Global Table Multicast with BGP-MVPN

draft-zzhang-l3vpn-mvpn-global-table-mcast

London, 89th IETF

Summary

- Original draft targeted for Mboned and presented in 86th IETF (in L3VPN)
- Re-homed for L3VPN; re-structured -01 version presented in 88th IETF
- Requesting adoption in L3VPN WG
  - Presenting in Mboned/PIM WG for comments
    - Slides borrowed from Eric Rosen's 88th presentation

Background/Motivation

- Service providers currently using and/or actively deploying BGP control plane (per MVPN RFCs/I-Ds) to:
  - carry customer multicast control information, and
  - multiplex customer multicast flows onto “P-tunnels” that travel through the SP “backbone”
- Procedures designed for use in VPN context
- SPs also have non-VPN multicast flows that have to be signaled and tunneled over the backbone
- Wouldn’t it be nice to use the same protocol and procedures for non-VPN multicast?

Why Would It Be Nice?

- By handling non-VPN multicast “just like” VPN multicast:
  - Same functionality,
  - Same tools,
  - Same training,
  - Same troubleshooting methodology,
  - Ability to aggregate VPN and non-VPN flows into the same tunnel
  - New features will apply to both, without having to do them twice
  - Etc.
- Purpose of draft-zzhang:
  - show how to apply MVPN procedures to non-VPN multicast
  - systematic attention to the few places where adaptation of the procedures is necessary or desirable

Global Table instead of VRF

- Basic approach: “use the MVPN protocols unchanged, just apply them to the Global Table instead of to a VRF”
  - “global table” is a routing table that is not specific to any VPN
  - GTM sometimes called “Internet multicast”, but:
    - the global tables don’t necessarily have Internet routes,
    - the “global” multicast flows aren’t necessarily going to or from the “Internet”
    - global just means “not VPN”
- No new SAFIs, NLRI formats, BGP path attributes
- No new semantics for existing messages
  - MVPN protocols use Route Distinguishers (RDs) to identify VRFs, but there is no use of RD 0
  - So let RD 0 identify the global table
  - Then just do everything the same ☺

Just a Few Details to Work Out

- Implementors need a little more detail than “do MVPN, but in the context of global table rather than VRF”
- MVPN procedures rely on Route Targets, but global tables don’t usually have route targets. Some adaptation is needed.
- MVPN procedures require egress PE to determine the ingress PE and the “upstream multicast hop” (UMH) for a given multicast flow. This is done by looking at MVPN-specific Extended Communities attached to VPN-IP routes. Some adaptation is needed.
- Is there anything needed for MVPN that isn’t also needed for GTM? Maybe a few things can be left out …
- Vice versa?
- As usual, there are a few special scenarios that some SPs would like to optimize for …

A Note on Terminology

- "PE" is well-established term in VPN context for routers that delimit the "SP backbone" and that attach directly to customer/subscriber routers (CEs)
- In GTM scenarios, the routers that delimit the backbone don't attach to subscribers, aren't necessarily "provider edge"
- So we use a new term "Protocol Boundary Router" (PBR) to denote those routers that play the same role in GTM procedures that PEs play in MVPN procedures
  - Any given multicast flow has its ingress PBR and its egress PBRs
  - MVPN-based BGP control plane used among the PBRs
  - The PBR interfaces that face away from the core (analogous to VRF or PE-CE interfaces) most likely use PIM to transfer multicast routing info. But we don't rule out the use of BGP, IGMP, whatever.
  - As in MVPN, the tunnels through the core may be of a variety of technologies

AFI/SAFI's needed for GTM/MVPN

- Always two AFI/SAFIs needed:
  - **UMH-eligible routes (RPF routes): r**outes to the multicast sources, used for finding upstream neighbor and ingress PE/PBR :
    - MVPN: SAFI 128 (labeled VPN unicast) or 129 (VPN multicast-UMH determination): NLRI specifies RD+prefix
    - GTM: SAFI 1 (unicast), 2 (multicast RPF-determination), or 4 (labeled unicast): NLRI specifies prefix but no RD
    - For MVPN UMH-eligible routes required to carry *VRF Route Import* and *Source AS* EC
      - To do GTM like MVPN, GTM UMH-eligible routes should have same requirement – but we will discuss a few scenarios where these can be omitted (at some compromise to the overall goals)
  - "**MCAST Routes**": used for disseminating multicast routing information, for assigning multicast flow to tunnels, and sometimes for joining and leaving tunnels (BGP C-multicast routes and BGP A-D routes)
    - SAFI 5, for both GTM and MVPN

Use of Route Targets

- GTM **requires**, like MVPN, IP-address-specific RTs on the MCAST C-multicast Join routes and the MCAST Leaf A-D routes.
  - These routes are always “targeted” to a single router
  - That router is identified by the RT
  - BGP may distribute those routes to other routers -- the RT is the only way a router knows whether it is the “target” of a Join router or a Leaf A-D route
  - The RT also identifies the “target” VRF, for GTM that’s always VRF zero.
- Do other MCAST routes need RTs?
  - Yes, if you don’t want every GTM route to be distributed to every PBR
  - Useful to configure global tables with import/export RTs (like VRFs), so that MCAST route distribution can be constrained (with same tools used for constraining distribution of MVPN routes)

Finding the “Upstream PBR”

- Standard method (from MVPN specs):
  - UMH-eligible route matching a multicast source/RP carries VRF Route Import EC and Source AS EC
  - VRF Route Import EC identifies “upstream PBR” (ingress PBR) for flows from that source/RP (remember: upstream PBR not necessarily the next hop)
    - This info is used for targeting Joins and Leaf A-D routes
  - Source AS needed for multi-AS procedures
  - For MVPN, “upstream RD” is also inferred from this EC,
- Same exact procedure will work for GTM
  - Of course, RD is always zero
- But – whereas MVPN UMH-eligible routes are always originated into BGP by ingress PE, and distributed by BGP to egress PEs, that’s not always the case in GTM
  - Non-VPN UMH-eligible routes may not be originated by ingress PBR and/or distributed by BGP

Alternative Methods of Finding the “Upstream PBR”

- If UMH-eligible routes are not already BGP-distributed:
  - Have ingress PBR redistribute routes into BGP as SAFI-2, attach MVPN ECs
    - Multicast works “normally”, unicast routing not impacted, no other special procedures needed
  - If backbone is fully meshed with TE tunnels,
    - When egress PBR looks up route to source/RP, next hop interface will be TE tunnel
    - Select as ingress PBR the remote endpoint of that tunnel
    - Assume ingress PE in same AS as egress PE
    - Applicability restrictions
  - May be other deployment and/or implementation-specific methods that can be used, such as consulting IGP database
    - anything that works is allowed *optionally*, but beware interop problems

Another Alternative Method for Determining the “Upstream PBR”

- Next Hop
  - If:
    - every UMH-eligible route is originated by its ingress PBR, and
    - the ingress PBR puts itself as the next hop, and
    - the next hop never changes while the route is being distributed,
  - Then:
    - the ingress PBR can be determined from the next hop.
  - **Only works if the BGP speakers distributing the UMH-eligible routes never do “next hop self”,** e.g., if routes distributed by “Service Route Reflector”

One More Alternative Method for Determining the “Upstream PBR”

- Scenario:
  - Source (S)---Attachment Router (AR)---I-PBR--- …. ---E-PBR
  - S is multicast source, AR is BGP speaker that without BGP MCAST support
  - AR talks PIM to I-PBR
  - AR distributes route to S, but doesn’t attach MVPN extended communities (doesn’t know about them)
  - The BGP-distributed route to S has AR as the next hop
- Finding the Upstream PBR by Recursive NH Resolution
  - I-PBR distributes in BGP a route to AR, with I-PBR as NH
  - I-PBR attaches VRF Route Import and Source AS ECs to those routes
  - When E-PBR looks up route to S:
    - it finds AR as the next hop
    - then it looks up route to AR, and finds I-PBR as the next hop
    - the route to AR has a VRF Route Import EC, so E-PBR knows that I-PBR is the upstream PBR for flows from S

Next Steps

- Requesting adoption in L3VPN WG
- Calling for review/comments in PIM/Mboned WGs