BGP Classful Transport Planes


IETF 113

Kaliraj Vairavakkalai

Juniper Networks
Mar 23, 2022
Agenda

• Recap: Problem statement.
• Why a new address-family?
• Problems with including Color in LPM lookup key
• Explain mechanics of BGP-CT.
• Expressing Intent in BGP-CT, using Mapping Community.
• Current status, executive summary.
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 LU – Inter-AS Option-C network**

- Gold, Bronze, Best-effort tunnels exist in AS1 domain (e.g. RSVP, Flex, SRTE)
- Only one of them can be advertised to other domains.
- All SLA cannot be preserved across domains
Why new address-family?

Why not re-use/hack existing families like LU, SRTE or L3VPN?

• With ‘extending LU’ approach,
  • it is not possible to get end-to-end SLA guarantee. Because a LU node without extensions will re-advertise the route even if it doesn’t satisfy the SLA. So even with a new Capability, ingress cannot be sure SLA is really met end-to-end.

  • Add-path-ID is per-session scope, doesn’t help with identifying originator of route. RD is an end-to-end distinguisher

• Further overloading L3VPN (service family) with transport-routes is not good. As route-propagation path is different for service vs transport routes.

• Carrying ‘Color’ as attribute (RT) makes more sense, instead of in the NLRI. More on this on next slide.

• Use of RT allows for RTC like mechanisms, providing ODN.. If we didn’t use well-known RT ext-comm for route-leaking, this is not possible

• Thus, new SAFI 76. A Transport family that can signal transport classes.
### Problems with including Color in LPM lookup key

<table>
<thead>
<tr>
<th>Functionality</th>
<th>IP-Prefix:Color (CAR)</th>
<th>Color:IP-Prefix (SRTE)</th>
<th>IP-Prefix Transport-RIB (CT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPM on EP : C1 route, Fallback on best-effort (without color) tunnels</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>LPM on EP : C1 route, Fallback on EP : C2 tunnel route</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>LPM lookup on non-host Prefix-length route EP : C1 (e.g. SRv6 locator route)</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Carry best-effort (without color) tunnels</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
BGP-CT: 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 : 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 “Topology Slice” by using appropriate Mapping community (e.g. Color extended community).
Expressing intent in BGP-CT. Using Mapping Community

• R1 wants the intent : C1 primary, fallback to best-effort
• R2 wants the intent : C1 primary, fallback to C2
• R3 wants the intent : C1 primary, fallback to C3

This is achieved using Mapping Communities, as below:

• R1 advertised with M1 (Color:0:<C1>),
  • maps to Resolution-Scheme1: Transport classes {C1, best-effort}
• R2 advertised with M2, (Color:0:<C1C2>)
  • maps to Resolution-Scheme2: Transport classes {C1, C2}
• R3 advertised with M3, (Color:0:<C1C3>)
  • maps to Resolution-Scheme3: Transport classes {C1, C3}
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.
BGP-CT: Current status, executive summary

• Draft submitted March 2020.

• Thanks for the WG discussion, feedback and support so far.

• Juniper Implementation available since Junos21.1R1. Uses IANA allotted code-points.

• Very interested customers.

• Requested WG adoption.
Related drafts

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

• Seamless SR – use cases.

• SRv6 and MPLS interop.

• MPLS namespaces: signaled via BGP

• Generic RTC
Thank you.