Visualizing BGP RIB Changes into Forwarding Plane by Leveraging BGP Monitoring Protocol and IPFIX

Master Thesis Presentation, collaboration with Swisscom

Livio Sgier, 30th October, 2020

Problem Statement

- Hundreds of thousands of networking devices in ISP networks/large-scale data centers
- Huge complexities
  - Many protocols on multiple layers
  - No inherent support for transparency and monitoring

➔ Impaired network visibility

➔ Delay in detection & repair

- Monitoring for **forwarding plane** exists (e.g., IPFIX for IP traffic), not for **control plane**

  ➔ Introduction of **BGP Monitoring Protocol (BMP)**

Thesis Goals

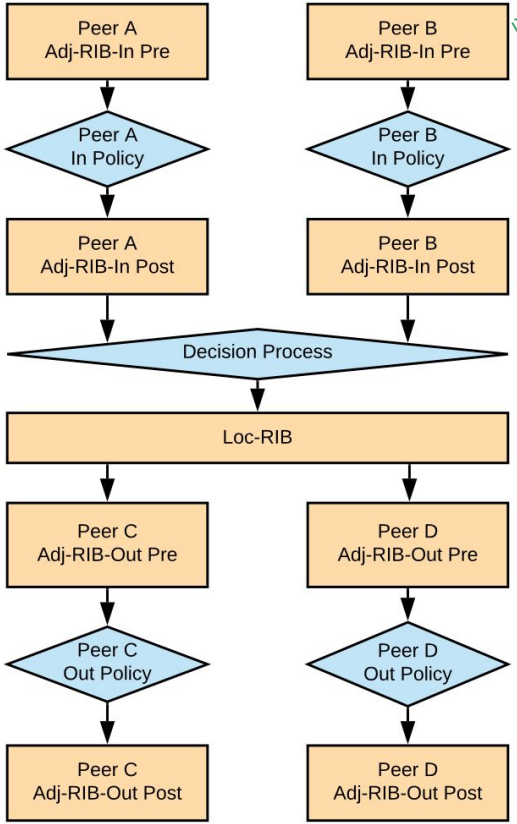
- Investigate BMP capabilities for **different vendors** and in the context of a Lab network (BGP/MPLS IP Virtual Private Networks)
- Build **end-to-end** pipeline
- Correlation on **VPN level** and in **time** between **control** (BMP) and **forwarding** plane (IPFIX)
    - Enable root cause analysis, (real time) performance monitoring

Before BMP: Traditional way to monitor control plane

- Active BGP Peerings
  - Establish BGP peering & receive BGP updates

- Separate active/passive protocols
  - **Screen-scraping**: Log-in to the router via SSH/Telnet and issue commands like ***show ip bgp***
  - BGP **Looking Glass**: Access public routing information from large companies/ISPs

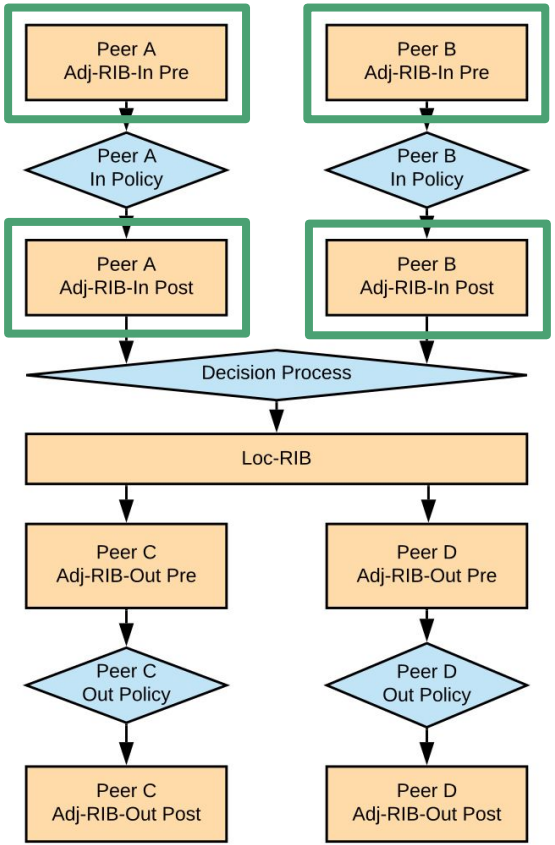
Border Gateway Protocol (BGP)




```
Border Gateway Protocol - UPDATE Message
  Marker: ffffffffffffffffffffffffffffffff
  Length: 79
  Type: UPDATE Message (2)
  Withdrawn Routes Length: 0
  Total Path Attribute Length: 51
▾ Path attributes
  ▸ Path Attribute - ORIGIN: IGP
  ▸ Path Attribute - AS_PATH: 65539 65536 65537
  ▸ Path Attribute - NEXT_HOP: 192.0.2.171
  ▸ Path Attribute - COMMUNITIES: 64496:299 64496:1001
▾ Network Layer Reachability Information (NLRI)
  ▸ 203.0.113.252/31
```

RIB Entry: Prefix with attached attributes

BGP Monitoring Protocol (BMP)

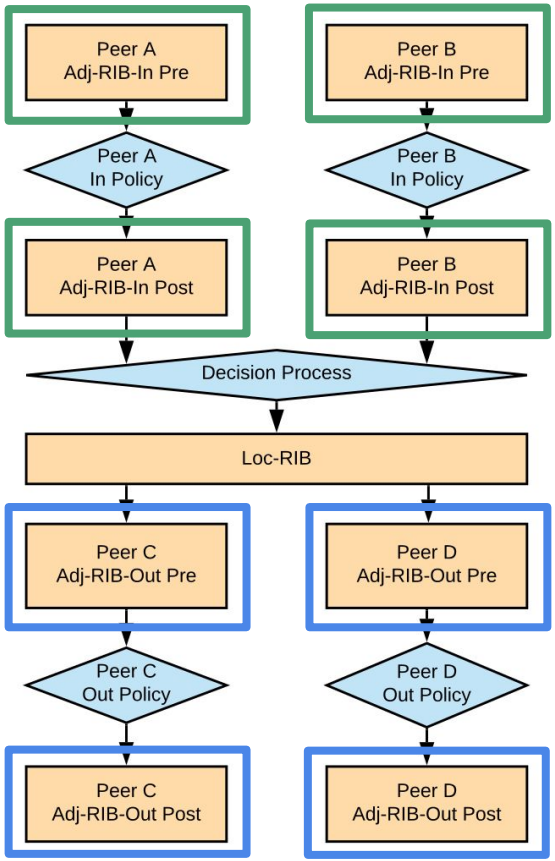


- Access to Adj-RIB-In [RFC 7854]

Enabled use cases:

➔ Access to **incoming** BGP updates

BGP Monitoring Protocol (BMP)

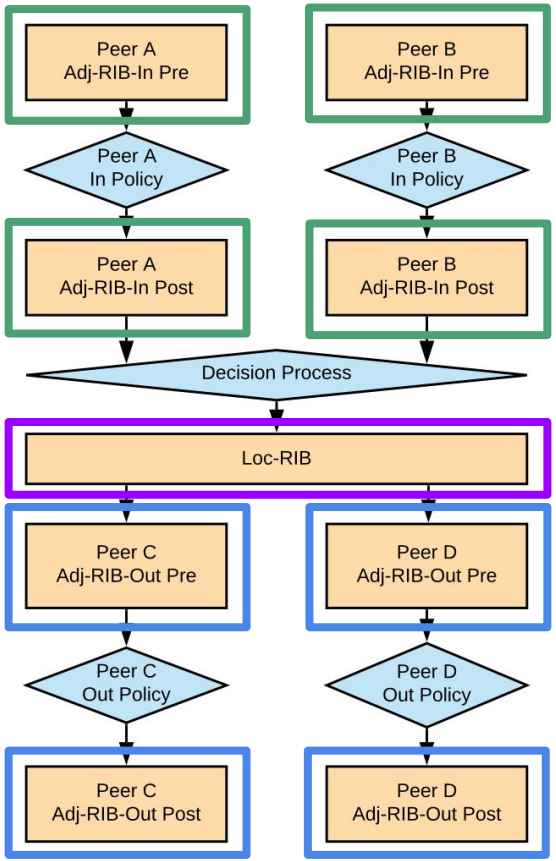


- Access to Adj-RIB-In [RFC 7854]
- Access to Adj-RIB-Out [RFC 8671]

Enabled use cases:

➔ Access to **outgoing** BGP updates

BGP Monitoring Protocol (BMP)

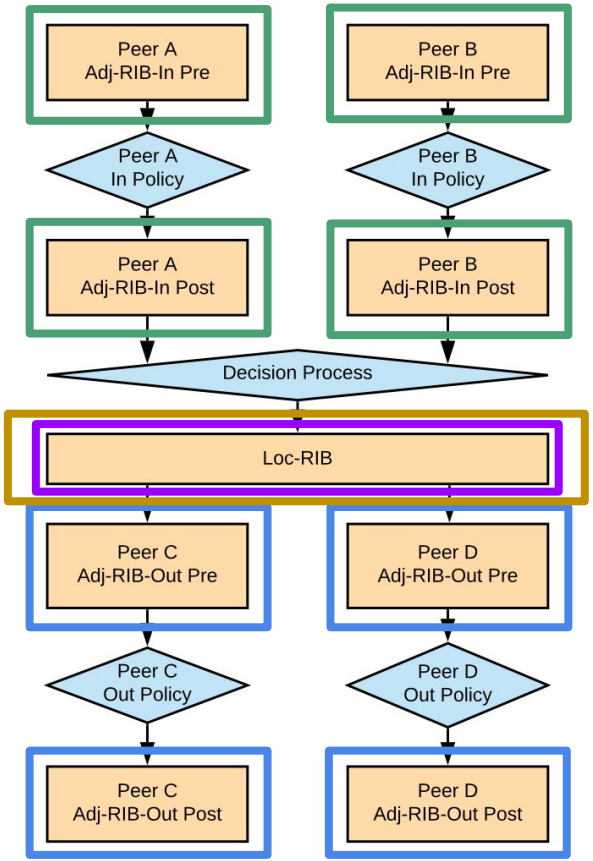


- Access to Adj-RIB-In [RFC 7854]
- Access to Adj-RIB-Out [RFC 8671]
- Access to Loc-RIB [draft]

Enabled use cases:

➔ Access to BGP updates **installed** in local RIB

BGP Monitoring Protocol (BMP)

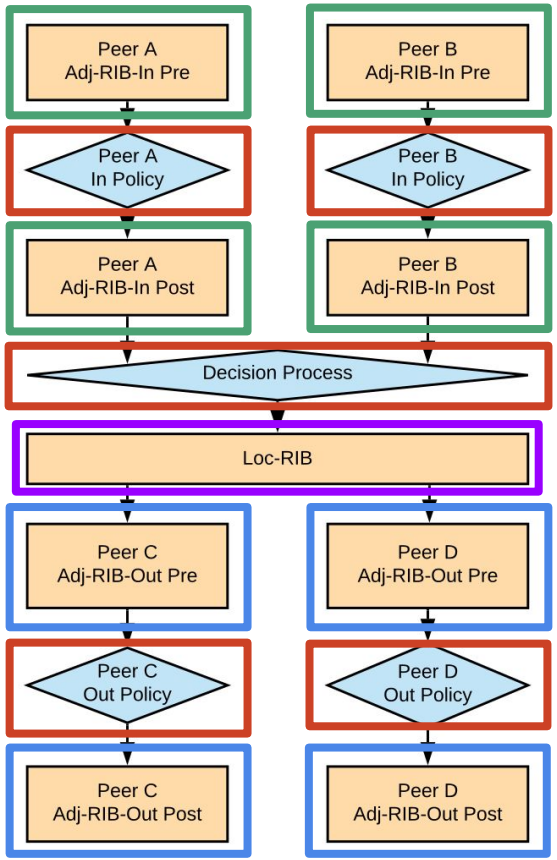


- Access to Adj-RIB-In [RFC 7854]
- Access to Adj-RIB-Out [RFC 8671]
- Access to Loc-RIB [draft]
- Route Monitoring Path Marking [draft]

Enabled use cases:

➔ Access to **how** BGP updates are installed in the local RIB

BGP Monitoring Protocol (BMP)



- Access to Adj-RIB-In [RFC 7854]
- Access to Adj-RIB-Out [RFC 8671]
- Access to Loc-RIB [draft]
- Route Monitoring Path Marking [draft]
- Route Policy and Attr. Tracing [draft]

Enabled use cases:

➔ Information about triggered route policies

Vendors vs. BMP Support

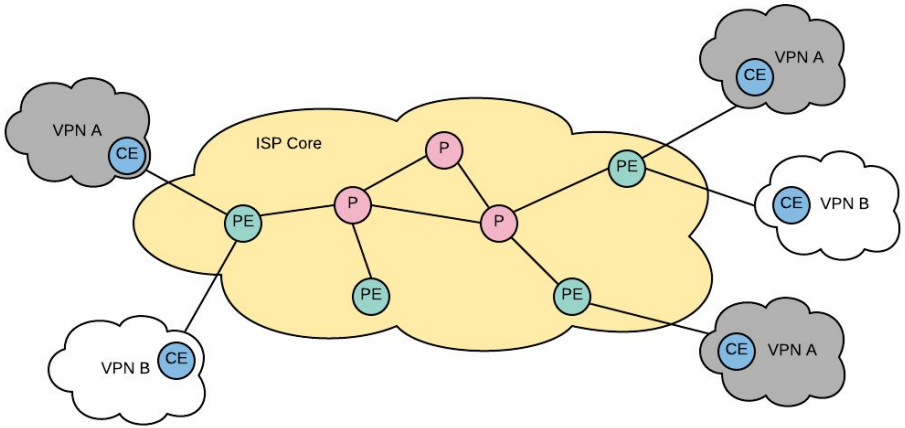
	Adj-RIB-In	Adj-RIB-Out	Loc-RIB	Route Policy and Attr. Tracing	Path Marking
Huawei	✓	✓	✓	✓	✓
Cisco XRv9000	✓	✗	✗	✗	✗
Cisco CSR1000V	✓	✗	✗	✗	✗
Juniper	✓[1]	✓[1]	✓[1]	✗	✗
FRRouting	✓	✗	✗	✗	✗

[1] Only non-VRF RIB entries are supported.

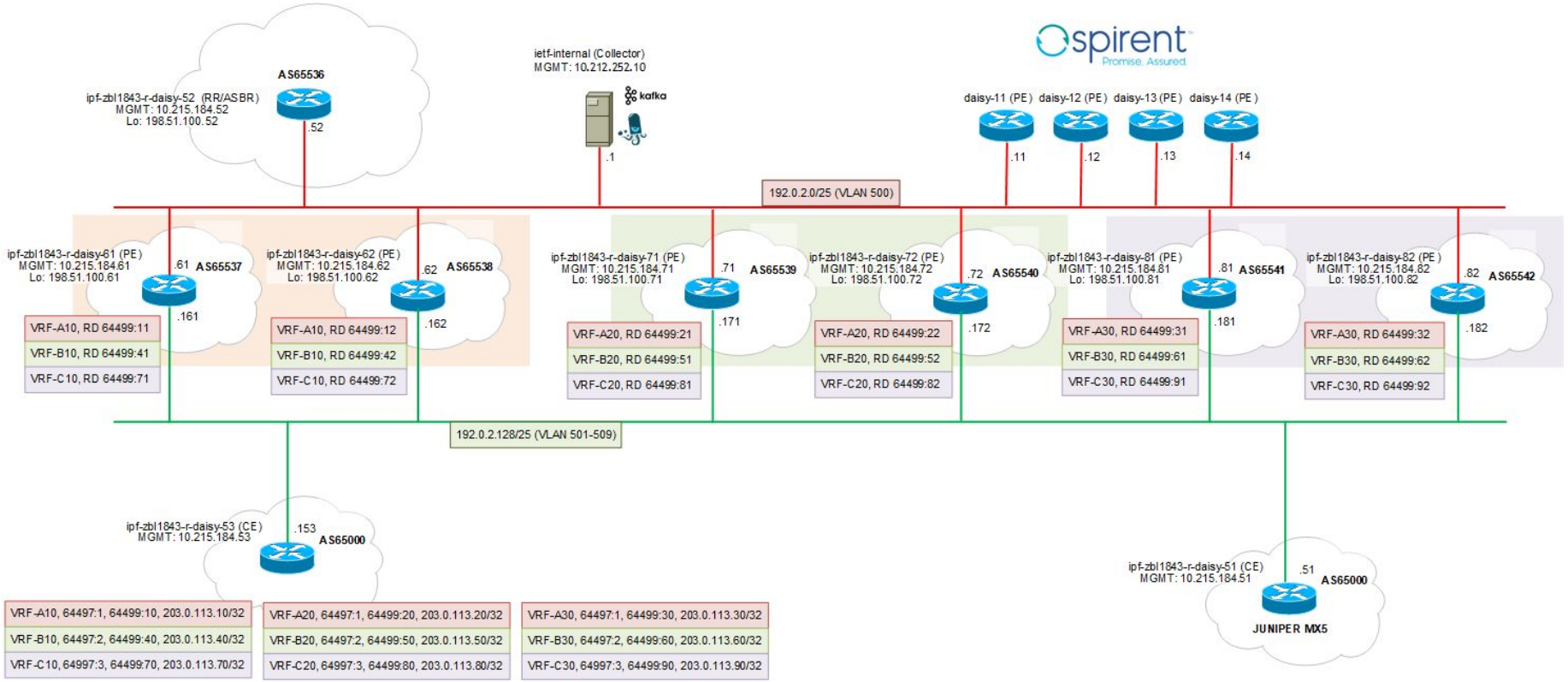
Lab Network Type

- MPLS/BGP IP Virtual Private Networks [RFC 4364]
- Common architecture for ISPs: Enabling **multiple VPNs** over same physical infrastructure
- Entities: Customer Edge (**CE**), Provider Edge (**PE**), Provider Core (**P**)
- Virtual routing and forwarding (**VRF**) instances per VPN to isolate routing tables
- New address family: **VPNv4/6** (Route Distinguisher + IPv4/6, e.g., 0:64499:6:203.0.113.10/32)

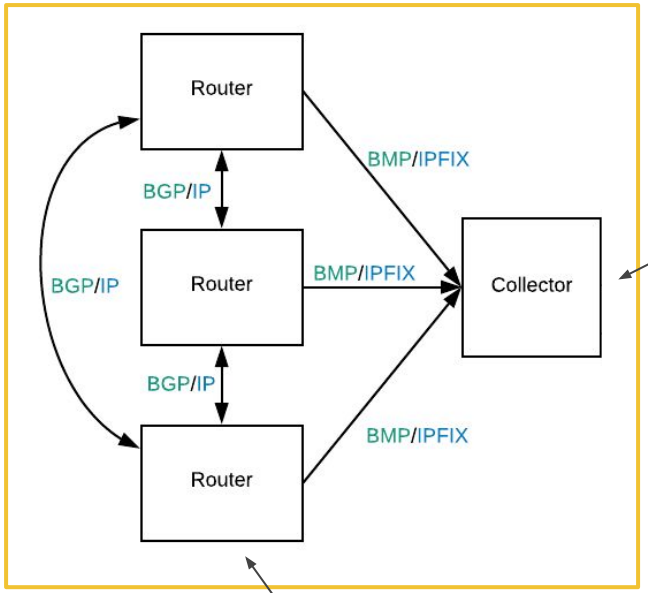
RD IPv4/6



BGP/MPLS IP VPN Lab Network



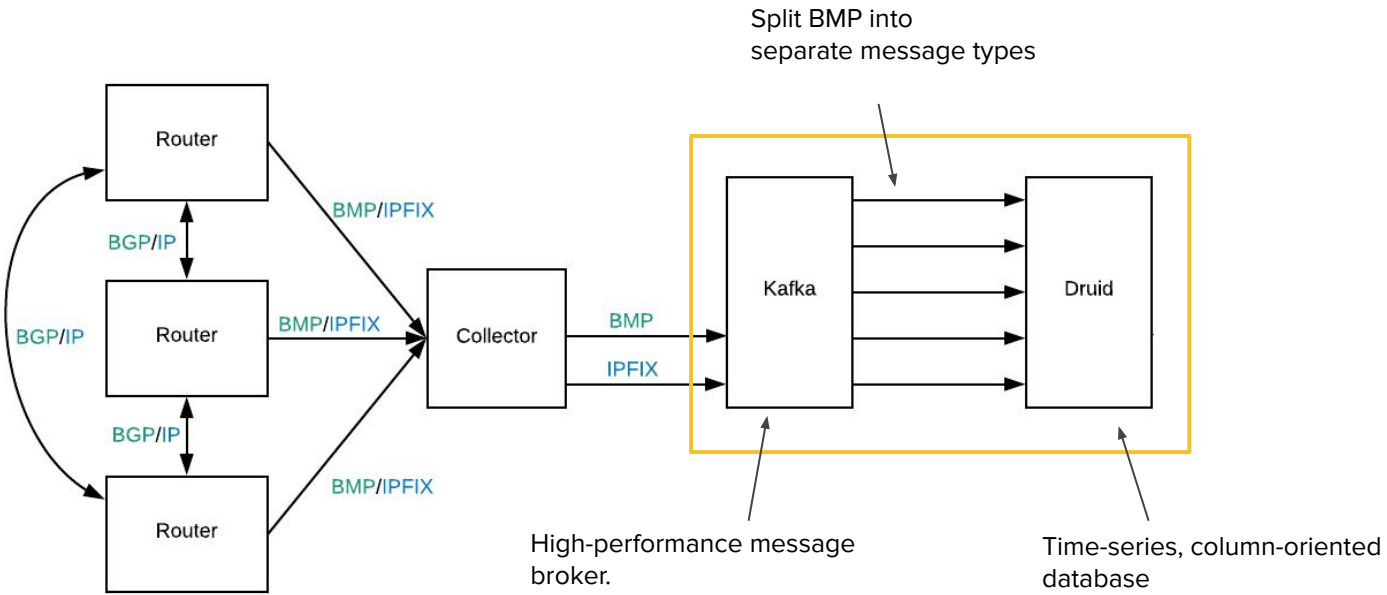
End-to-end Pipeline Design: Export & Collection



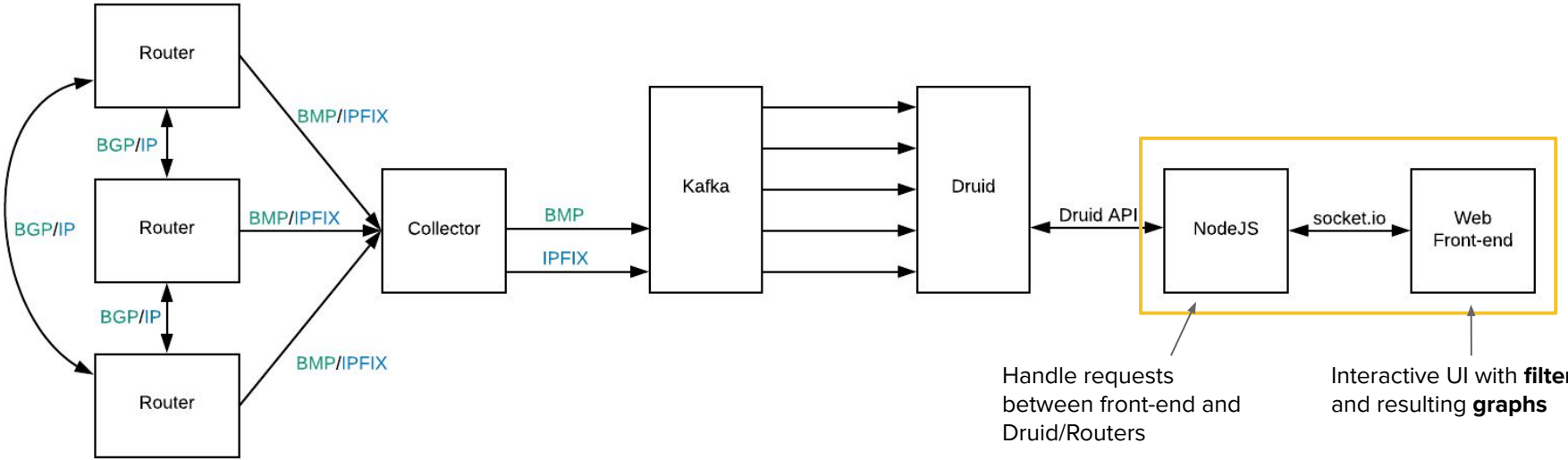
Collect **IPFIX** and **BMP**
➜ correlate & enrich

Which **RIBs**?
Which **address familes**?

End-to-end Pipeline Design: Processing & Storage

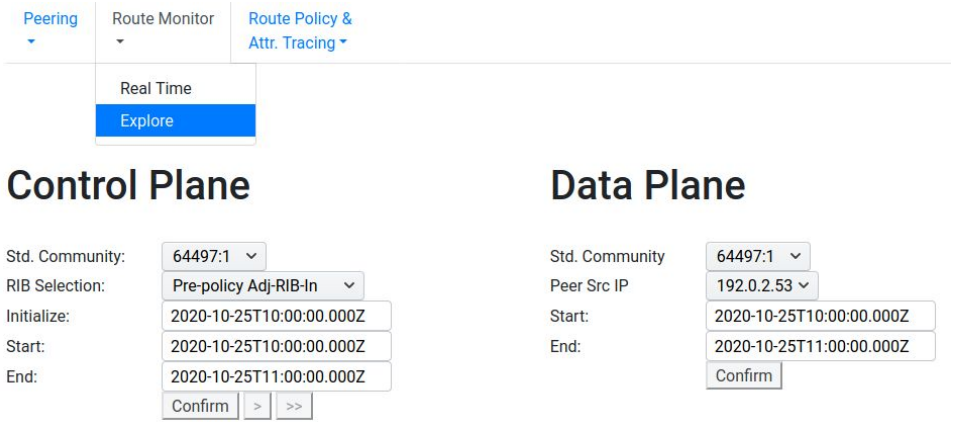


End-to-end Pipeline Design: Visualization



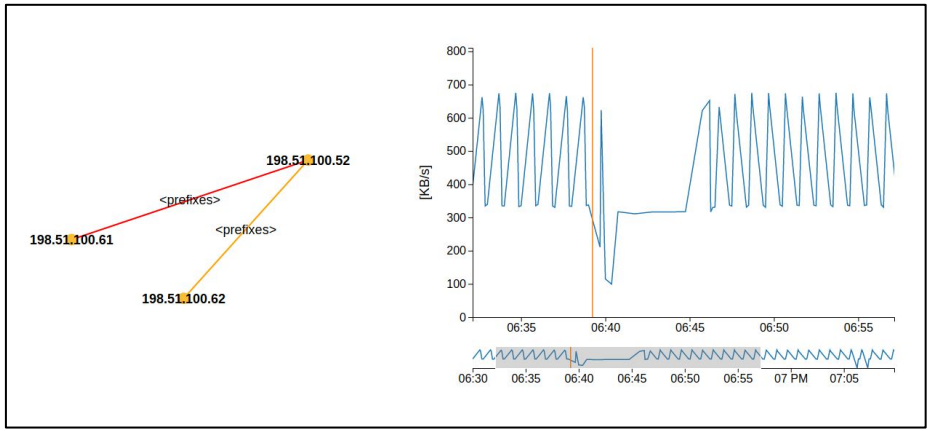
Visualization/Web Front End

- Goals
  - Support **exploratory** and **real time** queries for **Route Monitoring**, **Peering**, **Route Policy**
  - **Correlation** on **VPN level** and in **time** between control- and forwarding-plane
    ➜ Make cause and effect visible across planes

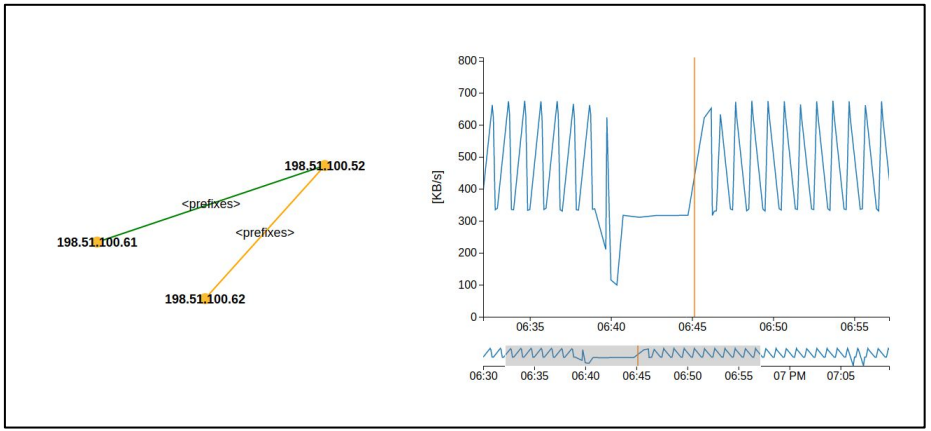


Visualization: Example Query (1)

“For **VPN A**, show me the **Adj-RIB-In Pre-Policy** and its **correlation** to the **forwarding plane** before and after enabling an interface on a **VPN level**.”



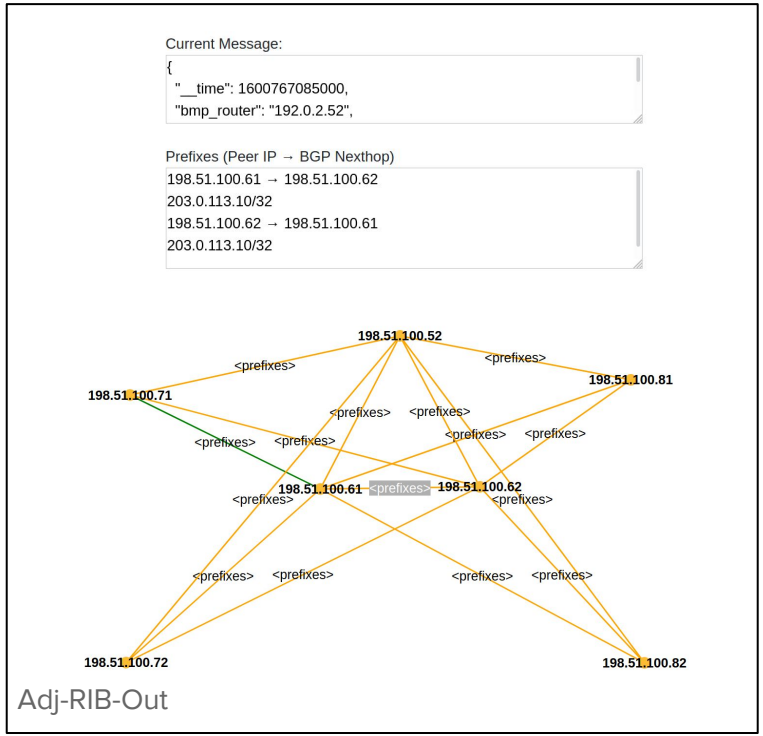
Removing Interface



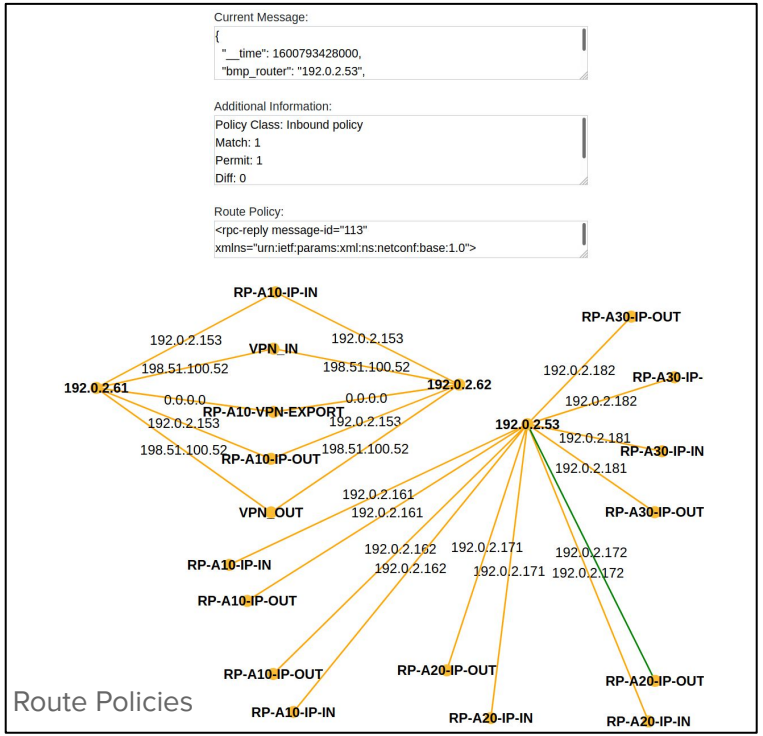
Re-enabling Interface

Visualization: Example Query (2)

“For re-enabled **prefix X** on **VPN A**, show a **live view** of **Adj-RIB-Out** and all triggered **route policies**”



Adj-RIB-Out



Route Policies

Outlook/Future Work

- Ongoing/Future work on BMP
  - BMP delay & loss detection and prevention
  - Additional BGP information in form of TLVs (e.g., capabilities) [draft]
  - Preventing I/O bound protocol ➔ compression on demand [draft]

- Tested successfully in production at Swisscom

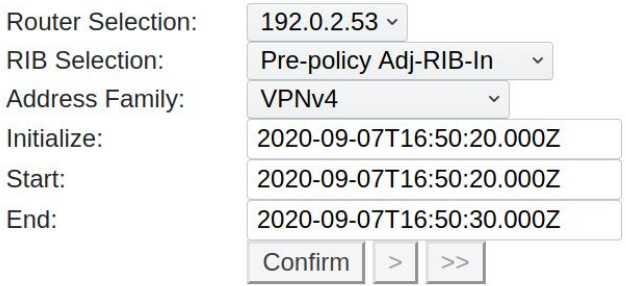
Backup Slides

Mapping IP address and ingress/egress interface to Route Distinguisher

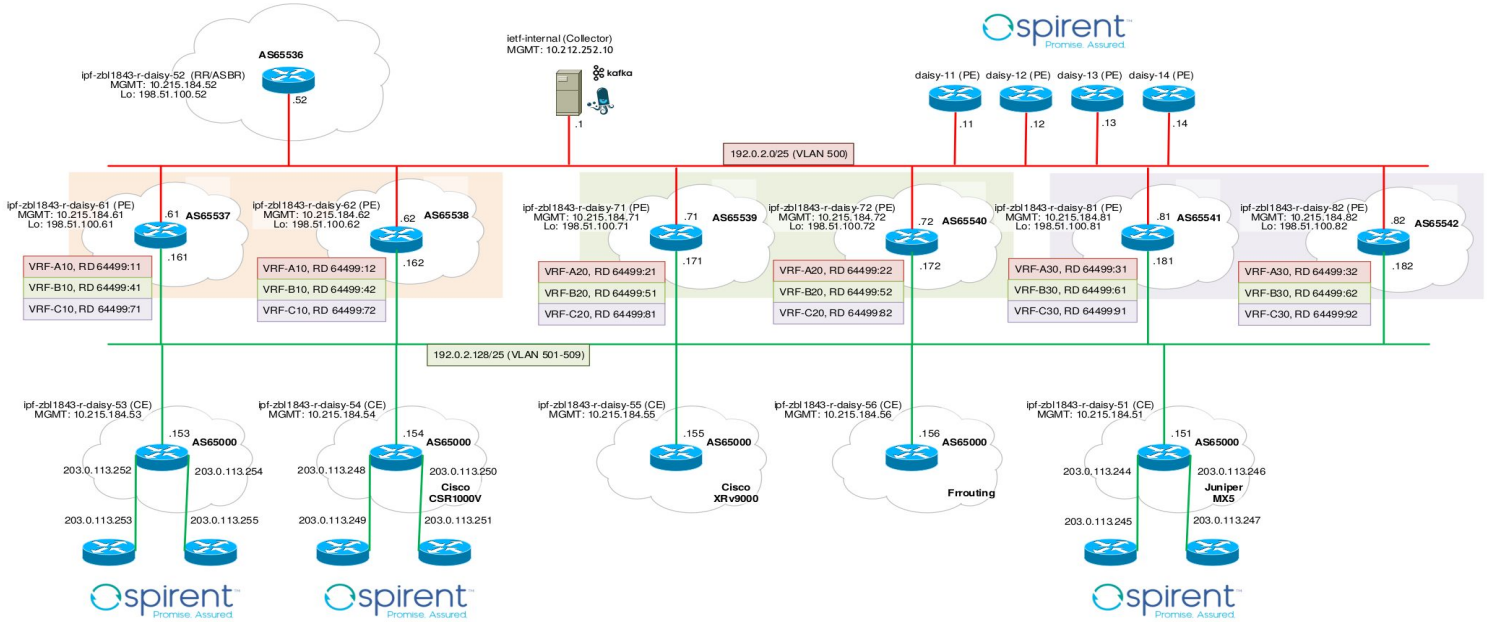
```
!192.0.2.72

id=0:64499:22 ip=192.0.2.72 in=47
id=0:64499:22 ip=192.0.2.72 out=47
id=0:64499:52 ip=192.0.2.72 in=49
id=0:64499:52 ip=192.0.2.72 out=49
id=0:64499:82 ip=192.0.2.72 in=50
id=0:64499:82 ip=192.0.2.72 out=50
id=0:0:0 ip=192.0.2.72 in=4
id=0:0:0 ip=192.0.2.72 out=4

!192.0.2.52

id=0:0:0 ip=192.0.2.52 in=42
id=0:0:0 ip=192.0.2.52 out=42
id=0:0:0 ip=192.0.2.52 in=4
id=0:0:0 ip=192.0.2.52 out=4

...
```

UI Filter Example

BMP Message type: **Peer up**



Updated lab network



VRF-A10, 64497:1, 64499:10, 203.0.113.10/32	VRF-A10, 64497:1, 64499:11, 203.0.113.11/32
VRF-B10, 64497:2, 64499:40, 203.0.113.40/32	VRF-B10, 64497:2, 64499:41, 203.0.113.41/32
VRF-C10, 64997:3, 64499:70, 203.0.113.70/32	VRF-C10, 64997:3, 64499:71, 203.0.113.71/32
VRF-A20, 64497:1, 64499:20, 203.0.113.20/32	VRF-A20, 64497:1, 64499:21, 203.0.113.21/32
VRF-B20, 64497:2, 64499:50, 203.0.113.50/32	VRF-B20, 64497:2, 64499:51, 203.0.113.51/32
VRF-C20, 64997:3, 64499:80, 203.0.113.80/32	VRF-C20, 64997:3, 64499:81, 203.0.113.81/32
VRF-A30, 64497:1, 64499:30, 203.0.113.30/32	VRF-A30, 64497:1, 64499:31, 203.0.113.31/32
VRF-B30, 64497:2, 64499:60, 203.0.113.60/32	VRF-B30, 64497:2, 64499:61, 203.0.113.61/32
VRF-C30, 64997:3, 64499:90, 203.0.113.90/32	VRF-C30, 64997:3, 64499:91, 203.0.113.91/32

VRF-A10, 64497:1, 64499:12, 203.0.113.12/32	VRF-A10, 64497:1, 64499:13, 203.0.113.13/32	64497:1, 64499:14, 203.0.113.14/32
VRF-B10, 64497:2, 64499:42, 203.0.113.42/32	VRF-B10, 64497:2, 64499:43, 203.0.113.43/32	64497:2, 64499:44, 203.0.113.44/32
VRF-C10, 64997:3, 64499:72, 203.0.113.72/32	VRF-C10, 64997:3, 64499:73, 203.0.113.73/32	64997:3, 64499:74, 203.0.113.74/32
VRF-A20, 64497:1, 64499:22, 203.0.113.22/32	VRF-A20, 64497:1, 64499:23, 203.0.113.23/32	64497:1, 64499:24, 203.0.113.24/32
VRF-B20, 64497:2, 64499:52, 203.0.113.52/32	VRF-B20, 64497:2, 64499:53, 203.0.113.53/32	64497:2, 64499:54, 203.0.113.54/32
VRF-C20, 64997:3, 64499:82, 203.0.113.82/32	VRF-C20, 64997:3, 64499:83, 203.0.113.83/32	64997:3, 64499:84, 203.0.113.84/32
VRF-A30, 64497:1, 64499:32, 203.0.113.32/32	VRF-A30, 64497:1, 64499:33, 203.0.113.33/32	64497:1, 64499:34, 203.0.113.34/32
VRF-B30, 64497:2, 64499:62, 203.0.113.62/32	VRF-B30, 64497:2, 64499:63, 203.0.113.63/32	64497:2, 64499:64, 203.0.113.64/32
VRF-C30, 64997:3, 64499:92, 203.0.113.92/32	VRF-C30, 64997:3, 64499:93, 203.0.113.93/32	64997:3, 64499:94, 203.0.113.94/32

Link Network between PE and CE

A10, 501, 192.0.11.x, 2001:DB8:11::x
A20, 502, 192.0.12.x, 2001:DB8:12::x
A30, 503, 192.0.13.x, 2001:DB8:13::x
B10, 504, 192.0.21.x, 2001:DB8:21::x
B20, 505, 192.0.22.x, 2001:DB8:22::x
B30, 506, 192.0.23.x, 2001:DB8:23::x
C10, 507, 192.0.31.x, 2001:DB8:31::x
C20, 508, 192.0.32.x, 2001:DB8:32::x
C3,0 509, 192.0.33.x, 2001:DB8:33::x

Sample list of BMP statistics

- Number of prefixes rejected by inbound policy.
- Number of (known) duplicate prefix advertisements.
- Number of (known) duplicate Withdraws.
- Number of updates invalidated due to CLUSTER_LIST loop
- Number of updates invalidated due to AS_PATH loop.
- Number of routes in per-AFI/SAFI Adj-RIB-In. The value is structured as: 2-byte Address Family Identifier (AFI), 1-byte Subsequent Address Family Identifier (SAFI), followed by a 64-bit Gauge.
- Number of routes in Adj-RIBs-In.