Chapter 17

Multicast Data MDT Overview

Protocol Independent Multicast (PIM) version 2 supports multicast over Layer 3 virtual private networks (VPNs) based on RFC 2547, *BGP/MPLS VPNs*, and Section 2 (Multicast Domains) of Internet draft draft-rosen-vpn-mcast-07.txt, *Multicast in MPLS/BGP IP VPNs*. This implementation does not require the provider (P) routers to maintain any VPN-specific PIM information, but this lack of VPN-specific information is not optimal. The issue is that a single multicast group is defined for each VPN to carry multicast control and data traffic inside the provider core and all VPNs are mapped to this single group in the provider's space. This mapping results in the delivery of packets to each provider edge (PE) router attached to the P router even if the PE router has no receivers for traffic from a multicast group in that VPN. Each PE router must process the encapsulated VPN traffic even if the multicast packets are then dropped. This is a waste of resources, especially in environments characterized by low bandwidth links in the core or a multicast source in the VPN sending a very high volume of information (for example, high-definition television [HDTV] packets) through the core.

A data multicast distribution tree (MDT), based on section 7 of Internet draft draft-rosen-vpn-mcast-07.txt, *Multicast in MPLS/BGP IP VPNs,* solves the problem of P routers flooding unnecessary multicast information to PE routers that have no interested receivers for a particular VPN multicast group. The multicast data MDT solution requires the creation of a new tunnel by the PE router if the source exceeds a configured rate threshold parameter. All other PE routers join the new tunnel only if the PE router has receivers in the VPN for that multicast group.

This chapter provides the following information about data MDTs:

Data MDT Creation Overview

Initially, the PE routers discover each other in a VPN routing and forwarding (VRF) instance using the default MDT. Each PE router configuration includes in its VRF instance various parameters to control the creation of a data MDT, such as when the source traffic in the VRF instance exceeds the configured threshold rate. The PE router monitors the rate during its periodic statistics-collection cycles. If the source locally attached to the PE router in the VPN exceeds this limit, the source PE advertises the new data MDT group and new MDT with a User Datagram Protocol (UDP) type-length-vector (TLV) packet called an *MDT join TLV*. The MDT join TLV describes the source and group pair (S,G) in the VRF instance and the new data MDT group address used in the provider space. The source PE periodically announces the MDT join TLV over the default MDT for that VRF instance as long as the source is active.

All PE routers receive the MDT join TLV because it is sent over the default MDT. Only the PE routers with receivers in the VRF instance for that multicast group can join the new group, and the PE routers must join the new group to receive the multicast traffic now sent over the new MDT by the source PE. PE routers without interested receivers listed in the VRF instance ignore the MDT join TLVs.

When remote PE routers join the new data MDT group, they send a PIM join message for the new group in the provider space. The PIM join message for the new group is sent directly to the source PE router from the remote PE routers by means of PIM source-specific multicast (SSM). SSM using (S,G) is possible with data MDT instead of the default MDT any-source multicast (ASM) (*,G) because the source address is known from the UDP signaling used with data MDT.

The source PE router starts encapsulating the multicast traffic for the entries in the VRF instance using the new data MDT group after 3 seconds, allowing time for the remote PE routers to switch to the new group. The source PE router then halts the flow of multicast packets over the default MDT, and the packet flow for the entries in the VRF instance source shifts to the newly created data MDT, joined only by PE routers with interested receivers.

When the preconfigured conditions, such as the rate threshold, are no longer met by the source because the source stops sending or the rate falls below the threshold, the source PE stops announcing the MDT join TLVs and the PE router switches to sending on the default MDT for that VRF instance again.

Data MDT Characteristics

The maximum number of data MDTs for all VPNs on a PE router is limited to 8000, and the maximum number of data MDTs for a VRF instance is 1024. The configuration of a VRF instance can limit the number of MDTs possible. No new MDTs can be created after this limit is reached in the VRF instance, and all traffic for other sources that exceed the configured limits is still sent on the default MDT.

Creation of data MDTs depends on the monitoring of the multicast source data rate. This rate is checked once per minute, so the creation of data MDTs can be delayed up to 1 minute after a source exceeds a configured limit. In the same way, if the source data rate falls below the configured value, data MDT deletion can be delayed for up to 1 minute until the next statistics monitoring collection cycle.

Changes to the configured MDT limit value do not affect existing tunnels that exceed the new limit. MDTs that are already active remain in place until the threshold conditions are no longer met.

To remove active MDTs no longer included in a newly configured group address range, you must restart the PIM routing instance. This restart clears all remnants of the former group addresses but disrupts routing and therefore requires a maintenance window for the change.

Multicast tunnel (mt) interfaces created because of exceeded thresholds are not recreated if the routing process crashes. Therefore, graceful restart does not automatically reinstate the data MDT state. However, as soon as the periodic statistics collection reveals that the threshold condition is still exceeded, the tunnels are quickly re-created.