# BFD going down… from BGP timers expiring?

**Louis Jarasius**
Senior Network Engineer
5G Networks
AusNOG 2025

# About me

- 6 years of making and breaking networks
- Some of the ASNs involved:
  - Connected Australia (AS132113)
  - Over the Wire (AS9268)
  - NEXTDC (AS56263)
  - Aussie Broadband - Corporate (AS136943)
  - VicTrack (AS132581)
  - TAFE Queensland (AS135396)
  - Gigafy (AS136972)
  - 5G Networks (AS63956)

# About 5G Networks

- IP transit
- Wavelengths
- Ethernet
- Data Centres
- Dark Fibre
- Last-mile Access Networks
- Cloud
- Managed Services
- Cyber Security


https://5gnetworks.au

https://bgp.tools/as/63956

https://lg.5gn.com.au

# What is the problem being solved?

5G Networks uses 3 primary sites to deliver internal and customer compute services, one in Sydney (SDC), one in Melbourne (MDC), and one in Adelaide (ADC). These networks are critical for business continuity, however are in need of a refresh after running successfully for many years.

- Ageing compute network infrastructure
- Spanning tree
- Lots of manual configuration for MACs
- MC-LAGs using virtual chassis
- Limited visibility
- Inconsistent design across facilities

# Our solution

We decided that it would be best to go back to the drawing board, rather than try and keep the existing network. This allows us to tackle all of the key issues, whilst also taking advantage of newer protocols and faster interface speeds.

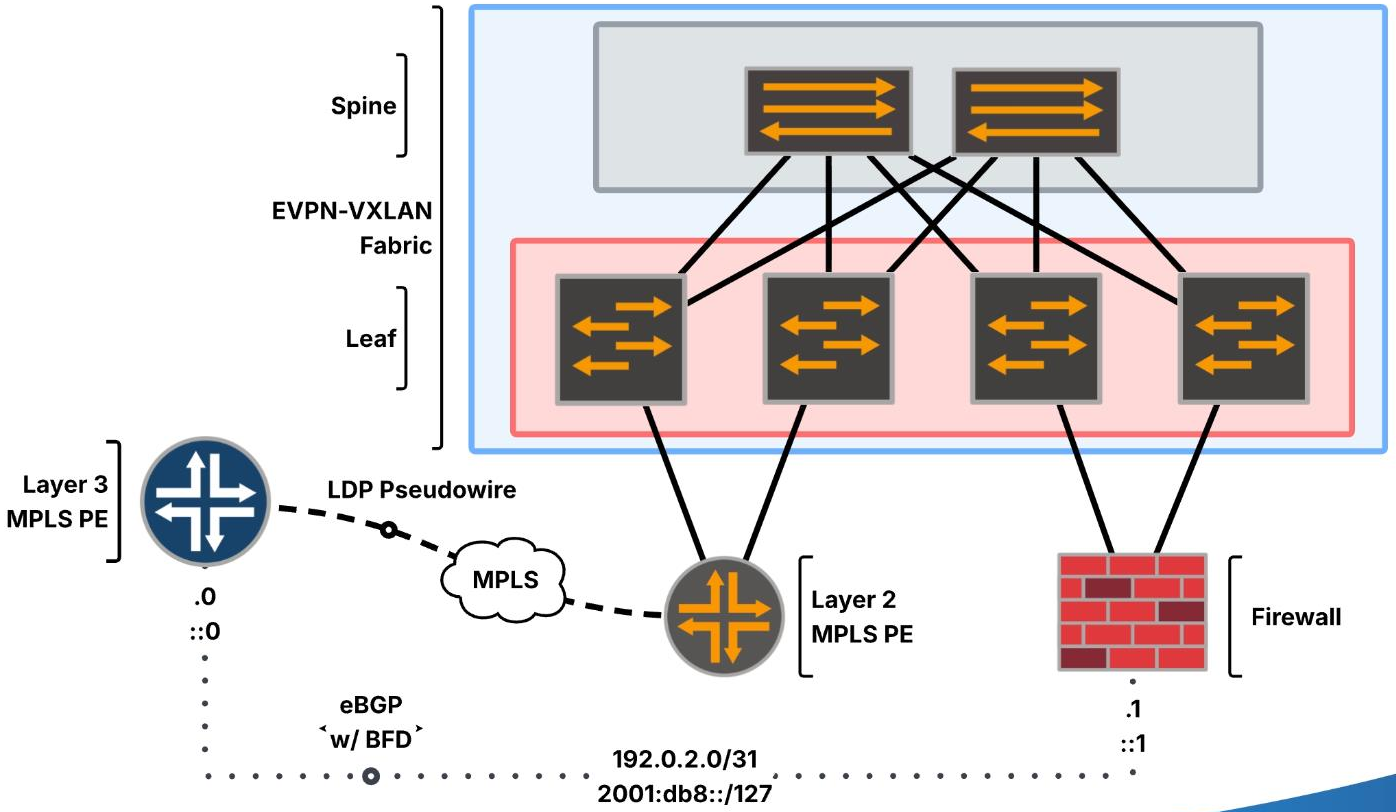
- EVPN-VXLAN
- Standardising the design
- Hardware refresh
- ESI
- Orchestration tooling
- Telemetry streaming

# Notes about VXLAN and hardware

Devices:
https://www.juniper.net/documentation/us/en/software/apstra6.0/apstra-user-guide/topics/topic-map/devices-qualified.html

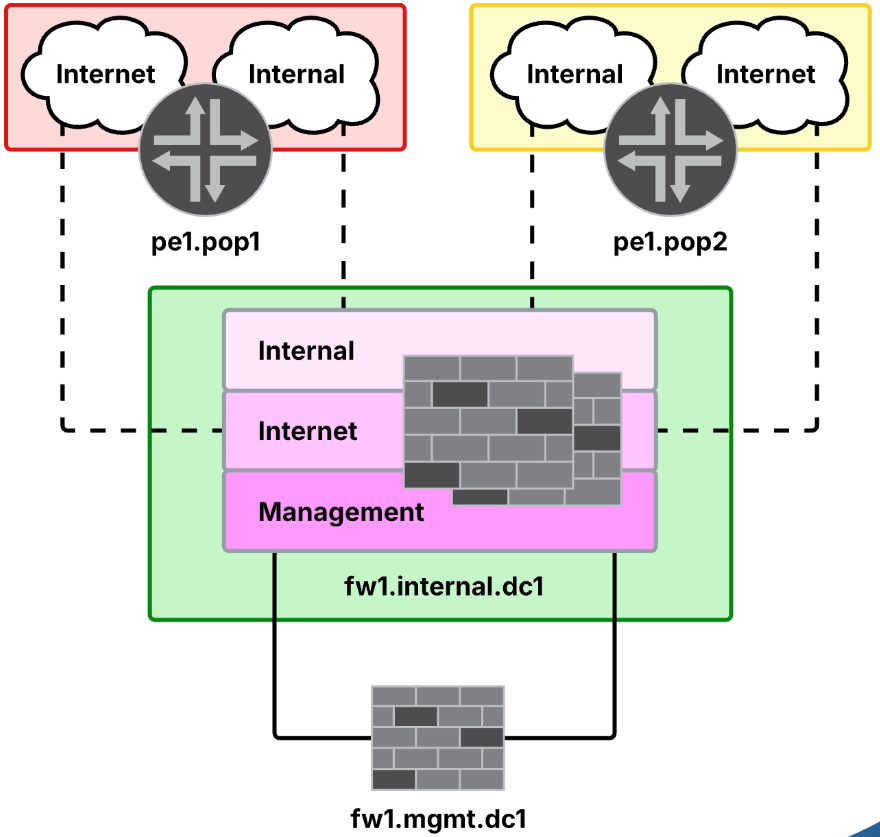
Chipsets:
https://www.juniper.net/documentation/us/en/software/apstra6.0/apstra-user-guide/topics/topic-map/evpn-support-addendum.html

# Upgrades people, upgrades



# Implementation

- Rack and stack
- OOB connectivity
- Stand up migration VRF
- Build new links
- Initial configuration
- Software upgrades
- Validation



# "BGP: Bloody Good Protocol" - Mark Turner

- I like dynamic routing and I cannot lie
- You other methods can’t deny

BGP with some tighter BFD timers is our solution of choice here. eBGP sessions, with private ASNs works well for our use case.

Any public IP allocations that get advertised back into our internet VRF will not have the private ASNs advertised anywhere else, as we nail down our aggregates using BGP communities and only export on a match in the policy.

# Testing

- Remove VLAN from Apstra
- Simulate a configuration mishap
- Transparent to the devices at either end

What could go wrong?


NO CONNECTIVITY?
imgflip.com

# Hello? Hello? Hello?

```
louisj@pe1.pop1> show bfd session address 192.0.2.1 extensive
                                                Detect   Transmit
Address                  State     Interface    Time     Interval  Multiplier
192.0.2.1                Up        irb.1        0.300     0.100        3
 Client BGP, TX interval 0.100, RX interval 0.100
 Session up time 00:03:59
 Local diagnostic None, remote diagnostic None
 Remote state Up, version 1
 Session type: Single hop BFD
 Min async interval 0.100, min slow interval 1.000
 Adaptive async TX interval 0.100, RX interval 0.100
 Local min TX interval 0.100, minimum RX interval 0.100, multiplier 3
 Remote min TX interval 0.100, min RX interval 0.100, multiplier 3
 Local discriminator 52, remote discriminator 37
 Echo TX interval 0.000, echo detection interval 0.000
 Echo mode disabled/inactive Session ID: 0×5f4

1 sessions, 1 clients
Cumulative transmit rate 10.0 pps, cumulative receive rate 10.0 pps
```

It's… still up?

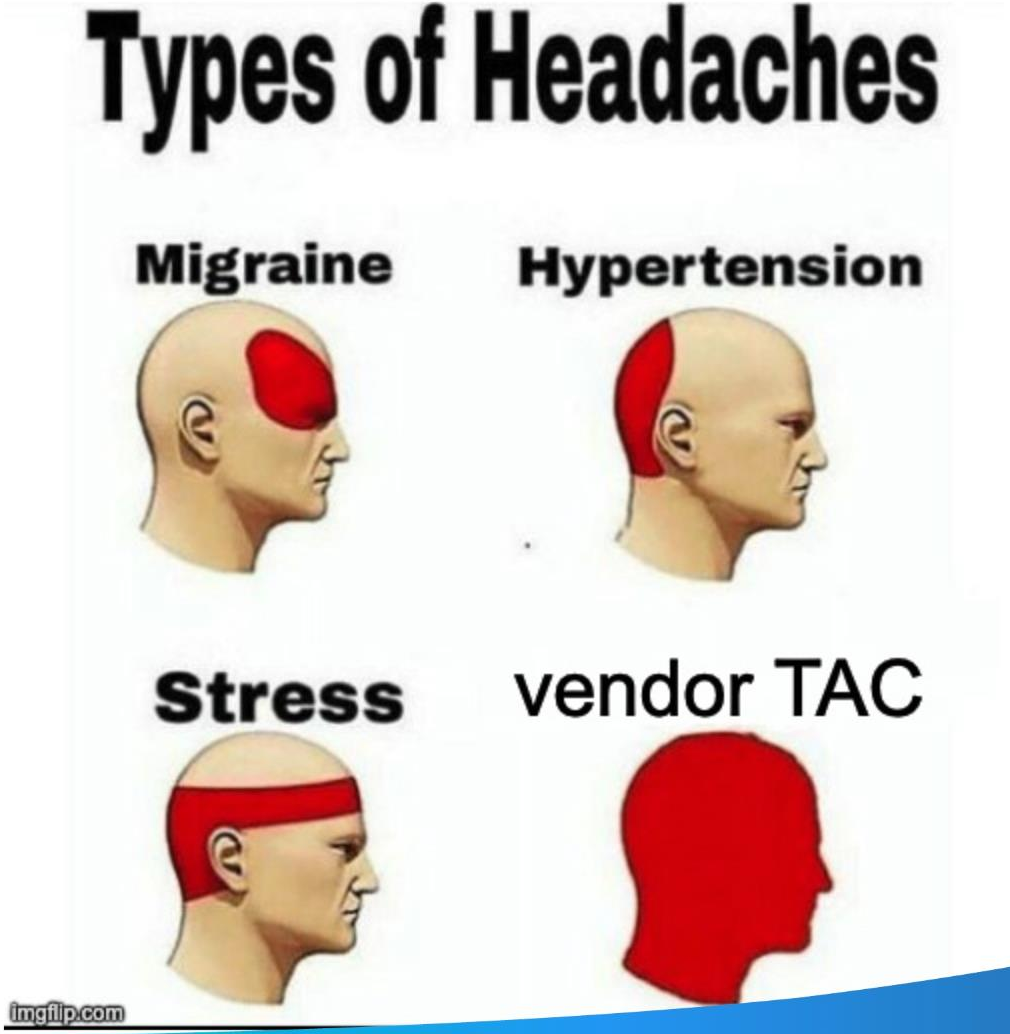
# Hello? Hello? Hello?

```
louisj@pe1.pop1> show bfd session address 192.0.2.1 extensive

0 sessions, 0 clients
Cumulative transmit rate 0.0 pps, cumulative receive rate 0.0 pps
```

Well, that's definitely a lot longer than 300ms.

# Types of Headaches



# Vendor TAC fun

```
rpd[30554]: BGP_IO_ERROR_CLOSE_SESSION: BGP peer 192.0.2.1 (External AS 65032):
Error event Operation timed out(60) for I/O session - closing it (instance {{ vrf }})

bfdd[29629]: BFDD_STATE_UP_TO_DOWN: BFD Session 192.0.2.1 (IFL 393) state Up → Down
LD/RD(46/19) Up time:00:03:12 Local diag: AdminDown Remote diag: None Reason: Received Upstream Destroy Session.

bfdd[29629]: BFDD_TRAP_SHOP_STATE_DOWN: local discriminator: 46, new state: down,
interface: irb.1, peer addr: 192.0.2.1
```

That order doesn't seem right to me. If BGP is taking down my BFD session, there's no real point in having it turned on. Surely a bug, yes?


VENDOR TAC
BOTTOM TEXT
imgflip.com

# Vendor TAC fun

```
Received Downstream RcvPkt (19) len 110:
  IfIndex (3) len 4: 393
  Protocol (1) len 1: BFD
  SrcAddr (5) len 8: 192.0.2.1
  Data (9) len 24: (hex) {{ data }}
  PktError (26) len 4: 0
  RtblIdx (24) len 4: 16
  MultiHop (64) len 1: (hex) 00
  Seamless (245) len 1: 0
  Unknown (213) len 1: (hex) 00
  Unknown (261) len 4: (hex) 00 00 0e c8
  Unknown (168) len 1: (hex) 00
  Authenticated (121) len 1: (hex) 0
```

TAC: "Our MX side's BFD is single hop, Fortigate bfd is multi hop."

Me: …


OL' RELIABLE
"PLEASE ESCALATE"
imgflip.com

# Vendor TAC fun

```
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD packet from 192.0.2.1 (IFL 393, rtbl 16, ttl 255) absorbed
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD packet from 192.0.2.1 (IFL 393, rtbl 16, ttl 255) absorbed
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD packet from 192.0.2.1 (IFL 393, rtbl 16, ttl 255) absorbed
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
PPM Trace: BFD periodic xmit to 192.0.2.1 (IFL 393, rtbl 16, single-hop port)
```

PR1845344 raised. Making some progress.


we
have
replicated
your
bug
imgflip.com

# Vendor TAC fun

> As per engineering team,it seems to be a limitation.
>
> l2interface of the IRB - VPLS routing instance is "lsi.1048576" pseudo interface does not have the FPC address associated!
> A single hop BFD session over IRB interface would not have pfe addr, if the VPLS instance the IRB belongs to has only LSI interfaces bound to VPLS pseudowires and has no local non-tunnel attachment circuits.
>
> We can have dummy interface(pysical interface) added to such vpls routing-instance to make it work!

TL;DR - you need a physical attachment interface for this to work.

# Vendor TAC fun

Day `x` :

> Engineering team suggested to test below knob which will change all BFD session from distributed mode to centralized mode
>
> ```
> set routing-options ppm no-delegate-processing-irb
> ```

Day `x+1` :

> We checked with below knob on lab devices however issue still persisted.

Well that was short lived.

# Vendor TAC fun

> We had multiple session with engineering team and extensively worked on LAB device and we found that issue is resolved after "restart PPM" without any configuration addition

Kick the PPM daemon and it’s all good? That’s a little bit of a backflip from it not being supported at all.

# Vendor TAC fun

> As per multiple testing we performed, issue is seen when device is upgraded but issue gets cleared when device is once rebooted or PPM restart after the upgrade and not seen again until device is again upgraded to same version.
>
> Also as requested by engineering team, we tested on latest DCB and issue is not seen.

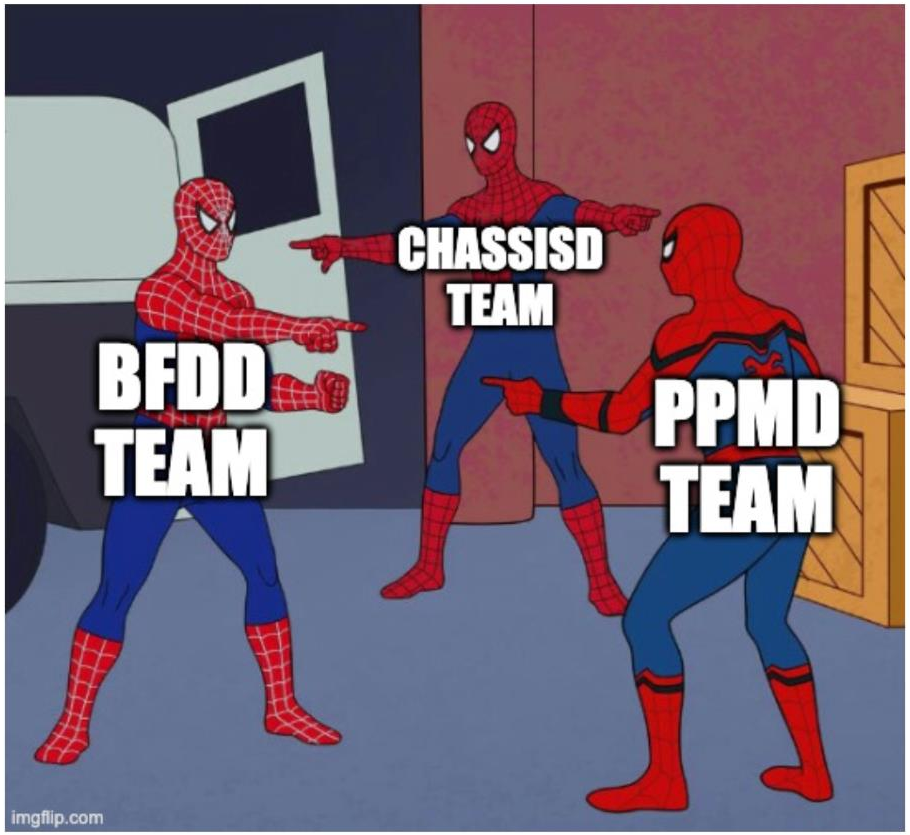
Wait, it’s a single restart of the PPM daemon and it’s fixed for good? Sounds like the bug has a fix somewhere too in the latest engineering builds.


TAC
PPMD
REBOOT
imgflip.com

# Vendor TAC fun

> As per engineering team, they are not able to find PR or RLI through which fix is added on latest releases.

That’s not very reassuring, but at least it is fixed.


CHASSISD
TEAM
BFDD
TEAM
PPMD
TEAM
imgflip.com

# "How issue is fixed in newer release is unidentified"

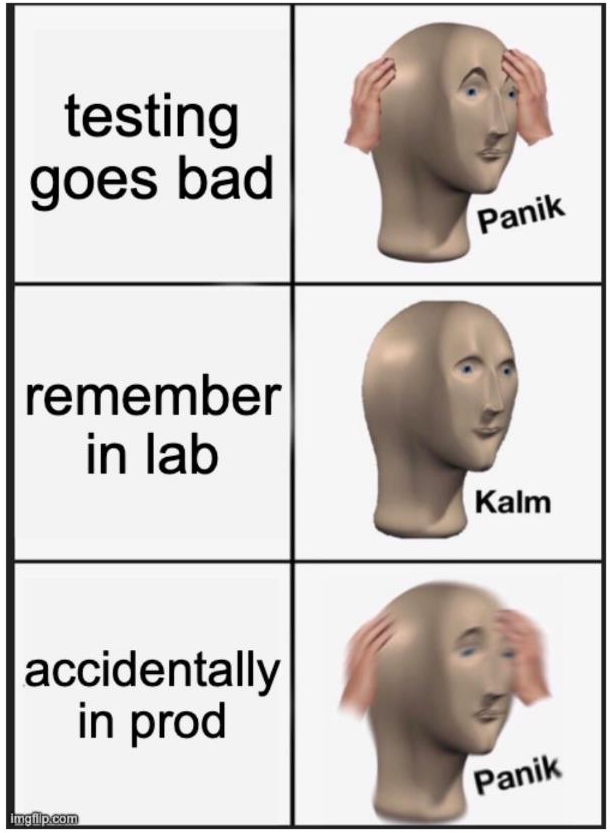
KB97136 - Juniper Networks Knowledge Base

https://supportportal.juniper.net/s/article/Singlehop-BFD-session-for-BGP-client-from-IRB-interface-of-VPLS-with-no-CE-facing-physical-interface-remain-up-after-connectvity-lost


first boot
should be the same
as every other boot
CHANGE MY MIND
imgflip.com

# Lessons learned

- Test as early in the cycle as possible
- Expect things to go wrong, and go looking to prove otherwise
- Simulate real failure conditions
- Engage with industry peers for new perspectives



# Questions?

ausnog@louis.jarasi.us or louisj@5gn.com.au

https://github.com/ljarasius/ausnog-2025

Special thanks to:

- Sohan @ Juniper
- Alain @ Juniper
- Brad @ 5G Networks
- Chris @ 5G Networks
- Dan @ 5G Networks
- Kenny @ 5G Networks
- The BBL crew