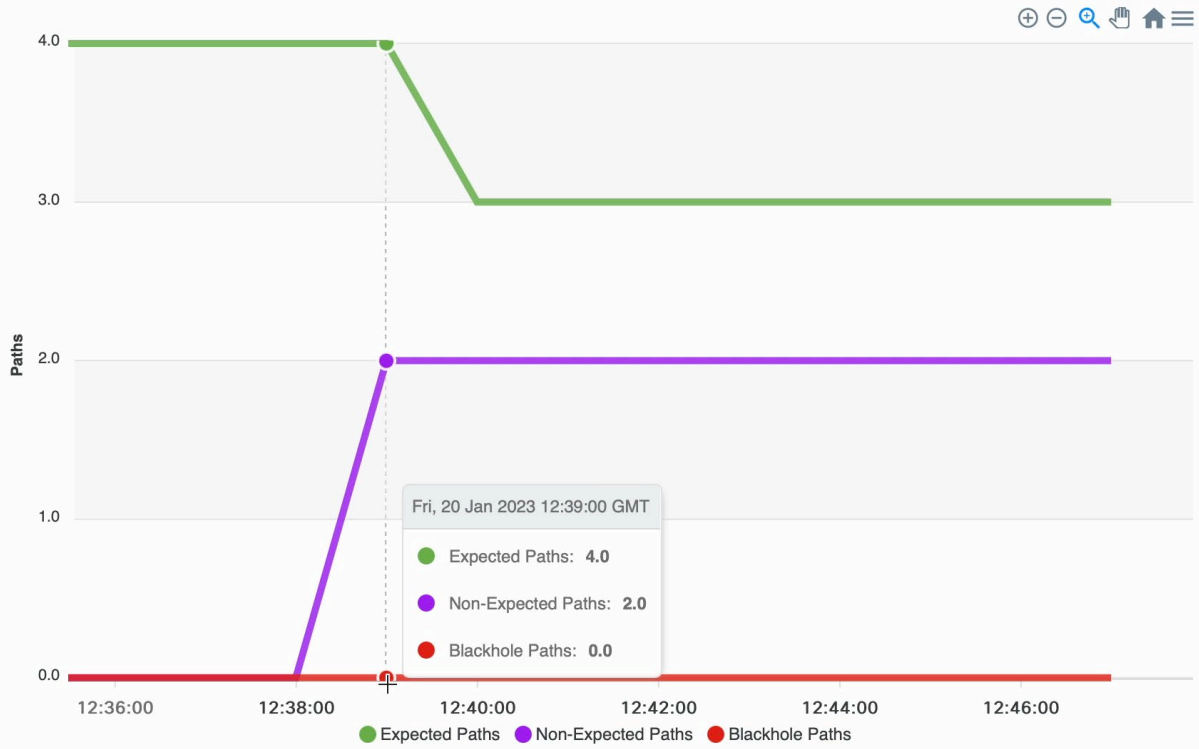
At 12:39



ECMP Analytics SFO -> NYC

Cisco Path Tracing

ECMP Paths monitoring



ECMP events

Fri, 20 Jan 2023 12:39:00 GMT	2 Non Expected paths.
Fri, 20 Jan 2023 12:31:00 GMT	0 Non Expected paths.
Fri, 20 Jan 2023 12:22:00 GMT	2 Non Expected paths.
Fri, 20 Jan 2023 12:22:00 GMT	End of an unstable topology period.
Fri, 20 Jan 2023 12:21:00 GMT	Start of an unstable topology period.
Fri, 20 Jan 2023 12:17:00 GMT	2 Non Expected paths.
Fri, 20 Jan 2023 12:17:00 GMT	End of an unstable topology period.
Fri, 20 Jan 2023 12:16:00 GMT	Start of an unstable topology period.
Fri, 20 Jan 2023 08:58:00 GMT	2 Non Expected paths.
Fri, 20 Jan 2023 08:54:00 GMT	End of an unstable topology period.
Fri, 20 Jan 2023 08:53:00 GMT	Start of an unstable topology period.
Fri, 20 Jan 2023 08:50:00 GMT	1 Blackhole paths.

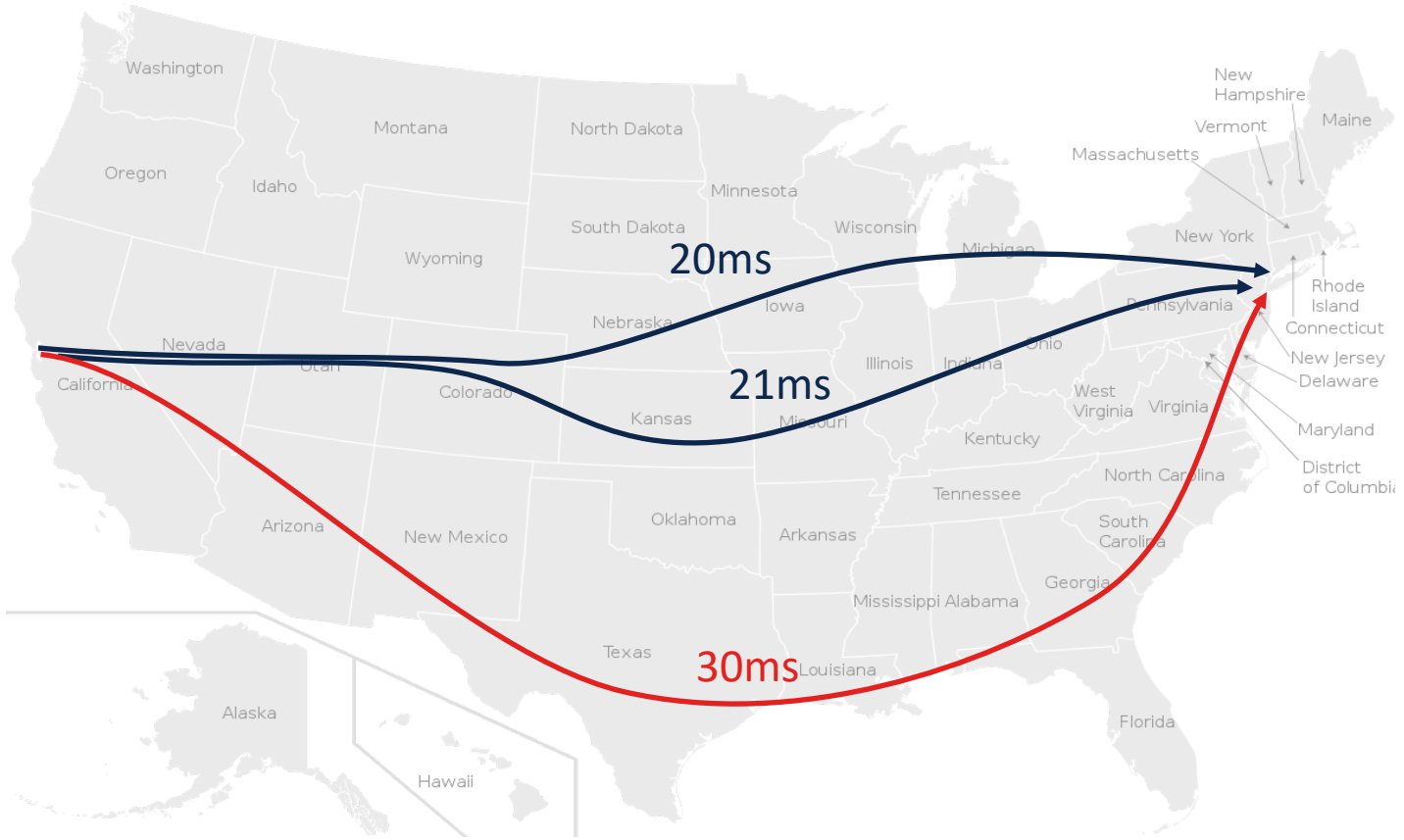
PT Demo use case 3 :
Non-homogeneous latency

Cisco SRservices

Path Tracing

Last period | From/To Time

Source SFO Destination NYC Submit



Conclusion

Conclusion

- SRv6 uSID integrated solution
  - End-to-end service creation
  - Measurements
  - Analytics
- Un-matched measurements hardware capability:
  - Un-matched performance
  - Un-matched accuracy
  - Un-matched economics (Silicon integration)
  - Un-matched coverage (per ECMP)
  - Un-matched visibility of the underlay fabric (per hop)
- Un-matched analytics:
  - Intelligent data. No point in raw/brute data.
  - Data correlated with routing information

The bridge to possible