Workgroup: Network Working Group
Internet-Draft:
draft-wang-bess-mvpn-upstream-df-selection-00
Published: 23 October 2021
Intended Status: Informational
Expires: 26 April 2022
Authors: H. Wang F. Duan
Huawei Technologies Huawei Technologies

Multicast VPN Upstream Designated Forwarder Selection

Abstract

This document defines Multicast Virtual Private Network (VPN) extensions and procedures that allow fast failover for upstream failures by allowing upstream Provider Edges (PEs) to determine a single forwarder for a VPN multicast flow, without the downstream PEs' duplication prevention. The fast failover is accomplished by using Virtual Router Redundancy Protocol (VRRP) [RFC5798] or similar technologies for the upstream PEs to determine a single desinated fowarder. Also, this document introduces a new BGP Extended Community called "Upstream Forwarder Selection", carried by BGP VPN route so that the upstream PEs can inform downstream PEs the election behavior. The downstream PEs, accordingly, send C-multicast routes to both the primary and standby upstream PEs and forward the multicast flow comming from both sides to the CEs.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

MVPN [RFC6513] and [RFC6514] defines the MVPN architecture and MVPN protocol specification which include the basic procedures for selecting the Upstream Multicast Hop. Further [RFC9026] defines some extension that allow fast failover for upstream failures by allowing downstream PEs to consider the status of Provider-Tunnels (P-tunnels) when selecting the Upstream PE for a VPN multicast flow. However, there are some problems when deploying the "hot root standby" mechanism described in [RFC9026].

First, all the ingress PEs, regardless of the primary or standby role, forward (C-S,C-G) flow to other PEs though a P-tunnel, forcing the egress PEs to discard all but one, which will cause the steady traffic redundancy throughout the backbone network.

Second, an efficient and accurate method for the downstream PEs to determine the "status" of a P-tunnel is required, which is somewhat complicated in some cases, as mentioned in Section 3.1.8 of [RFC9026]

This document proposes a different "warm root standby" procedure mentioned in Section 4.2 of [RFC9026]. The procedures include a) an upstream designated forwarder election between multi homing ingress PEs, and b) the downstream PEs' advertising Primary and Standby BGP C-multicast route and accepting traffic from any of both sides.

Section 3 describes procedures allowing multi homing ingress PEs to determine "locally" a single forwarder to avoid duplicate packets sending through the backbone, without the egress PEs' primary or standby UMH selection.

Section 4 describes an optional BGP Extended Community called "Upstream Forwarder Selection", which is carried by BGP VPN routes (SAFI 128 or 129), to inform the downstream PEs the selection behavior describes in Section 3.

Section 5 describes the downstream PEs' behavior in this case. The downstream PEs advertise C-multicast Source Tree Join route to both the primary and secondary Upstream PEs (carrying, as Route Target extended communities, the values of the VRF Route Import Extended Community of each VPN route from each Upstream PE). The Upstream Forwarder Selection Extended Community indicates that the packet duplication prevention will be accomplished by the upstream PEs and that any of the traffic from both the primary and secondary upstream PEs would be acceptable to be forwarded to the CEs.

2. Terminology

Readers of this document are assumed to be familiar with the terminology and concepts of the documents listed as Normative References.

3. Upstream Designated Forwarder Selection

Section 9.1.2 of [RFC6513] describes a "single forwarder selection" to ensure that duplicate packets not sending through the backbone. This document proposes a deployment of VRRP or some similar technology to enable dual or multi homing ingress PEs to determine a designated forwarder.

3.1. Upstream Designated Forwarder Selection by VRRP

VRRP specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers on a LAN. The VRRP router controlling the IPv4 or IPv6 address(es) associated with a virtual router is called the Master, and it forwards packets sent to these IPv4 or IPv6 addresses. Similarly, the role of the VRRP routers associated with a virtual router can also be that of the upstream PEs in MVPN dual homing upstream PEs deployment.

Virtual Router -- pair of dual homing upstream PEs

Virtual Router Master -- the primary upstream PE

Virtual Router Backup -- the standby upstream PE

The method of mapping the role of a VRRP router to that of a MVPN upstream PE is more likely an administrative measure and could be implemented as configurable policies. Both the primary and standby PEs install VRF PIM state corresponding to BGP Source Tree Join route and send C-Join messages to the CE toward C-S. Whereas only the primary upstream PE (Virtual Router Master according to VRRP) forwards (C-S,C-G) flow to downstream PEs through a P-tunnel.

3.2. Other Feasible Selection Technologies

VRRP is just an example of the feasible choices for the dual homing upstream PEs' single forwarder selection. Other private implementations or similar designated forwarder selection technologies could also be optional for further study. However, a feasible technology should has the ability of being deployed per VRF and being associated with one Multicast VPN instance.

4. Upstream Forwarder Selection Extended Community

This document defines a new BGP Extended Community called "Upstream Forwarder Selection".

The Upstream Forwarder Selection is an IP-address-specific Extended Community, of an extended type, and is transitive across AS boundaries [RFC4360].

An upstream PE constructs Upstream Forwarder Selection as follows, regardless of the role of the selection result:

The Global Administrator field of the Upstream Forwarder Selection SHOULD be set to a virtual IP address (or similar identity) of the upstream PEs (such as the VRRP Virtual IP address when using VRRP), which is identical between primary and standby PEs.

The Local Administrator field of the Upstream Forwarder Selection SHOULD be set to a master or backup status determined by the election which is different between primary and standby PEs.

Similar with the carrying of the VRF Route Import Extended Community imposed in Section 7 of [RFC6514], the multi homing PEs MUST also include in the BGP Updates message that carries the (unicast) VPN route the Upstream Forwarder Selection Extended Community that has the value of DF election result associated with this VRF.

5. Downstream PE Behavior

5.1. Standby C-multicast Route Advertisment

The Standby BGP C-multicast route advertisement described in Section 4 of [RFC9026] is still necessary. One downstream PE needs to determine a secondary UMH, originates and sends C-multicast routes with RTs that identify both the Primary and Standby Upstream PEs. However, because of the duplication prevention being accomplished by the upstream DF selection described above, carrying the new Standby PE BGP Communities with C-multicast routes is no longer a indispensable requirement.

5.2. Anycast Reverse Path Forwarding Checking

Multicast VPN specifications [RFC6513] impose that a downstream PE only forwards to CEs the packets coming from the expected Upstream PE (Section 9.1.1 of [RFC6513]).

When performing the UMH selection, if a route in the set of VPN-IP eligible UMH routes carries the Upstream Forwarder Selection Extended Community, the Upstream PE determined from the route should be considered a potentially valid Upstream PE. In most cases, there should be two of that routes for one (C-S,C-G) flow, indicateing the primary and standby upstream PEs. As a result, the downstream PE accepts the (C-S,C-G) flow from any of both sides and forward it to CEs. It is a kind of "anycast" reverse path forwarding (RPF) checking. Eventually, it is the upstream single forwarder selection mechanism that ensures the duplicate packets not passing through the backbone network, as described in Section 3.

6. Security Considerations

This document introduces no new security considerations beyond those already specified in [RFC6513] and [RFC6514].

7. IANA Considerations

This document contains no actions for IANA.

8. Acknowledgements

The authors wish to thank Jingrong Xie, for his reviews, comments and suggestions.

9. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[RFC5798] Nadas, S., Ed., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, DOI 10.17487/RFC5798, March 2010, <https://www.rfc-editor.org/info/rfc5798>.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC9026] Morin, T., Ed., Kebler, R., Ed., and G. Mirsky, Ed., "Multicast VPN Fast Upstream Failover", RFC 9026, DOI 10.17487/RFC9026, April 2021, <https://www.rfc-editor.org/info/rfc9026>.

Authors' Addresses

Heng Wang
Huawei Technologies

Email: wangheng21@huawei.com

Fanghong Duan
Huawei Technologies

Email: duanfanghong@huawei.com