

BMP

BGP Monitoring Protocol
GROW WG

IETF 108

July 20-24th, 2020

Virtual Hackathon



Hackathon - Plan

Functionality

- Test BMP BGP Local RIB to IPFIX metric correlation and interoperability between router and data-collection for peer and route monitoring for message type extensions defined in
 - [draft-ietf-grow-bmp-local-rib](#) (BGP Local RIB)
 - [draft-grow-bmp-tlv](#) (TLV support for BMP Route Monitoring and Peer Down Messages)
 - [draft-lucente-grow-bmp-tlv-ebit](#) (Support for Enterprise-specific TLVs)
 - [draft-cppy-grow-bmp-path-marking-tlv](#) (Path Marking TLV)
 - [draft-xu-grow-bmp-route-policy-attr-trace](#) (BGP Route Policy and Attribute Trace)

Performance

- Test performance impact of BMP on router CPU/Memory resources and BGP route propagation with YANG push.

Hackathon – Software

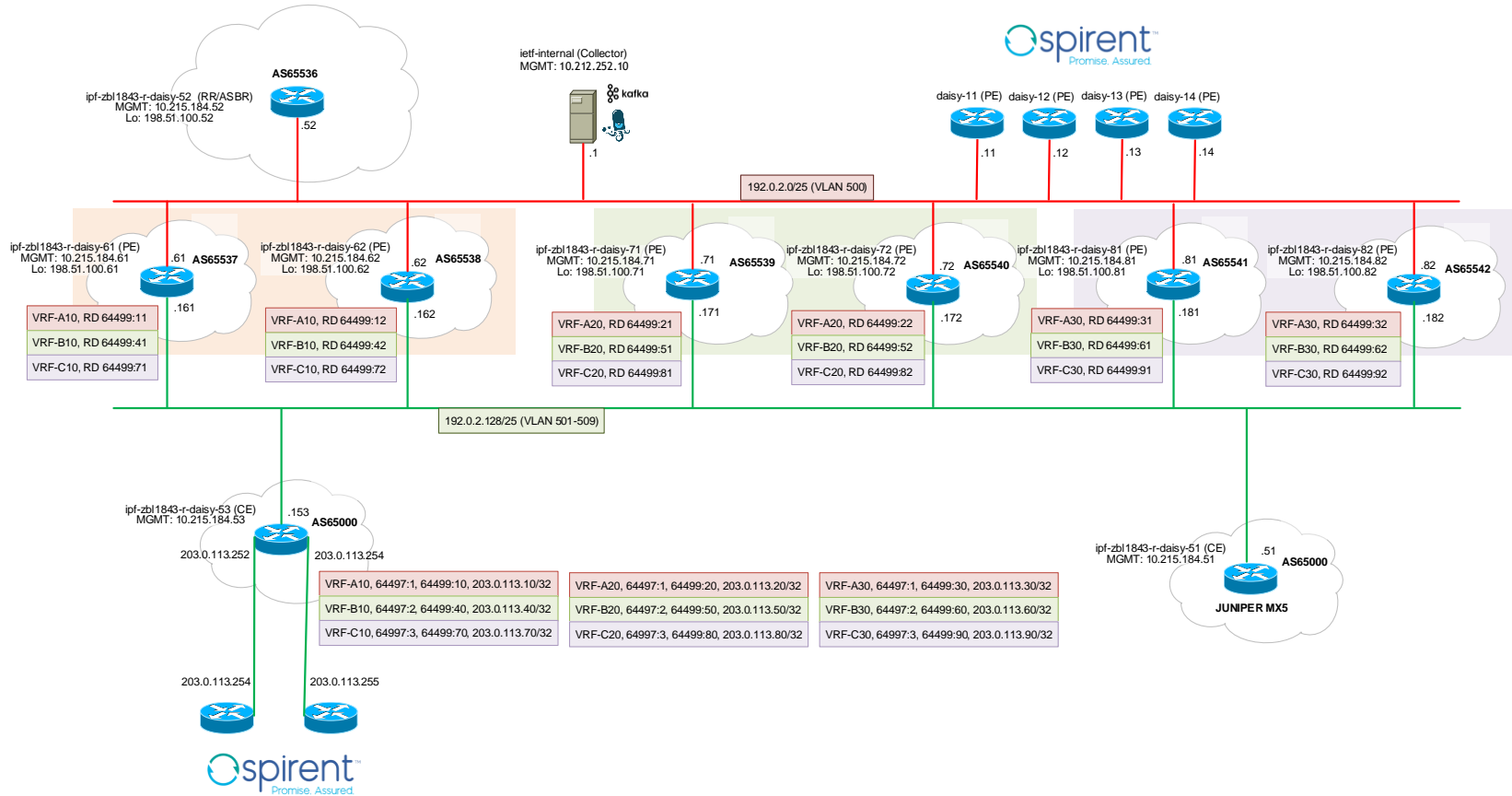
Software

- [pmacct](#) nfacctd for IPFIX and BMP data collection
- [pmacct](#) pmgrpcd for YANG push data collection
- Apache [Kafka](#) as message broker
- Apache [Druid](#) as timeseries DB
- [Pivot](#) as user interface
- Wireshark [BMP dissector](#) for packet analysis
- Spirent [Testcenter](#) for BGP VPnv4/6 route and IPV4/6 traffic generation

Tutorial

- <https://imply.io/post/add-bgp-analytics-to-your-imply-netflow-analysis>

Hackathon - Network



Swisscom – lab environment

Achievements

- Spirent Testcenter added for IPv4/6 traffic generation
- YANG push data collection for CPU and memory

Gaps Identified

- Test verification needs to be further automatized to improve efficiency

Next Steps

- BMP BGP RIB update flow delay heatmap to facilitate convergence delay RCA
- Improve testbed to measure the impact on network convergence with BMP
- Validate BGP router reset notification PDU for Adj-RIB In/Out and consequent action in correlator

Pmacct – nfacctd/pmbmpd

Achievements

- BMP BGP Local RIB to IPFIX correlation now works for prefixes with BGP route-distinguisher as well.
- 2 of 5 TLV's decoded of [draft-xu-grow-bmp-route-policy-attr-trace](#)

Gaps Identified

- Path Marking TLV could be optimized if contained paths would have been indexed.
Input for [draft-cppy-grow-bmp-path-marking-tlv-04](#)

<https://github.com/pmacct/pmacct/>

BMP BGP Local RIB with IPFIX Correlation

IPFIX Flow Aggregation

...

Q

Time

As Path

As Path Src

Bytes

Comms

Comms Src

Ecomms

Ecomms Src

Event Type

Forwarding Status

Iface In

Iface Out

Ip Dst

Ip Proto

FILTER

Latest 7 days

A Comms: 2 values

A Peer Ip Src: 192.0.2.61

A Mpls Vpn Rd: 0:64499:11

SHOW

A Peer Ip Src

A Mpls Vpn Rd

A Ip Src

A Ip Dst

A Comms

Table

Peer Ip Src, Mpls Vpn Rd, Ip Src, Ip Dst, Comms

Number of Events

Overall

3

192.0.2.61

3

0:64499:11

3

203.0.113.255

2

203.0.113.253

2

64496:299_64496:1001_64496:1033_64497:1_64499:10

2

203.0.113.253

1

203.0.113.255

1

64496:299_64496:1001_64496:1033_64497:1_64499:20

1

*UDP Testflow between two IPv4 Addresses with
BMP BGP Local RIB dimensions measured on MPLS PE in a VRF*

Huawei - VRP

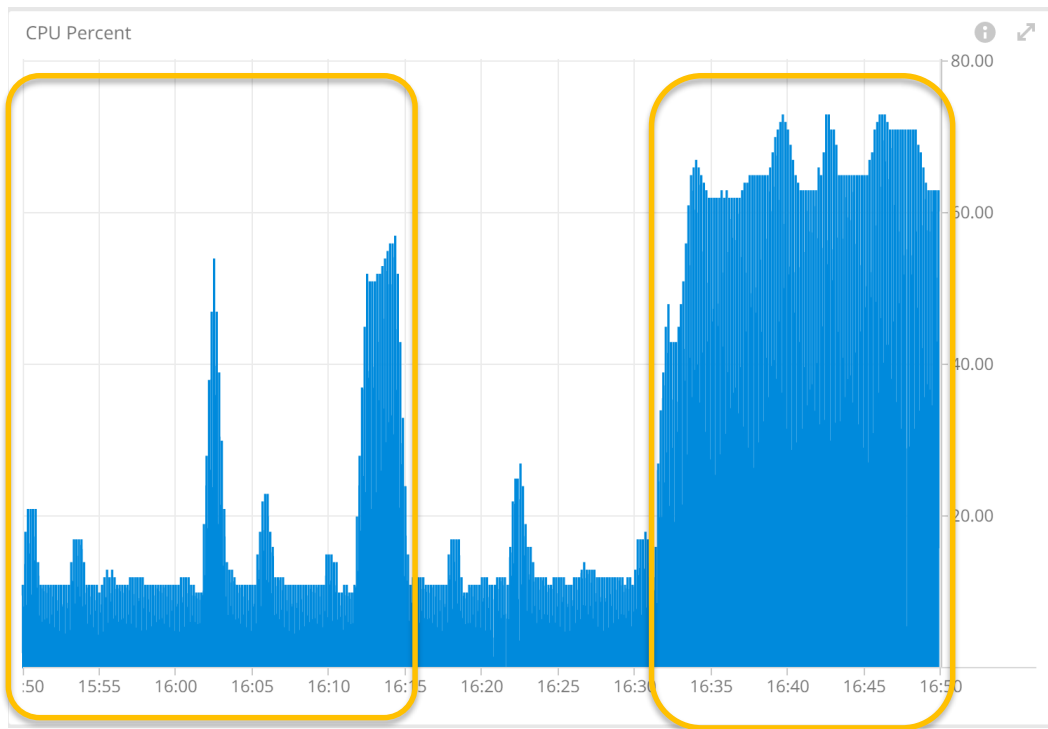
Achievements

- Supporting [draft-grow-bmp-tlv-00](#) and [draft-lucente-grow-bmp-tlv-ebit-00](#)
- Supporting path status of [draft-cppy-grow-bmp-path-marking-tlv-04](#) Supporting [draft-xu-grow-bmp-route-policy-attr-trace-04](#)
- Stress tests showing CPU and memory usage increase but no BGP propagation delay. CPU increase not to be realistic
- Wireshark dissector for route-policy tracing BMP message-type and route-monitoring path marking TLV

Next Steps

- Redo the BGP propagation delay tests with improved testbed
- Investigate BMP impact in CPU usage graph

BMP Stress Test – CPU usage



CPU usage monitoring of Router Reflector

Dataset:

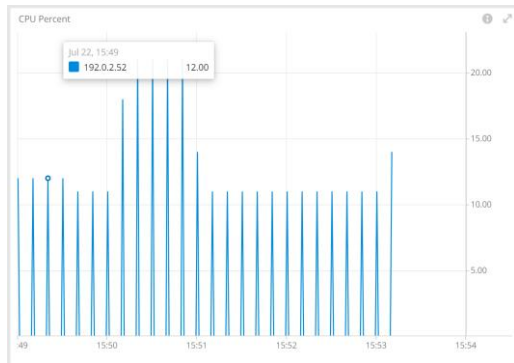
- Dataset 1: 100K routes from Spirent
- Dataset 2: 500K routes from Spirent
- Dataset 3: 1000K routes from Spirent

BMP disabled: 15:50 ~ 16:15

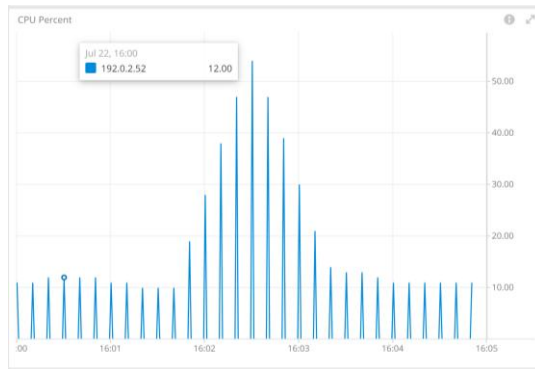
BMP enabled: 16:30 ~ 16:50

BMP Stress Test – CPU usage

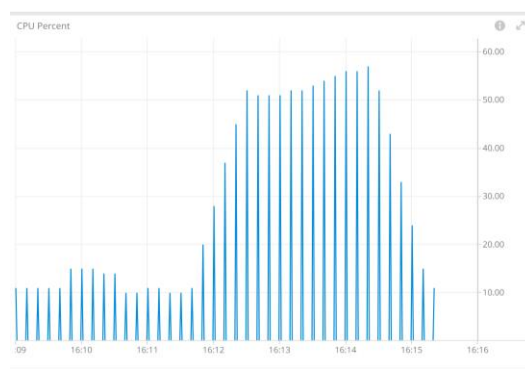
Before BMP enabled: 100K routes adv.



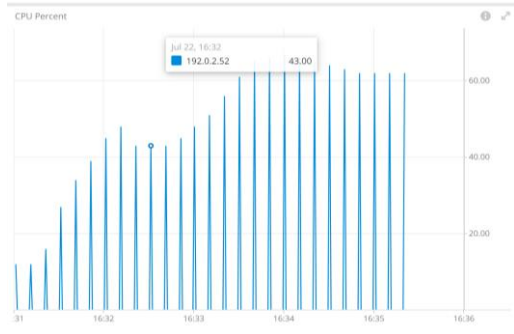
Before BMP enabled: 500K routes adv.



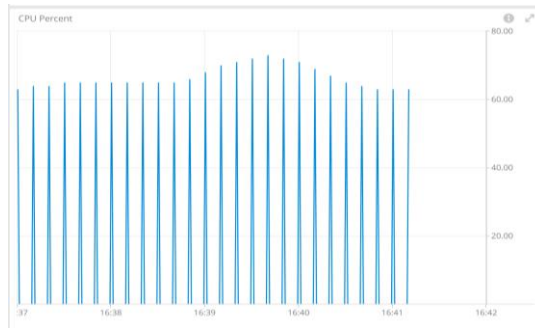
Before BMP enabled: 1000K routes adv.



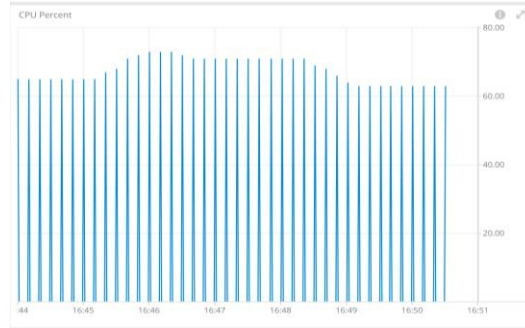
After BMP enabled: 100K routes adv.



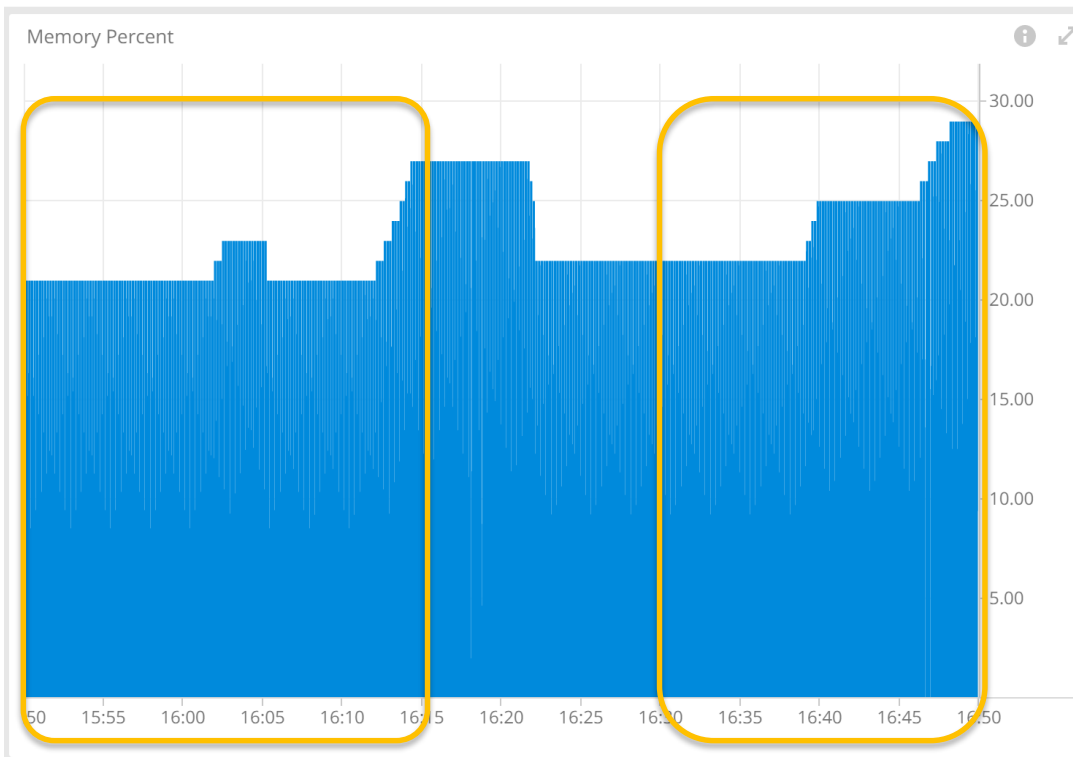
After BMP enabled: 500K routes adv.



After BMP enabled: 1000K routes adv.



BMP Stress Test – Memory Usage



Dataset:

- Dataset 1: 100K routes from Spirent
- Dataset 2: 500K routes from Spirent
- Dataset 3: 1000K routes from Spirent

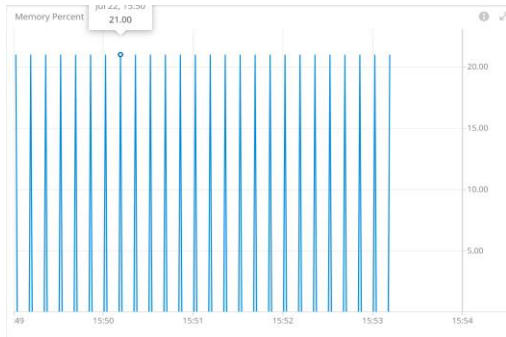
BMP disabled: 15:50 ~ 16:15

BMP enabled: 16:30 ~ 16:50

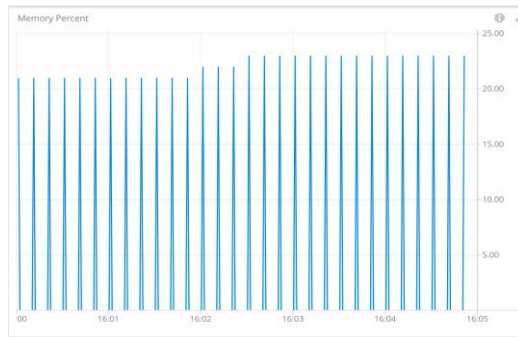
Memory usage monitoring of Router Reflector

BMP Stress Test – Memory Usage

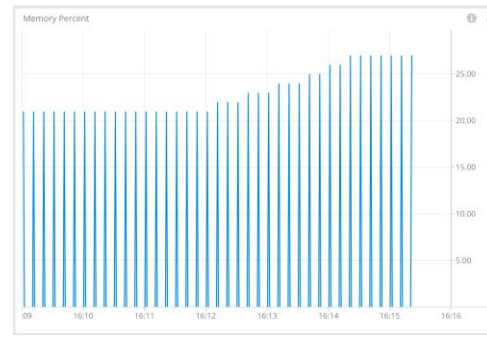
Before BMP enabled: 100K routes adv.



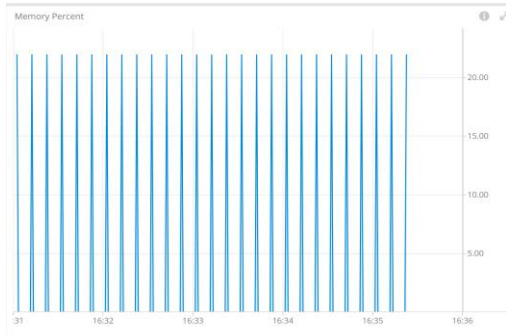
Before BMP enabled: 500K routes adv.



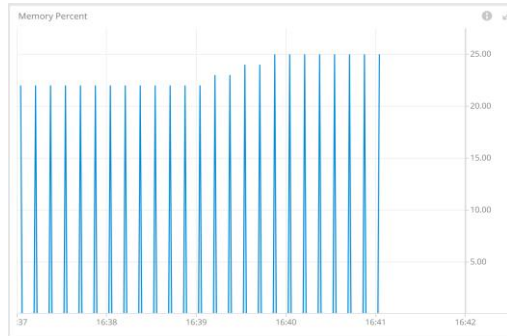
Before BMP enabled: 1000K routes adv.



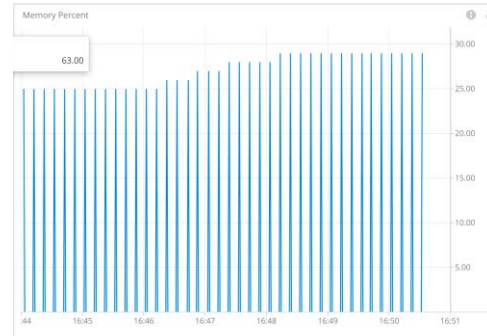
After BMP enabled: 100K routes adv.



After BMP enabled: 500K routes adv.



After BMP enabled: 1000K routes adv.



BMP Stress test – Convergence time

A very rough estimation of individual device RIB convergence time based on CPU stabilization

Dataset	Device	updates	Convergence time by clock (BMP disabled)	Convergence time by clock (BMP enabled)
Dataset 1:	RR: 10.215.184.52	100000	60 sec	60 sec
Dataset 2	RR: 10.215.184.52	500000	110 sec	120 sec
Dataset 3	RR: 10.215.184.52	1000000	220 sec	240 sec

BMP route-policy trace data visualization

Packet capture analysis tool showing a BMP route-policy trace. The packet list table is as follows:

No.	Time	Source	Destination	Protocol	Length	Info
34	4.274320	192.0.2.61	192.0.2.1	TCP	132	6312
35	4.274331	192.0.2.1	192.0.2.61	TCP	54	1790
36	5.184348	192.0.2.61	192.0.2.1	BMP ROFT Msg	241	BMP
37	5.184391	192.0.2.1	192.0.2.61	TCP	54	1790
38	5.184406	192.0.2.61	192.0.2.1	BMP ROFT Msg	1514	BMP
39	5.184418	192.0.2.1	192.0.2.61	TCP	54	1790
40	5.185120	192.0.2.61	192.0.2.1	TCP	651	6312

Frame 36: 241 bytes on wire (1928 bits), 241 bytes captured (1928 bits) on interface 0

Ethernet II, Src: HuaweiTe_ba:2c:e6 (20:65:8e:ba:2c:e6), Dst: VMWare_0e:d8:14 (00:0c:29:0e:d8:14)

Internet Protocol Version 4, Src: 192.0.2.61, Dst: 192.0.2.1

Transmission Control Protocol, Src Port: 63128, Dst Port: 1790, Seq: 9980, Ack: 1, Len: 187

BMP ROFT

BMP Version: 3

BMP MsgLength: 187

MsgType: BMP ROFT (100)

Data (181 bytes)

- Reserved: 0x00
- RD: 0x0000fb3000000015
- PrefixLen: 32
- zero-filled:
- prefix_ipv4: 203.0.113.20
- PeerRouterID: 0.0.0.0
- EventCount: 1
- EventTotalLen: 148
- SingleEventLen: 148
- EventIndex: 1
- Timestamp_Seconds: Jul 17, 2020 06:13:46.000838593 UTC
- PathIdentifier: 0
- PeerAFI: 1
- PeerSAFI: 1

VRF/Table Name TLV(11 bytes)

- TlvType: VRF/Table Name TLV (0)
- TlvLength: 7
- VrfID: 1
- VrfName: A10

Policy ID TLV(31 bytes)

Pre Policy Attribute TLV(80 bytes)

Post Policy Attribute TLV(4 bytes)

Optional TLV(4 bytes)

VRF/Table Name TLV(11 bytes)

TlvType: VRF/Table Name TLV (0)
TlvLength: 7
VrfID: 1
VrfName: A10

Policy ID TLV(31 bytes)

TlvType: Policy ID TLV (1)
TlvLength: 27
PolicyFlag: M = 0, P = 0, D = 0 (0x00)
PolicyCount: 0
PolicyClassification: VRF import (4)
zero-filled:
PeerAddress_ipv4: 0.0.0.0
PeerRouteID: 0.0.0.0
PeerAs: 65537

Pre Policy Attribute TLV(80 bytes)

TlvType: Pre Policy Attribute TLV (2)
TlvLength: 76

Path Attributes

- Path Attribute - ORIGIN:
 - Flags: 0x40, Transitive, Well-known, Complete (0x40)
 - Type Code: ORIGIN (1)
 - Length: 1
 - Type Code: IGP (0)
- Path Attribute - AS_PATH:
- Path Attribute - NEXT_HOP:
- Path Attribute - MED:
- Path Attribute - LOCAL PRE:
- Path Attribute - COMMUNITY:
- Path Attribute - EXTCOMMUNITY:

Post Policy Attribute TLV(4 bytes)

TlvType: Post Policy Attribute TLV (3)
TlvLength: 0

Optional TLV(4 bytes)

TlvType: Optional TLV (4)
TlvLength: 0

BMP path marking data visualization

Wireshark packet capture showing a BMP path marking message. The packet list shows a BMP path marking message (253 bytes) from 192.0.2.1 to 192.0.2.1. The packet details show the BMP non-roft_v3v4 structure, including the Per-Peer Header (42 bytes) and the Border Gateway Protocol - UPDATE Message (113 bytes).

Time	Source	Destination	Protocol	Length	Info
19.0008783	192.0.2.1	192.0.2.61	TCP	54	1790 → 63128 [ACK]
20.0009747	192.0.2.61	192.0.2.1	TCP	178	63128 → 1790 [PSH,]
21.0009756	192.0.2.1	192.0.2.61	TCP	54	1790 → 63128 [ACK]
22.4.190881	192.0.2.61	192.0.2.1	BMP pathmarking Msg	253	BMP pathmarking Msg
23.4.190910	192.0.2.1	192.0.2.61	TCP	54	1790 → 63128 [ACK]
24.4.191525	192.0.2.61	192.0.2.1	TCP	125	63128 → 1790 [PSH,]
25.4.191534	192.0.2.1	192.0.2.61	TCP	54	1790 → 63128 [ACK]
26.4.213714	192.0.2.61	192.0.2.1	BMP pathmarking Msg	257	BMP pathmarking Msg
27.4.213724	192.0.2.1	192.0.2.61	TCP	54	1790 → 63128 [ACK]

Frame 22: 233 bytes on wire (1864 bits), 233 bytes captured (1864 bits) on interface
Ethernet II, Src: HuaweiTe_ba:2c:e6 (20:65:8e:ba:2c:e6), Dst: VMware_0e:d8:14 (00:0c:29:0e:d8:14)
Internet Protocol Version 4, Src: 192.0.2.61, Dst: 192.0.2.1
Transmission Control Protocol, Src Port: 63128, Dst Port: 1790, Seq: 9091, Ack: 1, Len: 179
BMP_non-roft_v3v4
BMP Version: 4
BMP MsgLength: 179
MsgType: Route Monitoring (0)
Per-Peer Header(42 bytes)
Border Gateway Protocol - UPDATE Message(113 bytes)
Prefix Info TLV

Per-Peer Header(42 bytes)

Type: Unknown (3)

Flag: 1000 0000 = Flags: 0x80, Pre, In, IPv6 (0x80)

RD: 0x0000fbf300000029

peer address: ::

ASN: 65537

BGP ID: 192.0.2.61

Timestamp(sec): Jul 17, 2020 06:13:42.000000000 UTC

Timestamp(msec): 0

Border Gateway Protocol - UPDATE Message(113 bytes)

Marker: 000000000000000000000000

Length: 113

Type: UPDATE Message (2)

Withdrawn Routes Length: 0

Total Path Attribute: 90

Path Attribute

NLRI

Prefix Info TLV

tlv: Ip Prefix Info TLV (0x0000)

tlv Len: 14

Count: 1

Path Marking TLV

tlv: Path Marking IANA TLV (0x0001)

tlv Len: 8

PathStatusE: best, primary (0x0000000a)

ReasonCodeE: (0xffffffff)

Wireshark – BMP Dissector

Achievements

- Supporting [draft-xu-grow-bmp-route-policy-attr-trace-04](#) in latest [code commit](#)

Next Steps

- Support [draft-grow-bmp-tlv-00](#) and [draft-grow-bmp-tlv-ebit-00](#)
- Support [draft-cppy-grow-bmp-path-marking-tlv-04](#)

ETHZ – Livio Sgier

Achievements

- Setting up of end-to-end export/collection/visualization pipeline based on time-series database Druid
- D3.js visualization front-end for quick prototyping

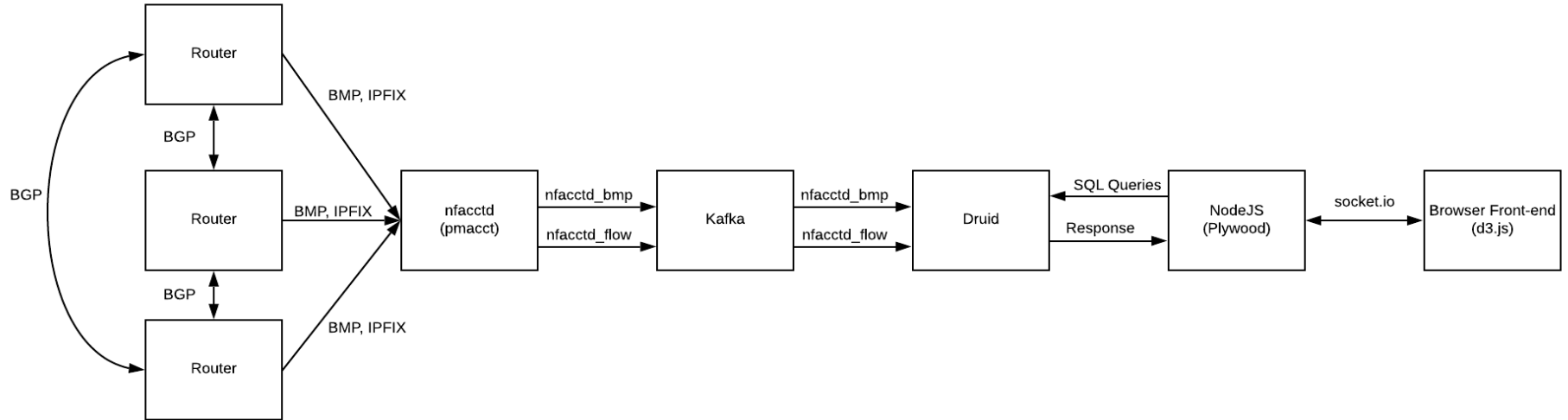
Next Steps

- Testing new visualization use-cases (L3 topology, VPN abstraction, control/data plane correlation, incorporating data from new drafts supplied by pmacct)

D3.js Front-end



ETHZ – Livio Sgier



End-to-End export/collection/visualization pipeline

What we learned

- Good

- Being virtual makes the BMP project more accessible to people
- Newcomers bring a fresh mindset and wonderful ideas into the team
 - BFD correlation to BMP peer_up/down message type
- YANG push CPU and memory with a 10 second, BMP with a second granularity improved insights into the performance impact

- Bad

- The missing beers and cocktails after 😊

Thanks to...

- Anurag Prakash - Ciena
- Hongwei Li - HPE
- Kian Jones - CENGN
- Alexis La Goutte – Wireshark
- Livio Sgier - ETHZ
- Yunan Gu - Huawei
- Binyang Huang - Huawei
- Paolo Lucente - NTT
- Heng Cui - Swisscom
- Matthias Arnold - Swisscom
- Thomas Graf - Swisscom

...[ImPLY](#) and Swisscom Time Analytics Platform team for providing us the big data and Huawei for the network environment.