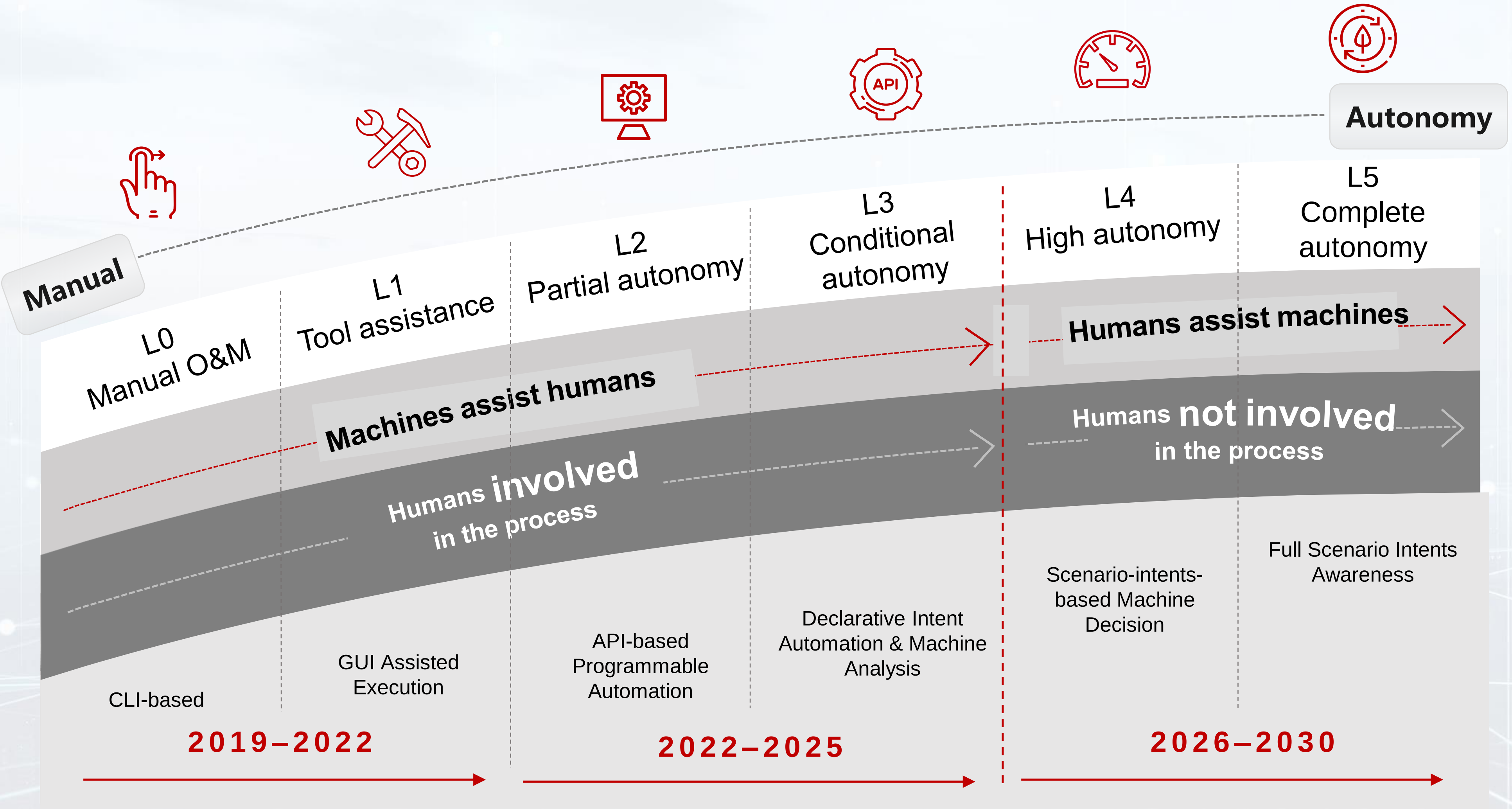
Measuring Network Latency – Latest News from the IETF

Benoît Claise
Data Communication Intelligent Operations & Management CTO
Huawei

Understanding Latency, Dec 2024

Latency, why?

Realize the Autonomous Driving Networks (ADN) Vision



Autonomous Driving Networks, Why?
Networks are too Expensive to Operate

- **It’s always the network fault!**
- **Even when it’s not**

The (network) complexity (we loved) is (on the verge of) turning against us

- **Business issue: how are the customer services impacted?**

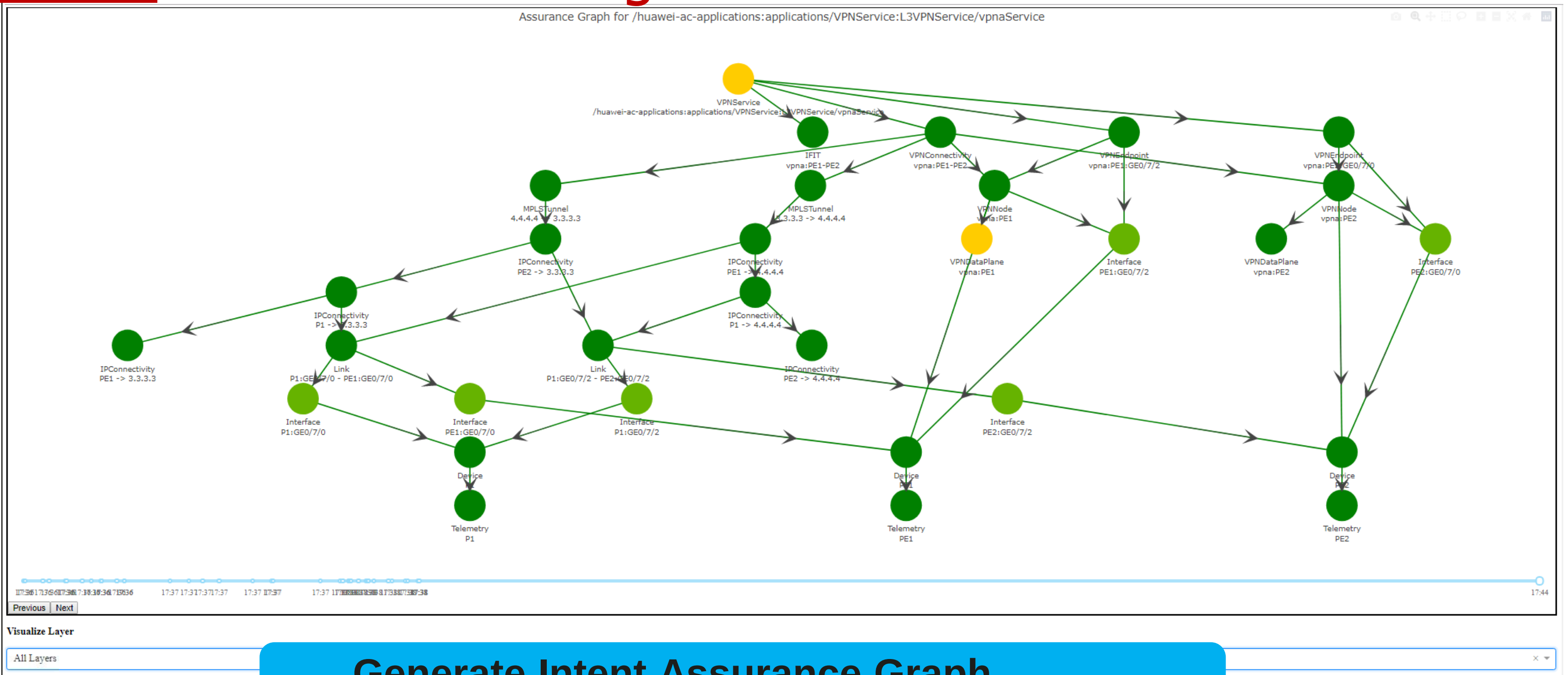
Next Generation Architecture:
Service Assurance for Intent-based Networking

Key challenge

- AI/ML without network intent does not work for ADN
- How to determine the intent?
- How to map between the intent/service and collected metric?

Value

- When a service degrades, where are the faults, what are the symptoms?
- What services are impacted by a network fault?



Generate Intent Assurance Graph

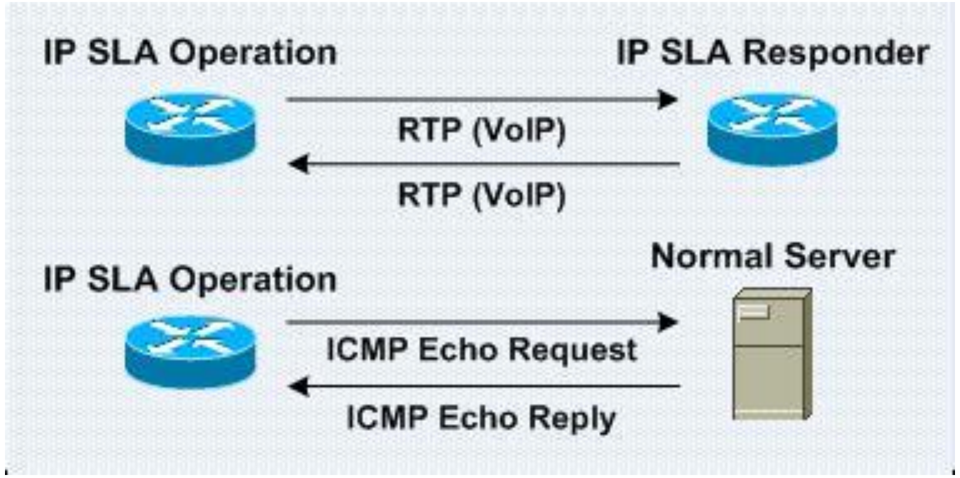
- Static intent assurance graph generated from info from controller(s) + domain knowledge
- Dynamic part generated from info from the network + domain knowledge

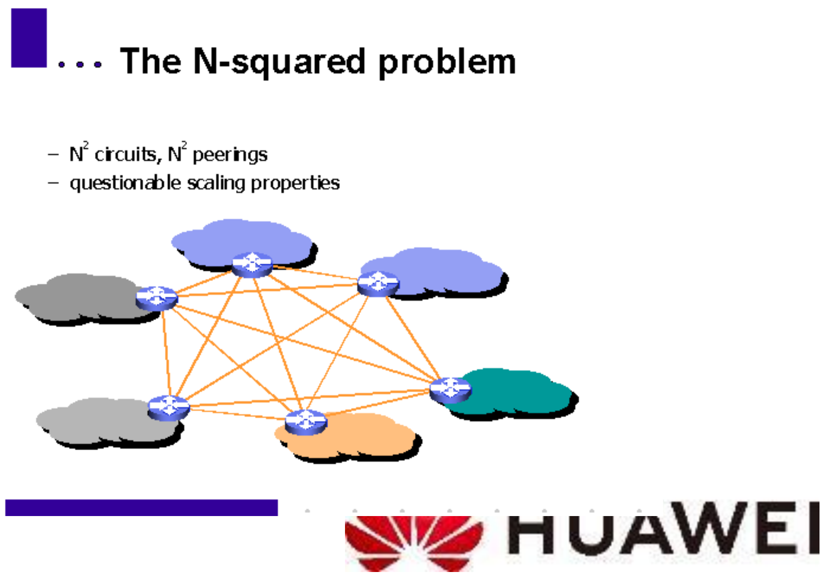
- RFC 9417 Service Assurance for Intent-Based Networking Architecture
- RFC 9418 A YANG Data Model for Service Assurance

HUAWEI

Measuring Networking Latency – Active Probing

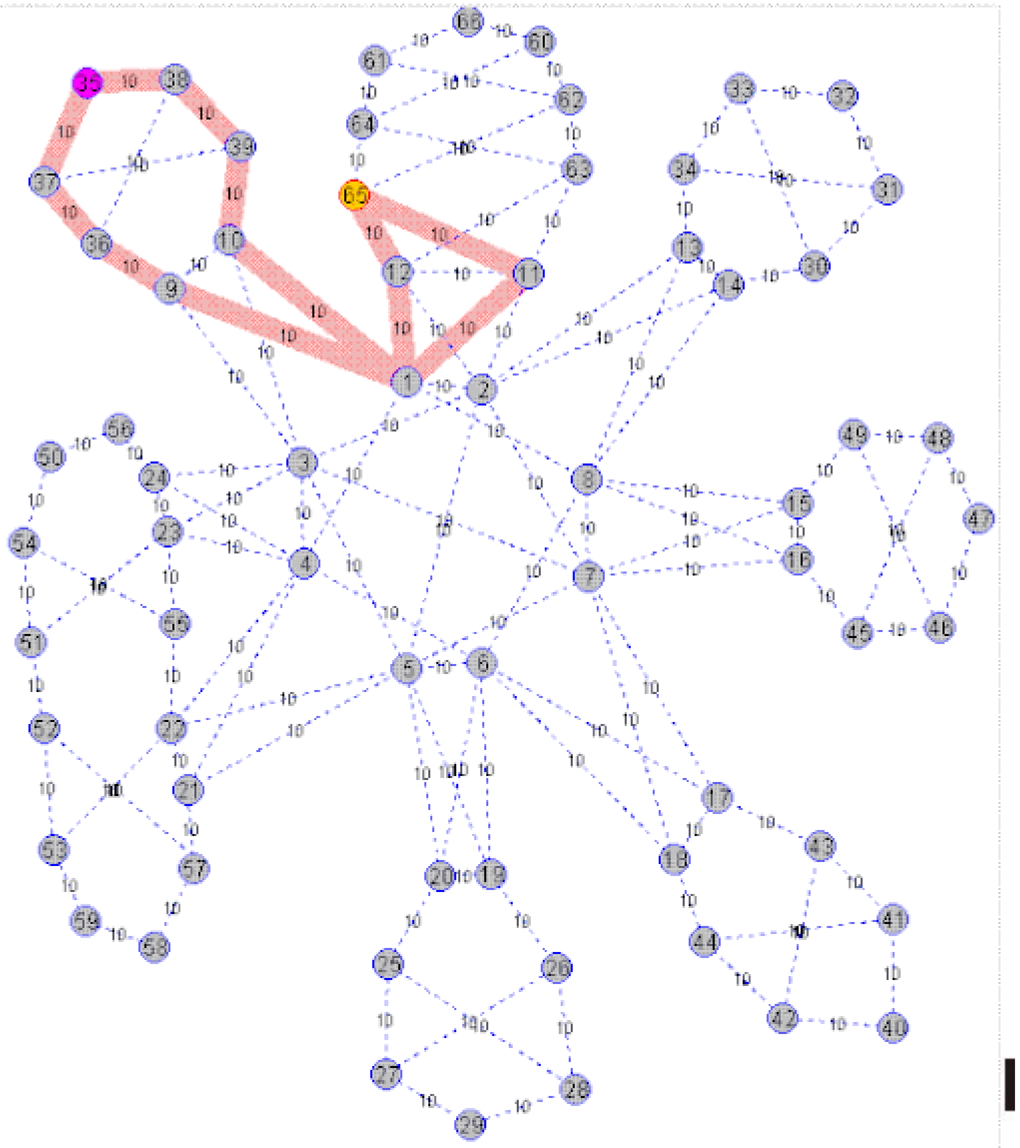
- IP SLA (Cisco Systems)
  - Sending synthetic traffic
- Issues:
  - Is this really the customers/service/application traffic?
  - N-square problems (From each PoP, PE router, etc)
  - Heavy from an operational point of view
  - Accuracy? Process switching in router
- Created shadow routers
  - Dedicated for IP SLA, for each PoP
  - That helped a little but …





The Biggest Issue with Active Probing: ECMP

- Equal Cost Multiple Path
- Which path is taken?
  - By the customer traffic
  - By the active probing
- Unless we know the hash function, we never really know!



IP Performance Metric IETF Working Group

- https://datatracker.ietf.org/wg/ippm/charter/

- Metrics:
  - A One-Way Delay Metric for IP Performance Metrics (RFC 7679)
  - But also one packet delay variation, loss, etc.

- OWAMP & TWAMP
  - A One-way Active Measurement Protocol (OWAMP) (RFC 4656)
  - A Two-way Active Measurement Protocol (TWAMP) (RFC5357)
  - Same issues as IP SLA

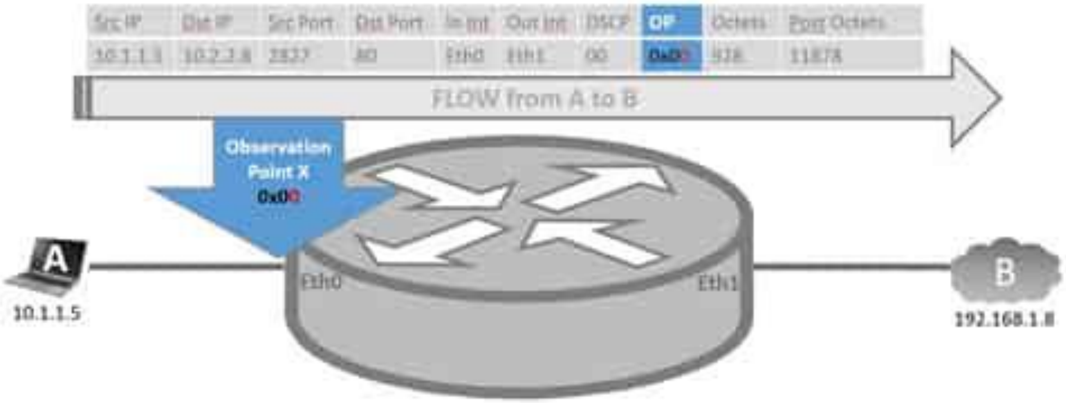
- Active and Passive Metrics and Methods (with Hybrid Types in-Between) RFC 7799
    - Some methods may use a subset of active and passive attributes => we refer to those as "Hybrid Methods"

HUAWEI

IP Flow Information eXport (IPFIX)

- NetFlow (RFC 3954) to IPFIX (RFC7011), for data plane monitoring

Netflow vs IPFIX

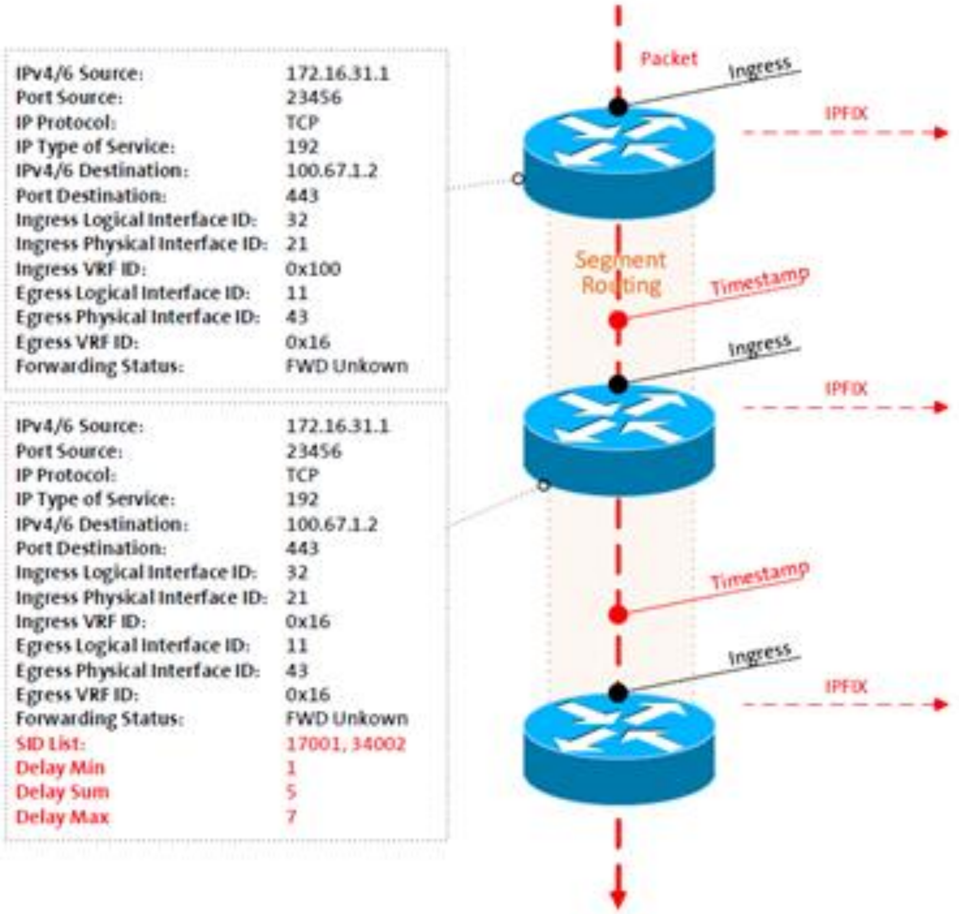


- Ingress vs egress, sampling or not, technology specific fields, etc.

What’s Missing? Delay in IPFIX, with ALT-Marking (RFC9341)

Inband Telemetry IPFIX Flow-Aggregation – Add Delay Dimensions

Minimum, Average and Maximum



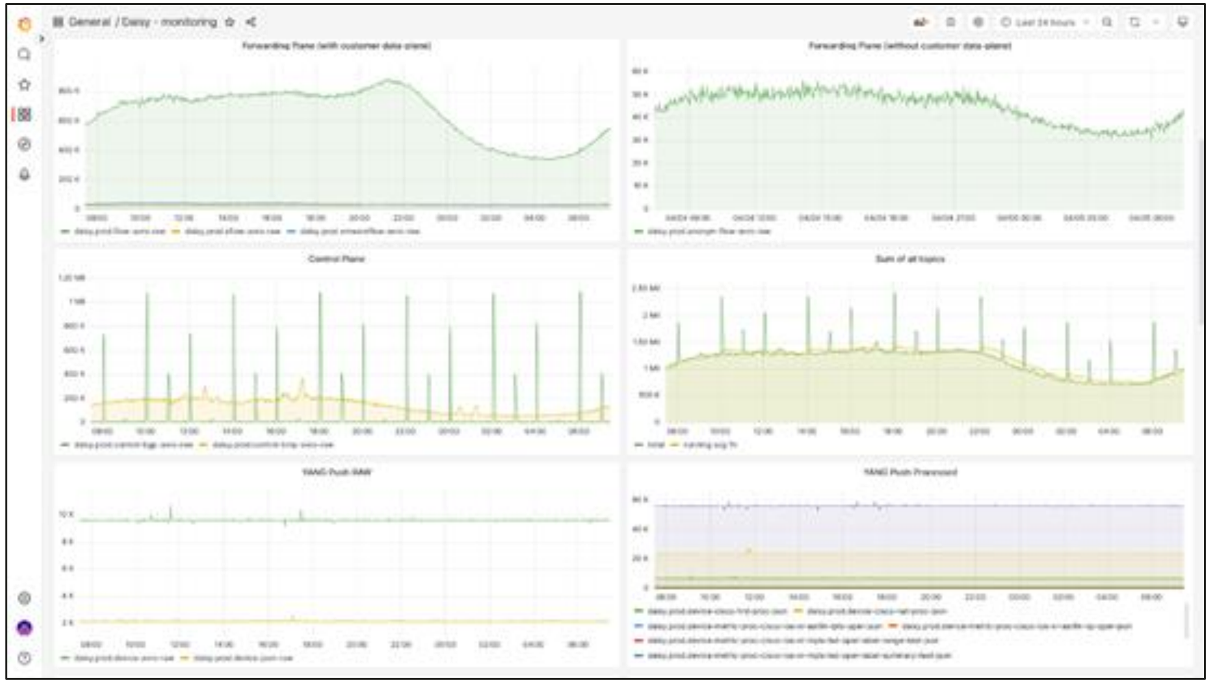
- IPFIX flow sampler, access-list matching TOS bit or probe configuration determine wherever packet should be traced or not.
- Timestamp and direct export flag is set in the packet when enters Segment Routing domain and forwards with the customer packet.
- Triggered by the direct export flag, each node measures at ingress the delay to the timestamp and populates it in three additional delay dimensions into the IPFIX flow cache for each flow ID.
- The delay_sum is **summed** with every packet being added to the flow ID.
- The delay_min is updated if a packet has a **lower** delay than previously measured for a flow ID.
- The delay_max is updated if a packet has a **higher** delay than previously measured for a flow ID.
- The flow-aggregation (node or/and data-collection) averages delay_sum at export.

HUAWEI

draft-ietf-opsawg-ipfix-on-path-telemetry, test images available

Accuracy and Scale?

- IPFIX on every ingress interface, of every single router
- Measuring the delay for <u>specific</u> flows at ingress and in the core
- It scales: peaking at 2.5 million kafka messages per second with an average close to 1.5 million



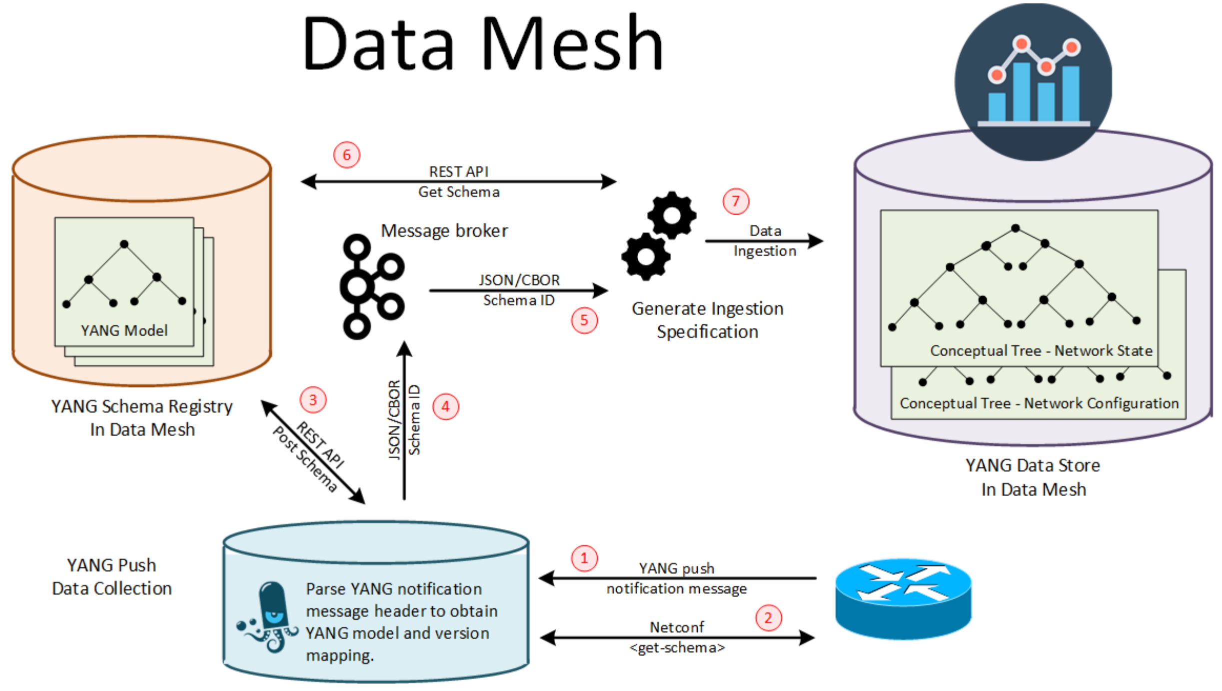
- NTP Interleaved Nodes, just approved
- <u>draft-ietf-ntp-interleaved-modes/</u>

```
Technical Summary

   This document extends the specification of Network Time Protocol
   (NTP) version 4 in RFC 5905 with special modes called the NTP
   interleaved modes, that enable NTP servers to provide their clients
   and peers with more accurate transmit timestamps that are available
   only after transmitting NTP packets.  More specifically, this
   document describes three modes: interleaved client/server,
   interleaved symmetric, and interleaved broadcast.
```

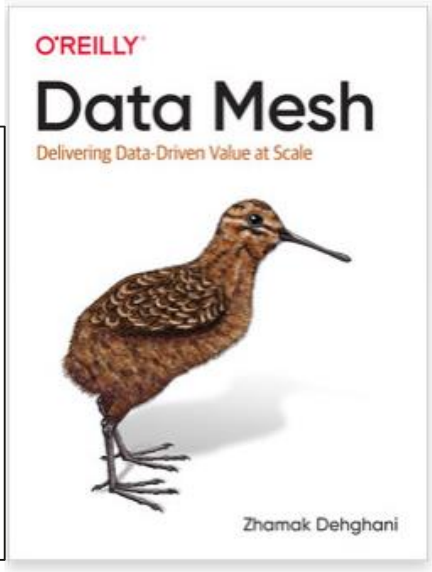
HUAWEI

Next-gen Network Analytics Collection, based on the Data Mesh Principles



- For IPFIX, for data plane
- For YANG-Push, for management plane
- BGP Monitoring Protocol, for control plane

Data Mesh is founded in four principles:
- domain-driven ownership of data,
- data as a product,
- self-serve data platform
- federated computational governance.



https://www.claise.be/yang-push-apache-kafka-semantic-network-visibility-for-analytics/
https://github.com/graf3net/draft-daisy-kafka-yang-integration/blob/main/draft-daisy-kafka-yang-integration-05.md

End to End? Segment Routing V6

IPFIX Covering Segment Routing

For MPLS-SR and SRv6

Segment Routing coverage in IPFIX bring visibility for:

- Which routing protocol provided the label in MPLS-SR.
- The IPv6 Segment where the packet is forwarded to in SRv6.
- The IPv6 Segments where the packet is going to be forwarded through in SRv6.

- **Export of MPLS Segment Routing Label Type Information in IPFIX**

  RFC 9160
- **Export of Segment Routing IPv6 Information in IPFIX**

  RFC 9487

RFC9160 released in December 2021. Complement the existing MPLS data-plane with Segment Routing dimensions. draft-tgraf-opsawg-ipfix-srv6-srh version 00 submitted in January 2022 presented IETF 113. Complement the existing IPv6 data-plane with Segment Routing dimensions.

```
> Frame 527: 182 bytes on wire (1456 bits), 182 bytes captured (1456 bits)
> Ethernet II, Src: Cisco_ea:ad:1c (00:32:17:ea:ad:1c), Dst: Vmware_21:95:d2 (00:0c:29:21:95:d2)
> Internet Protocol Version 4, Src: 138.187.57.63, Dst: 138.187.58.13
> User Datagram Protocol, Src Port: 44542, Dst Port: 9991
v Cisco NetFlow/IPFIX
    Version: 9
    Count: 1
    SysUptime: 516154.381000000 seconds
  > Timestamp: Feb 23, 2020 13:57:18.000000000 W. Europe Standard Time
    FlowSequence: 23685
    SourceId: 0
  v FlowSet 1 [id=313] (1 flows)
      FlowSet Id: (Data) (313)
      FlowSet Length: 120
      [Template Frame: 9]
    v Flow 1
      > MPLS-Label1: 17002 exp-bits: 0
      > MPLS-Label2: 24622 exp-bits: 0 bottom-of-stack
      > MPLS-Label3: 0 exp-bits: 0
      > MPLS-Label4: 0 exp-bits: 0
      > MPLS-Label5: 0 exp-bits: 0
      > MPLS-Label6: 0 exp-bits: 0
        InputInt: 87
        OutputInt: 111
        Octets: 216000
        Packets: 2000
      > [Duration: 5.753000000 seconds (switched)]
        TopLabelAddr: 138.187.57.13
        SrcAddr: ::
        DstAddr: ::
        ipv6FlowLabel: 0
        IPv6 Extension Headers: 0x00000000
        SrcAddr: 10.248.4.236
        DstAddr: 10.248.4.222
        SrcPort: 0
        DstPort: 2048
        MPLS Top Label Prefix Length: 32
        TopLabelType: LDP (5)
      > Forwarding Status
        Direction: Ingress (0)
        IP ToS: 0x00
        Protocol: ICMP (1)
      > TCP Flags: 0x00
        SamplerID: 1
        Ingress VRFID: 1610612736
        Egress VRFID: 1610612736
      Padding: 0000
```

Conclusion: Measuring Network Latency

- Believe in
  - Hybrid methods (active and passive)
  - IPFIX: Exporting the delay directly from the routers
  - ALT Marking
  - Linking the information with the customer service
  - SRv6
- As key building block to solve the Autonomous Driving Networks

Measuring Network Latency – Latest News from the IETF

Benoît Claise
Data Communication Intelligent Operations & Management CTO
Huawei

Understanding Latency, Dec 2024