BGP Classful Transport Planes

https://tools.ietf.org/html/draft-kaliraj-idr-bgp-classful-transport-planes-07

IETF 110

Kaliraj Vairavakkalai

Juniper Networks

March 09, 2021

Agenda

- Recap BGP CT; problem, solution, advantages, presented at IETF-108.
- Share changes to the draft – since last presentation.
- Share learnings from implementation, qualification.
- Introduce related drafts.
- Next steps.

BGP-CT: Problem

•A domain has intra-AS tunnels with varying TE characteristics (gold, silver, bronze).

•There could be multiple tunnels to the same destination. And different tunneling protocols creating those tunnels.

•These tunnels may need to be extended inter-domain, while preserving their TE characteristics end-to-end.

•Different Service routes want to resolve (put traffic) over intra/inter-domain tunnels of a certain TE characteristic, with an option to fallback on tunnels belonging to a different TE characteristic, including best-effort tunnels.

•How to extend BGP to signal these pieces of information, and get the job done.

•Solution agnostic of transport (RSVP, SRTE, Flex, IP-tunnels, etc..) and service layer (L3VPN, IPv6, Flowspec, Static, L2VPN, EVPN, etc..). i.e. works with any of these protocols in service and transport-layer.

BGP-CT: Solution constructs.

- Transport Class: collects tunnels with same TE characteristics (gold, silver, etc). Identifier: 32-bit Color.

- BGP-CT is a new BGP transport layer address-family (SAFI: 76, "Classful Transport") that follows RFC-4364 procedures.

- Ingress routes collected in a TC are advertised in BGP-CT family, to other BGP speakers.
  - With "Route Distinguisher:TunnelEndpoint" as the NLRI.
  - And "Transport Class Route Target" that identifies the TC it belongs to.

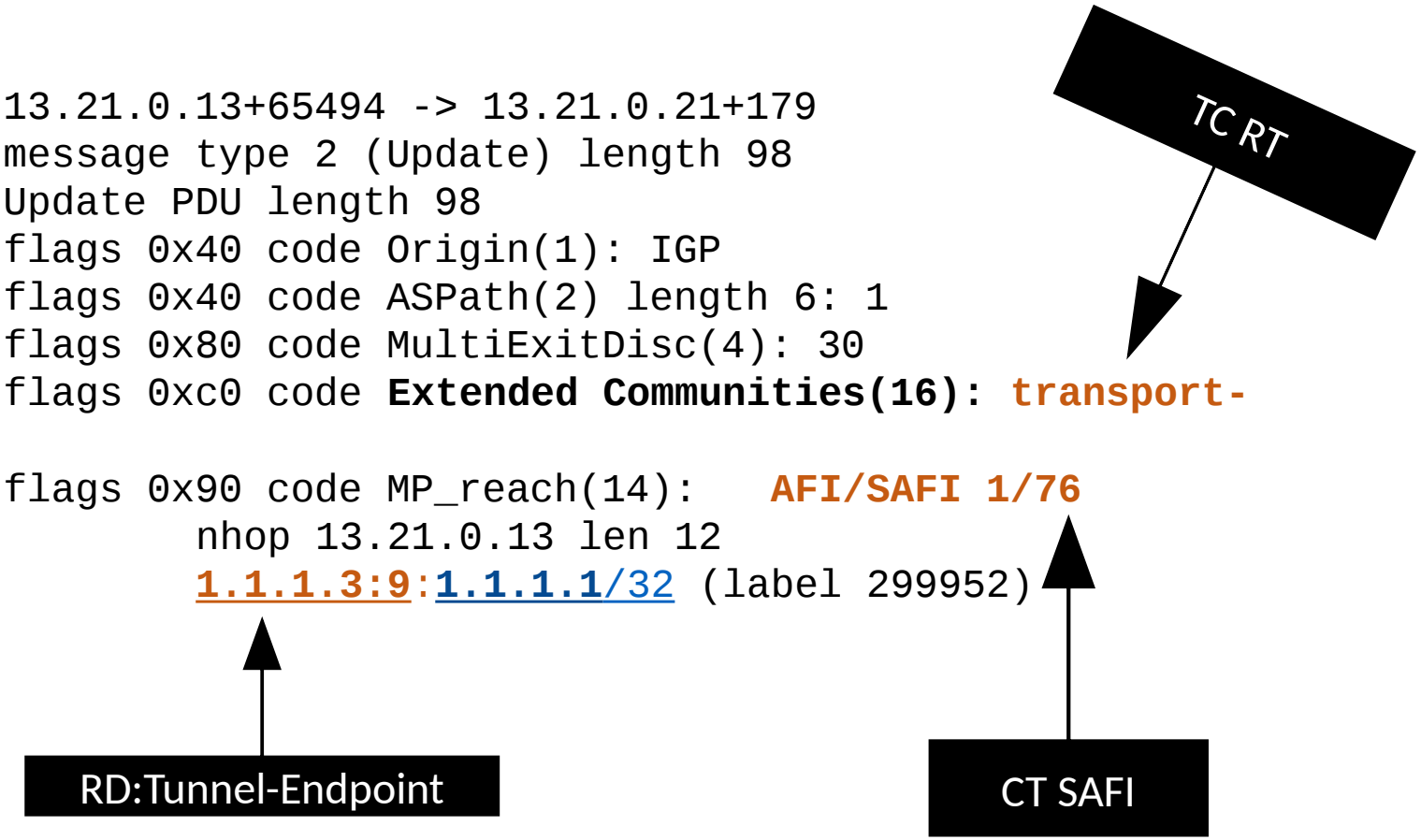
- BGP-CT extends the tunnel across inter-domain boundaries, while preserving the same Transport class end-to-end.
  - Resolve BGP NH using tunnels belonging to the same Transport class.
  - Follow RFC-4364 option-C style procedures, to create swap-routes on domain boundaries.
  - Works in conjunction with option-A, option-B scenarios as-well.

- Service routes want to resolve using a Resolution scheme asper user intent *(e.g.. use tunnels of a certain Transport class, with an option to fallback on Best-effort or another Transport class)*.

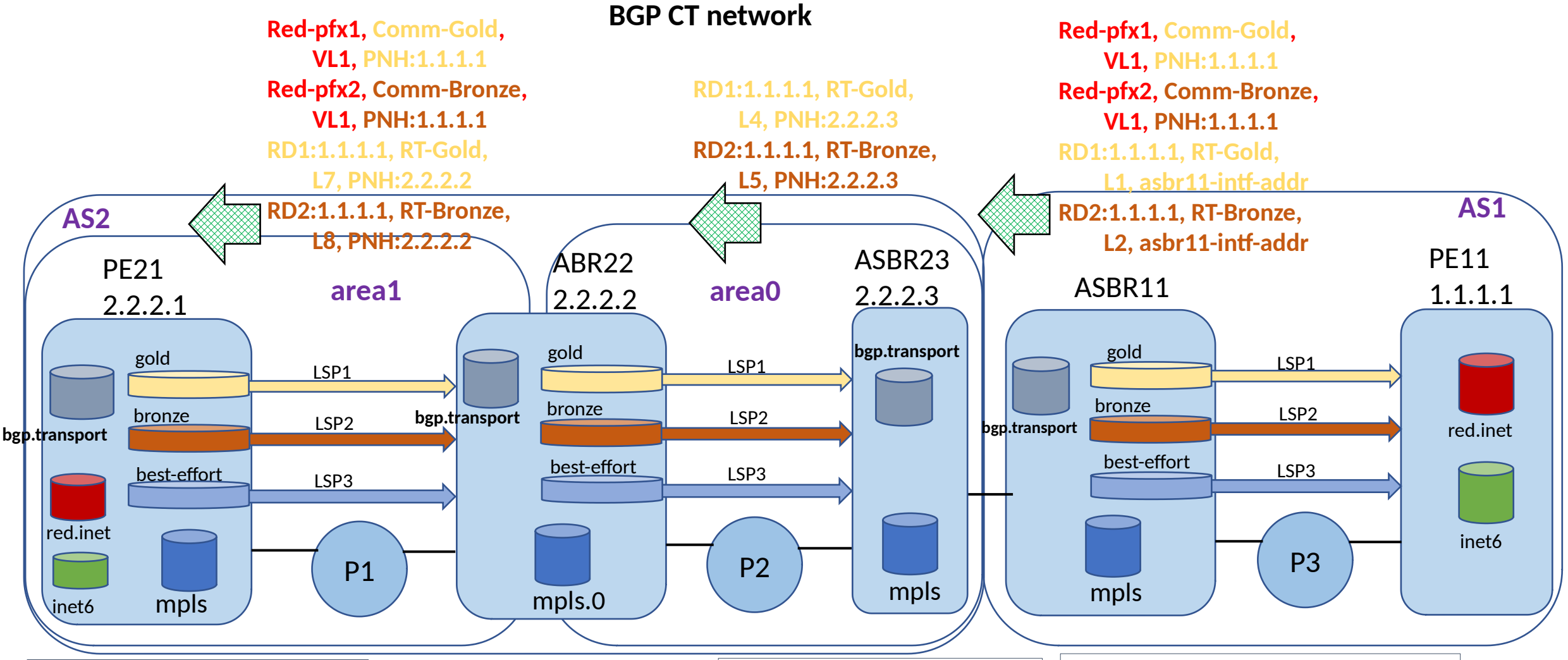
- Desired Resolution scheme is signaled via "Mapping community" on BGP route. E.g:
  - Color:0:<n> on the service-route. Resolves over Color "n" tunnels, with fallback on 'best-effort' tunnels.
  - Transport-Target on BGP-CT route. Resolves strictly over Color "n" tunnels.

BGP CT: pcap sneak peak

```
Nov 10 22:00:51.708561 BGP SEND 13.21.0.13+65494 -> 13.21.0.21+179
Nov 10 22:00:51.708563 BGP SEND message type 2 (Update) length 98
Nov 10 22:00:51.708572 BGP SEND Update PDU length 98
Nov 10 22:00:51.708574 BGP SEND flags 0x40 code Origin(1): IGP
Nov 10 22:00:51.708580 BGP SEND flags 0x40 code ASPath(2) length 6: 1
Nov 10 22:00:51.708581 BGP SEND flags 0x80 code MultiExitDisc(4): 30
Nov 10 22:00:51.708596 BGP SEND flags 0xc0 code Extended Communities(16): transport-
target:0:100
Nov 10 22:00:51.708605 BGP SEND flags 0x90 code MP_reach(14):   AFI/SAFI 1/76
Nov 10 22:00:51.708611 BGP SEND         nhop 13.21.0.13 len 12
Nov 10 22:00:51.708631 BGP SEND         1.1.1.3:9:1.1.1.1/32 (label 299952)
```



BGP CT network



PE21: red.inet fib
- **Pfx1** -> Push **VL1**, L7, LSP1
- **Pfx2** -> Push **VL1**, **L8**, **LSP2**

PE21:inet6
- **Pfx3** -> Push **2**, L7, LSP1

ABR22: mpls fib
- L7 -> Swap L4, Push LSP1
- **L8** -> Swap **L5**, Push **LSP2**
- **L9** -> Swap **L6**, Push **LSP3**

ASBR23: mpls fib
- L4 -> Swap L1, asbr11-Intf
- **L5** -> Swap **L2**, asbr11-Intf
- **L6** -> Swap **L3**, asbr11-Intf

ASBR11: mpls fib
- L1 -> Pop, Push LSP1 Labels
- **L2** -> Pop, Push **LSP2 Labels**
- **L3** -> Pop, Push **LSP3 Labels**

BGP-CT: advantages of reusing 4364 encoding

- Using RFC-4364 style “Route Distinguisher” allows advertising multiple tunnels to the same destination
  - Avoids using multiple loopbacks on Egress-PE,
  - Avoids path-hiding when transiting RR/ASBRs,
  - Allows unambiguously identifying the originating PE, for debugging.
  - Supports TunnelEndpoint being an Anycast-address participating in multiple domains.
  - Allows path-selection after stripping RD, when necessary. Helpful for faster convergence.

  Basically, RD is an identifier of convenience. Use it when needed, Strip it when not needed. Preserved end-to-end.

- Using RFC-4364 style “Route Target” to propagate Transport-Class allows:
  - Forming Venn diagrams of color domains as desired.
  - Core network having more fine-grained colors than Access networks.
  - Other creative use-cases possible in future, e.g. Hub and Spoke Color domains..?
- Treating “Color” as an attribute (adjective), rather than part of NLRI (noun)
  - Helps in cases where domains have different numbering of color values. Attribute rewrites is easier than rewriting NLRI.

- ODN using Route Target Constrain procedures.
  - Service-routes can have a clean API with Transport-layer, to request for only the BGP-CT routes required by service-routes.

- Re-uses the time tested, well deployed, RFC-4364 machinery. That cuts down implementation, testing time. Improves reliability of the solution, and time to deploy. Preserves ROI.
- Mantra of 21st century technologies is “reduce, re-use, recycle”. From Software perspective: “re-use, reduce, extend”.

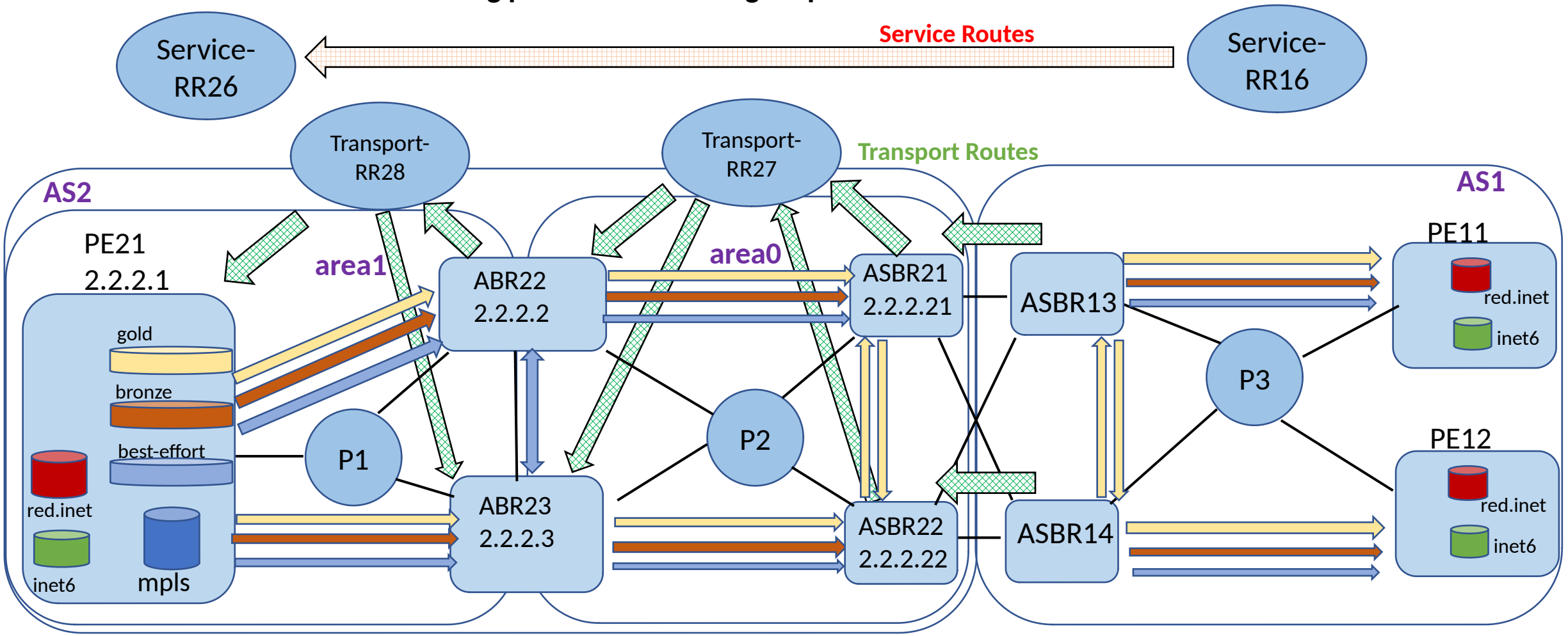
Updates since IETF-108

- Added illustration with example topology, MPLS OAM section.
- Documented how CT helps with Redundant ABRs scenario. Where RR is configured with nexthop-self. We will discuss this one in next two slide.
- Added Scaling considerations section.
  - Recommend RFC-8212 as default behavior for BGP-CT family.
  - Route-Target Filter usage for BGP-CT to provide ODN.
  - MPLS namespaces. A new concept can be applied to both LU or CT networks to deal with scaling.
- Added ‘Applicability to Network-Slicing’ section: Transport Class is the “Topology Slice” part of Transport slice (Transport slice = Topology slice + Resources)
- Welcome co-authors: VZ, Cox, Alibaba, Google.
- Status of Implementation: Code shipping on Junos 21.1R1

Redundant ABRs (RRs with NHS) in a BGP network

- Such topologies have possibility of forwarding loop forming between BGP-LU ABRs, because of RFC-4456 (it's focus is pure-RR functionality), which don't tie-break on Cluster-List before Router-ID.
- IGP-metrics need to be carefully chosen to avoid ABR choosing each other as best-path instead of ASBR.
- In some implementations LDP sets flat IGP-metric of 1 (perhaps for this reason). But when using L-ISIS or ISIS-Flex, IGP-metric makes a difference.
- Implementations may provide a way to put Cluster-List step before Router-ID step in path-selection, at "Forwarding RR-nodes doing nexthop-self".
- BGP-CT provides an easier deployment alternative: don't provision ABR to ABR colored-tunnels. This avoids possibility of any loops, without having to playing with IGP-metric or BGP path-selection.

Avoiding possible forwarding loop between redundant ABRs



ABR23: best-effort transport rib
- **PE11: PNH=ABR22**
- **PE11: PNH=ASBR21**
- **PE11: PNH=ASBR22**

ABR22: b-e transport rib
- **PE11: PNH=ABR23**
- **PE11: PNH=ASBR21**
- **PE11: PNH=ASBR22**

ABR23: gold transport rib
- **PE11: PNH=ASBR22**
- **PE11: PNH=ASBR21**
- **PE11: PNH=ABR22 (Hidden)**

ABR22: gold transport rib
- **PE11: PNH=ASBR21**
- **PE11: PNH=ASBR22**
- **PE11: PNH=ABR23 (Hidden)**

Related drafts

- PCEP RSVP Color
  draft-rajagopalan-pcep-rsvp-color-00

- Seamless SR – use cases.

https://datatracker.ietf.org/doc/draft-hegde-spring-mpls-seamless-sr/

- SRv6 and MPLS interop.

https://datatracker.ietf.org/doc/draft-bonica-spring-srv6-end-dtm/

- MPLS namespaces: signaled via BGP

https://datatracker.ietf.org/doc/draft-kaliraj-bess-bgp-sig-private-mpls-labels/

- Generic RTC

https://datatracker.ietf.org/doc/draft-zzhang-idr-bgp-rt-constrains-extension/

Thank you.