BGP Classful Transport Planes

https://datatracker.ietf.org/doc/draft-kaliraj-idr-bgp-classful-transport-planes/11/

IETF 111

Kaliraj Vairavakkalai

Juniper Networks

July 27, 2021

Agenda

- Recap problem, solution.
- Explain mechanics of BGP-CT, with focus on some key questions from WG chairs.
- Share changes to the draft – since last presentation.
- Current status, executive summary.

BGP-CT recap: Problem statement.

- 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. ***So, doing 'Intent driven Service-mapping' is the problem.***

- Solution should be 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.

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

BGP-CT recap: Solution constructs.

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

- BGP-CT is a new BGP transport layer address-family (SAFI: 76, "Classful Transport") that follows RFC-4364 procedures and RFC-8277 encodings.

- 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. aka *Transport-Target.*

- BGP-CT extends the tunnel across inter-domain boundaries, while preserving the same Transport class end-to-end.
  - Resolve BGP-CT route's NH using tunnels belonging to the same Transport class, as specified by Transport-Target on the route.
  - 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: answers to questions from WG chairs

- *In the Update packet carrying BGP routes with intent, where does your mechanism carry the "intent"?*

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

- *When a route needs to be sent from one intent domain to another, what is the mechanism that the intent mapping is updated for the receiving domain?*
  - Community/Route-Target rewrite. If the domains have different route-target values to represent an intent, then receiving domain-BN rewrites the received Transport-target to the appropriate Transport-target value for the local domain. This is similar to how L3VPN domains do route-target rewrite on AS boundaries today. Or, how Color:0:<n> communities are rewritten today in such scenarios.

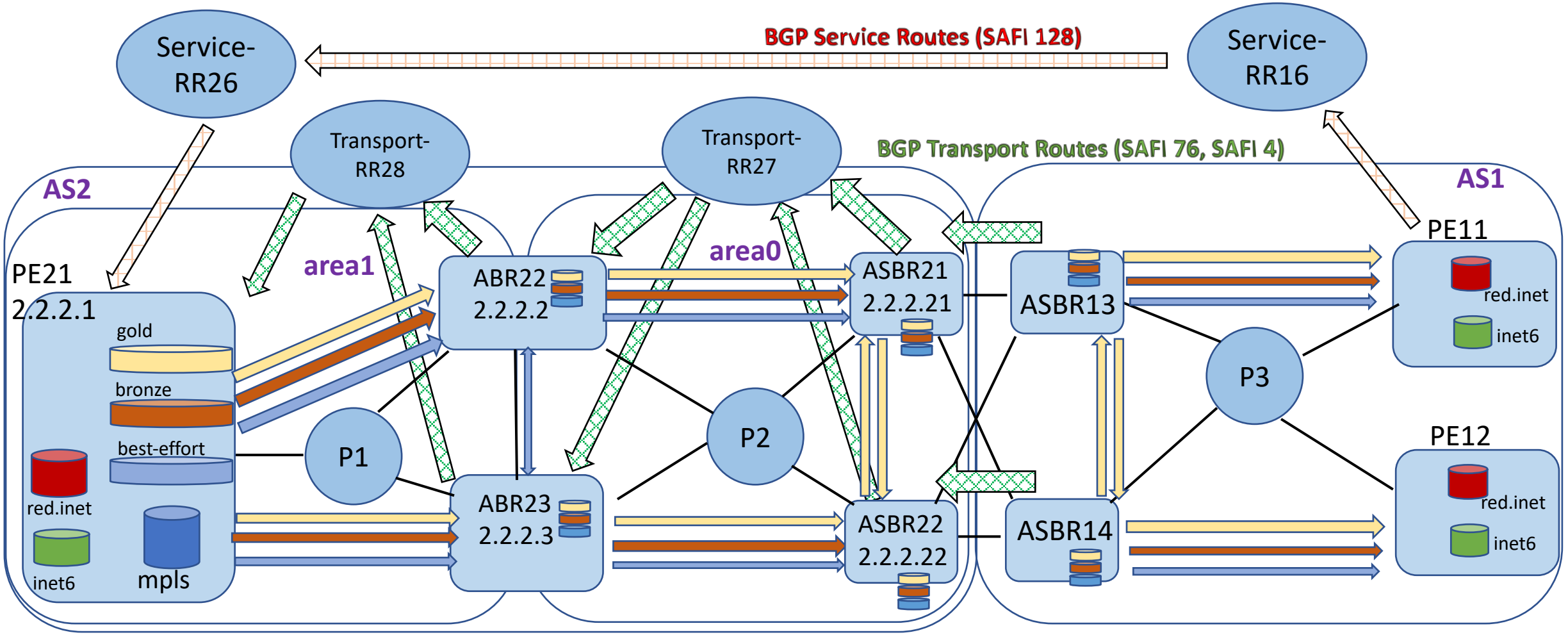
- *What procedure is used to carry routes with intent through BGP route selection "pinch points" such as route reflectors? A pinch point is where two routes that are considered equivalent for route selection purposes based on its NLRI key field may select one from many for propagation.*

  - "Route Distinguisher" is used to distinguish between different Transport-class routes for the same TunnelEndpoint.
  - Besides Route-Distinguisher, Add-path ID is also used, when crossing redundant domain-BNs and subsequent RR.
  - RD allows to uniquely identify the originating PE, across a multiple domains, which is helpful in troubleshooting.

- *How are the forwarding semantics associated with the intent carried in your mechanism? E.g. labels, SIDs, etc.*

  The BGP CT spec does not specify any changes to how label or SIDs are carried. It follows already existing mechanisms. E.g.
  - RFC 8277 specified encoding is used to carry label.
  - SIDs are carried Prefix-SID attr, as defined in mechanisms specified by SR specs.
  - Any changes in these mechanisms will also work for BGP CT, as long as those changes are backward-compatible to these existing standards.

BGP CT : Transport Class based Network Slicing



- Transport Class (e.g. gold, bronze, best-effort) provides the "**Topology Slice**" in Network Slicing
- Intra-domain Transport routes are populated in Transport class RIBs by tunneling protocols (e.g. RSVP, Flex, SRTE).
- Inter-domain Transport routes are populated in Transport class RIBs by BGP-CT family (SAFI 76).
- Service-routes (e.g. L3VPN, Internet) map to a "Toplogy Slice" by using appropriate Mapping community (e.g. Color extended community).

BGP-CT: advantages of reusing 4364 encoding

- Using RFC-4364 style “Route Distinguisher”.
  - 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.
  - RD is not used when doing per-prefix-label allocation, thus confining ripple of link/node failures local to the region where failure happened.

  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.
  - E.g. Core network having more fine-grained colors than Access networks.

- 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-using the time tested, well deployed, RFC-4364 machinery:
  - Cuts down implementation, testing time. Improves reliability of the solution, and time to deploy.
  - **Protects the investment operators have made in operational training, tooling, and procedures. Inventing new things just for fun, creates new OpEx**
- BGP-CT preserves ROI of existing deployments, by supporting all transport-tunneling protocols including RSVP.

Updates since IETF-110

- Clarified that BGP-CT routes can send/receive cross-family nexthop-type (e.g. IPv6-nexthop for AFI=1 BGP-CT routes) without negotiating RFC-5549 extended-nexthop capability.

- Added reference to draft-rajagopalan-pcep-rsvp-color-00, this was missing.

- Added ‘SRv6 support’ section.

- Corrected some typo errors. i.e. minor editorial changes.

BGP-CT: Current status, executive summary

- Draft submitted March 2020. Five IETFs ago.
- Thanks for the WG discussion, feedback and support so far.
- Juniper Implementation available since Junos21.1R1. Uses IANA allotted code-points.
- Very interested customers.

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/html/draft-salih-spring-srv6-inter-domain-sids/

- MPLS namespaces: signaled via BGP

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

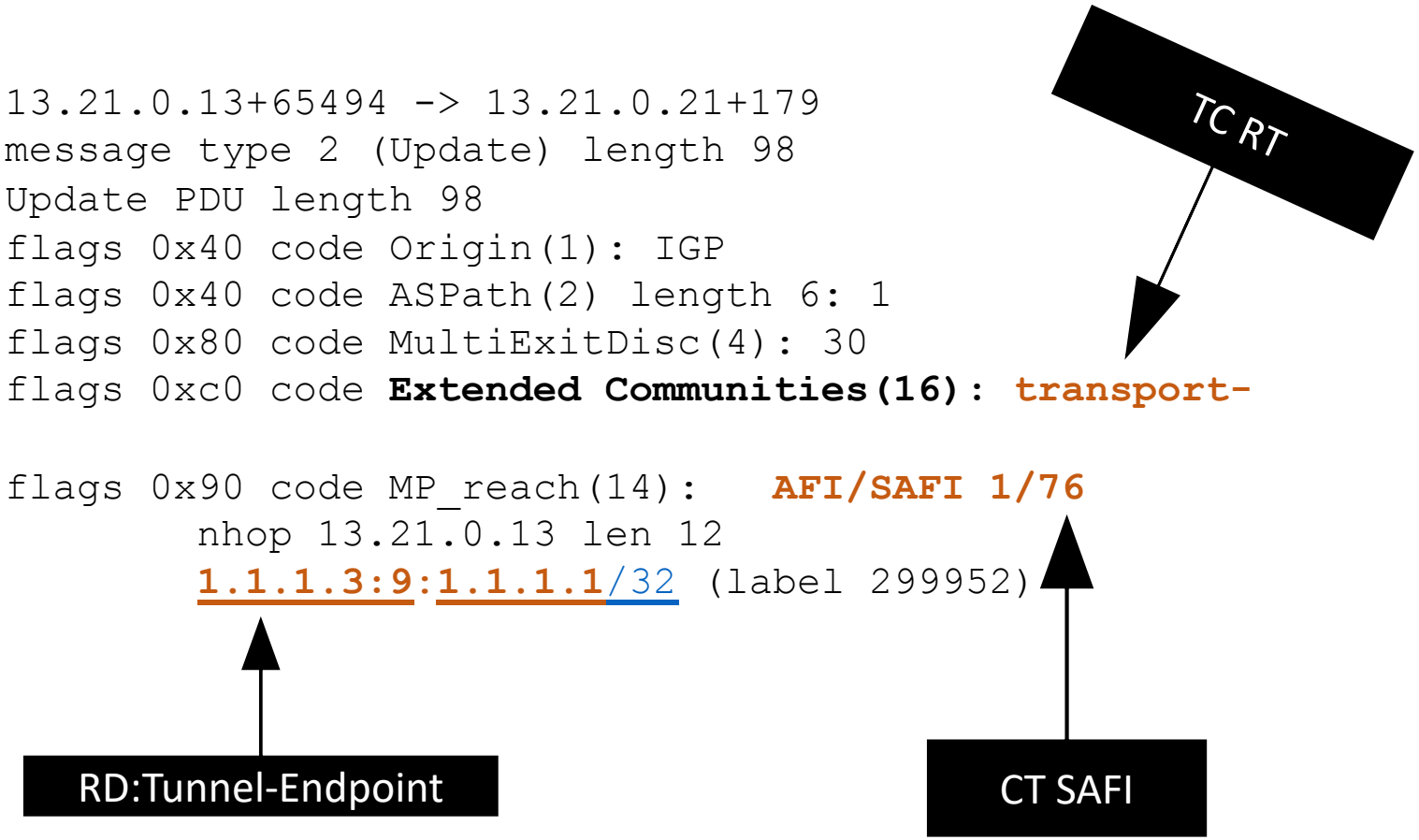
- Generic RTC

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

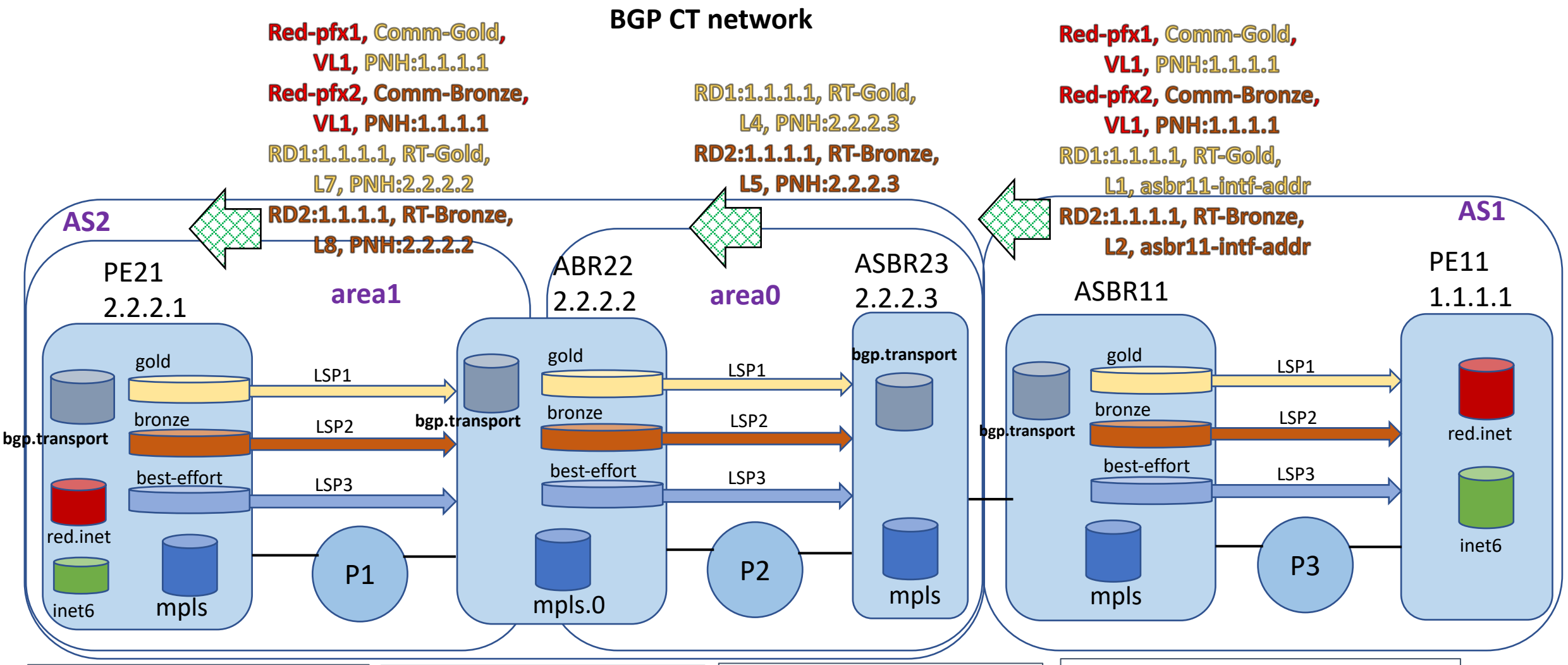
Thank you.

Backup slides: 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