BMP

BGP Monitoring Protocol
GROW WG

IETF 108

July 20-24th, 2020

Virtual Hackathon

Photo NASA

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

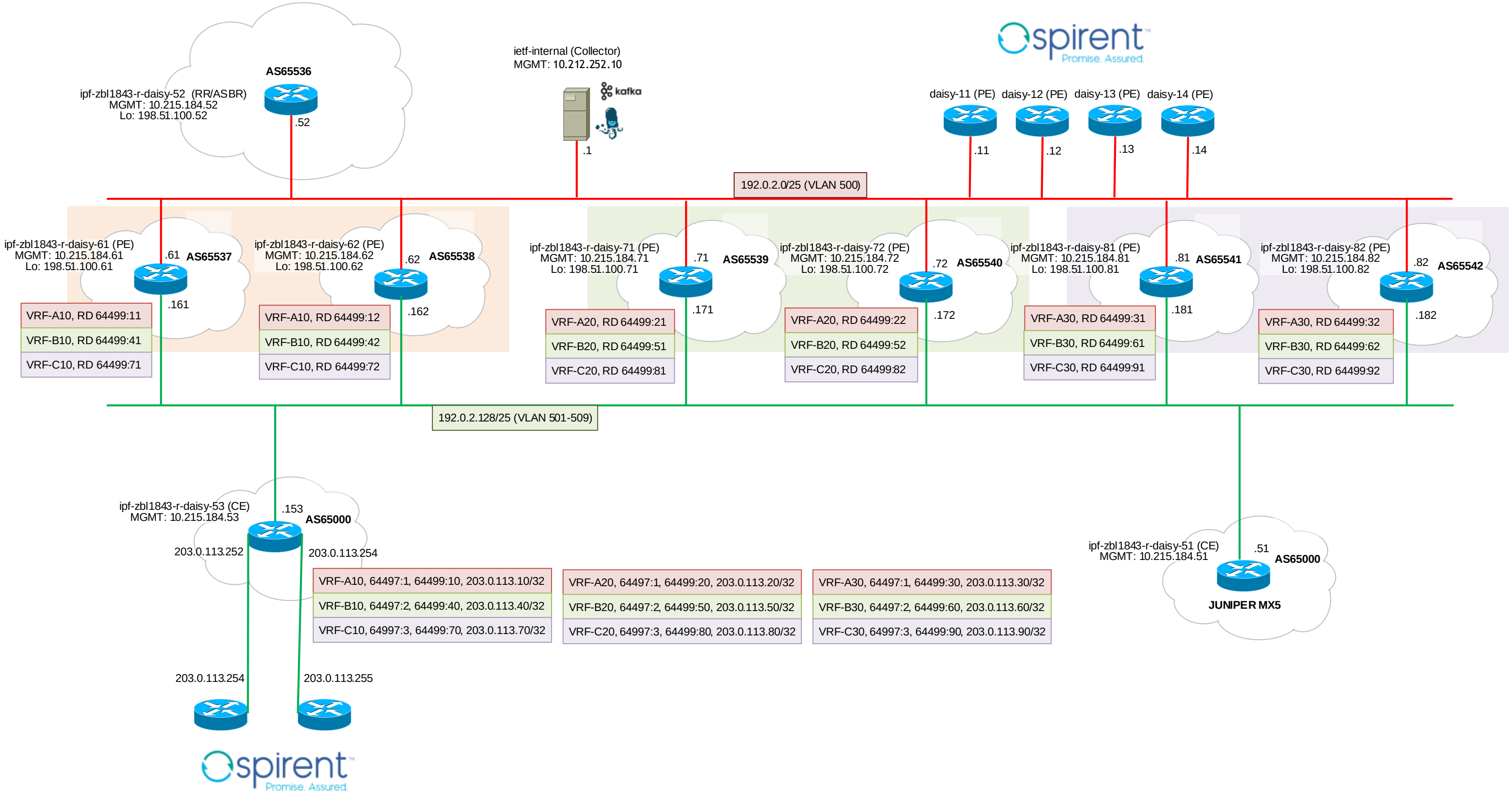
Software

- 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



AS65536

ipf-zbl1843-r-daisy-52 (RR/ASBR)
MGMT: 10.215.184.52
Lo: 198.51.100.52

.52

ietf-internal (Collector)
MGMT: 10.212.252.10

kafka

.1

spirent
Promise. Assured.

daisy-11 (PE) daisy-12 (PE) daisy-13 (PE) daisy-14 (PE)

.11 .12 .13 .14

192.0.2.0/25 (VLAN 500)

ipf-zbl1843-r-daisy-61 (PE)
MGMT: 10.215.184.61
Lo: 198.51.100.61

.61 AS65537

.161

VRF-A10, RD 64499:11
VRF-B10, RD 64499:41
VRF-C10, RD 64499:71

ipf-zbl1843-r-daisy-62 (PE)
MGMT: 10.215.184.62
Lo: 198.51.100.62

.62 AS65538

.162

VRF-A10, RD 64499:12
VRF-B10, RD 64499:42
VRF-C10, RD 64499:72

ipf-zbl1843-r-daisy-71 (PE)
MGMT: 10.215.184.71
Lo: 198.51.100.71

.71 AS65539

.171

VRF-A20, RD 64499:21
VRF-B20, RD 64499:51
VRF-C20, RD 64499:81

ipf-zbl1843-r-daisy-72 (PE)
MGMT: 10.215.184.72
Lo: 198.51.100.72

.72 AS65540

.172

VRF-A20, RD 64499:22
VRF-B20, RD 64499:52
VRF-C20, RD 64499:82

ipf-zbl1843-r-daisy-81 (PE)
MGMT: 10.215.184.81
Lo: 198.51.100.81

.81 AS65541

.181

VRF-A30, RD 64499:31
VRF-B30, RD 64499:61
VRF-C30, RD 64499:91

ipf-zbl1843-r-daisy-82 (PE)
MGMT: 10.215.184.82
Lo: 198.51.100.82

.82 AS65542

.182

VRF-A30, RD 64499:32
VRF-B30, RD 64499:62
VRF-C30, RD 64499:92

192.0.2.128/25 (VLAN 501-509)

ipf-zbl1843-r-daisy-53 (CE)
MGMT: 10.215.184.53

.153 AS65000

203.0.113.252 203.0.113.254

VRF-A10, 64497:1, 64499:10, 203.0.113.10/32	VRF-A20, 64497:1, 64499:20, 203.0.113.20/32	VRF-A30, 64497:1, 64499:30, 203.0.113.30/32
VRF-B10, 64497:2, 64499:40, 203.0.113.40/32	VRF-B20, 64497:2, 64499:50, 203.0.113.50/32	VRF-B30, 64497:2, 64499:60, 203.0.113.60/32
VRF-C10, 64997:3, 64499:70, 203.0.113.70/32	VRF-C20, 64997:3, 64499:80, 203.0.113.80/32	VRF-C30, 64997:3, 64499:90, 203.0.113.90/32

ipf-zbl1843-r-daisy-51 (CE)
MGMT: 10.215.184.51

.51 AS65000

JUNIPER MX5

203.0.113.254 203.0.113.255

spirent
Promise. Assured.

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

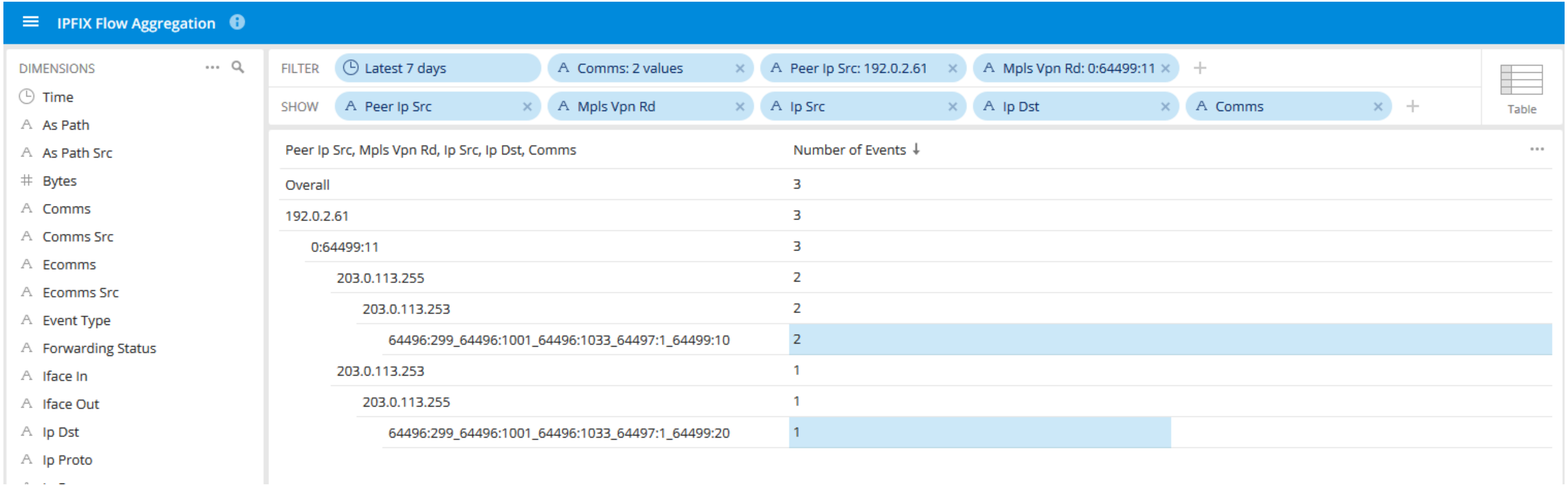
- BMP BGP Local RIB to IPFIX correlation now works for prefixes with BGP route-distinguisher as well.
- 2 of 5 TLV's decoded of draft-xu-grow-bmp-route-policy-attr-trace

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



FILTER: Latest 7 days | Comms: 2 values | Peer Ip Src: 192.0.2.61 | Mpls Vpn Rd: 0:64499:11

SHOW: Peer Ip Src | Mpls Vpn Rd | Ip Src | Ip Dst | Comms

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

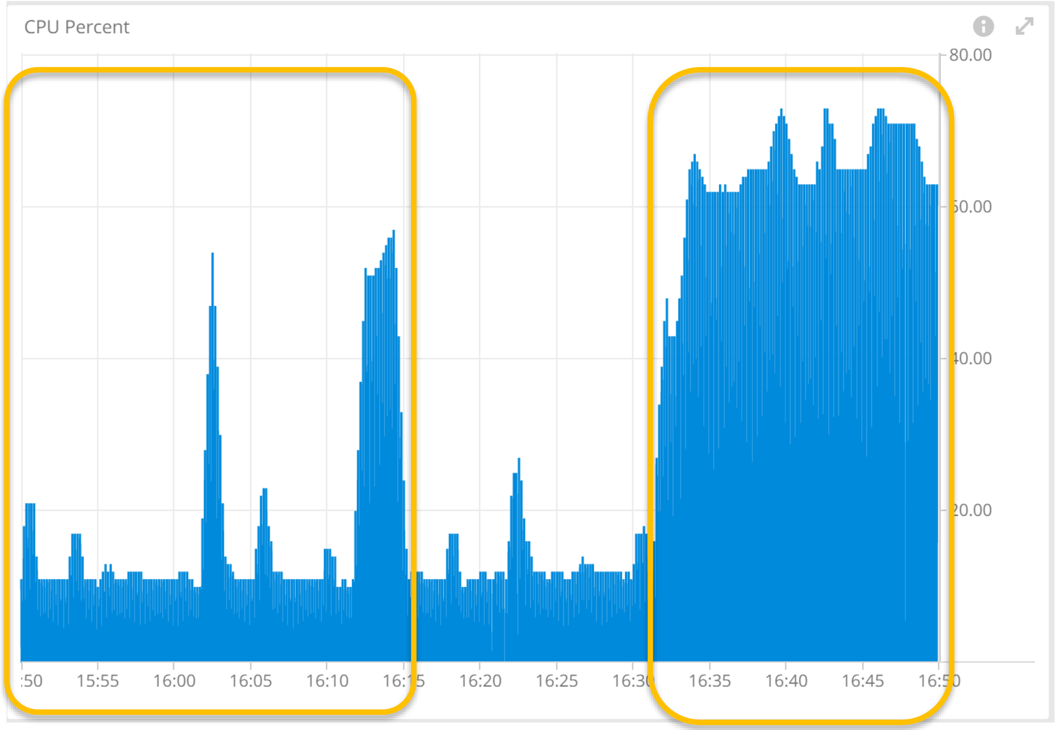
Achievements

- 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 expected CPU and memory usage increase but no BGP propagation delay.
- 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.

BMP Stress Test – CPU usage



CPU usage monitoring of Router Reflector

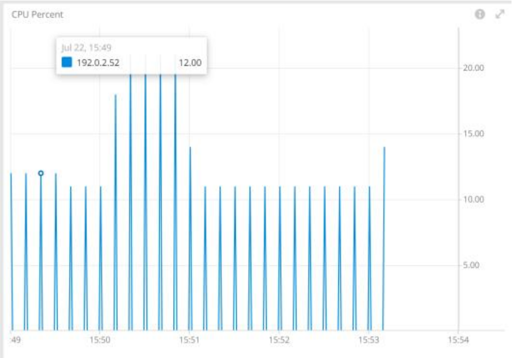
Dataset:

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

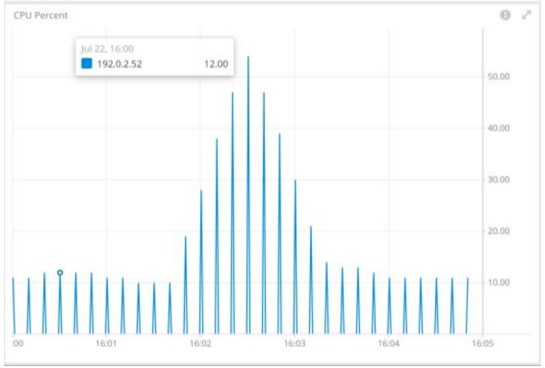
BMP disabled: 15:50 ~ 16: 15
BMP enabled: 16:30 ~ 16:50

BMP Stress Test – CPU usage

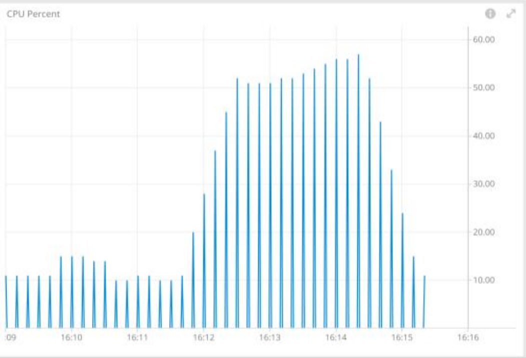
Before BMP enabled: 100K routes adv.



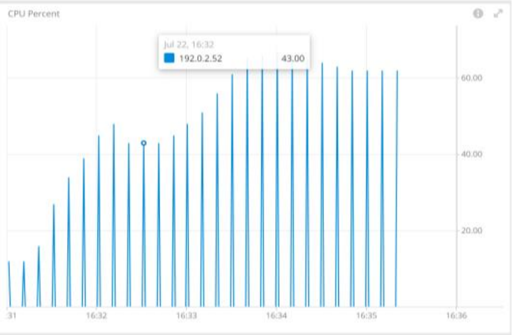
Before BMP enabled: 500K routes adv.



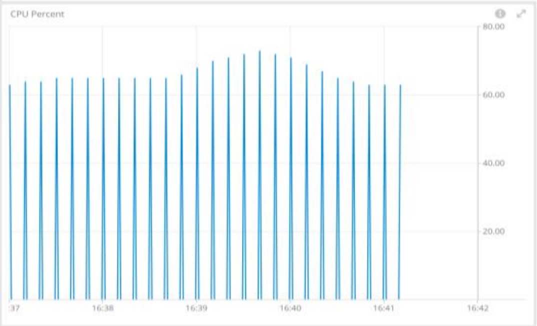
Before BMP enabled: 1000K routes adv.



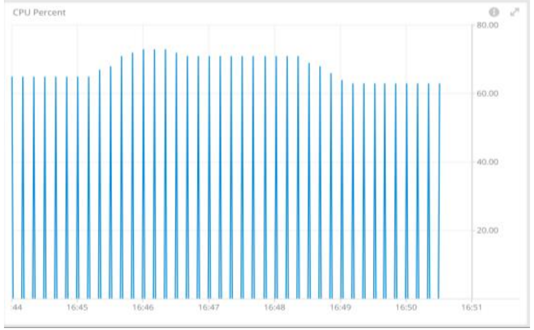
After BMP enabled: 100K routes adv.



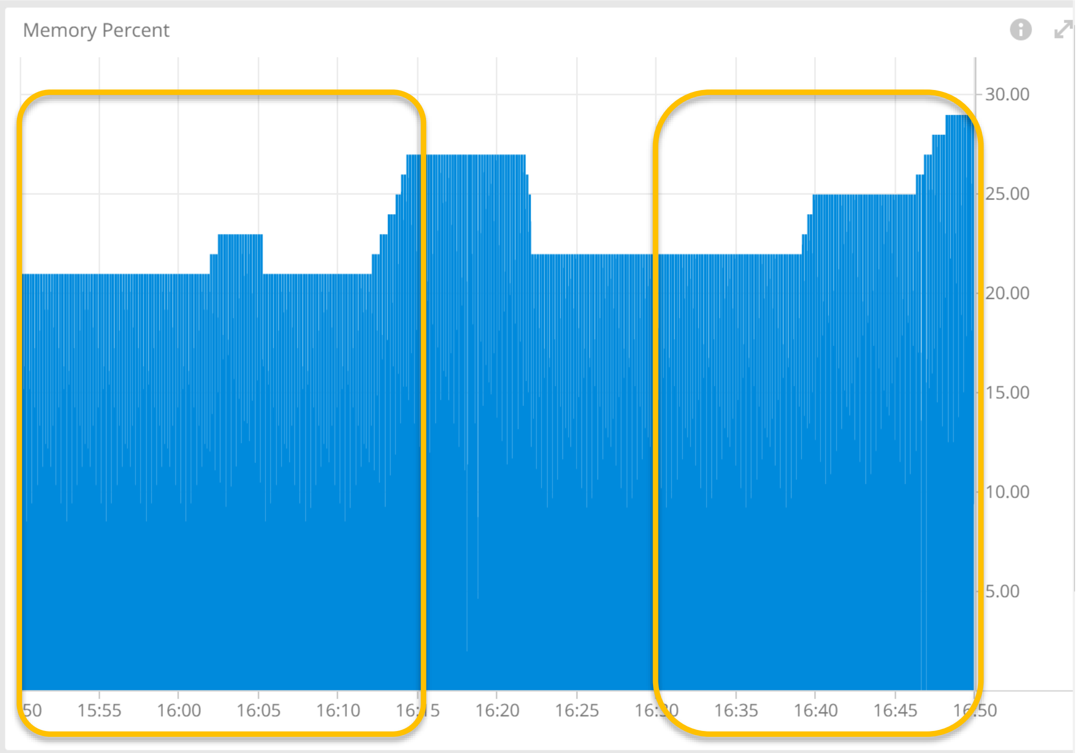
After BMP enabled: 500K routes adv.



After BMP enabled: 1000K routes adv.



BMP Stress Test – Memory Usage



Memory usage monitoring of Router Reflector

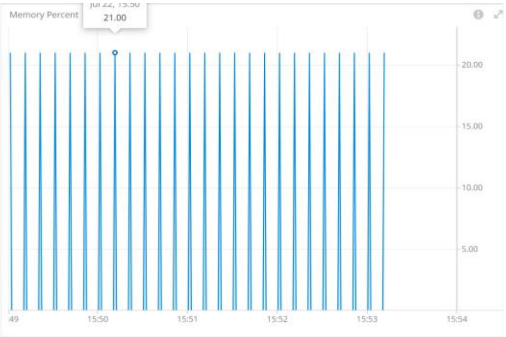
Dataset:

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

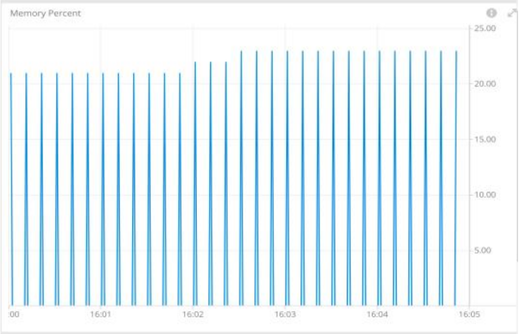
BMP disabled: 15:50 ~ 16: 15
BMP enabled: 16:30 ~ 16:50

BMP Stress Test – Memory Usage

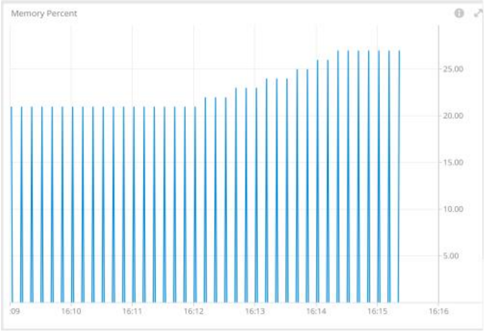
Before BMP enabled: 100K routes adv.



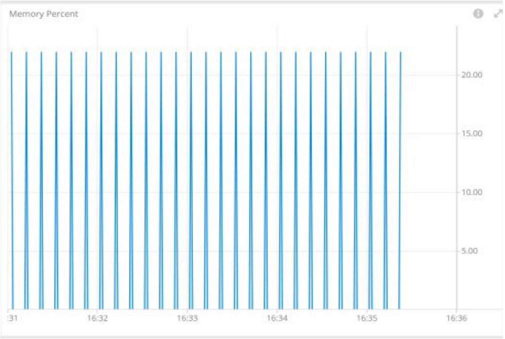
Before BMP enabled: 500K routes adv.



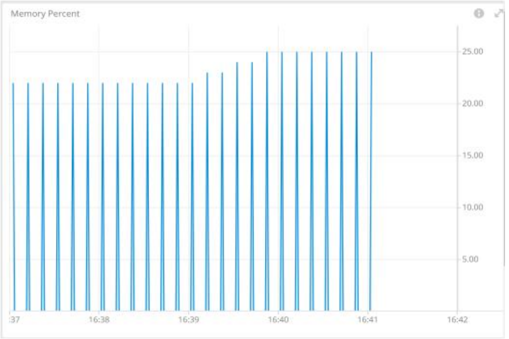
Before BMP enabled: 1000K routes adv.



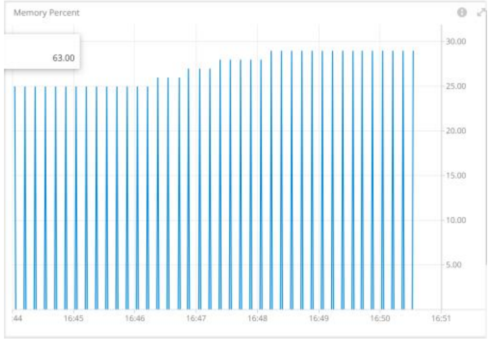
After BMP enabled: 100K routes adv.



After BMP enabled: 500K routes adv.



After BMP enabled: 1000K routes adv.

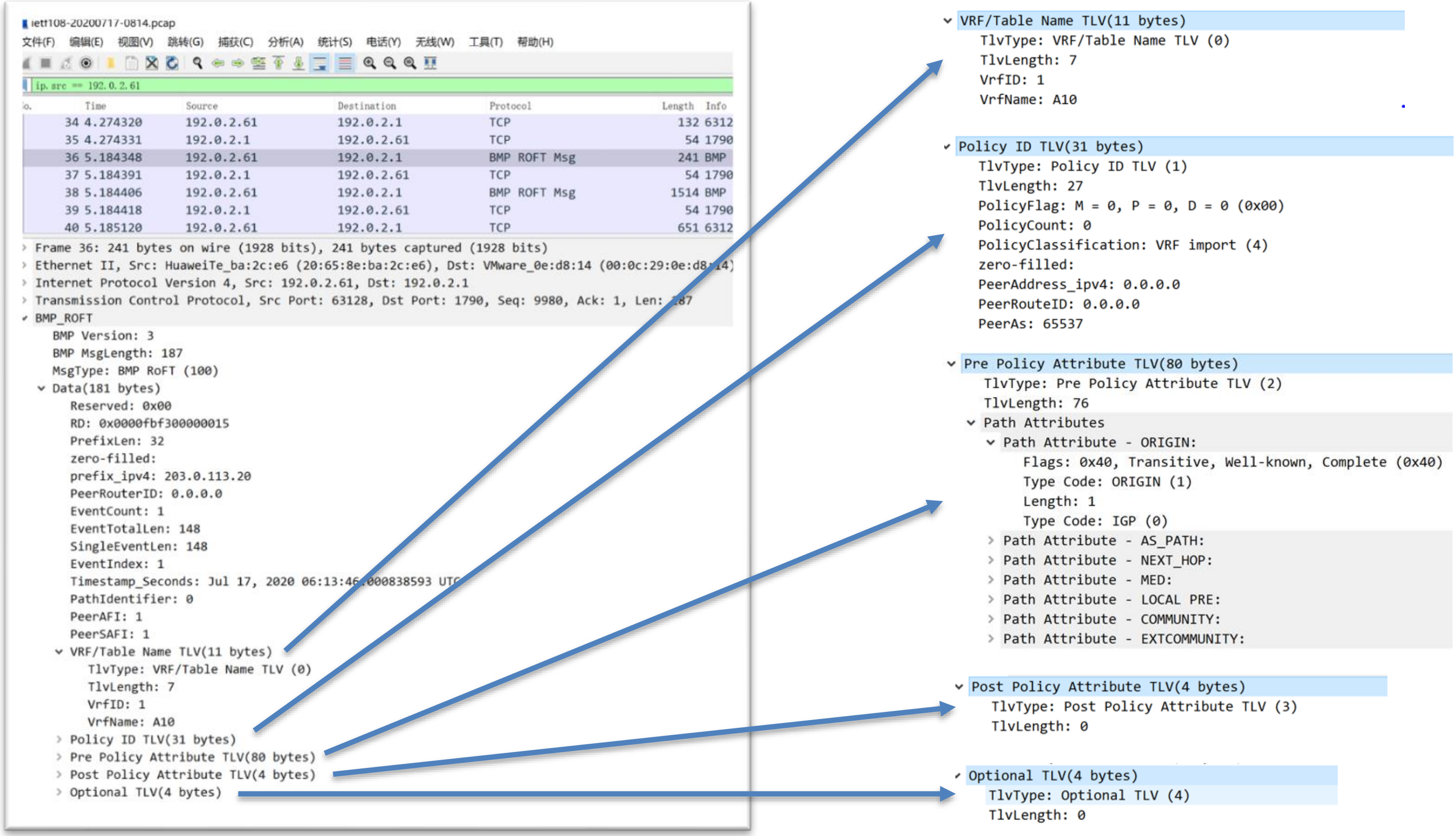


BMP Stress test – Convergence time

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



ietf108-20200717-0814.pcap
文件(F) 编辑(E) 视图(V) 跳转(G) 捕获(C) 分析(A) 统计(S) 电话(Y) 无线(W) 工具(T) 帮助(H)

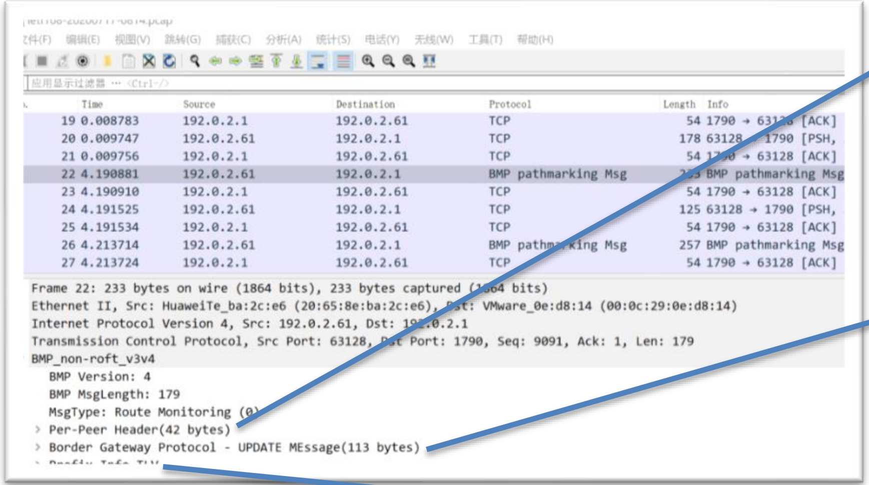
```
ip.src == 192.0.2.61
```

```
o.    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)
> 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
v BMP_ROFT
    BMP Version: 3
    BMP MsgLength: 187
    MsgType: BMP RoFT (100)
  v Data(181 bytes)
      Reserved: 0x00
      RD: 0x0000fbf300000015
      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
    v 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)
```

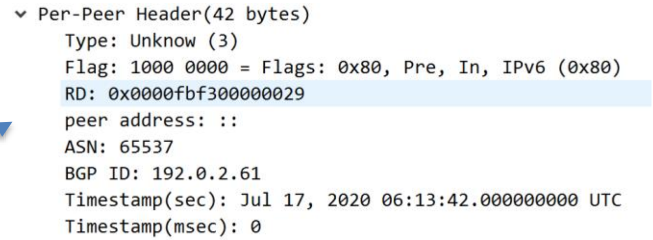
```
v VRF/Table Name TLV(11 bytes)
    TlvType: VRF/Table Name TLV (0)
    TlvLength: 7
    VrfID: 1
    VrfName: A10

v 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

v Pre Policy Attribute TLV(80 bytes)
    TlvType: Pre Policy Attribute TLV (2)
    TlvLength: 76
  v Path Attributes
    v 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:

v Post Policy Attribute TLV(4 bytes)
    TlvType: Post Policy Attribute TLV (3)
    TlvLength: 0

v Optional TLV(4 bytes)
    TlvType: Optional TLV (4)
    TlvLength: 0
```

BMP path marking data visualization

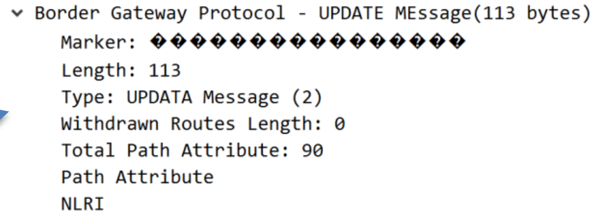


文件(F) 编辑(E) 视图(V) 跳转(G) 捕获(C) 分析(A) 统计(S) 电话(Y) 无线(W) 工具(T) 帮助(H)
应用显示过滤器 … <Ctrl-/>

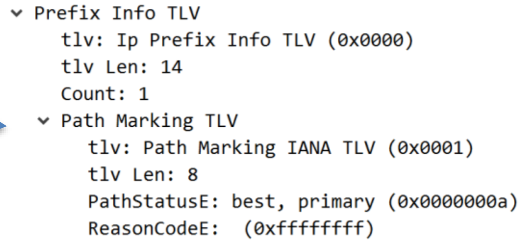
```
   Time       Source       Destination   Protocol              Length Info
19 0.008783   192.0.2.1    192.0.2.61    TCP                      54 1790 → 63128 [ACK]
20 0.009747   192.0.2.61   192.0.2.1     TCP                     178 63128 → 1790 [PSH,
21 0.009756   192.0.2.1    192.0.2.61    TCP                      54 1790 → 63128 [ACK]
22 4.190881   192.0.2.61   192.0.2.1     BMP pathmarking Msg     233 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)
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: Unknow (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: ����������������
    Length: 113
    Type: UPDATA 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

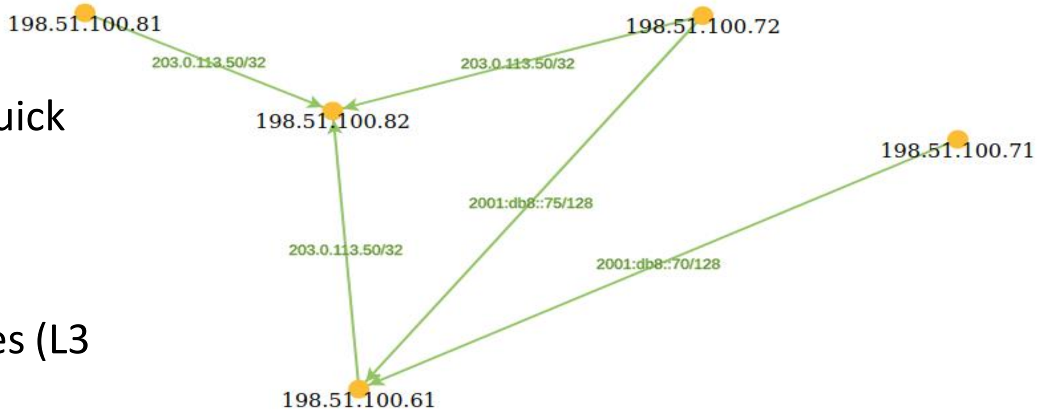
Achievements

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

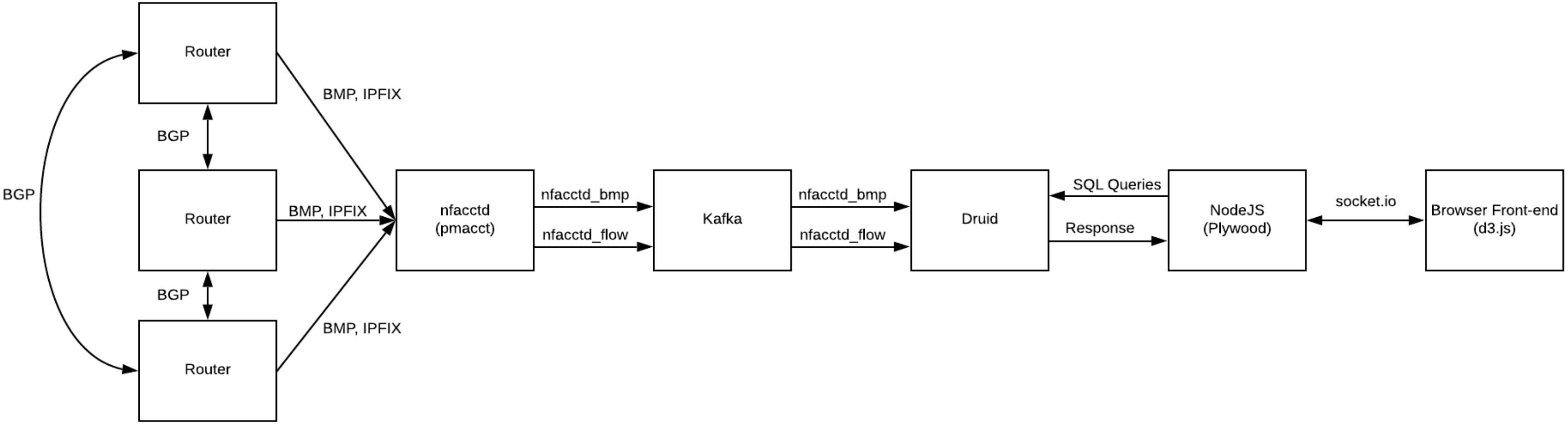
Next Steps

- 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…

- Prakash Anurag - 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.