Network Working Group
Internet-Draft
Updates: 6514 (if approved)
Intended status: Standards Track
Expires: December 5, 2013

Zhang
Rekhter
Juniper Networks
Dolganow
Alcatel-Lucent

June 3, 2013

Simulating "Partial Mesh of MP2MP P-Tunnels" with Ingress Replication
draft-zzhang-l3vpn-mvpn-bidir-ingress-replication-00.txt

Abstract

RFC 6513 described a method to support bidirectional C-flow using "Partial Mesh of MP2MP P-Tunnels". This document describes how partial mesh of MP2MP P-Tunnels can be simulated with Ingress Replication, instead of a real MP2MP tunnel.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

Section 11.2 of RFC 6513, "Partitioned Sets of PEs", describes two methods of carrying bidirectional C-flow traffic over a provider core without using the core as RPL or requiring Designated Forwarder election.

With these two methods, all PEs of a particular VPN are separated into partitions, with each partition being all the PEs that elect the same PE as the UMH wrt the C-RPA. A PE must discard bidirectional C-flow traffic from PEs that are not in the same partition as the PE itself.

In particular, Section 11.2.3 of RFC 6513, "Partial Mesh of MP2MP P-Tunnels", guarantees the above discard havavior without using an extra PE Distinguisher label by having all PEs in the same partition join a single MP2MP tunnel dedicated to that partition and use it to transmit traffic. All traffic arriving on the tunnel will be from PEs in the same partition, so it will be always accepted.

RFC 6514 specifies BGP encodings and procedures used to implement MVPN as specified in RFC 6513, while the details related to MP2MP tunnels are specified in [draft-ietf-l3vpn-mvpn-bidir-05].

[draft-ietf-l3vpn-mvpn-bidir-05] assumes that an MP2MP P-tunnel is realized either via PIM-Bidir, or via MP2MP mLDP. Each of them would require signaling and state not just on PEs, but on the P routers as well. This document describes how the MP2MP tunnel can be simulated with a mesh of P2P or MP2P LSPs, i.e. Ingress Replication. The advantage is that existing P2P/MP2P LSPs created for unicast can be used for multicast as well w/o introducing additional signaling or state in the core. While there may be concerns with traffic replication in the core, in many situations the traffic could be low-rate and/or sporadic and the advantage of signaling and state savings will outweight the concerns with traffic replication, making Ingress Replication an applicable and attractive alternative.

This documentation specifies the BGP signaling and procedures used to simulate "Partial Mesh of MP2MP P-Tunnels" with Ingress Replication.

1.1. Terminology

This document uses terminology from [RFC6513], [RFC6514], and [draft-ietf-l3vpn-mvpn-bidir-05]. In particular, the following new term is defined:

- C-G-BIDIR: A C-G where G is a Bidir-PIM group.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Operation

3.1. Control State

If a PE, say PEx, is connected to a site of a given VPN, and that site hosts the C-RPA for some Bidir-PIM groups, i.e., the route to the C-RPA is through a local PE-CE interface, then PEx MUST advertises a (C-*,C-BIDIR) S-PMSI A-D route, regardless of whether it has any local Bidir-PIM join states corresponding to the C-RPA learned from its CEs. It MAY also advertise a (C-*,C-G-BIDIR) S-PMSI A-D route, just like how any other S-PMSI A-D routes are triggered (e.g, when the (C-*,C-G-BIDIR) traffic rate goes above a threshold). Here the C-G-BIDIR refers to a C-G where G is a Bidir-PIM group, and the corresponding C-RPA is in the site that the PEx connects to.

The S-PMSI A-D routes include a Provider Tunnel Attribute (PTA) with tunnel type set to Ingress Replication, with Leaf Information Required flag set, and with a downstream allocated MPLS label that other PEs in the same partition MUST use when sending relevant C-bidir flows to this PE.

If some other PE, PEy, receives and imports into one of its VRFs such a (C-*,C-BIDIR) S-PMSI A-D route, and the VRF has any local Bidir-PIM join state that PEy has received from its CEs, and if PEy chooses PEx as its UMH wrt the C-RPA for those states, PEy MUST advertise a Leaf A-D route in response. Or, if PEy has received and imported into one of its VRFs a (C-*,C-BIDIR) S-PMSI A-D route from PEx before, then upon receiving in the VRF any local Bidir-PIM join state from its CEs with PEx being the UMH for those states' C-RPA, PEy MUST advertise a Leaf A-D route.

The encoding of the Leaf A-D route is as specified in RFC 6514, except that the Route Targets are set to the same value as in the corresponding S-PMSI A-D route so that the Leaf A-D route will be imported by all VRFs that import the corresponding S-PMSI A-D route. This is irrespective of whether from a receiving PE, PEz's perspective PEx (oiginator of the S-PMSI A-D route) is the UMH PE or not. The label in the PTA of the Leaf A-D route originated by PEy MUST be allocated specifically for PEx, so that when traffic arrives with that label, the traffic can associated with the partition (represented by the PEx).

With PEy advertising Leaf A-D route only if it chooses the originator of the S-PMSI A-D route as its UMH, it won't receive traffic from PEs in other partitions, so the label is actually useful only when PEy switches to a different UMH - it will stop accepting traffic before sending PEs stop sending it traffic (upon the receipt of its Leaf A-D route withdrawl). To speed up convergency (so that PEy starts

receiving traffic from its new UMH immediately instead of waiting until the new Leaf A-D route corresponding to the new UMH is received by sending PEs), PEy MAY advertise a Leaf A-D route even if does not choose PEx as its UMH wrt the C-RPA. With that, it will receive traffic from all PEs, but some will arrive with the label corresponding to its choice of UMH while some will arrive with a different label, and the traffic in the latter case will be discarded.

Similar to the (C-*,C-BIDIR) case, if PEy receives and imports into one of its VRFs such a (C-*,C-G-BIDIR) S-PMSI A-D route, and PEy chooses PEx as its UMH wrt the C-RPA, and it has corresponding local (C-*,C-G-BIDIR) join state that it has received from its CEs in the VRF, PEy MUST advertise a Leaf A-D route in response. Or, if PEy has received and imported into one of its VRFs a (C-*,C-G-BIDIR) S-PMSI A-D route before, then upon receiving its local (C-*,C-G-BIDIR) join state from its CEs in the VRF, it MUST advertise a Leaf A-D route.

The encoding of the Leaf A-D route is as specified in RFC 6514, except that the Route Targets are set to the same as in the corresponding S-PMSI A-D route so that the Leaf A-D route will be imported by all VRFs that import the corresponding S-PMSI A-D route. This is irrespective of whether from the receiving PE, PEz's perspective PEx (oiginator of the S-PMSI A-D route) is the UMH PE or not. The label in the PTA of the Leaf A-D route originated by PEy MUST be allocated specifically for PEx, so that when traffic arrives with that label, the traffic can associated with the partition (represented by the PEx).

Whenever the (C-*,C-BIDIR) or (C-*,C-G-BIDIR) S-PMSI A-D route is withdrawn, or if PEy no longer chooses the originator PEx as its UMH wrt C-RPA and PEy only advertises Leaf A-D routes in response to its UMH's S-PMSI A-D route, or if relevant local join state is pruned, PEy MUST withdraw the corresponding Leaf A-D route.

3.2. Forwarding State

The following specification regarding forwarding state matches the "When an S-PMSI is a 'Match for Transmission'" and "When an S-PMSI is a 'Match for Reception'" rules for "Flat Partitioning" method in [draft-ietf-l3vpn-mvpn-bidir-05], except that the rules about (C-*,C-*) are not applicable, because this document requires that (C-*,C-BIDIR) S-PMSI A-D routes are always originated for a VPN that supports C-Bidir flows.

For the (C-*,C-G-BIDIR) S-PMSI A-D route that a PEy receives and imports into one of its VRFs from its UMH wrt the C-RPA, or if PEy itself advertises the S-PMSI A-D route in the VRF, PEy maintains a

(C-*,C-G-BIDR) forwarding state in the VRF, with the Ingress Replication provider tunnel leaves being the originators of the S-PMSI A-D route and all relevant Leaf-A-D routes. The relevant Leaf A-D routes are the routes whose Route Key field contains the same information as the MCAST-VPN NLRI of the (C-*, C-G-BIDIR) S-PMSI A-D route advertised by the UMH.

For the (C-*,C-BIDIR) S-PMSI A-D route that a PEy receives and imports into one of its VRFs from its UMH wrt a C-RPA, or if PEy itself advertises the S-PMSI A-D route in the VRF, it maintains appropriate forwarding states in the VRF for the ranges of bidirectional groups for which the C-RPA is responsible. The provider tunnel leaves are the originators of the S-PMSI A-D route and all relevant Leaf-A-D routes. The relevant Leaf A-D routes are the routes whose Route Key field contains the same information as the MCAST-VPN NLRI of the (C-*, C-BIDIR) S-PMSI A-D route advertised by the UMH. This is for the so-called "Sender Only Branches" where a router only has data to send upstream towards C-RPA but no explicit join state for a particular bidirectional group. Note that the traffic must be sent to all PEs (not just the UMH) in the partition, because they may have specific (C-*,C-G-BIDIR) join states that this PEy is not aware of, while there is no corresponding (C-*,C-G-BIDIR) S-PMSI A-D and Leaf A-D routes.

For a (C-*,C-G-BIDIR) join state that a PEy has received from its CEs in a VRF, if there is no corresponding (C-*,C-G-BIDIR) S-PMSI A-D route from its UMH in the VRF, PEy maintains a corresponding forwarding state in the VRF, with the provider tunnel leaves being the originators of the (C-*,C-BIDIR) S-PMSI A-D route and all relevant Leaf-A-D routes (same as the above Sender Only Branch case). The relevant Leaf A-D routes are the routes whose Route Key field contains the same information as the MCAST-VPN NLRI of the (C-*, C-BIDIR) S-PMSI A-D route originated by the UMH. If there is no (C-*,C-BIDIR) S-PMSI A-D route from its UMH either, then the provider tunnel has an empty set of leaves and PEy does not forward relevant traffic across the provider network.

4. Security Considerations

This document raises no new security issues. Security considerations for the base protocol are covered in [RFC6514].

5. IANA Considerations

This document has no IANA considerations.

This section should be removed by the RFC Editor prior to final publication.

6. Acknowledgements

7. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC6513] Rosen, E. and R. Aggarwal, "Multicast in MPLS/BGP IP VPNs", RFC 6513, February 2012.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, February 2012.

[RFC5015] Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", RFC 5015, October 2007.

[I-D.ietf-l3vpn-mvpn-bidir] Rosen, E., Wijnands, I., Cai, Y., and A. Boers, "MVPN: Using Bidirectional P-Tunnels", draft-ietf-l3vpn-mvpn-bidir-05 (work in progress), April 2013.

Authors' Addresses

Jeffrey Zhang
Juniper Networks
10 Technology Park Dr.
Westford, MA 01886
US

Email: zzhang@juniper.net

Yakov Rekhter
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
US

Email: yakov@juniper.net

Andrew Dolganow
Alcatel-Lucent
600 March Rd.
Ottawa, ON K2K 2E6
CANADA

Email: andrew.dolganow@alcatel-lucent.com