Requirement for Multicast MPLS/BGP VPN Partitioning

`draft-mnapierala-mvpn-part-reqt-01`

IETF 71 March 10 2008

Maria Napierala

Why Multicast VPN Partitioning

- Enterprises and firms typically “segment” their IP VPN's by data center or hub locations in order to meet specific performance, security, or application access requirements
- This “segmentation” consists in partitioning of VPN sites into disjoined groups that share the same routing policy
- Multicast transmission in a VPN should be subject to partitioning by multicast source or Rendezvous Point location
  - such partitioning is based on “anycast” souring (more on the next slide) and
  - allows different downstream PE's or even different mVRF’s choose different upstream PE's as the next-hops to the RP or the source

What is “Anycast” Sourcing

- In IP VPN the same routes can be injected at different VPN sites (typically at hub or data center locations)
  - these could be default or summary routes, or even specific routes
- If these are the routes to customer RP’s or customer sources, they are examples of "anycast" sourcing of multicast traffic in a VPN
- “Anycast" sourcing includes multi-homed sites with sources or RP's or RPA’s
- “Anycast" sourcing includes Anycast RPs

Main Requirement

- Support of "anycast" sourcing in MVPN without duplicate packets or packets from "wrong" upstream PE’s being sent to customer receivers, except during routing transients
- When preventing duplicate or “wrong” streams being sent to receivers, the solution should not waste provider's network resources by discarding, at network egress, already transmitted traffic
  - this includes PIM-SM streams: an egress PE should receive a PIM-SM stream either from the customer RP or directly from the source but never from both

Main Requirement – Cont.

- While supporting "anycast" sourcing, the MVPN solution should not impose any restriction on multicast VPN service offering
  - it cannot require a customer to outsource its RP functionality to the service provider or a service provider to run MSDP with the customer
- MVPN partitioning should be supported for the following PIM modes in customer domain: PIM-SM, PIM-SSM, and PIM-Bidir

When Supporting Anycast Sourcing, MVPN solution should

- Conform to customer's PIM-SM SPT-thresholds by
  - not triggering or retaining unexpected (S, G) states in customer's network
  - this includes preserving shared trees in customer network if CE's do not switch traffic to SPT's

Support of Anycast RP

- VPN customer Anycast RP should be supported in the following two ways:
  - based on provider's network routing cost
    - receivers join the closest Anycast RP, according to the routing in the SP backbone
  - based on VPN customer routing policy
    - partitioning of receivers by Anycast RP location is determined by VPN routing policy
    - allows multicast VPN customer to define its own Anycast RP selection, based on other criterion than the provider’s network closest distance

Support of PIM-Bidir in MVPN

- Many enterprises use multicast applications that scale or even operate correctly only with PIM-Bidir
- PIM-Bidir is already deployed in many of these networks and its support in MVPN context is required
  - this is a change from MVPN generic requirements document (RFC 4834) where PIM-Bidir support on PE-CE interfaces is only recommended
- MVPN solution for PIM Bidir should prevent any packet looping and should support source-only branches

Progress on draft-mnapierala-mvpn-rev-03 and -04

- Changed the title to be more specific - “Segmented Multicast MPLS/BGP VPNs”
- New work includes:
  - Clarification of PIM-SM inter-PE procedures
  - Explanation of S-PMSI Aggregation
  - Included support for source-specific host reports in PIM-SM

Objective of the Proposal

- Support "anycast" sourcing in MVPN without duplicate packets or packets from "wrong" upstream PE’s being sent to egress PE’s

while

- Not triggering or retaining unexpected (C-S, C-G) states in customer's network

Highlights of the Solution

(differences with the current specification)

- Every multicast stream whether (C-S, C-G) or (C-*, C-G) is carried in an S-PMSI
- PIM sparse mode C-stream, only if not carried in (C-S, C-G) S-PMSI, it is carried in (C-*, C-G) S-PMSI
- C-source discovery method uses PIM control messages but it is not based on customer-initiated RPT-to-SPT switchover or on outsourced C-RP model or on running MSDP to CE

Results in …

- Partitioning of an MVPN into sets of mVRF's with overlapping routing policies
- Each partition being served by distinct set of P-Multicast Distribution Trees (P-tunnels)

Summary of PIM-SM Support – (C-*, C-G) S-PMSI

- Join (C-*, C-G) message when received by a PE that has a VRF interface which is next hop to the C-RP, causes that PE to generate S-PMSI announcement for (C-*, C-G) traffic
- Different downstream mVRF’s can choose different upstream PE’s to reach the same C-RP and hence join different P-tunnels announced by different ingress PE’s

Summary of PIM-SM Support - (C-C, C-G) S-PMSI

- Active C-sources are discovered by observing Join (C-S, C-G) messages from direction of C-RP, which triggers Source Active Advertisements
- Source Active Advertisement when received by a PE that has a VRF interface which is next hop to C-S, causes that PE to generate S-PMSI announcement for (C-S, C-G) traffic
- Traffic from sources such that Join (C-S, C-G) is never received from a CE that is next hop to C-RP as well as traffic from C-sources attached to the same PE as the C-RP, both stay on (C-*, C-G) S-PMSI

Summary of PIM-SM Support – Switching between C-SPT and C-RPT

- If a (C-S, C-G) stream is carried in an S-PMSI, and for the same C-G, the (C-*,C-G) stream is carried in an S-PMSI, then the (C-S, C-G) traffic is <u>not</u> carried in the (C-*, C-G)'s S-PMSI
- If the source C-S carried in its own S-PMSI becomes inactive, PE attached to C-RP switches back to receiving C-S traffic on the shared tree, by triggering a Join(C-S,C-G,rpt) towards the CE

Summary of PIM-Bidir Support – DF-PE and (C-*, C-G) S-PMSI

- Different mVRF's in a given VPN might have different next-hop PE's towards C-RPA (i.e., different *Designated Forwarder*-PE's) due to different routing policies or they might have temporarily different next-hop PE's to C-RPA due to routing transients
- Join (C-*, C-G) message when received by a DF-PE causes that PE to generate S-PMSI announcement for (C-*, C-G) traffic
  - –whether S-PMSI announcements are used depends on P-tunnel technology and on traffic aggregation requirements (see next slide)
- S-PMSI's are instantiated by MP2MP tunnels

Support of S-PMSI Aggregation without Duplicates to Egress PE’s

- A single unidirectional P-tunnel rooted a particular PE can aggregate, in a given MVPN, traffic from all C-RP’s and C-sources attached to this PE
- A single bidirectional P-tunnel can aggregate all Bidir traffic per DF-PE in a given MVPN
- If those P-tunnels are P2MP or MP2MP LDP LSPs, they can be algorithmically and uniquely chosen by the egress mVRFs and don’t need to be announced
- (C-S, C-G) S-PMSI’s can aggregate only congruent flows

Supporting Source-Specific Host Reports in PIM-SM

- PE can receive Join (C-S, C-G) for a sparse mode group even if no PE in MVPN has ever received Join (C-*, C-G)
- Useless S-PMSI creation for idle C-sources operating in sparse groups is prevented by:
  - announcing (C-S, C-G) S-PMSI only when the 1st packet is received on the (C-S, C-G) state, which is indicated by (C-S, C-G) "SPTbit“