February 2020

OPTIMIZED TRAFFIC FORWARDING IN MULTICAST VIRTUAL PRIVATE NETWORK (VPN) CORE

Sameer Gulrajani

Rishabh Parekh

Arvind Venkateswaran

Swadesh Agrawal

OPTIMIZED TRAFFIC FORWARDING IN MULTICAST VIRTUAL PRIVATE NETWORK (VPN) CORE

AUTHORS:

Sameer Gulrajani
Rishabh Parekh
Arvind Venkateswaran
Swadesh Agrawal

ABSTRACT

Presented herein are techniques for leveraging the global Border Gateway Protocol (BGP) multicast virtual private network (mVPN) view (using mVPN context discovery) on the route reflector to optimize the distribution of the traffic forwarded in a BGP C-multicast mVPN core where receives are only located behind the provider edge (PE) routers (PEs).

DETAILED DESCRIPTION

For customer flows using PIM Sparse Mode in a BGP mVPN deployment, the standards (e.g., RFC 6514) mandate that the PE hosting the site containing the PIM customer RP continue to join the source tree, even if all the receiver PEs are receiving the traffic from the source PE. This means that the customer traffic traverses the core to the RP PE, even if it is not required. In certain cases, the RP PE turns around and sends this traffic back to all the receiver Pes, even when the receiver PEs are already receiving the same traffic from the source PE.

Consider the topology shown in Figure 1, below, where PE1 is the source PE, PE2 is the RP PE, and PE3 and PE4 are each receiver Pes.

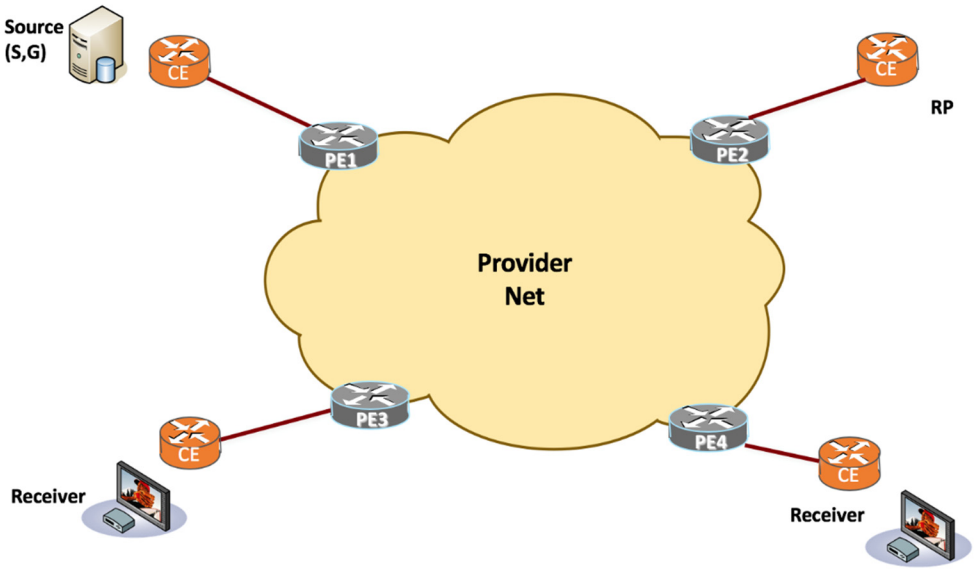


Figure 1

The normal sequence operations with BGP mVPNs, as defined in RFC 6514, is:

- Receiver PEs (PE3, PE4) will send a Type 6 (*,G) join targeted towards PE2.
- When the source becomes active, the first hop router will send a PIM register to the RP. This causes the RP to join the source tree.
- The RP PE will originate a Type 7 (S,G) join targeted towards PE1 to join the source tree and also join the mVPN core tree rooted at the source PE.
- Upon receiving the Type 7 join, PE1 will originate a Type 5 Source Active route.
- When the receiver PEs process this Type 5 route, they will each join the mVPN core tree rooted at the source PE to pull the customer traffic.

Incase all the receiver sites have SPT threshold infinity, the receiver PEs will not originate a Type 7 route. In such cases, it is mandatory for the RP PE to continue to advertise its Type 7 route to keep the source tree alive.

If one or more of the receiver sites have SPT Threshold 0, then the receiver PE will originate a Type 7 Route to join the source tree. In this case, the RP PE does not need to pull traffic towards itself and can let that receiver PE keep the source tree alive. However, the RP PE does not have the visibility of the sites connected to the receiver PEs to determine if any receiver PE has originated a Type 7 route.

At this point, all receiver PEs are receiving traffic from the source tree and the RP PE does not need to pull traffic towards itself. However, the RP PE does not have the visibility of the sites connected to the receiver PEs to determine if they have joined the source tree or not. Therefore, the RP PE has to default to the behavior of originating a Type 7 route to keep the source tree alive. The negative consequences of this is that the RP PE will always pull traffic from the source PE, even if there no receives located the RP PE. In certain scenarios, the RP PE even turns around this traffic to the egress PEs. The net effect is that, in the worst case, each multicast packet traverses the core three (3) times.

Proposed herein is the use of a route reflector or a centralized controller (collectively and generally referred to herein as a route reflector), which has a global visibility of all the routes of a Multicast VPN. The route reflector can keep track of which sites that have originated a Type 6 and a Type 7 route. As long as the route reflector sees one or more Type 7 routes from the set of PEs that have also originated a Type 6 route [matching the same group address], then the route reflector can notify the RP PE to prune itself from the source tree by withdrawing its own Type 7 route.

In general, there are three aspects/parts to the techniques presented herein, each of which is described further below.

Part 1: VPN Context Discovery:

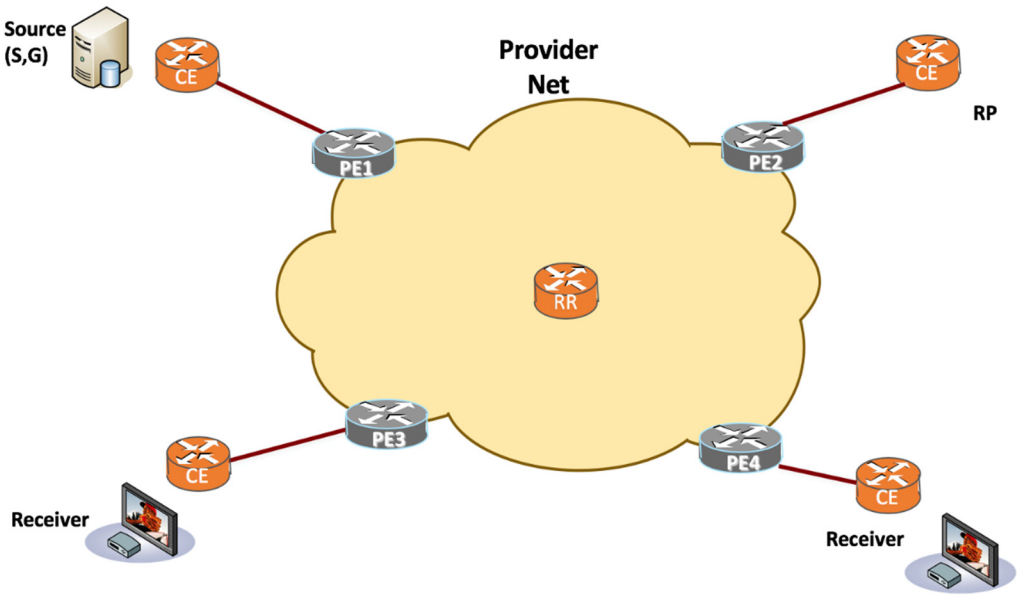
The route reflector does not have the context of the VPNs configured on each PE. The techniques presented herein require the route reflector to associate the BGP mVPN routes to a particular VPN context (Type 5/6/7 routes). Below are possible approaches to achieve this:

- Configure the same VPN ID on all PE devices that participate in a given VPN. This VPN ID is attached as an optional transitive community to all Type 5,6,7 routes originated by the PE for that VPN.
- A 3rd party module that allows the route reflector to derive the context from the BGP mVPN routes. For example, the route reflector could be configured with all the import RTs of all the PEs that participate in a given MVPN.

It is to be appreciated that, in addition to the above, there could also be other mechanisms used to allow the route reflector to associate the BGP mPVN routes to a given VPN.

Part 2: Tracking Type 6 and 7's from the Receiver PEs:

Figure 2, below, illustrates that, for a given Group address [G], the route reflector maintains a set of PEs that have originated a Type 6 route, referred to herein as a "receiver set." When the first PEs from this receiver set originate a Type 7 route for a [S,G], the route reflector will notify the RP PE to prune itself off from the source tree, as described in Part 3 below]



When the last Type 7 route is withdrawn from a PE in the receiver set and this set is not empty, the route reflector must ensure that the source tree is not pruned off. The route reflector may accomplish this by notifying the RP PE to join the source tree and not withdrawing the Type 7 route to the source PE. Eventually, the route reflector will receive a Type 7 from the RP PE to keep the source tree alive.

Part 3: Notifying RP PE:

RFC 6514 uses the Type 5 route to signal to the RP PE to prune a given source from the shared tree and thereby withdraw its Type 7 route. However, as described above, the RP PE defaults to always advertising a Type 7 route in order to keep the source tree alive. In the techniques presented herein, the Type 5 route is augmented with some information that allows the route reflector to notify whether or not the RP can prune the source off the shared tree. One mechanism to do this is to allow the route reflector to attach a PMSI Tunnel attribute [PTA] to the Type 5 route and use 1 bit in this flags field of this attribute to notify the RP. The RP PE will not withdraw the Type 7 route upon receiving a Type 5 route that contains the new PTA if it has local interest in the multicast flow.