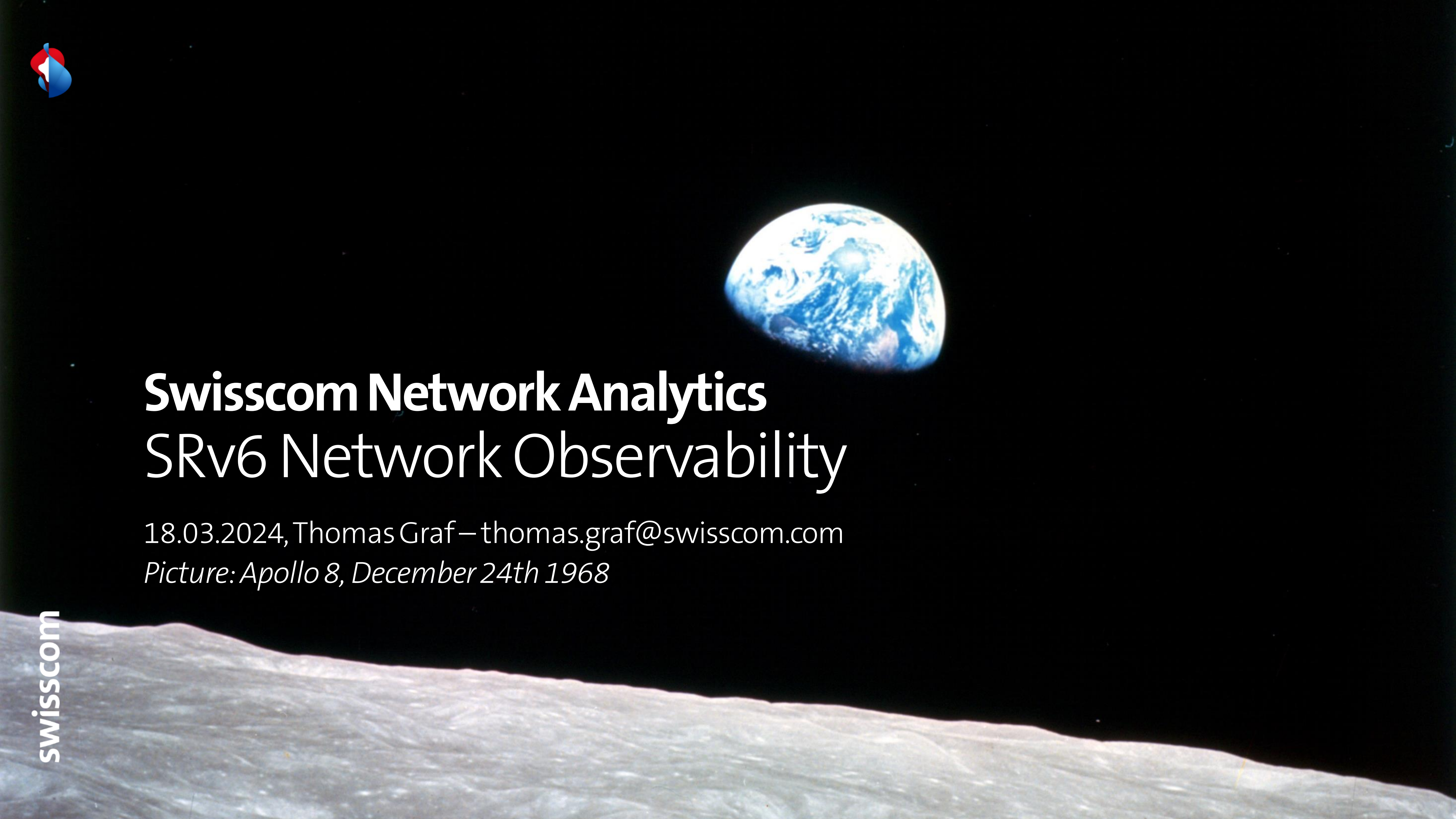


Swisscom Network Analytics

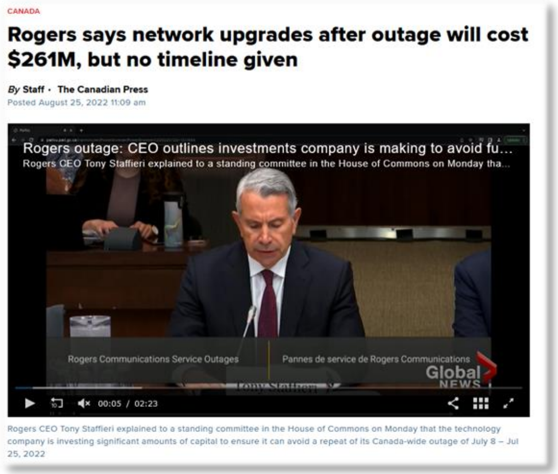
SRv6 Network Observability

18.03.2024, Thomas Graf – thomas.graf@swisscom.com

Picture: Apollo 8, December 24th 1968

Nationwide Network Outages everywhere

Increasing in impact and duration - hinting Network Visibility deficiencies



CANADA

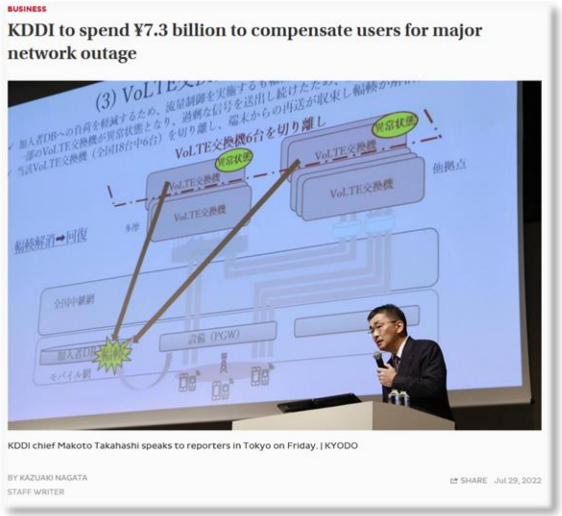
Rogers says network upgrades after outage will cost $261M, but no timeline given

By Staff · The Canadian Press
Posted August 25, 2022 11:09 am

Rogers outage: CEO outlines investments company is making to avoid fu...
Rogers CEO Tony Staffieri explained to a standing committee in the House of Commons on Monday tha...

Rogers Communications Service Outages
Pannes de service de Rogers Communications
Global NEWS
00:05 / 02:23

Rogers CEO Tony Staffieri explained to a standing committee in the House of Commons on Monday that the technology company is investing significant amounts of capital to ensure it can avoid a repeat of its Canada-wide outage of July 8 – Jul 25, 2022



BUSINESS

KDDI to spend ¥7.3 billion to compensate users for major network outage

KDDI chief Makoto Takahashi speaks to reporters in Tokyo on Friday. | KYODO

BY KAZUAKI NAGATA
STAFF WRITER

SHARE Jul 29, 2022

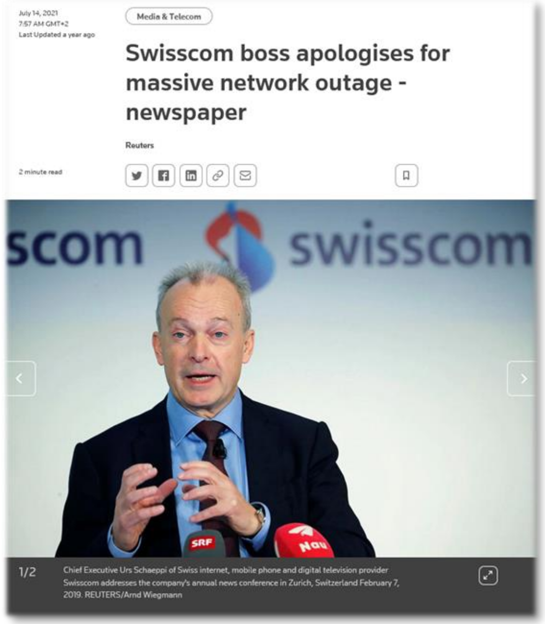
ORANGE FRANCE UNDER FIRE FOR MISHANDLING NETWORK OUTAGE

Posted by Harry Baldock | Jul 22, 2021 | Subsea, INFRASTRUCTURE, Satellite, Towers, COMPANY NEWS, Governance, Data Centres, Networks, Wholesale, Virtualisation, Europe, Middle East & Africa, News

Optus: Telecom boss Kelly Bayer Rosmarin quits after Australian outage

6 days ago

The firm has come under fire following a nationwide network outage this month



July 14, 2021
7:57 AM GMT+2
Last Updated a year ago

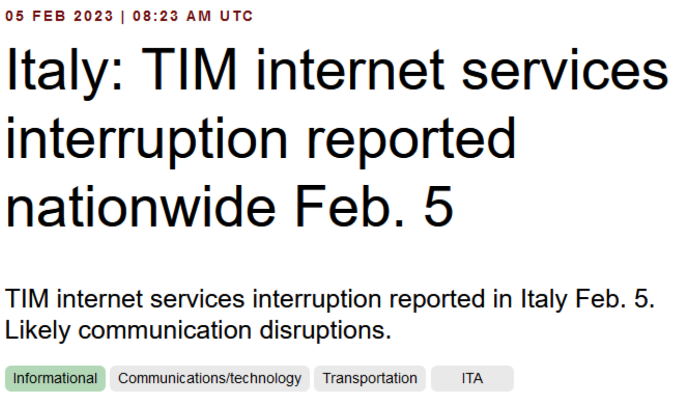
Media & Telecom

Swisscom boss apologises for massive network outage - newspaper

Reuters

2 minute read

1/2 Chief Executive Urs Schaeppi of Swiss internet, mobile phone and digital television provider Swisscom addresses the company's annual news conference in Zurich, Switzerland February 7, 2019. REUTERS/Arnd Wiegmann



05 FEB 2023 | 08:23 AM UTC

Italy: TIM internet services interruption reported nationwide Feb. 5

TIM internet services interruption reported in Italy Feb. 5. Likely communication disruptions.

Informational | Communications/technology | Transportation | ITA

Facebook outage: what went wrong and why did it take so long to fix after social platform went down?

Billions of users were unable to access Facebook, Instagram and WhatsApp for hours while the social media giant scrambled to restore services

Facebook, Instagram and WhatsApp all went down, and reappeared online after a six-hour global outage. Photograph: Anadolu Agency/Getty Images

“ Swisscom would not have chosen SRv6 if data plane visibility wasn't implemented at day 1 of the production deployment “

Network Analytics Transforms Swisscom DevOps Mindset

From device monitoring to network analytics with closed loop operation

2015-2016

Flow Aggregation Proof of Concept
Internet Distribution Core and TV 2.0

2017-2018

Swisscom Big Data onboarded,
Meerkat Anomaly Detection Feasibility
10 active users. 9 platforms. 87 nodes. 250'000 metrics per seconds.

2019

BGP Monitoring Protocol and YANG Push
IETF Engagement started
40 active users. 17 platforms. 233 nodes. 1'200'000 metrics per second.

2020

Pivot Migration, Druid Scale Out,
IETF collaboration established
160 active users. 34 platforms. 2500 nodes. 3'000'000 metrics per second. Active probing with 1'500'000 broadband subscribers.

2021

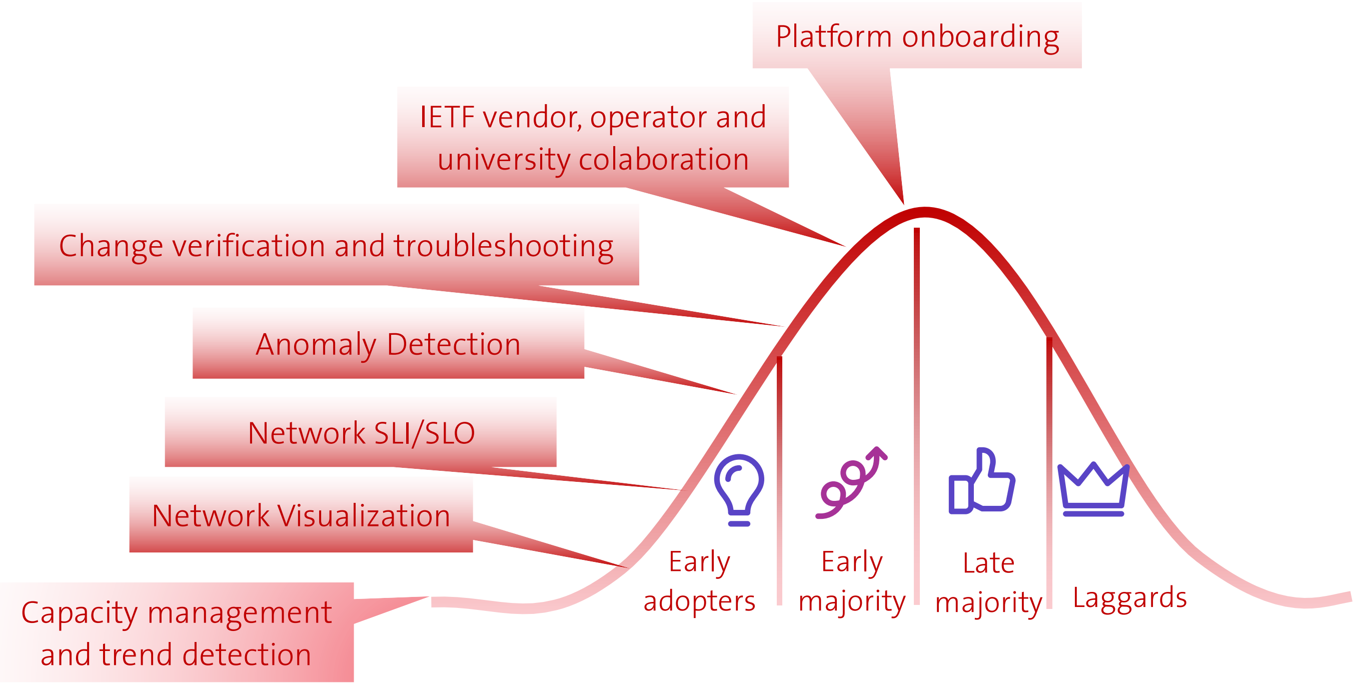
Taking over end to end Daisy Chain Responsibility
215 active users. 40 platforms. 2700 nodes. 20'000'000 metrics per second.

2022

L3 VPN Anomaly Detection Development started
400 active users. 47 platforms. 7000 nodes. 25'000'000 metrics per second.

2023

L3 VPN Anomaly Detection PoC started,
First SRv6 network onboarded
500 active users. 51platforms. 25000 nodes. 30'000'000 metrics per second.


Platform onboarding

IETF vendor, operator and university colaboration

Change verification and troubleshooting

Anomaly Detection

Network SLI/SLO

Network Visualization

Capacity management and trend detection

Key Points

- From bottom up **to mainstream**. From IETF **to Swisscom DevOps teams**.
- From network verification and troubleshooting **to visualization with Anomaly Detection and Network SLI/SLO**
- From capacity management **to trend detection**
- From network automation **to closed loop operation**
- **MPLS-SR data plane visibility since 2021, SRv6 since 2023.**

IPFIX Covering Segment Routing

For MPLS-SR, SRv6 and On-path Delay

SRv6 is commonly standardized, network vendors implementations are available and network operators are at various stages in their deployments, missing data-plane visibility though.

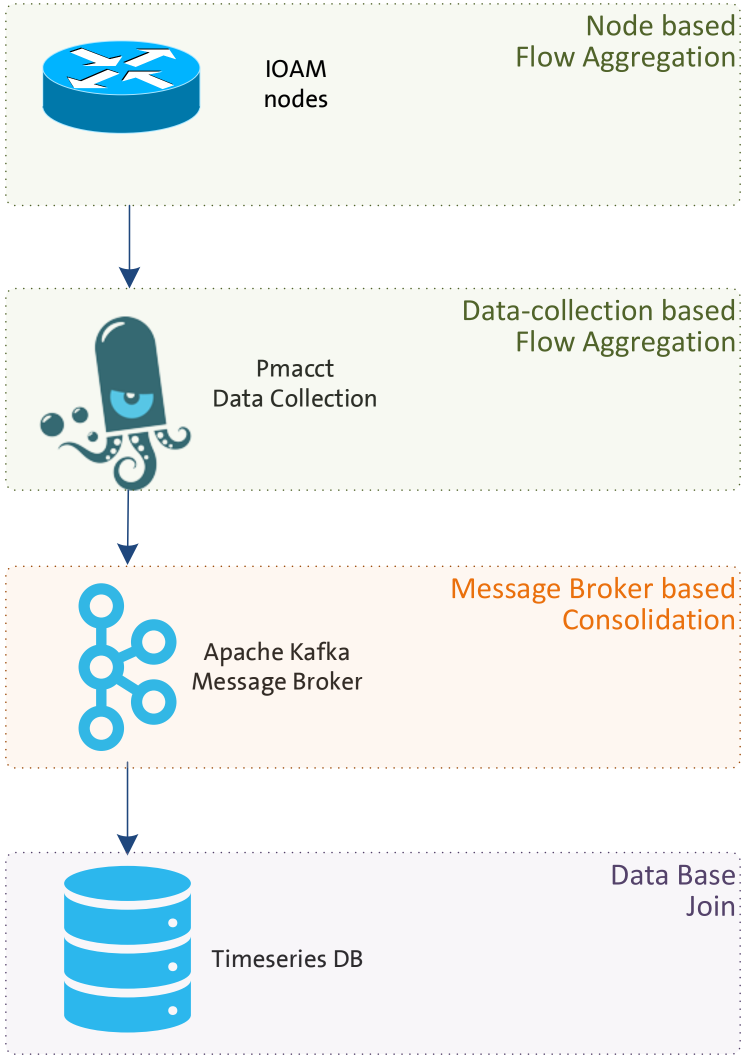
Segment Routing coverage in IPFIX brings visibility for:

- Which routing protocol provided the label or IPv6 Segment in the SR domain.
- The active Segment where the packet is forwarded to in the SRv6 Domain.
- The Segment List where the packet is going to be forwarded throughout the SRv6 Domain.
- The Endpoint Behavior describing how the packet is being forwarded in the SRv6 Domain.
- The Min, Max and Average On-path delay at each hop in the SR domain.

Export of MPLS Segment Routing Label Type Information in IPFIX
https://datatracker.ietf.org/doc/html/rfc9160

Export of Segment Routing IPv6 Information in IPFIX
https://datatracker.ietf.org/doc/html/rfc9487

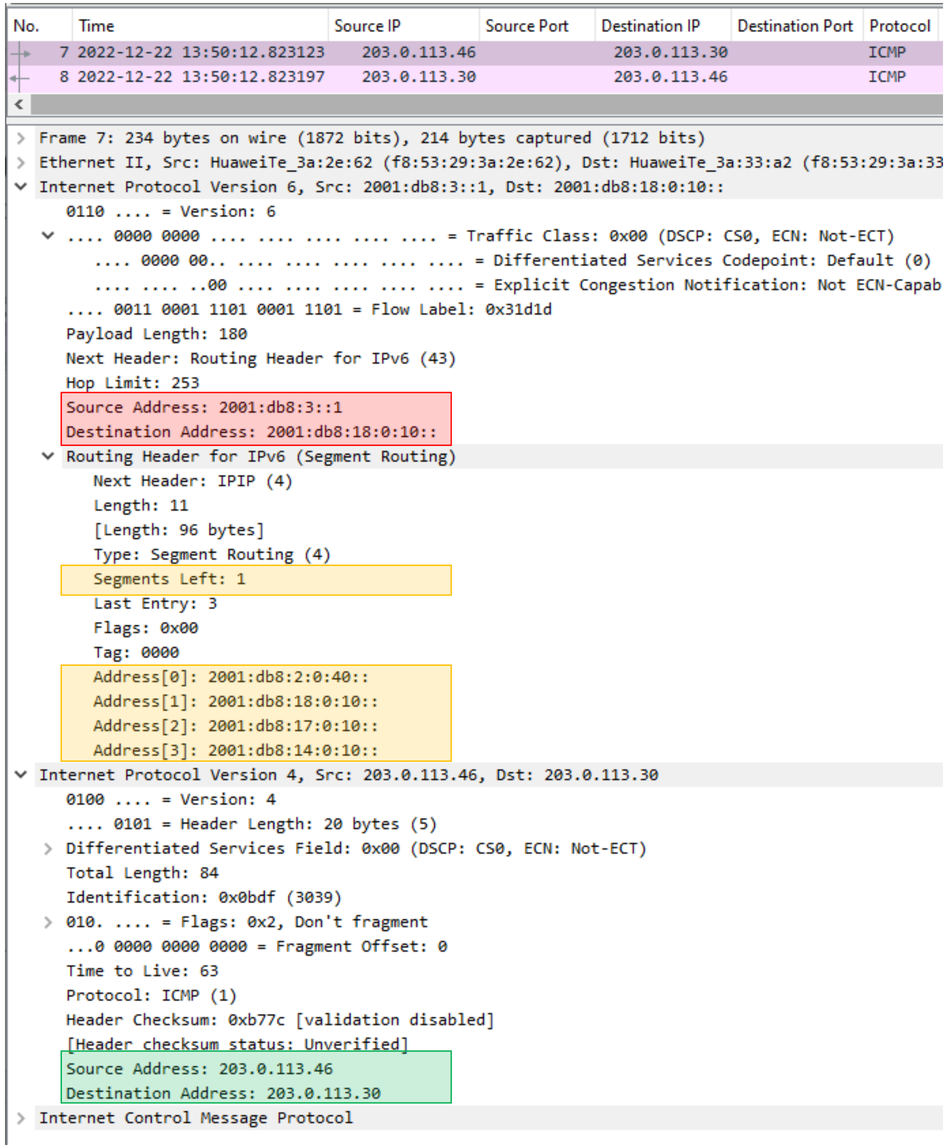
Export of Forwarding Path Delay in IPFIX
https://datatracker.ietf.org/doc/html/draft-ietf-opsawg-ipfix-on-path-telemetry



Node based Flow Aggregation
IOAM nodes
Data-collection based Flow Aggregation
Pmacct Data Collection
Message Broker based Consolidation
Apache Kafka Message Broker
Data Base Join
Timeseries DB

Segment Routing IPv6 Encapsulation

3 headers, one more then MPLS



No.	Time	Source IP	Source Port	Destination IP	Destination Port	Protocol
7	2022-12-22 13:50:12.823123	203.0.113.46		203.0.113.30		ICMP
8	2022-12-22 13:50:12.823197	203.0.113.30		203.0.113.46		ICMP

```
> Frame 7: 234 bytes on wire (1872 bits), 214 bytes captured (1712 bits)
> Ethernet II, Src: HuaweiTe_3a:2e:62 (f8:53:29:3a:2e:62), Dst: HuaweiTe_3a:33:a2 (f8:53:29:3a:33
v Internet Protocol Version 6, Src: 2001:db8:3::1, Dst: 2001:db8:18:0:10::
    0110 .... = Version: 6
  v .... 0000 0000 .... .... .... .... .... = Traffic Class: 0x00 (DSCP: CS0, ECN: Not-ECT)
      .... 0000 00.. .... .... .... .... .... = Differentiated Services Codepoint: Default (0)
      .... .... ..00 .... .... .... .... .... = Explicit Congestion Notification: Not ECN-Capab
    .... 0011 0001 1101 0001 1101 = Flow Label: 0x31d1d
    Payload Length: 180
    Next Header: Routing Header for IPv6 (43)
    Hop Limit: 253
    Source Address: 2001:db8:3::1
    Destination Address: 2001:db8:18:0:10::
  v Routing Header for IPv6 (Segment Routing)
      Next Header: IPIP (4)
      Length: 11
      [Length: 96 bytes]
      Type: Segment Routing (4)
      Segments Left: 1
      Last Entry: 3
      Flags: 0x00
      Tag: 0000
      Address[0]: 2001:db8:2:0:40::
      Address[1]: 2001:db8:18:0:10::
      Address[2]: 2001:db8:17:0:10::
      Address[3]: 2001:db8:14:0:10::
v Internet Protocol Version 4, Src: 203.0.113.46, Dst: 203.0.113.30
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 84
    Identification: 0x0bdf (3039)
  > 010. .... = Flags: 0x2, Don't fragment
    ...0 0000 0000 0000 = Fragment Offset: 0
    Time to Live: 63
    Protocol: ICMP (1)
    Header Checksum: 0xb77c [validation disabled]
    [Header checksum status: Unverified]
    Source Address: 203.0.113.46
    Destination Address: 203.0.113.30
> Internet Control Message Protocol
```

Provider data-plane

Divided into an IPv6 and Segment Routing Header.

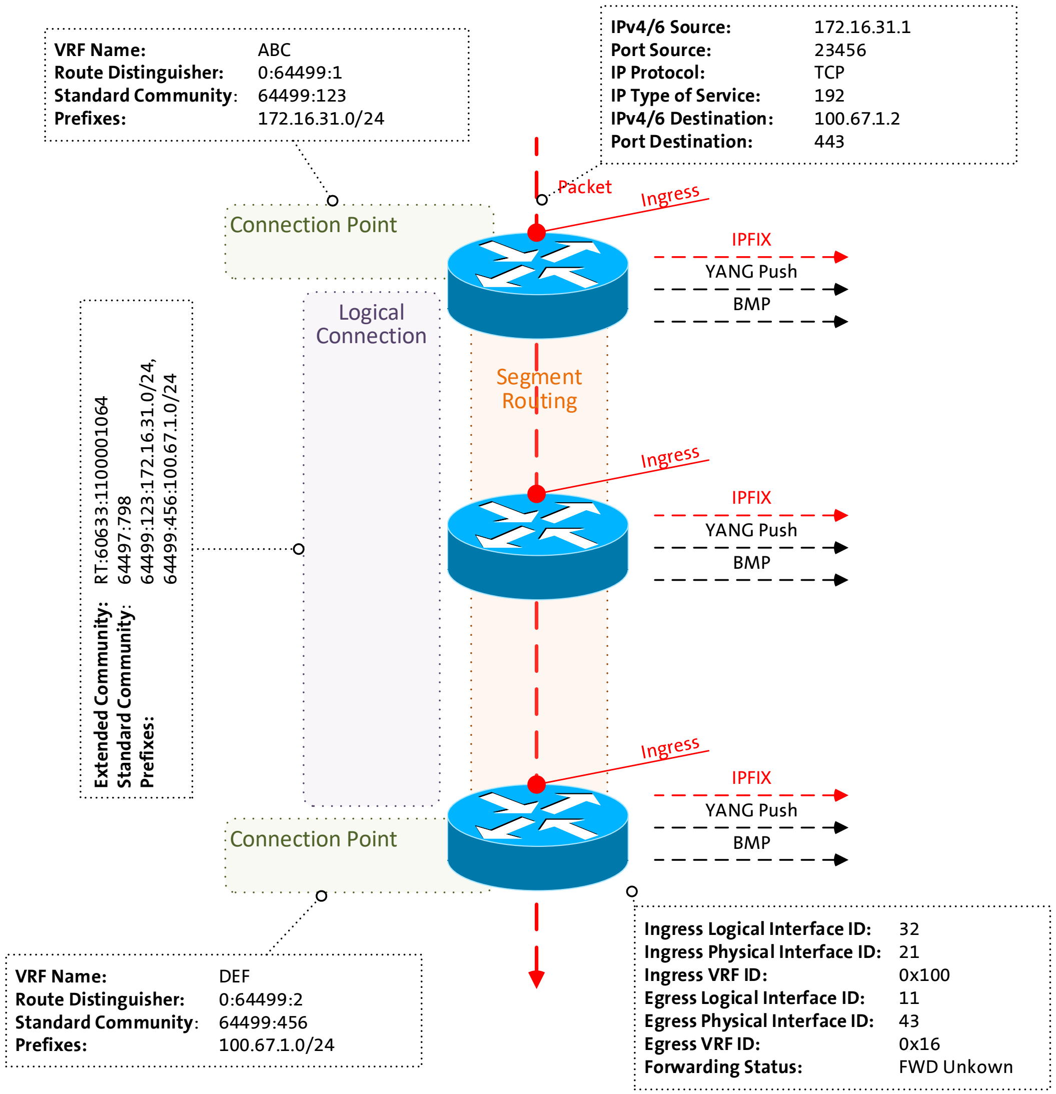
The IPv6 header shows from which PE to which next-hop it is being forwarded. The Segment Routing Header the list of segments this packet needs to pass through and points to the active segment.

Customer data-plane

This is what we receive from the customer and encapsulate for transport through the SRv6 core.

Monitoring L3 SRv6 VPN's with IPFIX and BGP Monitoring Protocol

From L3 VPN Inventory to Realtime Network Analytics



VRF Name:	ABC
Route Distinguisher:	0:64499:1
Standard Community:	64499:123
Prefixes:	172.16.31.0/24

IPv4/6 Source:	172.16.31.1
Port Source:	23456
IP Protocol:	TCP
IP Type of Service:	192
IPv4/6 Destination:	100.67.1.2
Port Destination:	443

Extended Community:	RT:60633:1100001064
Standard Community:	64497:798
Prefixes:	64499:123:172.16.31.0/24, 64499:456:100.67.1.0/24

VRF Name:	DEF
Route Distinguisher:	0:64499:2
Standard Community:	64499:456
Prefixes:	100.67.1.0/24

Ingress Logical Interface ID:	32
Ingress Physical Interface ID:	21
Ingress VRF ID:	0x100
Egress Logical Interface ID:	11
Egress Physical Interface ID:	43
Egress VRF ID:	0x16
Forwarding Status:	FWD Unkown

- From an inventory perspective, **Connection Points are connected through Logical Connections**.
- **From a BGP control-plane perspective, IPv4/6 unicast prefixes in VRF's are tagged with BGP standard communities.**
  - One BGP standard community to identify the Logical Connection. One BGP standard community to identify each Connection Point.
  - When IPv4/6 prefixes are exported from VRF's, a BGP route-distinguisher, BGP extended community route-targets, a SRv6 VPN SID for the IPv6 next-hop is allocated.
- **From a forward-plane perspective,** when IPv4/6 unicast traffic is received from the edge at the SRv6 PE, a lookup is performed, the SRv6 VPN SID **is obtained and** IPv6 next-hop **is added when forwarded to the core.**
- Daisy **collects SRv6 provider data-plane, IPv4/6 unicast customer data-plane in IPFIX** and **at provider edge BGP VPNv4/6 unicast** to perform real-time data correlation.

Trace Path and Measure Delay in IPv6 Data Plane

Use Case Overview

Traffic to Measure

Active probing addresses what-if and Hybrid Type 1 passive customer packet delay and loss SLI.

IPv6 Data Plane Applicability

Applied to IPv6 Destination or Hop-by-Hop options header and observable on transit and encapsulation or decapsulation node only.

Delay Measurement

Measure on-path or round-trip delay observed on network node or outside network node.

Data Aggregation

Delay measurement and trace dimensions are control and management plane aggregated on network node (postcard) or accumulated in the IPv6 packet across the forwarding path (passport) and aggregated outside network node.

Network Dimensions

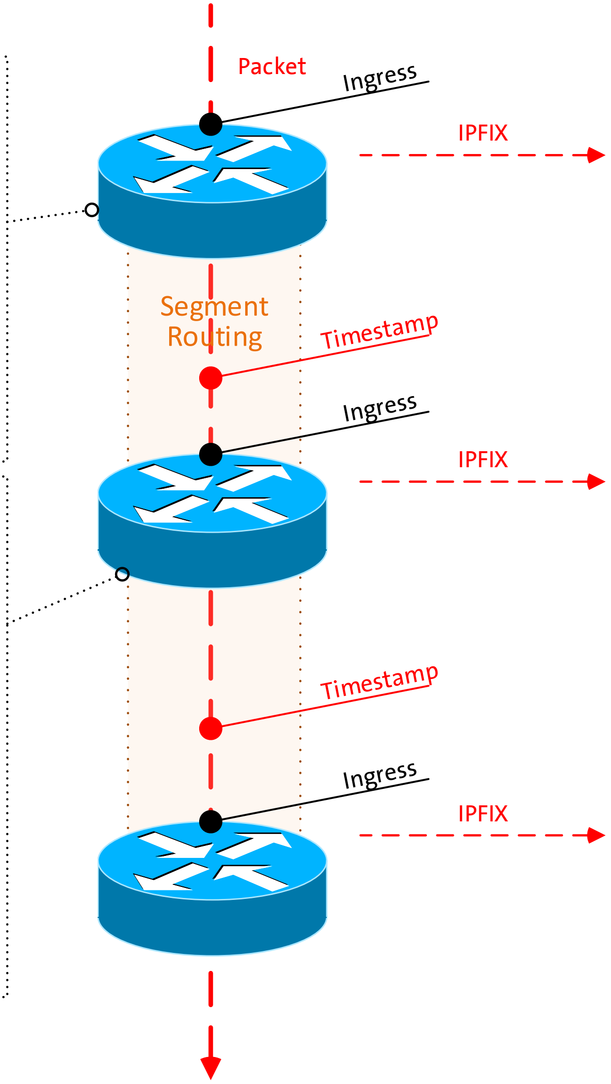
Ability to trace packet forwarding path to describe which interfaces, queues, nodes and domains the flow was forwarded through due to which IP next-hop, top MPLS label or active SRv6 SID. Ability to trace a single or group of flows with same properties.

Measure delay and give network context

Enabling a statistical network delay view

IPv4/6 Source:	172.16.31.1
Port Source:	23456
IP Protocol:	TCP
IP Type of Service:	192
IPv4/6 Destination:	100.67.1.2
Port Destination:	443
Ingress Logical Interface ID:	32
Ingress Physical Interface ID:	21
Ingress VRF ID:	0x100
Egress Logical Interface ID:	11
Egress Physical Interface ID:	43
Egress VRF ID:	0x16
Forwarding Status:	FWD Unkown

IPv4/6 Source:	172.16.31.1
Port Source:	23456
IP Protocol:	TCP
IP Type of Service:	192
IPv4/6 Destination:	100.67.1.2
Port Destination:	443
Ingress Logical Interface ID:	32
Ingress Physical Interface ID:	21
Ingress VRF ID:	0x16
Egress Logical Interface ID:	11
Egress Physical Interface ID:	43
Egress VRF ID:	0x16
Forwarding Status:	FWD Unkown
SID List:	17001, 34002
Delay Min	1
Delay Sum	5
Delay Max	7



- Packets are captured ingress with an optional sampler, data-plane dimensions extracted, enriched with device and control-plane dimensions and **added with a unique flow ID to a flow cache on the node for aggregation.**
- The data-plane dimensions answers **which packet**. The control-plane **which service**. The device dimensions **where in the network**.
- In case of On-Path Delay Measurement, **a timestamp and optionally a direct export tag is added** to the packet header when entering the IOAM domain.
- Each subsequent packet for the same flow increases byte and packet count. Each new flow creates a new flow ID in the flow cache.
- In case of On-Path Delay Measurement, At each node in transit (postcard) or only at the last node (passport), **the delay is calculated by comparing the timestamp in the packet and when packet is received** on the node**. Delay is populated into the flow cache besides packet and byte count.**

Trace Path and Measure Delay in IPv6 Data Plane

Use Case Applicability

	STAMP TWAMP Light	Path Tracing	Alternate Marking	Enhanced Alternate Marking	IOAM Trace Option Type	IOAM Proof of Transit	IOAM Edge to Edge	IOAM Direct Export
Active (what if) measurement	x	x						
Hybrid Type 1 (Connectivity SLI) measurement			x	x	x	x	x	x
Measure on-path delay	x	x	x	x	x	x	x	x
Measure round trip delay	x							
Delay measured on network node	x			x			x	x
Delay measured outside network node		x	x	x	x	x		
Trace domains being forwarded through					x	x		x
Verifies that specified forwarding path is used						x		
Trace nodes and interfaces being forwarded trough		x	x	x	x	x		x
Trace next-hop, top MPLS label or active SRv6 SID			x	x				x
Ability to trace single flows				x				x
Applied to IPv6 Destination Options Header		x					x	
Applied to IPv6 HbH Options Header		x	x	x	x	x		x
Ability to aggregate on network node	x		x	x			x	x
In packet aggregate able		x			x	x		

Main Objectives

Between VPN endpoints we need to understand **which forwarding path is used.**

Between VPN endpoints we need to understand **where the delay is being accumulated and why.**

Network Observability in SRv6
Status, Summary and Next Steps

Status

BGP Monitoring Protocol is BGP Address Family agnostic. **Enables visibility in the SRv6 overlay BGP control plane.**

RFC 9487 adds SRv6 dimensions in IPFIX entities. **Enables visibility in the SRv6 data plane.**

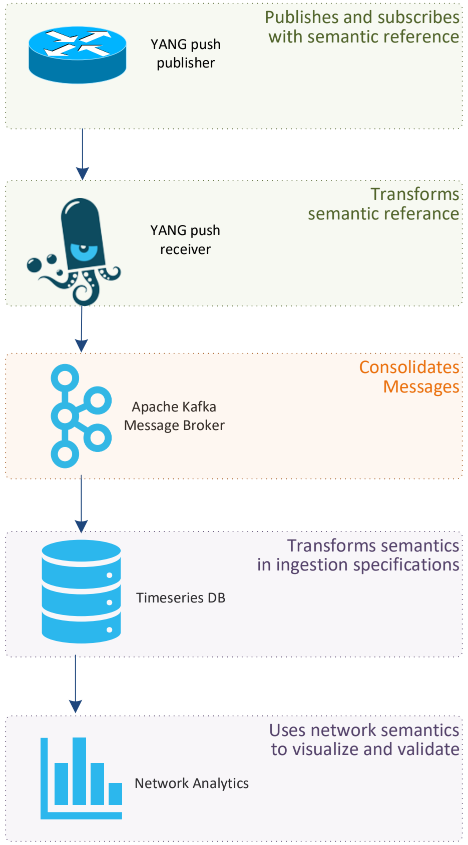
draft-ietf-opsawg-ipfix-on-path-telemetry adds delay measurements in IPFIX entities. **Enables visibility into the on-path delay.**

RFC 9378 defines IOAM. RFC 9341 defines Alternate Marking. draft-gfz-opsawg-ipfix-alt-mark and draft-spiegel-ippm-ioam-rawexport adds Alternate Marking and IOAM dimensions in IPFIX. **Enables tracing visibility in the packet forwarding.**

Summary

The key asset of SRv6 is that at the source of the packet generation the forwarding path can be decided. To validate source routing, on-path visibility is a necessity. However, the protocols involved **are still in development phase, major vendors did not implement, and no requirement document has been written by an operator yet**. SRV6OPS could contribute by not only outlining the use cases, objectives and specifying the requirements but also feedback on implementations.

-> Do you recognize that SRV6OPS should define requirements and use cases along with recomondations on protocol implementations to other working groups?



YANG push publisher

Publishes and subscribes with semantic reference

YANG push receiver

Transforms semantic referance

Apache Kafka Message Broker

Consolidates Messages

Timeseries DB

Transforms semantics in ingestion specifications

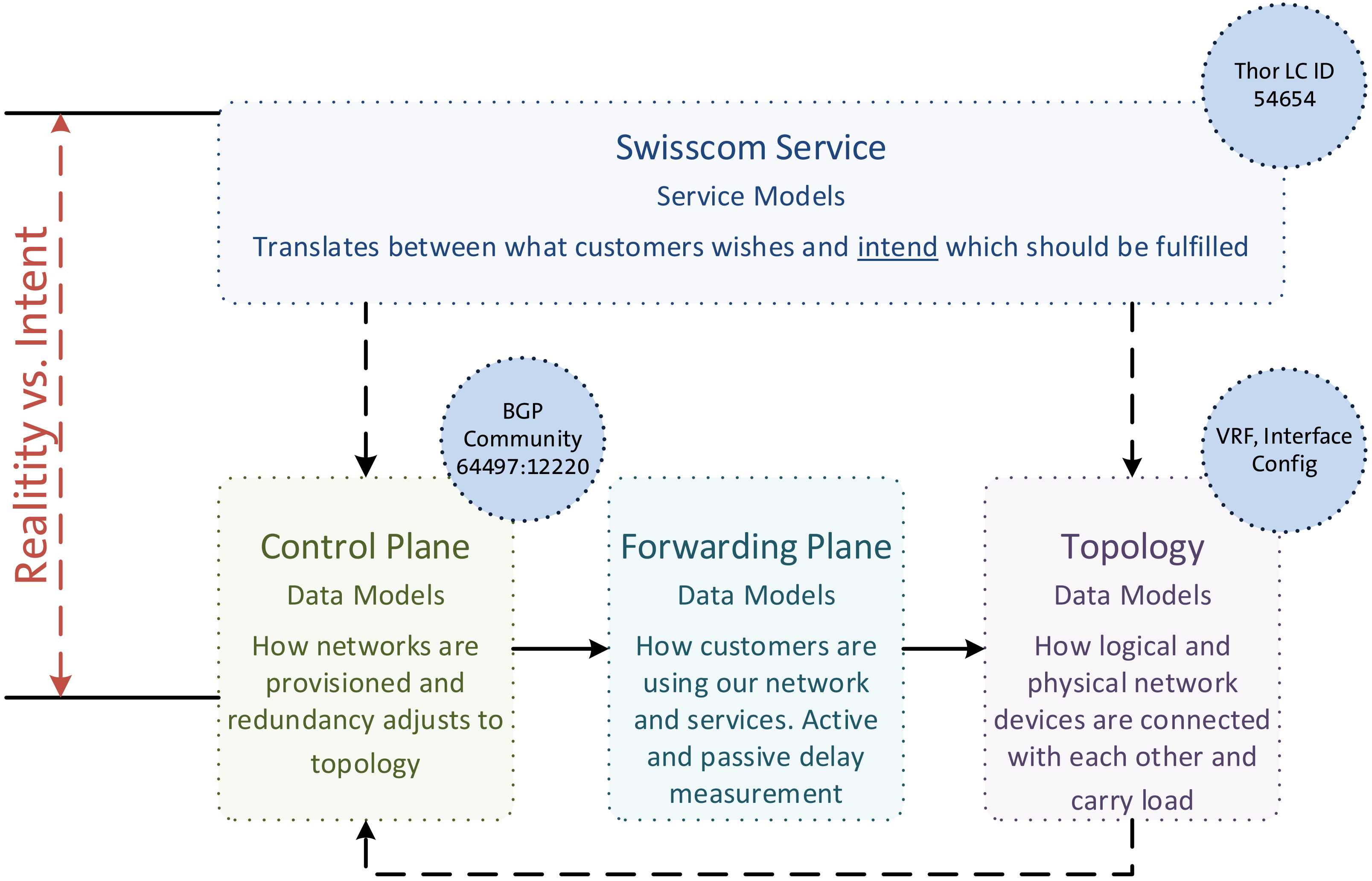
Network Analytics

Uses network semantics to visualize and validate

Backup

Network Data Collection with Network Telemetry

Structured metrics enable informed decision-making



Realitity vs. Intent

Thor LC ID 54654

Swisscom Service

Service Models

Translates between what customers wishes and <u>intend</u> which should be fulfilled

BGP Community 64497:12220

VRF, Interface Config

Control Plane

Data Models

How networks are provisioned and redundancy adjusts to topology

Forwarding Plane

Data Models

How customers are using our network and services. Active and passive delay measurement

Topology

Data Models

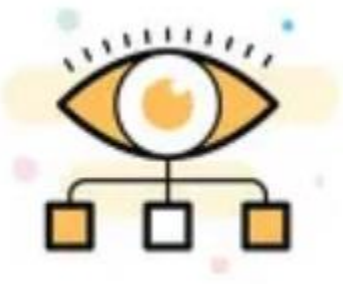
How logical and physical network devices are connected with each other and carry load

Network Telemetry:

- A data collection framework where the network device pushes its metrics to Big Data. Defined in RFC 9232.

Data Modelling:

- Key for Big Data correlation to understand and react in the right context
  - Are interface drops bad?
  - How should we react?



Network Anomaly Detection

For inventoried L2 or L3 VPNs, Network Anomaly Detection **constantly monitors and detects any network or device topology changes**, along with their associated forwarding consequences for customers. Notifications are sent to the Network Operation Center before the customer is aware of service disruptions. It offers operational metrics for in-depth analysis, allowing to understand on which platform the problem originates and facilitates problem resolution.

Answers

What changed and when, on which connectivity service, and how does it impact the customers?

Focuses

Provides meaningful connectivity service impact information before customer is aware of and support in root-cause analysis.

Data Mesh

Consumes real-time Forwarding-Plane, Control-Plane and Management-Plane metrics and produces analytical alerts.

Direction

From connectivity service to network platform.

Network Service Level Indicator and Objective

For inventoried L2 or L3 VPNs, **forwarding-plane loss and latency objectives (SLO) are established for each Quality of Service (QoS) class** in accordance with customer Service Level Agreement (SLA) criteria. Notifications of Service Level Objective (SLO) violations, along with the state of control-plane redundancy, management-plane interface and BGP peering, are raised when forwarding-plane loss or latency objectives are at risk of not being met. These notifications assist network reliability engineering and collaborating relevant platform team to quickly identify and resolve the issue.

Answers

How much budget is left on a connectivity service? Does redundancy exist on all connection points? Is it safe to perform a maintenance window?

Focuses

Service connectivity state and how far state objectives are being fulfilled or not.

Data Mesh

Consumes real-time Forwarding-Plane, Control-Plane and Management-Plane metrics and produces analytical SLI metrics and SLO alerts..

Direction

From connectivity service to network platform.

BMP Covering all RIB's

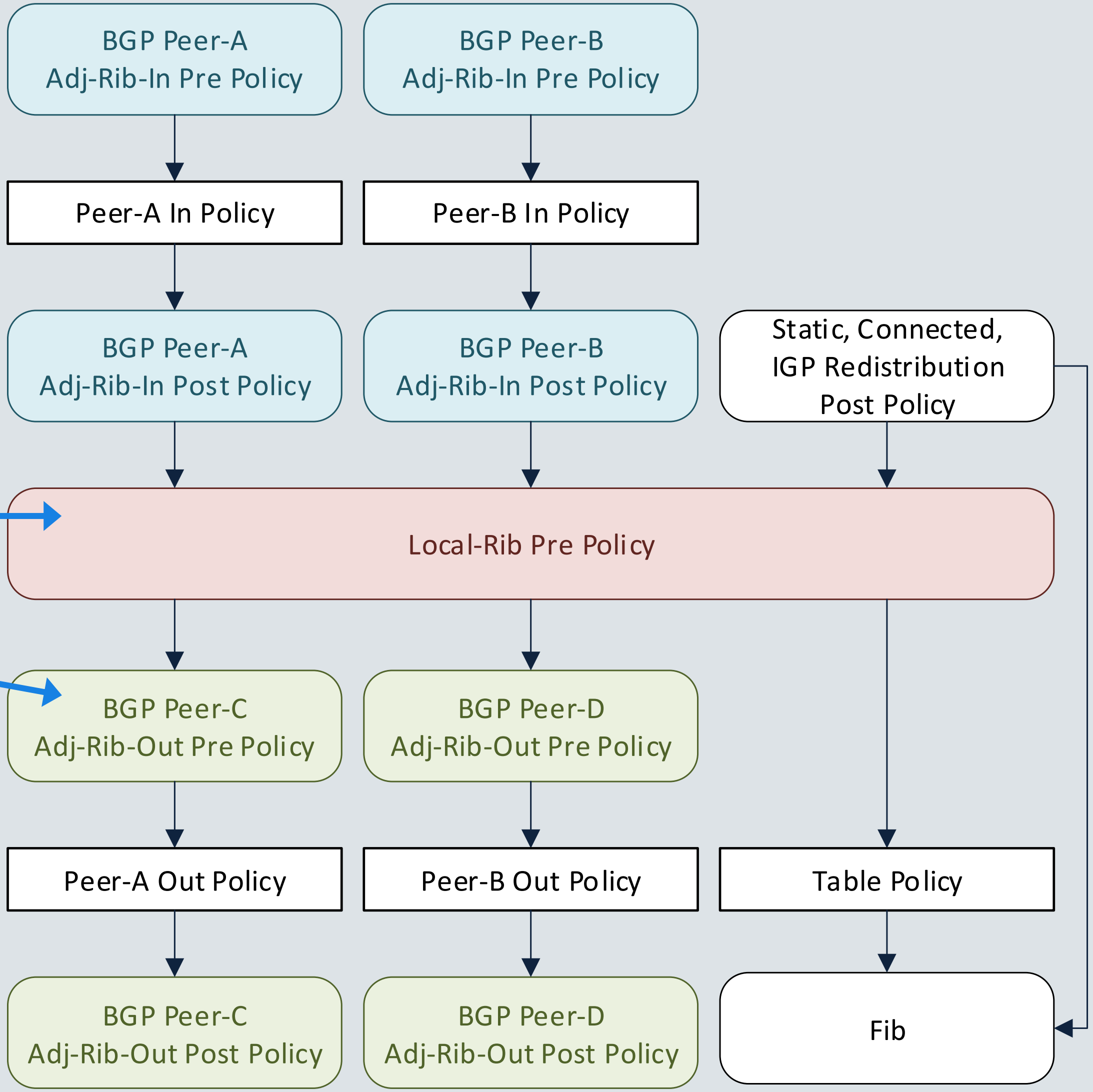
Extends much needed RIB coverage

BGP route exposure without BMP is a challenge of the first order:

- Only best path is exposed (missing best-external and ECMP routes)
- Next-hop attribute not preserved all the time
- Filtering between RIB's not visible

- **Support for Local RIB in BGP Monitoring Protocol** https://datatracker.ietf.org/doc/html/rfc9069
- **Support for Adj-RIB-Out in BGP Monitoring Protocol** https://tools.ietf.org/html/rfc8671

Adj-RIB-Out an RFC since November 2019. **Local RIB since February 2022**. Juniper, Huawei and Nokia have public releases available supporting both. Cisco has test code available but haven't released yet.



BGP Peer-A Adj-Rib-In Pre Policy
BGP Peer-B Adj-Rib-In Pre Policy
Peer-A In Policy
Peer-B In Policy
BGP Peer-A Adj-Rib-In Post Policy
BGP Peer-B Adj-Rib-In Post Policy
Static, Connected, IGP Redistribution Post Policy
Local-Rib Pre Policy
BGP Peer-C Adj-Rib-Out Pre Policy
BGP Peer-D Adj-Rib-Out Pre Policy
Peer-A Out Policy
Peer-B Out Policy
Table Policy
BGP Peer-C Adj-Rib-Out Post Policy
BGP Peer-D Adj-Rib-Out Post Policy
Fib

BMP with extended TLV support

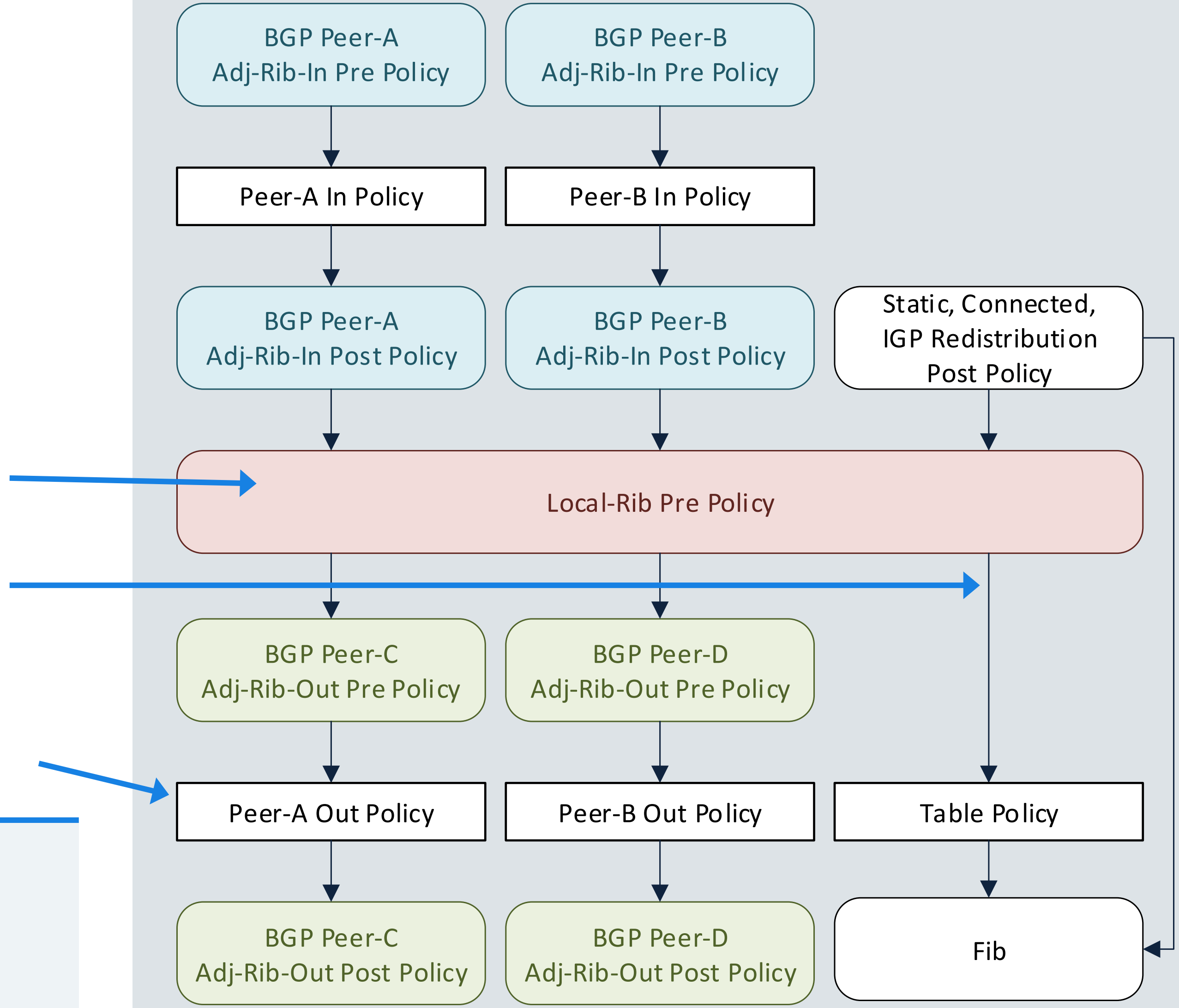
Brings visibility into FIB's and route-policies

Knowing all the routes in all the RIB's brings the new challenge

- That we don't know how they are being used in the FIB/RIB (which one is best, best-external, ECMP, backup)
- That we don't know which route-policy permitted/denied/changed which prefix/attribute

- **TLV support for BMP Route Monitoring and Peer Down Messages**
  https://tools.ietf.org/html/draft-ietf-grow-bmp-tlv
- **Support for Enterprise-specific TLVs in the BGP Monitoring Protocol**
  https://tools.ietf.org/html/draft-lucente-grow-bmp-tlv-ebit
- **BMP Extension for Path Marking TLV**
  https://tools.ietf.org/html/draft-cppy-grow-bmp-path-marking-tlv
- **Logging of routing events in BGP Monitoring Protocol**
  https://datatracker.ietf.org/doc/html/draft-ietf-grow-bmp-rel

For IETF 110 Hackathon, IETF lab network with Big Data integration has been further extended to collaborate development research with ETHZ, INSA, Cisco, Huawei and pmacct (open source data-collection by Paolo Lucente).


BGP Peer-A Adj-Rib-In Pre Policy
BGP Peer-B Adj-Rib-In Pre Policy
Peer-A In Policy
Peer-B In Policy
BGP Peer-A Adj-Rib-In Post Policy
BGP Peer-B Adj-Rib-In Post Policy
Static, Connected, IGP Redistribution Post Policy
Local-Rib Pre Policy
BGP Peer-C Adj-Rib-Out Pre Policy
BGP Peer-D Adj-Rib-Out Pre Policy
Peer-A Out Policy
Peer-B Out Policy
Table Policy
BGP Peer-C Adj-Rib-Out Post Policy
BGP Peer-D Adj-Rib-Out Post Policy
Fib