BESS Working Group Z. Zhang
Internet-Draft L. Giuliano
Intended status: Standards Track E. Rosen, Ed.
Expires: March 20, 2016 Juniper Networks, Inc.
K. Subramanian
Cisco Systems, Inc.
D. Pacella
Verizon
September 17, 2015

Global Table Multicast with BGP-MVPN Procedures
draft-ietf-bess-mvpn-global-table-mcast-03

Abstract

RFC6513, RFC6514, and other RFCs describe protocols and procedures which a Service Provider (SP) may deploy in order offer Multicast Virtual Private Network (Multicast VPN or MVPN) service to its customers. Some of these procedures use BGP to distribute VPN-specific multicast routing information across a backbone network. With a small number of relatively minor modifications, the very same BGP procedures can also be used to distribute multicast routing information that is not specific to any VPN. Multicast that is outside the context of a VPN is known as "Global Table Multicast", or sometimes simply as "Internet multicast". In this document, we describe the modifications that are needed to use the MVPN BGP procedures for Global Table Multicast.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

[RFC4364] specifies architecture, protocols, and procedures that a Service Provider (SP) can use to provide Virtual Private Network (VPN) service to its customers. In that architecture, one or more Customer Edge (CE) routers attach to a Provider Edge (PE) router. Each CE router belongs to a single VPN, but CE routers from several VPNs may attach to the same PE router. In addition, CEs from the same VPN may attach to different PEs. BGP is used to carry VPN-specific information among the PEs. Each PE router maintains a separate Virtual Routing and Forwarding table (VRF) for each VPN to which it is attached.

[RFC6513] and [RFC6514] extend the procedures of [RFC4364] to allow the SP to provide multicast service to its VPN customers. The customer's multicast routing protocol (e.g., PIM) is used to exchange multicast routing information between a CE and a PE. The PE stores a given customer's multicast routing information in the VRF for that customer's VPN. BGP is used to distribute certain multicast-related control information among the PEs that attach to a given VPN, and BGP may also be used to exchange the customer multicast routing information itself among the PEs.

While this multicast architecture was originally developed for VPNs, it can also be used (with a small number of modifications to the procedures) to distribute multicast routing information that is not specific to VPNs. The purpose of this document is to specify the way in which BGP MVPN procedures can be adapted to support non-VPN multicast.

Multicast routing information that is not specific to VPNs is stored in a router's "global table", rather than in a VRF; hence it is known as "Global Table Multicast" (GTM). GTM is sometimes more simply called "Internet multicast". However, we will avoid that term because it suggests that the multicast data streams are available on the "public" Internet. The procedures for GTM can certainly be used to support multicast on the public Internet, but they can also be used to support multicast streams that are not public, e.g., content distribution streams offered by content providers to paid subscribers. For the purposes of this document, all that matters is that the multicast routing information is maintained in a global table rather than in a VRF.

This architecture does assume that the network over which the multicast streams travel can be divided into a "core network" and one or more non-core parts of the network, which we shall call "attachment networks". The multicast routing protocol used in the attachment networks may not be the same as the one used in the core,

so we consider there to be a "protocol boundary" between the core network and the attachment networks. We will use the term "Protocol Boundary Router" (PBR) to refer to the core routers that are at the boundary. We will use the term "Attachment Router" (AR) to refer to the routers that are not in the core but that attach to the PBRs.

This document does not make any particular set of assumptions about the protocols that the ARs and the PBRs use to exchange unicast and multicast routing information with each other. For instance, multicast routing information could be exchanged between an AR and a PBR via PIM, IGMP, or even BGP. Multicast routing also depends on an exchange of routes that are used for looking up the path to the root of a multicast tree. This routing information could be exchanged between an AR and a PBR via IGP, via EBGP, or via IBGP ([RFC6368]). Note that if IBGP is used, the [RFC6368] "push/pop procedures" are not necessary.

The PBRs are not necessarily "edge" routers, in the sense of [RFC4364]. For example, they may be both be Autonomous System Border Routers (ASBR). As another example, an AR may be an "access router" attached to a PBR that is an OSPF Area Border Router (ABR). Many other deployment scenarios are possible. However, the PBRs are always considered to be delimiting a "backbone" or "core" network. A multicast data stream from an AR is tunneled over the core network from an Ingress PBR to one or more Egress PBRs. Multicast routing information that a PBR learns from the ARs attached to it is stored in the PBR's global table. The PBRs use BGP to distribute multicast routing and auto-discovery information among themselves. This is done following the procedures of [RFC6513], [RFC6514], and other MVPN specifications, as modified in this document.

In general, PBRs follow the same MVPN/BGP procedures that PE routers follow, except that these procedures are adapted to be applicable to the global table rather than to a VRF. Details are provided in subsequent sections of this document.

By supporting GTM using the BGP procedures designed for MVPN, one obtains a single control plane that governs the use of both VPN and non-VPN multicast. Most of the features and characteristics of MVPN carry over automatically to GTM. These include scaling, aggregation, flexible choice of tunnel technology in the SP network, support for both segmented and non-segmented tunnels, ability to use wildcards to identify sets of multicast flows, support for the Any Source Multicast (ASM), Single Source Multicast (SSM), and Bidirectional (bidir) multicast paradigms, support for both IPv4 and IPv6 multicast flows over either an IPv4 or IPv6 SP infrastructure, support for unsolicited flooded data (including support for BSR as RP-to-group mapping protocols), etc.

This document not only uses MVPN procedures for GTM, but also, insofar as possible, uses the same protocol elements, encodings, and formats. The BGP Updates for GTM thus use the same Subsequent Address Family Identifier (SAFI), and have the same Network Layer Reachability Information (NLRI) format, as the BGP Updates for MVPN.

Details for supporting MVPN (either IPv4 or IPv6 MVPN traffic) over an IPv6 backbone network can be found in [RFC6515]. The procedures and encodings described therein are also applicable to GTM.

The document [RFC7524] extends [RFC6514] by providing procedures that allow tunnels through the core to be "segmented" at ABRs within the core. The ABR segmentation procedures are also applicable to GTM as defined in the current document. In general, the MVPN procedures of [RFC7524], adapted as specified in the current document, are applicable to GTM.

The document [RFC7524] also defines a set of procedures for GTM. Those procedures are different from the procedures defined in the current document, and the two sets of procedures are not interoperable with each other. The two sets of procedures can co-exist in the same network, as long as they are not applied to the same multicast flows or to the same multicast group addresses. See Section 3 for more details.

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].

2. Adapting MVPN Procedures to GTM

In general, PBRs support Global Table Multicast by using the procedures that PE routers use to support VPN multicast. For GTM, where [RFC6513] and [RFC6514] talk about the "PE-CE interface", one should interpret that to mean the interface between the AR and the PBR. For GTM, where [RFC6513] and [RFC6514] talk about the "backbone" network, one should interpret that to mean the part of the network that is delimited by the PBRs.

A few adaptations to the procedures of [RFC6513] and [RFC6514] need to be made. Those adaptations are described in the following sub-sections.

2.1. Use of Route Distinguishers

The MVPN procedures require the use of BGP routes, defined in [RFC6514], that have a SAFI value of 5 ("MCAST-VPN"). We refer to these simply as "MCAST-VPN routes". [RFC6514] defines the Network

Layer Reachability Information (NLRI) format for MCAST-VPN routes. The NLRI field always begins with a "Route Type" octet, and, depending on the route type, may be followed by a "Route Distinguisher" (RD) field.

When a PBR originates an MCAST-VPN route in support of GTM, the RD field (for those routes types where it is defined) of that route's NLRI MUST be set to zero (i.e., to 64 bits of zero). Since no VRF may have an RD of zero, this allows "MCAST-VPN" routes that are "about" GTM to be distinguished from MCAST-VPN routes that are about VPNs.

2.2. Use of Route Targets

The MVPN procedures require all MCAST-VPN routes to carry Route Targets (RTs). When a PE router receives an MCAST-VPN route, it processes the route in the context of a particular VRF if and only if the route is carrying an RT that is configured as one of that VRF's "import RTs".

There are two different "kinds" of RT used in MVPN.

o One kind of RT is carried only by the following MCAST-VPN route types: C-multicast Shared Tree Joins, C-multicast Source Tree Joins, and Leaf A-D routes. This kind of RT identifies the PE router that has been selected by the route's originator as the "Upstream PE" or as the "Upstream Multicast Hop" (UMH) for a particular (set of) multicast flow(s). Per [RFC6514] and [RFC6515], this RT must be an IPv4-address-specific or IPv6-address-specific Extended Community (EC), whose "Global Administrator" field identifies the Upstream PE or the UMH. If the Global Administrator field identifies the Upstream PE, the "Local Administrator" field identifies a particular VRF in that PE.

The GTM procedures of this document require the use of this type of RT, in exactly the same situations where it is used in the MVPN specification. However, one adaptation is necessary: the "Local Administrator" field of this kind of RT MUST always be set to zero, thus implicitly identifying the global table, rather than identifying a VRF. We will refer to this kind of RT as an "upstream-node-identifying RT".

o The other kind of RT is the conventional RT first specified in [RFC4364]. It does not necessarily identify a particular router by address, but is used to constrain the distribution of VPN routes, and to ensure that a given VPN route is processed in the

context of a given VRF if and only if the route is carrying an RT that has been configured as one of that VRF's "import RTs".

Whereas every VRF must be configured with at least one import RT, there is heretofore no requirement to configure any RTs for the global table of any router. As stated above, this document makes the use of upstream-node-identifying RTs mandatory for GTM. This document makes the use of non-upstream-node-identifying RTs OPTIONAL for GTM.

The procedures for the use of RTs in GTM are the following:

o If the global table of a particular PBR is NOT configured with any import RTs, then a received MCAST-VPN route is processed in the context of the global table only if it is carrying no RTs, or if it is carrying an upstream-node-identifying RT whose Global Administrator field identifies that PBR.

o The global table in each PBR MAY be configured with (a) a set of export RTs to be attached to MCAST-VPN routes that are originated to support GTM, and (b) with a set of import RTs for GTM.

If the global table of a given PBR has been so configured, the PBR will process a received MCAST-VPN route in the context of the global table if and only if the route carries an RT that is one of the global table's import RTs, or if the route carries an upstream-node-identifying RT whose global administrator field identifies the PBR.

If the global tables are configured with RTs, care must be taken to ensure that the RTs configured for the global table are distinct from any RTs used in support of MVPN (except in the case where it is actually intended to create an "extranet" [MVPN-extranet] in which some sources are reachable in global table context while others are reachable in VPN context.)

The "RT Constraint" procedures of [RFC4684] MAY be used to constrain the distribution of MCAST-VPN routes (or other routes) that carry RTs that have been configured as import RTs for GTM. (This includes the upstream-node-identifying RTs.)

N.B.: If the "RT Constraint" procedures of [RFC4684] are deployed, but the MCAST-VPN routes are not carrying RTs, then proper operation requires the "default behavior" specified for the MCAST-VPN address family in Section 3 ("Default Behavior") of [RTC_without_RTs].

In [RFC6513], the UMH-eligible routes (see section 5.1 of [RFC6513], "Eligible Routes for UMH Selection") are generally routes of SAFI 128 (Labeled VPN-IP routes) or 129 (VPN-IP multicast routes), and are required to carry RTs. These RTs determine which VRFs import which such routes. However, for GTM, when the UMH-eligible routes may be routes of SAFI 1, 2, or 4, the routes are not required to carry RTs. This document does NOT specify any new rules for determining whether a SAFI 1, 2, or 4 route is to be imported into the global table of any PBR.

2.3. UMH-eligible Routes

[RFC6513] section 5.1 defines procedures by which a PE router determines the "C-root", the "Upstream Multicast Hop" (UMH), the "Upstream PE", and the "Upstream RD" of a given multicast flow. (In non-VPN multicast documents, the UMH of a multicast flow at a particular router is generally known as the "RPF neighbor" for that flow.) It also defines procedures for determining the "Source AS" of a particular flow. Note that in GTM, the "Upstream PE" is actually the "Upstream PBR".

The definition of the C-root of a flow is the same for GTM as for MVPN.

For MVPN, to determine the UMH, Upstream PE, Upstream RD, and Source AS of a flow, one looks up the C-root of the flow in a particular VRF, and finds the "UMH-eligible" routes (see section 5.1.1 of [RFC6513]) that "match" the C-root. From among these, one is chosen as the "selected UMH route".

For GTM, the C-root is of course looked up in the global table, rather than in a VRF. For MVPN, the UMH-eligible routes are routes of SAFI 128 or 129. For GTM, the UMH-eligible routes are routes of SAFI 1, SAFI 4, or SAFI 2. If the global table has imported routes of SAFI 2, then these are the UMH-eligible routes. Otherwise, routes of SAFI 1 or SAFI 4 are the UMH-eligible routes. For the purpose of UMH determination, if a SAFI 1 route and a SAFI 4 route contain the same IP prefix in their respective NLRI fields, then the two routes are considered by the BGP bestpath selection process to be comparable.

[RFC6513] defines procedures for determining which of the UMH-eligible routes that match a particular C-root is to become the "Selected UMH route". With one exception, these procedures are also applicable to GTM. The one exception is the following. Section 9.1.2 of [RFC6513] defines a particular method of choosing the Upstream PE, known as "Single Forwarder Selection" (SFS). This

procedure MUST NOT be used for GTM (see Section 2.3.4 for an explanation of why the SFS procedure cannot be applied to GTM).

In GTM, the "Upstream RD" of a multicast flow is always considered to be zero, and is NOT determined from the Selected UMH route.

The MVPN specifications require that when BGP is used for distributing multicast routing information, the UMH-eligible routes MUST carry the VRF Route Import EC and the Source AS EC. To determine the Upstream PE and Source AS for a particular multicast flow, the Upstream PE and Source AS are determined, respectively, from the VRF Route Import EC and the Source AS EC of the Selected UMH route for that flow. These ECs are generally attached to the UMH-eligible routes by the PEs that originate the routes.

In GTM, there are certain situations in which it is allowable to omit the VRF Route Import EC and/or the Source AS EC from the UMH-eligible routes. The following sub-sections specify the various options for determining the Upstream PBR and the Source AS in GTM.

The procedures in Section 2.3.1 MUST be implemented. The procedures in Section 2.3.2 and Section 2.3.3 are OPTIONAL to implement. It should be noted that while the optional procedures may be useful in particular deployment scenarios, there is always the potential for interoperability problems when relying on OPTIONAL procedures.

2.3.1. Routes of SAFI 1, 2 or 4 with MVPN ECs

If the UMH-eligible routes have a SAFI of 1, 2 or 4, then they MAY carry the VRF Route Import EC and/or the Source AS EC. If the selected UMH route is a route of SAFI 1, 2 or 4 that carries the VRF Route Import EC, then the Upstream PBR is determined from that EC. Similarly, if the selected UMH route is a route of SAFI 1, 2, or 4 route that carries the Source AS EC, the Source AS is determined from that EC.

When the procedure of this section is used, a PBR that distributes a UMH-eligible route to other PBRs is responsible for ensuring that the VRF Route Import and Source AS ECs are attached to it.

If the selected UMH-eligible route has a SAFI of 1, 2 or 4, but is not carrying a VRF Route Import EC, then the Upstream PBR is determined as specified in Section 2.3.2 or Section 2.3.3 below.

If the selected UMH-eligible route has a SAFI of 1, 2 or 4, but is not carrying a Source AS EC, then the Source AS is considered to be the local AS.

2.3.2. MVPN ECs on the Route to the Next Hop

Some service providers may consider it to be undesirable to have the PBRs put the VRF Route Import EC on all the UMH-eligible routes. Or there may be deployment scenarios in which the UMH-eligible routes are not advertised by the PBRs at all. The procedures described in this section provide an alternative that can be used under certain circumstances.

The procedures of this section are OPTIONAL.

In this alternative procedure, each PBR MUST originate a BGP route of SAFI 1, 2 or 4 whose NLRI is an IP address of the PBR itself. This route MUST carry a VRF Route Import EC that identifies the PBR. The address that appears in the Global Administrator field of that EC MUST be the same address that appears in the NLRI and in the Next Hop field of that route. This route MUST also carry a Source AS EC identifying the AS of the PBR.

Whenever the PBR distributes a UMH-eligible route for which it sets itself as next hop, it MUST use this same IP address as the Next Hop of the UMH-eligible route that it used in the route discussed in the prior paragraph.

When the procedure of his section is used, then when a PBR is determining the Selected UMH Route for a given multicast flow, it may find that the Selected UMH Route has no VRF Route Import EC. In this case, the PBR will look up (in the global table) the route to the Next Hop of the Selected UMH route. If the route to the Next Hop has a VRF Route Import EC, that EC will be used to determine the Upstream PBR, just as if the EC had been attached to the Selected UMH Route.

If recursive route resolution is required in order to resolve the next hop, the Upstream PBR will be determined from the first route with a VRF Route Import EC that is encountered during the recursive route resolution process. (The recursive route resolution process itself is not modified by this document.)

The same procedure can be applied to find the Source AS, except that the Source AS EC is used instead of the VRF Route Import EC.

Note that this procedure is only applicable in scenarios where it is known that the Next Hop of the UMH-eligible routes is not be changed by any router that participates in the distribution of those routes; this procedure MUST NOT be used in any scenario where the next hop may be changed between the time one PBR distributes the route and another PBR receives it. The PBRs have no way of determining

dynamically whether the procedure is applicable in a particular deployment; this must be made known to the PBRs by provisioning.

Some scenarios in which this procedure can be used are:

- all PBRs are in the same AS, or

- the UMH-eligible routes are distributed among the PBRs by a Route Reflector (that does not change the next hop), or

- the UMH-eligible routes are distributed from one AS to another through ASBRs that do not change the next hop.

If the procedures of this section are used in scenarios where they are not applicable, GTM will not function correctly.

2.3.3. Non-BGP Routes as the UMH-eligible Routes

In particular deployment scenarios, there may be specific procedures that can be used, in those particular scenarios, to determine the Upstream PBR for a given multicast flow.

Suppose the PBRs neither put the VRF Route Import EC on the UMH-eligible routes, nor do they distribute BGP routes to themselves. It may still be possible to determine the Upstream PBR for a given multicast flow, using specific knowledge about the deployment.

For example, suppose it is known that all the PBRs are in the same OSPF area. It may be possible to determine the Upstream PBR for a given multicast flow by looking at the link state database to see which router is attached to the flow's C-root.

As another example, suppose it is known that the set of PBRs is fully meshed via Traffic Engineering (TE) tunnels. When a PBR looks up, in its global table, the C-root of a particular multicast flow, it may find that the next hop interface is a particular TE tunnel. If it can determine the identify of the router at the other end of that TE tunnel, it can deduce that that router is the Upstream PBR for that flow.

This is not an exhaustive set of examples. Any procedure that correctly determines the Upstream PBR in a given deployment scenario MAY be used in that scenario.

2.3.4. Why SFS Does Not Apply to GTM

To see why the SFS procedure cannot be applied to GTM, consider the following example scenario. Suppose some multicast source S is homed to both PBR1 and PBR2, and suppose that both PBRs export a route (of SAFI 1, 2, or 4) whose NLRI is a prefix matching the address of S. These two routes will be considered comparable by the BGP decision process. A route reflector receiving both routes may thus choose to redistribute just one of the routes to S, the one chosen by the bestpath algorithm. Different route reflectors may even choose different routes to redistribute (i.e., one route reflector may choose the route to S via PBR1 as the bestpath, while another chooses the route to S via PBR2 as the bestpath). As a result, some PBRs may receive only the route to S via PBR1 and some may receive only the route to S via PBR2. In that case, it is impossible to ensure that all PBRs will choose the same route to S.

The SFS procedure works in VPN context as long as the following assumption holds: if S is homed to VRF-x in PE1 and to VRF-y in PE2, then VRF-x and VRF-y have been configured with different RDs. In VPN context, the route to S is of SAFI 128 or 129, and thus has an RD in its NLRI. So the route to S via PE1 will not have the same NLRI as the route to S via PE2. As a result, all PEs will see both routes, and the PEs can implement a procedure that ensures that they all pick the same route to S.

That is, the SFS procedure of [RFC6513] relies on the UMH-eligible routes being of SAFI 128 or 129, and relies on certain VRFs being configured with distinct RDs. Thus the procedure cannot be applied to GTM.

One might think that the SFS procedure could be applied to GTM as long as the procedures defined in [ADD-PATH] are applied to the UMH-eligible routes. Using the [ADD-PATH] procedures, the BGP speakers could advertise more than one path to a given prefix. Typically [ADD-PATH] is used to report the n best paths, for some small value of n. However, this is not sufficient to support SFS, as can be seen by examining the following scenario.

```
              AS-X  |   AS-Y  |    AS-Z
                    |         |
   S--PBR1---ASBR1--|--ASBR2--|---ASBR5
   |   \______/     |         |
   |   /      \     |         |
   |--PBR2---ASBR3--|--ASBR4--|---ASBR6
                    |         |
```

In AS-X, PBR1 reports to both ASBR1 and ASBR3 that it has a route to S. Similarly, PBR2 reports to both ASBR1 and ASBR3 that it has a route to S. Using [ADD-PATH], ASBR1 reports both routes to ASBR2, and ASBR3 reports both routes to ASBR4. Now AS-Y sees 4 paths to S. The AS-Z ASBRs will each see eight paths (four via ASBR2 and four via ASBR4). To avoid this explosion in the number of paths, a BGP speaker that uses [ADD-PATH] is usually considered to report only the n best paths. However, there is then no guarantee that the reported set of paths will contain at least one path via PBR1 and at least one path via PBR2. Without such a guarantee, the SFS procedure will not work.

2.4. Inclusive and Selective Tunnels

The MVPN specifications allow multicast flows to be carried on either Inclusive Tunnels or on Selective Tunnels. When a flow is sent on an Inclusive Tunnel of a particular VPN, it is sent to all PEs in that VPN. When sent on a Selective Tunnel of a particular VPN, it may be sent to only a subset of the PEs in that VPN.

This document allows the use of either Inclusive Tunnels or Selective Tunnels for GTM. However, any service provider electing to use Inclusive Tunnels for GTM should carefully consider whether sending a multicast flow to ALL its PBRs would result in problems of scale. There are potentially many more MBRs for GTM than PEs for a particular VPN. If the set of PBRs is large and growing, but most multicast flows do not need to go to all the PBRs, the exclusive use of Selective Tunnels may be a better option.

2.5. I-PMSI A-D Routes

2.5.1. Intra-AS I-PMSI A-D Routes

Per [MVPN-BGP}, there are certain conditions under which is it NOT required for a PE router implementing MVPN to originate one or more Intra-AS I-PMSI A-D routes. These conditions apply as well to PBRs implementing GTM.

In addition, a PBR implementing GTM is NOT required to originate an Intra-AS I-PMSI A-D route if both of the following conditions hold:

- The PBR is not using Inclusive Tunnels for GTM, and

- The distribution of the C-multicast Shared Tree Join and C-multicast Source Tree Join routes is done in such a manner that the next hop of those routes does not change.

Please see also the sections on RD and RT usage (Sections 2.1 and 2.2 respectively).

2.5.2. Inter-AS I-PMSI A-D Routes

There are no GTM-specific procedures for the origination, distribution, and processing of these routes, other than those specified in the sections on RD and RT usage.

2.6. S-PMSI A-D Routes

There are no GTM-specific procedures for the origination, distribution, and processing of these routes, other than those specified in the sections on RD and RT usage.

2.7. Leaf A-D Routes

There are no GTM-specific procedures for the origination, distribution, and processing of these routes, other than those specified in the sections on RD and RT usage.

2.8. Source Active A-D Routes

Please see the sections on RD and RT usage for information applies to the origination and distribution of Source Active A-D routes. Additional procedures governing the use of Source Active A-D routes are given in the sub-sections of this section.

2.8.1. Finding the Originator of an SA A-D Route

To carry out the procedures specified in [RFC6514] (e.g., in Section 13.2 of that document), it is sometimes necessary for an egress PE to determine the ingress PE that originated a given Source Active A-D route. The procedure used in [RFC6514] to find the originator of a Source Active A-D route assumes that no two routes have the same RD unless they have been originated by the same PE. However, this assumption is not valid in GTM, because each Source Active A-D route used for GTM will have an RD of 0, and all the UMH-eligible routes also have an RD of 0. So GTM requires a different procedure for determining the originator of a Source Active A-D route.

In GTM, the procedure for determining the originating PE of a Source Active A-D route is the following:

- o When a Source Active A-D route is originated, the originating PE MAY attach a VRF Route Import Extended Community to the route.

- When a Source Active A-D route is distributed by one BGP speaker to another, then

  * if the Source Active A-D route does not carry the VRF Route Import EC, the BGP speaker distributing the route MUST NOT change the route's next hop field;

  * if the Source Active A-D route does carry the VRF Route Import EC, the BGP speaker distributing the route MAY change the route's next hop field to itself.

- When an egress PE needs to determine the originator of a Source Active A-D route, then

  * if the Source Active A-D route carries the VRF Route Import EC, the originating PE is the PE identified in the Global Administrator field of that EC;

  * if the Source Active A-D route does not carry the VRF Route Import EC, the originating PE is the PE identified in the route's next hop field.

2.8.2. Optional Additional Constraints on Distribution

If some site has receivers for a particular ASM group G, then it is possible (by the procedures of [RFC6514]) that every PBR attached to a site with a source for group G will originate a Source Active A-D route whose NLRI identifies that source and group. These Source Active A-D routes may be distributed to every PBR. If only a relatively small number of PBRs are actually interested in traffic from group G, but there are many sources for group G, this could result in a large number of (S,G) Source Active A-D routes being installed in a large number of PBRs that have no need of them.

For GTM, it is possible to constrain the distribution of (S,G) Source Active A-D routes to those PBRs that are interested in GTM traffic to group G. This can be done using the following OPTIONAL procedures:

- If a PBR originates a C-multicast Shared Tree Join whose NLRI contains (RD=0,*,G), then it dynamically creates an import RT for its global table, where the Global Administrator field of the RT contains the group address G, and the Local Administrator field contains zero. (Note that an IPv6-address-specific RT would need to be used if the group address is an IPv6 address.)

- When a PBR creates such an import RT, it uses "RT Constraint" [RFC4684] procedures to advertise its interest in routes that carry this RT.

- When a PBR originates a Source Active A-D route from its global table, it attaches the RT described above.

- When the C-multicast Shared Tree Join is withdrawn, so is the corresponding RT constrain route, and the corresponding RT is removed as an import RT of its global table.

These procedures enable a PBR to automatically filter all Source Active A-D routes that are about multicast groups in which the PBR has no interest.

This procedure does introduce the overhead of distributing additional "RT Constraint" routes, and therefore may not be cost-effective in all scenarios, especially if the number of sources per ASM group is small. This procedure may also result in increased join latency.

2.9. C-multicast Source/Shared Tree Joins

Section 11.1.3 of [RFC6514] describes how to determine the IP-address-specific RT(s) that should be attached to a C-multicast route. The "upstream PE", "upstream RD", and "source AS" (as defined in Section 5 of [RFC6513]) for the NLRI of the C-multicast route are first determined. An IP-address-specific RT whose "global administrator" field carries the IP address of the upstream PE is then attached to the C-multicast route. This procedure applies as well to GTM, except that the "upstream PE" is actually an "upstream PBR".

Section 11.1.3 of [RFC6514] also specifies that a second IP-address specific RT be attached to the C-multicast route, if the source AS of the NLRI of that route is different than the AS of the PE originating the route. The procedure for creating this RT may be summarized as:

(a) determine the upstream PE, upstream RD,and source AS corresponding to the NLRI of the route;

(b) find the corresponding Inter-AS or Intra-AS I-PMSI A-D route based on (a);

(c) find the next hop of that A-D route;

(d) place the IP address of that next hop in the global administrator field of the RT.

However, for GTM, in scenarios where it is known a priori by a PBR that the next hop of the C-multicast Source/Shared Tree Joins does not change during the distribution of those routes, the second RT (the one based on the next hop of an I-PMSI A-D route) is not needed,

and should not be present. In other scenarios, the procedure of section 11.1.3 of [RFC6513] (as modified by this document's sections on "RD usage" and "RT usage") is applied by the PBRs.

3. Differences from other MVPN-like GTM Procedures

The document [RFC7524] also defines a procedure for GTM that is based on the BGP procedures that were developed for MVPN.

However, the GTM procedures of [RFC7524] are different than and are NOT interoperable with the procedures defined in this document.

The two sets of procedures can co-exist in the same network, as long as they are not applied to the same multicast flows or to the same ASM multicast group addresses.

Some of the major differences between the two sets of procedures are the following:

- The [RFC7524] procedures for GTM do not use C-multicast Shared Tree Joins or C-multicast Source Tree Joins at all. The procedures of this document use these C-multicast routes for GTM, setting the RD field of the NLRI to zero.

- The [RFC7524] procedures for GTM use Leaf A-D routes instead of C-multicast Shared/Source Tree Join routes. Leaf A-D routes used in that manner can be distinguished from Leaf A-D routes used as specified in [RFC6514] by means of the NLRI format; [RFC7524] defines a new NLRI format for Leaf A-D routes. Whether a given Leaf A-D route is being used according to the [RFC7524] procedures or not can be determined from its NLRI. (See [RFC7524] section "Leaf A-D Route for Global Table Multicast".)

- The Leaf A-D routes used by the current document contain an NLRI that is in the format defined in [RFC6514], NOT in the format as defined in [RFC7524]. The procedures assumed by this document for originating and processing Leaf A-D routes are as specified in [RFC6514], NOT as specified in [RFC7524].

- The current document uses an RD value of zero in the NLRI in order to indicate that a particular route is "about" a Global Table Multicast, rather than a VPN multicast. No other semantics are inferred from the fact that RD is zero. [RFC7524] uses two different RD values in its GTM procedures, with semantic differences that depend upon the RD values.

- In order for both sets of procedures to co-exist in the same network, the PBRs MUST be provisioned so that for any given IP

group address in the global table, all egress PBRs use the same set of procedures for that group address (i.e., for group G, either all egress PBRs use the GTM procedures of this document or all egress PBRs use the GTM procedures of [RFC7524].

4. IANA Considerations

This document has no IANA considerations.

5. Security Considerations

The security considerations of this document are primarily the security considerations of the base protocols, as discussed in [RFC6514], [RFC4601], and [RFC5294].

The protocols and procedures described in this document make use of a type of route (routes with the "MCAST-VPN" BGP SAFI) that was originally designed for use in VPN contexts only. The protocols and procedures described in this document also make use of various BGP path attributes and extended communities (VRF Route Import Extended Community, Source AS Extended Community, Route Target Extended Community) that were originally intended for use in VPN contexts. If these routes, attributes, and/or extended communities leak out into "the wild", multicast data flows may be distributed in an unintended and/or unauthorized manner.

When VPNs are provisioned, each VRF is configured with import RTs and export RTs. These RTs constrain the distribution and the import of the VPN routes, making it difficult to cause a route to be distributed to and imported by a VRF that is not authorized to import that route. Additionally, VPN routes do not go into the "global table" with the "ordinary Internet routes" (i.e., non-VPN routes), and non-VPN routes do not impact the flow of multicast data within a VPN. In GTM, some of these protections against improper distribution/import of the routes is lost -- import of the routes is not restricted to VRFs, and the RT infrastructure that controls the distribution of routes among the VRFs is not present when routes are exported from and imported into global tables.

Internet Service Providers often make extensive use of BGP extended communities, sometimes adding, deleting, and/or modifying a route's extended communities as the route is distributed throughout the network. Care should be taken to avoid deleting or modifying the VRF Route Import Extended Community and Source AS Extended Community. Incorrect manipulation of these extended communities may result in multicast streams being lost or misrouted.

The procedures of this document require certain BGP routes to carry IP multicast group addresses. Generally such group addresses are only valid within a certain scope. If a BGP route containing a group address is distributed outside the boundaries where the group address is meaningful, unauthorized distribution of multicast data flows may occur.

6. Additional Contributors

Jason Schiller
Google
Suite 400
1818 Library Street
Reston, Virginia 20190
United States
Email: jschiller@google.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China
Email: lizhenbin@huawei.com

Wei Meng
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing
China
Email: meng.wei2@zte.com.cn,vally.meng@gmail.com

Cui Wang
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing
China
Email: wang.cui1@zte.com.cn

Shunwan Zhuang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China
Email: zhuangshunwan@huawei.com

7. Acknowledgments

The authors and contributors would like to thank Rahul Aggarwal, Huajin Jeng, Hui Ni, Yakov Rekhter, and Samir Saad for their ideas and comments.

8. References

8.1. 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, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <http://www.rfc-editor.org/info/rfc4364>.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <http://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, <http://www.rfc-editor.org/info/rfc6514>.

[RFC6515] Aggarwal, R. and E. Rosen, "IPv4 and IPv6 Infrastructure Addresses in BGP Updates for Multicast VPN", RFC 6515, DOI 10.17487/RFC6515, February 2012, <http://www.rfc-editor.org/info/rfc6515>.

8.2. Informative References

[ADD-PATH]
Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", internet-draft draft-ietf-idr-add-paths-10, October 2014.

[MVPN-extranet]
Rekhter, Y., Rosen, E., Aggarwal, R., Cai, Y., and T. Morin, "Extranet Multicast in BGP/IP MPLS VPNs", internet-draft draft-ietf-bess-mvpn-extranet-02, May 2015.

[RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, DOI 10.17487/RFC4601, August 2006, <http://www.rfc-editor.org/info/rfc4601>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <http://www.rfc-editor.org/info/rfc4684>.

[RFC5294] Savola, P. and J. Lingard, "Host Threats to Protocol Independent Multicast (PIM)", RFC 5294, DOI 10.17487/RFC5294, August 2008, <http://www.rfc-editor.org/info/rfc5294>.

[RFC6368] Marques, P., Raszuk, R., Patel, K., Kumaki, K., and T. Yamagata, "Internal BGP as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 6368, DOI 10.17487/RFC6368, September 2011, <http://www.rfc-editor.org/info/rfc6368>.

[RFC7524] Rekhter, Y., Rosen, E., Aggarwal, R., Morin, T., Grosclaude, I., Leymann, N., and S. Saad, "Inter-Area Point-to-Multipoint (P2MP) Segmented Label Switched Paths (LSPs)", RFC 7524, DOI 10.17487/RFC7524, May 2015, <http://www.rfc-editor.org/info/rfc7524>.

[RTC_without_RTs] Rosen, E., Ed., Patel, K., Haas, J., and R. Raszuk, "Route Target Constrained Distribution of Routes with no Route Targets", internet-draft draft-ietf-idr-rtc-no-rt-01, June 2015.

Authors' Addresses

Zhaohui Zhang
Juniper Networks, Inc.
10 Technology Park Drive
Westford, Massachusetts 01886
US

Email: zzhang@juniper.net

Lenny Giuliano
Juniper Networks, Inc.
2251 Corporate Park Drive
Herndon, Virginia 20171
US

Email: lenny@juniper.net

Eric C. Rosen (editor)
Juniper Networks, Inc.
10 Technology Park Drive
Westford, Massachusetts 01886
US

Email: erosen@juniper.net

Karthik Subramanian
Cisco Systems, Inc.
170 Tasman Drive
San Jose, CA 95134
US

Email: kartsubr@cisco.com

Dante J. Pacella
Verizon
22001 Loudoun County Parkway
Ashburn, Virginia 95134
US

Email: dante.j.pacella@verizonbusiness.com