SONiC Readiness for SRv6 uSID deployment

April 2024

Eddie Ruan

Is SONiC ready for Routing deployment?

阿里云

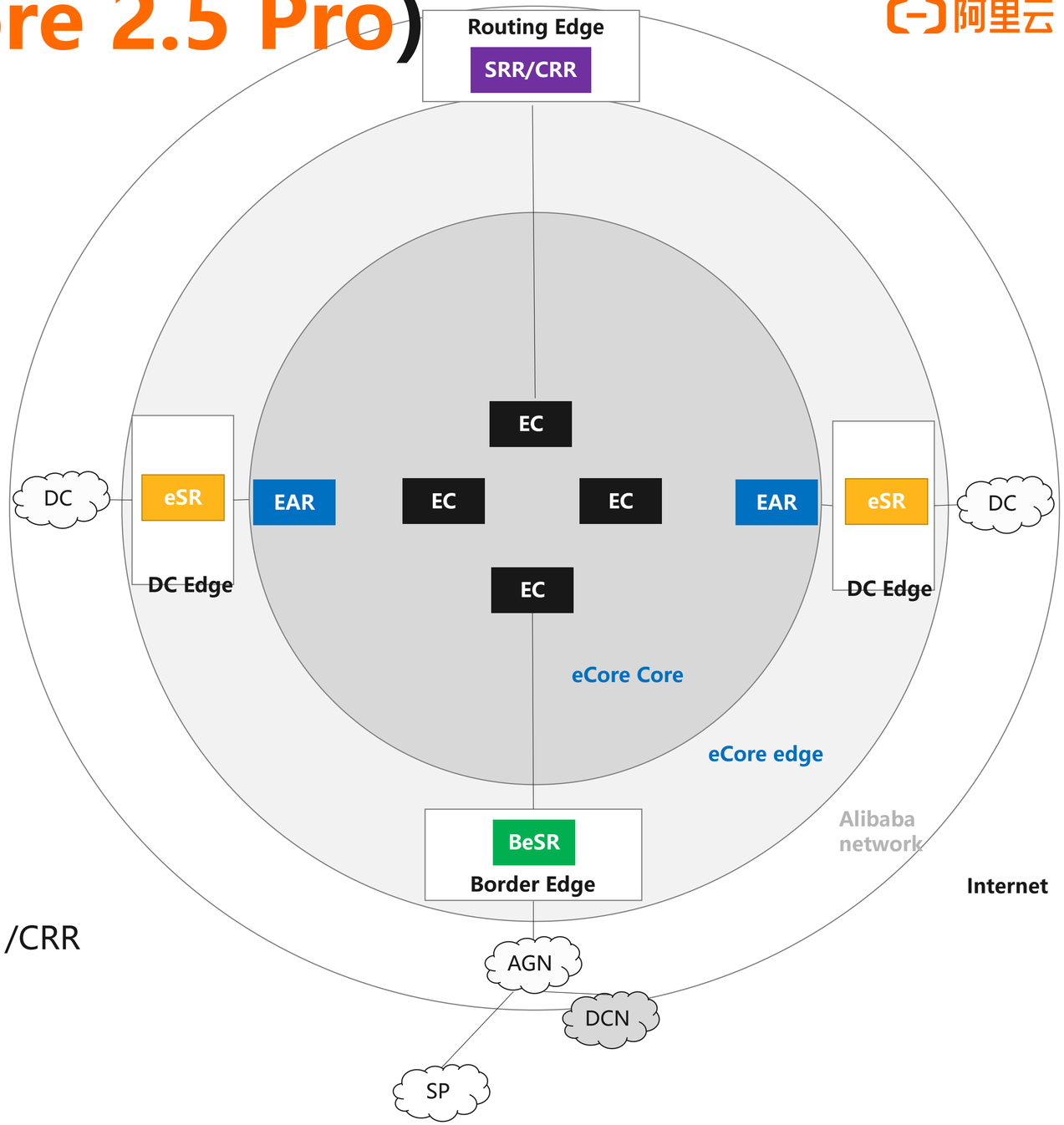
Assess this question from the following perspectives

- Use cases
- Required Feature readiness
- Infra readiness
- Existing features’ routing deployment readiness

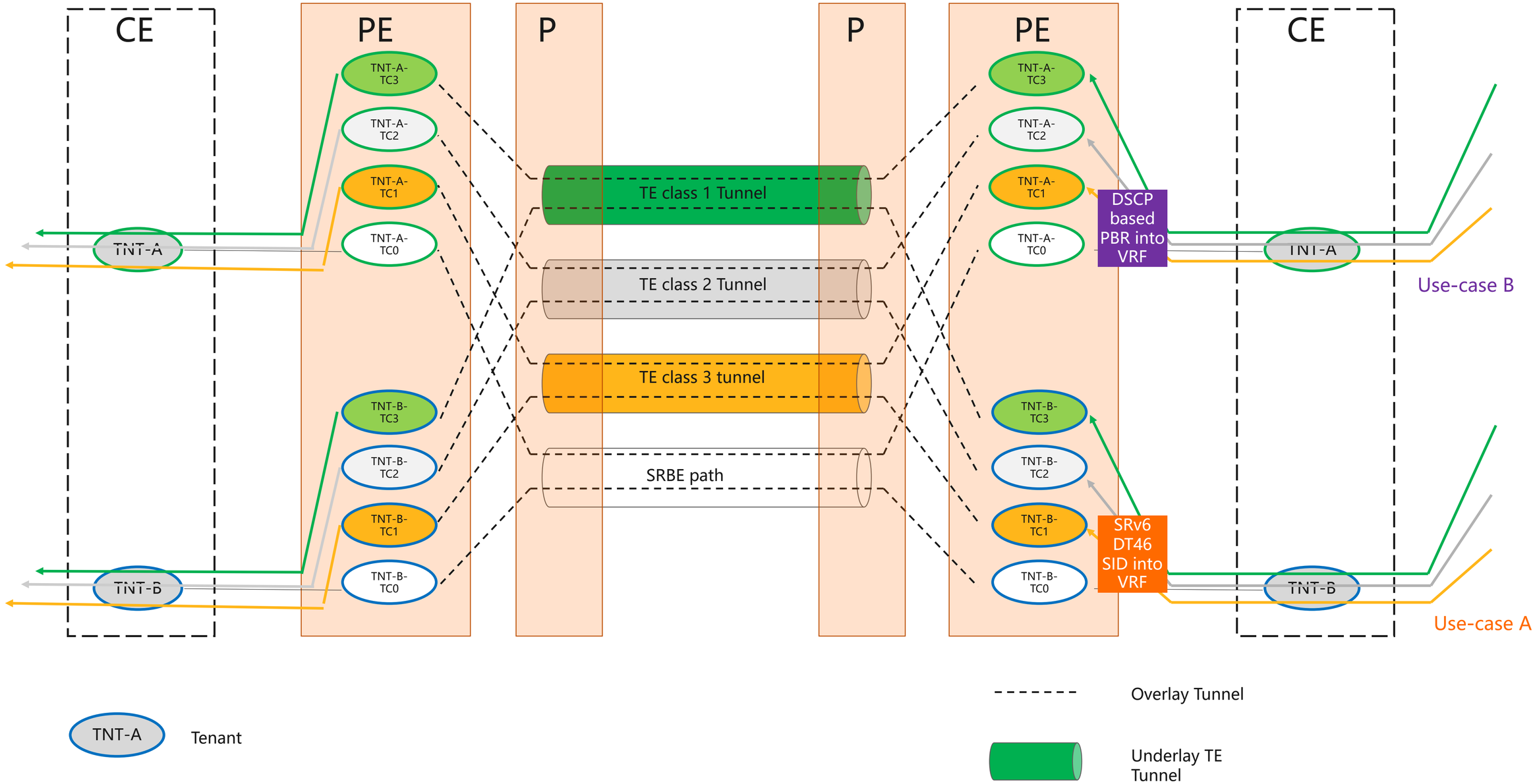
NG DCI – Outlook (eCore 2.5 Pro)

- Core
  - role
    - SRv6 VPN P router – underlay
    - Traffic-Engineering
  - layers
    - National Core - EC
    - Regional Core – EAR（collocated w/ ESR）

- Edge
  - role
    - SRv6 VPN PE router – overlay
  - categories
    - DC edge – eSR
    - Border edge – BeSR
    - Routing edge – SRR（service route reflector）/CRR



TE overview


CE
PE
P
P
PE
CE
TNT-A-TC3
TNT-A-TC2
TNT-A-TC1
TNT-A-TC0
TE class 1 Tunnel
TE class 2 Tunnel
TE class 3 tunnel
SRBE path
TNT-A
DSCP based PBR into VRF
TNT-A
Use-case B
TNT-B-TC3
TNT-B-TC2
TNT-B-TC1
TNT-B-TC0
SRv6 DT46 SID into VRF
TNT-B
TNT-B
Use-case A
TNT-A
Tenant
Overlay Tunnel
Underlay TE Tunnel

SRv6 uSID enabled SONiC

- 2021: SRv6 uSID in SONIC collaboration started
  - Very rich ecosystem (Alibaba, Cisco, Broadcom, Intel, others)
  - Several major milestones
- 2022: Alibaba Routing Platform
  - SONiC based
  - SRv6 uSID enabled
- 2023: Alibaba SRv6 uSID deployment
  - Alibaba and Cisco Blog
- 2024:
  - TE Policy
  - BFD offload
  - Infra enhancements

Features		SONiC Status
Underlay		
	SRv6 uSID	Available: SONiC 202211 [Link] [Link]
	TE Policy	Target: SONiC 202411
	BFD Offload	Target: SONiC 202411
Overlay		
	BGP VPN	Available: SONiC 202305 [Link]

SONiC Working Group on Routing sonic-wg-routing@lists.sonicfoundation.dev

Working Group on Routing Area -- Routing Performance/Scale etc.

Group Information

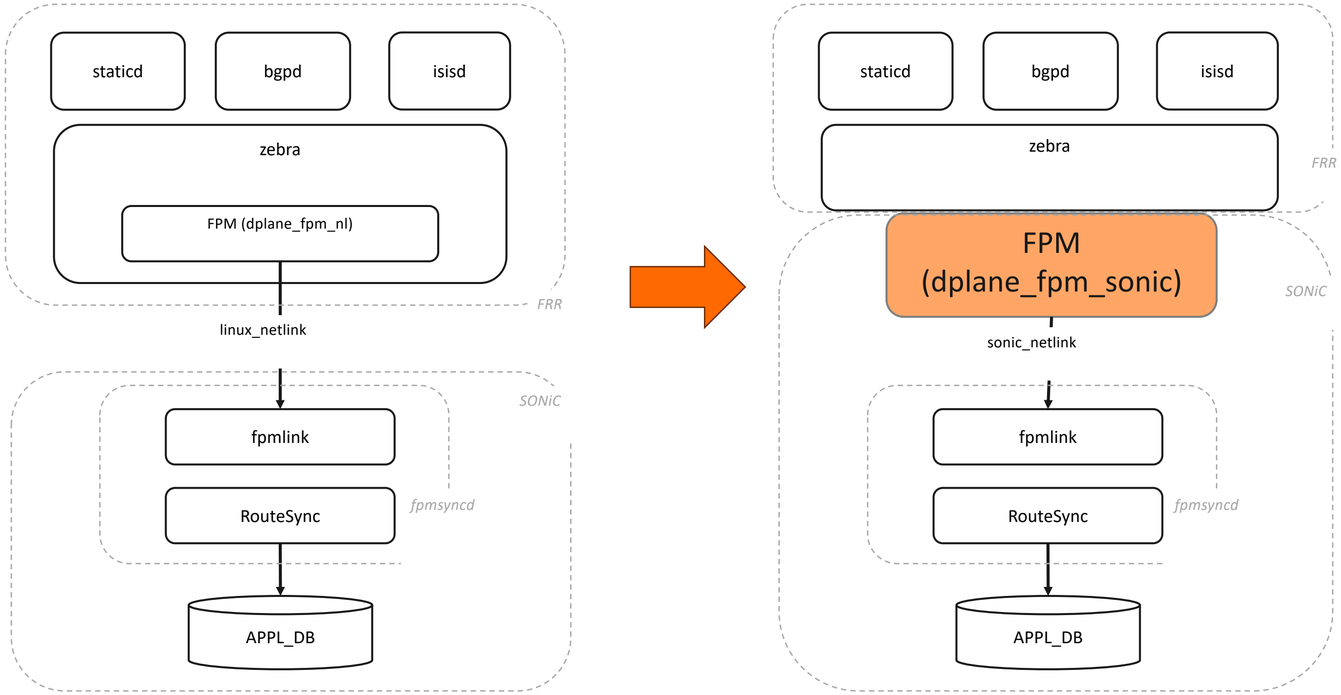
215 Members
133 Topics , Last Post: Feb 23
Started on 04/02/23
Feed

Gather contributors from both FRR and SONiC community to improve white box supports in routing.

Initial focus areas:

1. Communication among FRR, kernel and SONiC
2. Routing related features
3. Routes memory reduction
4. BGP loading time
5. Routes convergence time

Infra readiness - FRR-SONiC Communication Channel



- Decouple feature developments between Linux kernel and SONiC hardware.
  - Linux kernel is not suitable for large scale routes handling.

https://github.com/cscarpitta/SONiC/blob/frr-sonic-communication-channel-enhancements/doc/fpmsyncd/frr_sonic_protobuf_channel.md

Infra readiness – BGP Loading time Optimization

- Current SONiC Status
  - Alibaba eSR platform : 10K routes per second
  - Community SONiC : 3-6 K routes per second

- Phase 1 (202411 release)
  - 19K routes per second
    https://github.com/a114j0y/SONiC/blob/master/doc/bgp_loading_optimization/bgp-loading-optimization-hld.md

- Phase 2:
  - Further improvements needed.

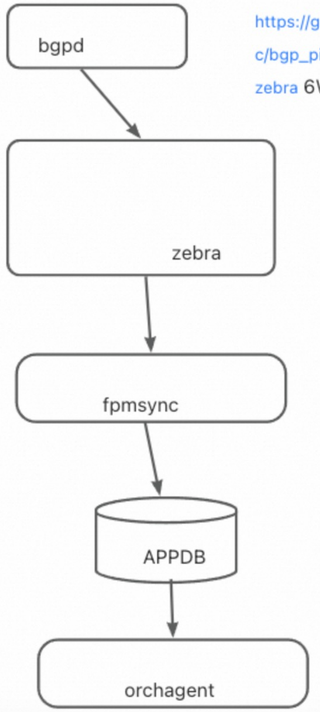
Infra readiness -Prefix Independent Convergence

阿里云

Prefix Independent Converngence

Arch Doc (Cisco, Broadcom, Nvidia and Alibaba)

https://github.com/sonic-net/SONiC/blob/master/doc/pic/bgp_pic_arch_doc.md

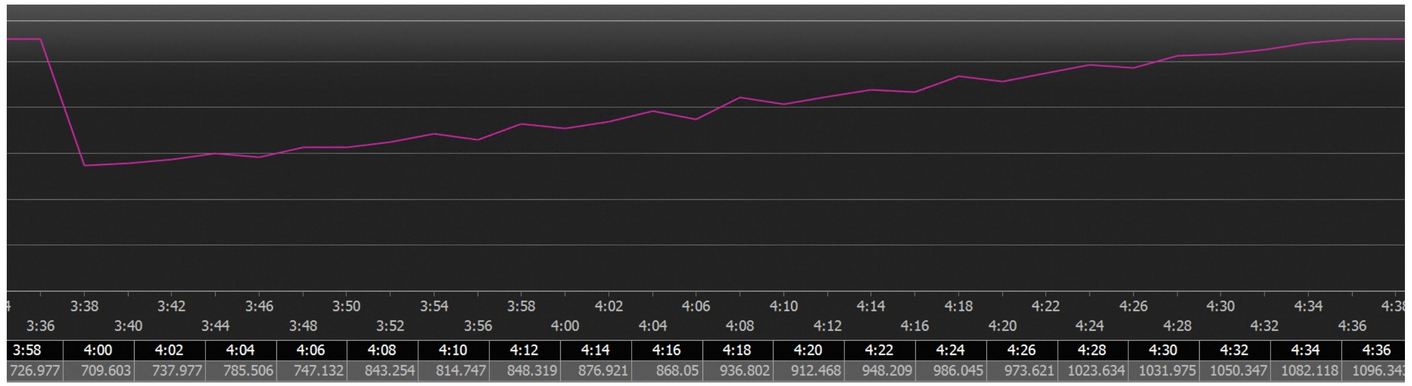


https://github.com/pguibert6WIND/SONiC/blob/proposal_bgp/doc/pic/bgp_pic_arch_doc.md#9-frrouting-ipc-messaging-from-bgp-to-zebra 6WIND

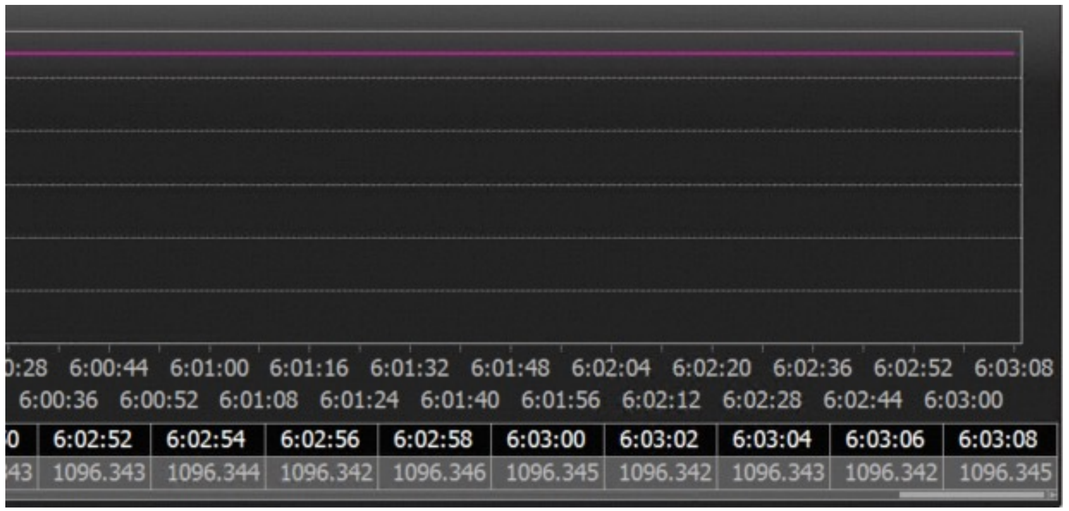
PIC core (Recursive routes support)- Alibaba, Accton, 6WIND, Cisco, Nvidia, Broadcom: https://github.com/eddieruan-alibaba/SONiC/blob/eruan-recursive/doc/recursive/recursive_route.md

PIC edge - Alibaba, Cisco, Nvidia: https://github.com/eddieruan-alibaba/SONiC/blob/eruan-pic/doc/pic/bgp_pic_edge.md

https://github.com/sonic-net/SONiC/blob/master/doc/pic/hld_fpmsyncd.md - NTT, Broadcom



Packet loss lasts for about 1 minute without PIC for 400k Routes



Packet loss for about 2ms with PIC for 400K routes

https://lists.sonicfoundation.dev/g/sonic-wg-routing/wiki/35908

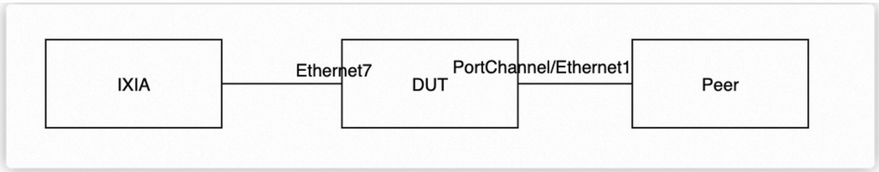
Existing features' routing deployment readiness

SONiC team/LAG issue - 1

https://github.com/jpirko/libteam/issues/75

shutdown one port, all LAGs go to down because of LACP timeout #75

Open shuaishang opened this issue on Dec 6, 2023 · 0 comments



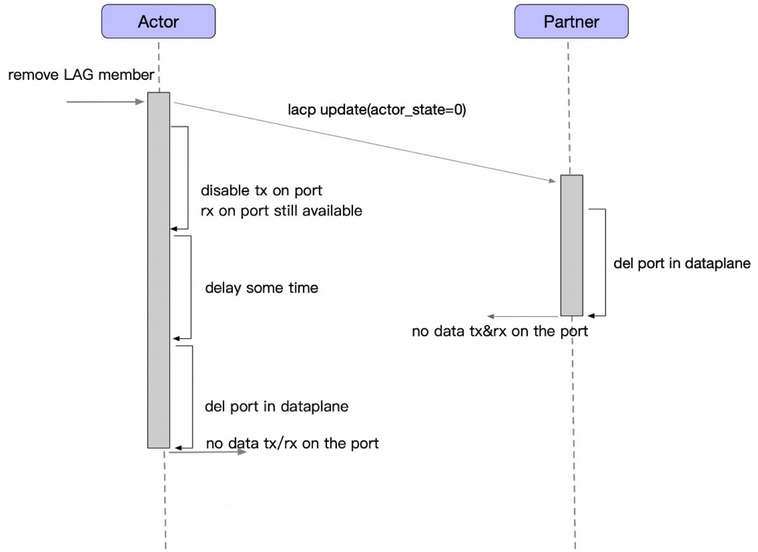
- **reproduce step**
  - DUT connected to Peer with LAG(member Ethernet1)
  - DUT connected to IXIA with Ethernet7, with 16 subports on Ethernet7
  - Each subport has one BGP session with IXIA
  - Each IXIA BGP session advertise 8K routes to DUT
  - 16 IXIA BGP session advertise same 8K routes
  - Shutdown Ethernet7, LAG will timeout and down
- **Reason & solution**
  - Processing of Ethernet7 down message takes a long time（more than 3s）
  - Teamd is a single threaded process, LACP timer handler get blocked
  - Solution: ignore netlink event about Ethernet7 since it is not my LAG member

This issue is due to large scale routes handling impact from kernel

SONiC team/LAG issue - 3

Packet loss when removing a LAG member dynamically

- **Solution - part2**
  - Implement a delay remove logic



This issue is to routing deployment requirements.

https://lists.sonicfoundation.dev/g/sonic-wg-routing/wiki/35760

SONiC vs Vender OS for SRv6 deployments

		SONiC	Vendor OS
Forwarding ASIC		Various vendors' ASIC including Cisco silicon one.	Cisco Silicon One
SRv6 Features		Limited SRv6 and SRv6 VPN features for BGP only	Rich SRv6 features for various services
Infra			
	Slow path	Linux kernel, which is not suitable for large scale routes handling.	Well designed slow path for large scale routes handling.
	Convergences	PIC ready	PIC ready
	Loading time	15K Routes per second	Over 20K routes per second
Other features		Limited features and need hardening efforts for routing deployments	Rich feature sets and ready for routing deployments.

Is SONiC ready for Routing deployment?

阿里云

- Ready for controlled deployment scenarios.
- There's still a significant journey ahead for general-purpose deployment scenarios.

Join the SONiC routing working group and collaborate to make it happen.

阿里云

Thank you