Network Working Group
Internet Draft
Intended status: Standards Track
Expires: June 2013

Huajin Jeng
AT&T

James Uttaro
AT&T

Luay Jalil
Verizon

Bruno Decraene
France Telecom

Yakov Rekhter
Juniper Networks

Rahul Aggarwal
Arktan

December 12 2012

Virtual Hub-and-Spoke in BGP/MPLS VPNs

draft-ietf-l3vpn-virtual-hub-04.txt

Status of this Memo

Abstract

With BGP/MPLS VPNs any-to-any connectivity among sites of a given Virtual Private Network would require each Provider Edge router that has one or more of these sites connected to it to hold all the routes of that Virtual Private Network. The approach described in this document allows to reduce the number of Provider Edge routers that have to maintain all these routes by requiring only a subset of these routers to maintain all these routes.

Furthermore, when Provider Edge routers use ingress replication to carry multicast traffic of VPN customers, the approach described in this document may under certain circumstances allow to reduce bandwidth inefficiency associated with ingress replication, and to redistribute the replication load among Provider Edge routers.

Table of Contents

1. Specification of requirements

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. Overview

With BGP/MPLS VPNs [rfc4364] providing any-to-any connectivity among sites of a given Virtual Private Network (VPN) is usually accomplished by requiring each Provider Edge router (PE) that has one or more of these sites connected to it to hold all the routes of that VPN. The approach described in this document allows to reduce the number of PEs that have to maintain all these routes by requiring only a subset of such PEs to maintain all these routes.

Consider a set of PEs that maintain VRFs of a given VPN. In the context of this VPN we designate a subset of these PEs as "Virtual Spoke" PEs (or just Virtual Spokes), while some other (non-overlapping) subset of these PEs will be "Virtual Hub" PEs (or just Virtual Hubs), with the rest of the PEs in the set will be "vanilla" PEs (PEs that implement the [rfc4364] procedures, but do not implement procedures specified in this document).

For the sake of brevity we will use the term "V-hub" to denote a Virtual Hub, and "V-spoke" to denote a Virtual Spoke.

For a given VPN, its set of V-hubs may include not only the PEs that have sites of that VPN connected to them, but also PEs that have no sites of that VPN connected to them. On such PEs the VRF associated with that VPN may import routes from other VRFs of that VPN, even if the VRF has no sites of that VPN connected to it.

Note that while in the context of one VPN a given PE may act as a V-hub, in the context of another VPN the same PE may act as a V-spoke, and vice versa. Thus a given PE may act as a V-hub only for some, but not all the VPNs present on that PE. Likewise, a given PE may act as a V-spoke only for some, but not all the VPNs present on that PE.

For a given VPN each V-spoke of that VPN is "associated" with one or more V-hubs of that VPN (one may use two V-hubs for redundancy to avoid a single point of failure). Note that a given V-hub may have no V-spokes associated with it. For more on how a V-spoke and a V-hub become "associated" with each other see section "Routing Information Exchange".

Consider a set of V-spokes that are associated with a given V-hub, V-

hub-1. If one of these V-spokes is also associated with some other V-hub, V-hub-2, then other V-spokes in the set need not be associated with the same V-hub, V-hub-2, but may be associated with some other V-hubs (e.g., V-hub-3, V-hub-4, etc...)

This document defines a VPN-IP default route as a VPN-IP route whose VPN-IP prefix contains just an RD.

A PE that acts as a V-hub of a given VPN maintains all the routes of that VPN (such a PE imports routes from all other V-hubs and V-spokes, as well as from "vanilla" PEs of that VPN). A PE that acts as a V-spoke of a given VPN needs to maintain only the routes of that VPN that are originated by the sites of that VPN connected to that PE, plus one or more VPN-IP default route originated by the V-hub(s) associated with that V-spoke (such a PE need to import only VPN-IP default route from certain V-hubs). This way, only a subset of PEs that maintain VRFs of a given VPN, namely only the PEs acting as V-hubs of that VPN, have to maintain all the routes of that VPN. PEs acting as V-spokes of that VPN need to maintain only a (small) subset of the routes of that VPN.

This document assumes that a given V-hub and associated with it V-spoke(s) are in the same Autonomous Systems. However, if PEs that maintain VRFs of a given VPN span multiple Autonomous Systems, this document does not restrict all the V-hubs of that VPN to be in the same Autonomous System - the V-hubs may be spread among these Autonomous Systems.

One could model the approach defined in this document as a two-level hierarchy, where the top level consists of V-hubs and the bottom level consists of V-spokes. Generalization of this approach to more than two levels of hierarchy is outside the scope of this document.

When PEs use ingress replication to carry multicast traffic of VPN customers, the approach described in this document may under certain circumstances allow to reduce bandwidth inefficiency associated with ingress replication, and to redistribute the replication load among the PEs. This is because a PE that acts as a V-spoke of a given VPN would need to replicate multicast traffic only to other V-hubs (while other V-hubs would replicate this traffic to the V-spokes associated with these V-hubs), rather than to all the PEs of that VPN. Likewise, a PE that acts as a V-hub of a given VPN would need to replicate multicast traffic to other V-hubs, and the V-spokes, but only to the V-spokes associated with that V-hub, rather than replicating the traffic to all the PEs of that VPN. Limiting replication could be especially beneficial if the V-spoke PEs have limited replication capabilities and/or have links with limited bandwidth.

3. Routing Information Exchange

Routing information exchange among all the PEs of a given VPN is subject to the following rules.

A PE that has sites of a given VPN connected to it has to retain routing information received from these sites. This is irrespective of whether this PE acts as a V-hub or a V-spoke of that VPN, and follows the rules specified in [rfc4364].

A PE that has sites of a given VPN connected to it follows the rules specified in [rfc4364] when exporting (as VPN-IP routes) the routes received from these sites. This is irrespective of whether this PE acts as a V-hub or a V-spoke of that VPN.

In addition, a V-hub of a given VPN MUST export a VPN-IP default route for that VPN. This route MUST be exported to only the V-spokes of that VPN that are associated with that V-hub.

To enable a V-spoke of a given VPN to share its outbound traffic load among the V-hubs associated with that V-spoke, each V-hub of that VPN, when originating a VPN-IP default route MUST use a distinct RD (per V-hub, per VPN). Use of Type 1 RDs may be an attractive option for such RDs.

If a V-spoke imports several VPN-IP default routes, each originated by its own V-hub, and these routes have the same preference, then traffic from the V-spoke to other sites of that VPN would be load shared among the V-hubs.

Following the rules specified in [rfc4364], a V-hub of a given VPN imports all the non-default VPN-IP routes originated by all other PEs that have sites of that VPN connected to them (irrespective of whether these other PEs act as V-hubs or V-spokes or just "vanilla" PEs for that VPN, and irrespective of whether these V-spokes are associated with the V-hub or not).

A V-hub of a given VPN MUST NOT import a VPN-IP default route unless this is the Internet VPN-IP default route (for the definition of the "Internet VPN-IP default route", and how to distinguish between a VPN-IP default route and the Internet VPN-IP default route see section "Internet Connectivity").

Within a given VPN, a V-spoke MUST import all VPN-IP default routes that have been originated by the V-hubs associated with that V-spoke.

In addition, a V-spoke of a given VPN MAY import VPN-IP routes for that VPN that have been originated by some other V-spokes of that

VPN, but only by the V-spokes that are associated with the same V-hub(s) as the V-spoke itself.

The above rules are realized using Route Target (RT) extended communities [rfc4360] and VRF export/import policies based on these RTs. This document defines the following procedures for realizing the above rules.

Consider a "vanilla" any-to-any VPN. This document assumes that all the PEs of that VPN (or to be more precise, all the VRFs of that VPN) are provisioned with the same export and import RT - we will refer to this RT as RT-VPN (of course, for a given VPN service provider each VPN would use its own RT-VPN, distinct from RT-VPNs used by other VPNs).

To evolve this VPN into V-hubs and V-spokes all the PEs (or to be more precise all the VRFs) that are designated as either V-hubs or V-spokes of that VPN keep the same export RT-VPN. This RT-VPN is attached to all the VPN-IP routes originated by these PEs. Also, all the V-hubs keep the same import RT-VPN.

In addition, each V-hub of a given VPN is provisioned with its own export RT, called RT-VH. This RT-VH MUST be different from the export RT (RT-VPN) provisioned on that V-hub. Furthermore, for a given VPN service provider no two VPNs can use the same RT-VH.

A given V-spoke becomes associated with a given V-hub by virtue of provisioning the V-spoke to import only the VPN-IP route(s) that carry RT-VH provisioned on the V-hub (thus associating a new V-spoke with a given V-hub requires provisioning only on that V-spoke - no provisioning changes are required on the V-hub).

To avoid the situation where within a given VPN all the V-spokes would be associated with every V-hub (in other words, to partition V-spokes among V-hubs), different V-hubs within that VPN MAY use different RT-VHs. At one extreme every V-hub may use a distinct RT-VH. Use of IP-address specific RTs may be an attractive option for this scenario. However, it is also possible for several V-hubs to use the same RT-VH, in which case all these V-hubs would be associated with the same set of V-spokes.

When a V-hub originates a (non-Internet) VPN-IP default route, the V-hub MUST attach RT-VH to that route (the case where a V-hub originates the Internet VPN-IP default route is covered in section "Internet Connectivity"). Thus this route is imported by all the V-spokes associated with the V-hub.

A V-spoke MAY be provisioned to export VPN-IP routes not just to the

V-hubs, but also to the V-spokes that import the same VPN-IP default route(s) as the V-spoke itself. The V-spoke accomplishes this by adding its import RT-VH(s) to the VPN-IP routes exported by the V-spoke.

4. Forwarding Considerations

This section describes changes/modifications to the forwarding procedures specified in [rfc4364].

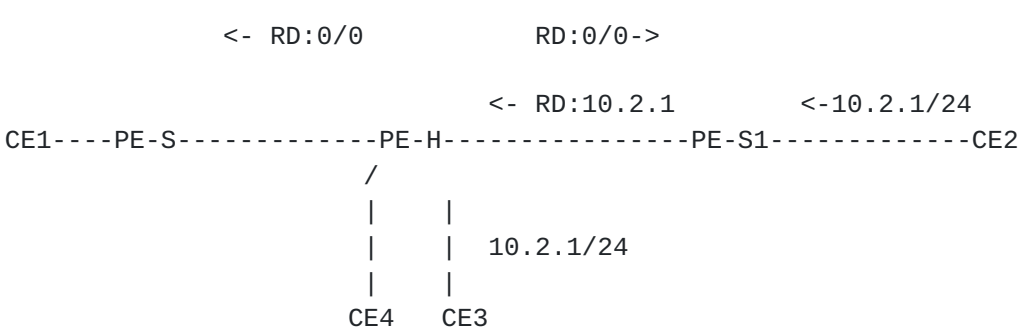
For a given VPN, the MPLS label that a V-hub of that VPN advertises with a VPN-IP default route MUST be the label that is mapped to an NHLFE that identifies the VRF of the V-hub. As a result, when the V-hub receives a packet that carries such label, the V-hub pops the label and determines further disposition of the packet based on the lookup in the VRF.

Note that this document does not require to advertise labels mapped to an NHLFE that identifies a VRF for routes other than the VPN-IP default route.

When a V-hub of a given VPN originates a VPN-IP default route for that VPN, the V-hub MUST NOT install in its VRF of that VPN a default route, unless this route has been originated either (a) as a result of the V-hub receiving an IP default route from one of the CEs of that VPN connected to it, or (b) as a result of the V-hub receiving (and importing) the Internet VPN-IP default route from some other PE (for the definition of the "Internet VPN-IP default route" see section "Internet Connectivity"), or (c) the VRF being provisioned with a default route pointing to the routing table that maintains the Internet routes.

When a multi-homed site is connected to a V-hub and a V-spoke, then the V-hub uses the following OPTIONAL procedures to support IBGP/EBGP load balancing for the site's inbound traffic that has been originated by some other V-spoke associated with the V-hub. When the V-hub receives from some other PE a packet that carries an MPLS label that the V-hub advertised in the VPN-IP default route, then the V-hub uses the label to identify the VRF that should be used for further disposition of the packet. If (using the information present in the VRF) the V-hub determines that the packet has to be forwarded using a non-default route present in the VRF, and this route indicates that the packet's destination is reachable either over one of the VRF attachment circuit (for the definition of "VRF attachment circuits" see [rfc4364]) or via some other (V-spoke) PE, the V-hub forwards the packet either over this attachment circuit, or via that other PE. The choice between the two is a local matter to the V-hub.

To illustrate this consider the following example:



A multi-homed site (not shown in the above figure) is connect via CE2 and CE3. Thus both CE2 and and CE3 advertise a route to 10.2.1/24. CE2 advertises this route (route to 10.2.1/24) to PE-S1, which in turn originates a VPN-IP route RD:10.2.1. CE3 advertises this route to PE-H.

PE-H is a V-hub, while PE-S and PE-S1 are V-spokes associated with that V-hub. Thus PE-H originates a VPN-IP default route (RD:0/0), and both PE-S and PE-S1 import that route.

PE-H receives from PE-S1 a VPN-IP route to RD:10.2.1 and from CE3 a plain IP route to 10.2.1. Thus the VRF entry on PE-H has two possible next hops for 10.2.1: CE3 and PE-S1 (the latter is a next hop that is not directly connected to PE-H).

Now consider what happens when CE1 originates a packet destined to 10.2.1.1. When PE-S receives this packet, PE-S (following the VPN-IP default route) forwards the packet to PE-H. The MPLS label in the packet is the label that PE-H advertised to PE-S in the VPN-IP default route. Thus, following the rule specified above, PE-H may forward the packet either via CE3 or via PE-S1 (with PE-S1 subsequently forwarding the packet to CE2), resulting in IBGP/EBGP load balancing.

Likewise, if CE4 originates a packet destined to 10.2.1.1, PE-H may forward the packet either via CE3 or via PE-S1 (with PE-S1 subsequently forwarding the traffic to CE2), resulting in IBGP/EBGP load balancing.

Note however, that if there is some other CE, CE5, connected to PE-S1, and that CE5 sends a packet to 10.2.1.1, then (due to the IP longest match rule) PE-S1 will always forward this packet to CE2. Thus for a mult-homed site connected to a V-hub and a V-spoke IBGP/EBGP load balancing will be available for some, but not all the traffic destined to that site. Specifically, IBGP/EBGP load balancing will not be available for the traffic destined to that site if this

traffic has been originated within some other site that is connected to the same V-spoke.

Moreover, if CE3 advertises 10.2.1.0/25 and 10.2.1/24, while CE2 advertises 10.2.1.128/25 and 10.2.1/24 (which is yet another form of load balancing for a multi-homed site), then when CE5 sends a packet to 10.2.1.1, then (due to the IP longest match rule) PE-S1 will always forward this packet to CE2, even though the VPN customer would expect this traffic to flow via CE3.

This document proposes two options to address the issues raised in the previous two paragraphs. The first option is to disallow for a given VPN to provision PEs that have multi-homed sites of that VPN connected to them as V-spokes (such PEs could be provisioned as either V-hubs, or as plain "vanilla" PEs). The second option is for the V-spoke, when it receives an IP route from a CE, not to install this route in its forwarding table, but just re-advertise this route as a VPN-IP route, together with an MPLS label. The NHLFE [rfc3031] associated with that label MUST specify the CE that advertises the IP route as the next hop. As a result, when the PE receives data that carries that label, the PE performs no IP lookup on the data, and just forwards the data to the CE. Note that doing this would result in forcing the traffic between a pair of sites connected to the same V-spoke to go through the V-hub of that V-spoke.

An implementation that supports IBGP/EBGP load balancing, as specified above, SHOULD support the second option. If the implementation does not support the second option, then deploying this implementation to support IBGP/EBGP load balancing, as specified above, would either (a) restrict the set of PEs that could be provisioned as V-spokes (any PE that has a multi-homed site connected to it can not be provisioned as a V-spoke), or (b) IBGP/EBGP load balancing may not be available for certain scenarios (the scenarios that the second option is intended to cover).

5. Internet Connectivity

This document specifies two possible alternatives of providing Internet connectivity for a given VPN.

The first alternative is when a PE that maintains Internet routes also maintains a VRF of a given VPN. In this case the Internet connectivity for that VPN MAY be provided by provisioning in the VPN's VRF on that PE a default route pointing to the routing table on that PE that maintains the Internet routes. This PE MUST NOT be provisioned as a V-spoke for that VPN (this PE may be provisioned as either a V-hub, or as a "vanilla" PE). If this PE is provisioned as a

V-hub, then this PE MUST originate a VPN-IP default route. The route MUST carry both RT-VPN and RT-VH of the V-hub (see section "Routing Information Exchange" for the definition of "RT-VPN" and "RT-VH"). Thus this route will be imported by all the V-spokes associated with the V-hub, as well as by other V-hubs. An implementation MUST support the first alternative.

The second alternative is when a site of a given VPN has connection to the Internet, and a CE of that site advertises an IP default route to the PE connected to that CE. This alternative has two sub-cases: (a) PE is provisioned as a V-hub, and (b) PE is provisioned as a V-spoke. An implementation MUST support the sub-case (a). An implementation MAY support the sub-case (b).

If a PE is provisioned as a V-hub, then the PE re-advertises this IP default route as a VPN-IP default route, and install in its VRF an IP default route with the next hop specifying the CE(s) that advertise the IP default route to the PE. Note, that when re-advertising the VPN-IP default route, the route MUST carry both RT-VPN and RT-VH of the V-hub (see section "Routing Information Exchange" for the definition of "RT-VPN" and "RT-VH"). Thus this route will be imported by all the V-spokes associated with the V-hub, as well as by other V-hubs.

If a PE is provisioned as a V-spoke, then receiving a default route from a CE MUST NOT cause the V-spoke to install an IP default route in its VRF. The V-spoke MUST originate a VPN-IP default route with a (non-null) MPLS label. The route MUST carry only RT-VPN (as a result, this route is not imported by any of the V-spokes, but is imported by V-hubs). The packet's next hop of the Next Hop Label Forwarding Entry (NHLFE) [rfc3031] associated with that label MUST specify the CE that advertises the IP default route. As a result, when the V-spoke receives data that carries that label, the V-spoke performs no IP lookup on the data, and just forwards the data to the CE. Note that in this case the VRF on the V-spoke is going to have an IP default route, but this route would be created as a result of receiving a VPN-IP default route from one of the V-hubs associated with that V-spoke (and not as a result of receiving the IP default route from the CE). Note also that if this V-spoke has other sites of that VPN connected to it, then traffic from these sites to the Internet would go to that V-spoke, then to the V-hub selected by the V-spoke, then from that V-hub back to the V-spoke, and then to the CE that advertises IP default route to the V-spoke.

If a PE is provisioned as a V-spoke of a given VPN, and if a CE of that VPN advertises an IP default route to the PE (as the CE belongs to the site that provides the Internet connectivity for the VPN), then the PE MUST NOT advertise an IP default route back to that CE.

Yet, the CE has to specify that PE as the next hop for all the traffic to other sites of that VPN. A way to accomplish this is to require the V-spoke to implement procedures of section "Further Refinements".

In all of the above scenarios described in this section we refer to the originated VPN-IP default route as the "Internet VPN-IP default route". Specifically, the Internet VPN-IP default route is a VPN-IP default route originated by a PE (this PE could be either a V-hub or a V-spoke) as a result of either (a) receiving an IP default route from a CE, or (b) of PE maintaining Internet routes, and by provisioning in the VPN's VRF on that PE a default route pointing to the routing table on that PE that maintains the Internet routes.

The difference between the Internet VPN-IP default route and a non-Internet VPN-IP default route originated by a V-hub is in the RTs carried by the route - for a given VPN and a given V-hub of that VPN the Internet VPN-IP default route carries both RT-VPN and RT-VH of that V-hub, the non-Internet VPN-IP default route carries just RT-VH of that V-hub.

When a V-hub originates the Internet VPN-IP default route, the V-hub MUST withdraw the non-Internet VPN-IP default route that has been originated by the V-hub. When a V-hub withdraws the Internet VPN-IP default route that has been originated by the V-hub, the V-hub MUST originate a non-Internet VPN-IP default route. That is, at any given point in time a given V-hub originates either the Internet VPN-IP default route, or a non-Internet VPN-IP default route.

As a result of the rules specified above, if a V-hub originates the Internet VPN-IP default route, then all the V-spokes associated with that V-hub MUST import that route. In addition (and in contrast with a non-Internet VPN-IP default route), other V-hubs MAY import that route. A V-hub MAY also import the Internet VPN-IP default routes originated by V-spoke(s). A V-spoke MUST NOT import the Internet VPN-IP default route originated by any other V-spoke. Such a route MAY be imported only by V-hubs.

If the Internet VPN-IP default route originated by a V-hub has the same preference as the (non Internet) VPN-IP default route originated by some other V-hub, then a V-spoke that imports VPN-IP default routes originated by both of these V-hubs would load share the outgoing Internet traffic between these two V-hubs (and thus some of the outgoing Internet traffic from that V-spoke will first be routed to the V-hub that does not originate the Internet VPN-IP default route, and then from that V-hub to the V-hub that does originate the Internet VPN-IP default route).

If taking an extra-hub hop for the Internet traffic is viewed as undesirable, then it is RECOMMENDED that the Internet VPN-IP default route be of higher preference than a (non-Internet) VPN-IP default route originated by some other V-hub. However, in this case the traffic from the V-spokes to other sites of that VPN will not be load shared between these two V-hubs.

6. Deployment Considerations

For a given VPN a V-hub and a set of V-spokes associated with that V-hub should be chosen in a way that minimizes the additional network distance/latency penalty, given that VPN geographic footprint.

For a given VPN some/all of its V-spokes could be grouped into geographically-based clusters (V-spokes within a given cluster be in close geographical proximity to each other) with any-any connectivity within each cluster. Note that the V-spokes within a given cluster need not be associated with the same V-hub(s). Likewise, not all V-spokes associated with a given V-hub need to be in the same cluster. A use case for this would be VPN for large retail chain in which data traffic is hub/spoke between each store and centralized data-centers but there is need for direct VoIP traffic between stores within same geographical area.

Use of RT Constrains [rfc4684] may further facilitate/optimize routing exchange in support of V-hubs and V-spokes.

Introducing V-spoke PE in a VPN may introduce the following change for the customer of that VPN:

+ traceroute from a CE connected to a V-spoke may report an additional hop: the V-hub PE.

+ latency for traffic sent from a CE connected to a V-spoke may increase, depending on the location of the V-hub in the layer 3 and layer 1 network topology of the SP.

7. Multicast Considerations

This section describes procedures for supporting MVPN in the presence of Virtual Hub-and-Spoke. The procedures rely on MVPN specifications as defined in [rfc6513], [rfc6514], [rfc6625].

The procedures assume that for the purpose of ensuring non-duplication both V-hubs and V-spokes can discard packets from a "wrong" PE, as specified in section 9.1.1 of [rfc6513]. The existing procedures for S-PMSI A-D routes ([rfc6513], [rfc6514], [rfc6625]) are sufficient to discard packets coming from a "wrong" PE for all types of P-tunnels specified in [rfc6514] (including Ingress Replication). The existing procedures for I-PMSI A-D routesa ([rfc6513], rfc6514]) are sufficient to discard packets coming from a "wrong" PE for all types of P-tunnels specified in [rfc6514], except for Ingress Replication. When Ingress Replication is used for I-PMSI P-tunnels section "Use of Ingress Replication with I-PMSI A-D route" specifies changes to the procedures in [rfc6514] to enable discard packets from a "wrong" PE.

The V-hub/V-spoke architecture, as specified in this document, affects certain multicast scenarios. In particular, it affects multicast scenarios where the source of a multicast flow is at a site attached to a V-hub, and a receiver of that flow is at a site attached to a V-spoke that is not associated with that same V-hub. It also affects multicast scenarios where the source of a multicast flow is at a site attached to a V-spoke, a receiver of that flow is at a site attached to a different V-spoke, and the set intersection between the V-hub(s) associated with the first V-spoke and the V-hub(s) associated with the second V-spoke is empty. It may also affect multicast scenarios where the source of a multicast flow is at a site connected to a V-spoke, a receiver of that flow is at a site attached to a different V-spoke, and the set intersection between the V-hub(s) associated with the first V-spoke and the V-hub(s) associated with the second V-spoke is non-empty (the multicast scenarios are affected if the I-PMSI/S-PMSI A-D routes originated by the first V-spoke are not imported by the second V-spoke).

Use of Virtual Hub-and-Spoke in conjunction with seamless MPLS multicast [seamless-MPLS-multicast] is outside the scope of this document.

7.1. Terminology

We will speak of a P-tunnel being "bound" to a particular I-PMSI/S-PMSI A-D route if the P-tunnel is specified in that route's PMSI Tunnel Attribute.

When Ingress Replication is used, the P-tunnel bound to a particular I-PMSI/S-PMSI A-D route is actually a set of unicast tunnels. The PE originating the route uses these unicast tunnels to carry traffic to the PEs that import the route. When we say that traffic has been received by a PE on a P-tunnel "bound" to particular I-PMSI/S-PMSI A-D route received by that PE, we refer to the unicast tunnel that the PE originating the route uses to send traffic to the PE that receives the route.

7.2. Eligible Upstream Multicast Hop (UMH) routes

On a V-spoke the set of Eligible UMH routes consists of all the unicast VPN-IP routes received by the V-spoke, including the default VPN-IP routes received from its V-hub(s). Note that such routes MAY include routes received from other V-spokes. The routes received from other V-spokes could be either "vanilla" VPN-IP routes (routes using the IPv4 or IPv6 AFI, and SAFI set to 128 "MPLS-labeled VPN address" [IANA-SAFI]), or routes using the IPv4 or IPv6 AFI (as appropriate), but with the SAFI set to SAFI 129 "Multicast for BGP/MPLS IP Virtual Private Networks (VPNs)" [IANA-SAFI].

The default VPN-IP routes received from the V-hub(s) may be either Internet default VPN-IP routes, or non-Internet default VPN-IP routes.

7.3. Originating VPN-IP default route by a V-hub

When originating a VPN-IP default route, a V-hub, in addition to the procedures specified in section "Routing Information Exchange", also follows the procedures of sections 6 and 7 of [rfc6514] (see also section 5.1 of [rfc6513]). Specifically the V-hub MUST add the VRF Route Import extended community that embeds the V-hub's IP address. The route also MUST include the Source AS extended community.

7.4. Handling C-multicast routes

Origination of C-multicast routes follow the procedures of [rfc6514] (this is irrespective of whether these routes are originated by a V-hub or by a V-spoke).

When a V-spoke receives a C-multicast route, the V-spoke follows the [rfc6514] procedures.

When a V-hub receives a C-multicast route, the V-hub determines whether the C-RP/C-S of the route is reachable via one of its VRF interfaces, and if yes, then the V-hub follows the [rfc6514] procedures.

Otherwise, the C-RP/C-S of the route is reachable via some other PE (this is the case where the received route was originated by a V-spoke that sees the V-hub as the "upstream PE" for a given source, but the V-hub sees some other PE (either V-hub or V-spoke) as the "upstream PE" for that source). In this case the V-hub uses the type (Source Tree Join vs Shared Tree Join), the Multicast Source, and Multicast Group from the received route to construct a new route of the same type, with the same Multicast Source and Multicast Group. The hub constructs the rest of the new route following procedures specified in 11.1.3 of [rfc6514]. The hub also creates the appropriate (C-*, C-G)/(C-S, C-G) state in its MVPN-TIB.

7.5. Originating I-PMSI/S-PMSI/SA A-D routes by V-spoke

When a V-spoke originates an I-PMSI/S-PMSI/SA A-D route, the V-spoke follows the procedures of [rfc6514] (or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]), including the procedures for constructing RT(s) carried by the route. Note that as a result, such a route will be imported by the V-hubs. In the case of an I-PMSI/S-PMSI A-D route, the P-tunnel bound to this route is used to carry to these V-hubs traffic originated by the sites connected to the V-spoke.

If the V-spoke exports its (unicast) VPN-IP routes not just to the V-hubs, but also to some other V-spokes (as described in section "Routing Information Exchange"), then (as a result of following the procedures of [rfc6514], or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]) the I-PMSI/S-PMSI/SA A-D route originated by the V-spoke will be imported not just by the V-hubs, but also by these other V-spokes. This is because in this scenario the route will carry more than one RT; one of these RTs, RT-VPN, will result in importing the route by the V-hubs, while other RT(s) will result in importing the route by the V-spokes (the other RT(s) are

the RT(s) that the V-spoke uses for importing the VPN-IP default route). In this case the P-tunnel bound to this I-PMSI/S-PMSI A-D route is also used to carry traffic originated by the sites connected to the V-spoke that originates the route to these other V-spokes.

7.6. Originating I-PMSI/S-PMSI/SA A-D routes by V-hub

When a V-hub originates an I-PMSI/S-PMSI/SA A-D route, the V-hub follows the procedures of [rfc6514] (or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]), except that in addition to the RT(s) constructed following these procedures, the route MUST also carry the RT of the VPN-IP default route advertised by the V-hub (RT-VH). Note that as a result, such a route will be imported by other V-hubs, and also by the V-spokes, but only by the V-spokes that are associated with the V-hub (the V-spokes that import the VPN-IP default route originated by the V-hub). In the case of an I-PMSI/S-PMSI A-D route, the P-tunnel bound to this route is used to carry to these other V-hubs and V-spokes the traffic originated by the sites connected to the V-hub that originates the route.

In addition, if a V-hub originates an I-PMSI A-D route following the procedures of [rfc6514], the V-hub MUST originate another I-PMSI A-D route - we'll refer to this route as a "Associated-V-spoke-only I-PMSI A-D route". The RT carried by this route MUST be the RT that is carried in the VPN-IP default route advertised by the V-hub (RT-VH). Therefore, this route will be imported only by the V-spokes associated with the V-hub (the V-spokes that import the VPN-IP default route advertised by this V-hub). The P-tunnel bound to this route is used to carry to these V-spokes traffic originated by the sites connected to either (a) other V-hubs, or (b) other V-spokes, including the V-spokes that import the VPN-IP default route from the V-hub, or (c) "vanilla" PEs. More details on the use of this P-tunnel are described in section "Receiving I-PMSI/S-PMSI/SA A-D routes by V-hub".

As a result, a V-hub originates not one, but two I-PMSI A-D routes - one is a "vanilla" I-PMSI A-D route, and another is an Associated-V-spoke-only I-PMSI A-D route. Each of these routes MUST have a distinct RD.

When a V-hub receives traffic from one of the sites connected to the V-hub, and the V-hub determines (using some local policies) that this traffic should be transmitted using an I-PMSI, the V-hub forwards this traffic on the P-tunnel bound to the "vanilla" I-PMSI A-D route, but MUST NOT forward it on the P-tunnel bound to the Associated-V-spoke-only I-PMSI A-D route.

7.7. Receiving I-PMSI/S-PMSI/SA A-D routes by V-spoke

When a V-spoke receives an I-PMSI/S-PMSI/SA A-D route, the V-spoke follows the procedures of [rfc6514] (or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]). As a result, a V-spoke that is associated with a given V-hub (the V-spoke that imports the VPN-IP default route originated by that V-hub) will also import I-PMSI/S-PMSI/SA A-D routes originated by that V-hub. Specifically, the V-spoke will import both the "vanilla" I-PMSI A-D route and the Associated-V-spoke-only I-PMSI A-D route originated by the V-hub.

In addition, if a V-spoke imports the (unicast) VPN-IP routes originated by some other V-spokes (as described in section "Routing Information Exchange"), then the V-spoke will also import I-PMSI/S-PMSI/SA A-D routes originated by these other V-spokes.

7.8. Receiving I-PMSI/S-PMSI/SA A-D routes by V-hub

The following describe procedures that a V-hub MUST follow when it receives an I-PMSI/S-PMSI/SA A-D route.

7.8.1. Case 1

This is the case where a V-hub receives an I-PMSI/S-PMSI/SA A-D route, and one of the RT(s) carried in the route is the RT that the V-hub uses for advertising its VPN-IP default route (RT-VH).

In this case the receiving route was originated either

+ by a V-spoke associated with the V-hub (the V-spoke that imports the VPN-IP default route originated by the V-hub), or

+ by some other V-hub that uses the same RT as the receiving V-hub for advertising the VPN-IP default route.

In this case the received I-PMSI/S-PMSI/SA A-D route carries more than one RT. One of these RTs results in importing this route by the V-hubs. Another of these RTs is the RT that the V-hub uses when advertising its VPN-IP default route (RT-VH). This RT results in importing the received I-PMSI/S-PMSI/SA A-D route by all the V-spokes associated with the V-hub (the V-spokes that import the VPN-IP default route originated by the V-hub).

In handling such I-PMSI/S-PMSI/SA A-D route the V-hub follows the procedures of [rfc6514] (or in the case of a wildcard S-PMSI A-D

route the procedures of [rfc6625]). Specifically, in contrast to Case 2 below, the V-hub MUST NOT re-advertise this route.

The following specifies the rules that the V-hub MUST follow when handling traffic that the V-hub receives on a P-tunnel bounded to this I-PMSI/S-PMSI A-D route. The V-hub may forward this traffic to only the sites connected to that V-hub (forwarding this traffic to these sites follows the procedures specified in [rfc6514], or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]). The V-hub MUST NOT forward the traffic received on this P-tunnel to any other V-hubs or V-spokes, including the V-spokes that import the VPN-IP default route originated by the V-hub (V-spokes associated with the V-hub). Specifically, the V-hub MUST NOT forward the traffic received on the P-tunnel advertised in the received I-PMSI A-D route over the P-tunnel that the V-hub binds to its Associated-V-spoke-only I-PMSI A-D route.

7.8.2. Case 2

This is the case where a V-hub receives an I-PMSI/S-PMSI/SA A-D route, and the route does not carry the RT that the receiving V-hub uses when advertising its VPN-IP default route (RT-VH).

In this case the receiving I-PMSI/S-PMSI/SA A-D route was originated by either some other V-hub, or by a V-spoke. The I-PMSI/S-PMSI/SA A-D route is imported by the V-hub (as well as by other V-hubs), but not by any of the V-spokes associated with the V-hub (V-spokes that import the VPN-IP default route originated by the V-hub).

In this case the V-hubs follows the procedures of [rfc6514] (or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]) with the following additions.

Once a V-hub accepts an I-PMSI A-D route, when the V-hub receives data on the P-tunnel bound to that I-PMSI A-D route, the V-hub follows procedures in [rfc6513], [rfc6514] to determine whether to accept the data. If the data is accepted, then the V-hub further forwards the data over the P-tunnel bound to the Associated-V-spoke-only I-PMSI A-D route originated by the V-hub. Note that in deciding whether to forward the data over the P-tunnel bound to the Associated-V-spoke-only I-PMSI A-D route originated by the V-hub, the V-hub SHOULD take into account the (multicast) state present in its MVPN-TIB that has been created as a result receiving C-multicast routes from the V-spokes associated with the V-hub. If (using the information present in the MVPN-TIB) the V-hub determines that none of these V-spokes have receivers for the data, the V-hub SHOULD NOT forward the data over the P-tunnel bound to the Associated-V-spoke-

only I-PMSI A-D route originated by the V-hub.

Whenever a V-hub accepts an S-PMSI/SA A-D route, the V-hub, in contrast to Case 1 above, MUST re-advertise this route to its V-spokes. To accomplish this, the V-hub replaces the RT(s) carried in the route with the RT that the V-hub uses when originating its VPN-IP default route (RT-VH), changes the RD of the route to the RD that the V-hub uses when originating its Associated-V-spoke-only I-PMSI A-D route, and sets Next Hop to self. For S-PMSI A-D routes the V-hub also changes the Originating Router's IP address in the MCAST-VPN NLRI of the route to its own IP address (the same address as the one in the Next Hop). Moreover, before re-advertising S-PMSI A-D routes to V-spokes, the V-hub modifies the PMSI Tunnel attribute of these routes as appropriate (e.g., by replacing the P-tunnel rooted at the originator of these routes with a P-tunnel rooted at the V-hub).

Note that a V-hub of a given MVPN may receive and accept multiple (C-*, C-*) wildcard S-PMSI A-D routes [rfc6625], each originated by its own PE. Yet, even if the V-hub receives and accepts such multiple (C-*, C-*) S-PMSI A-D routes, the V-hub re-advertises just one (C-*, C-*) S-PMSI A-D route, thus aggregating the received (C-*, C-*) S-PMSI A-D routes. The same applies for (C-*, C-G) S-PMSI A-D routes.

Whenever a V-hub receives data on the P-tunnel bound to a received S-PMSI A-D route, the V-hub follows procedures of [rfc6513], [rfc6514] (or in the case of a wildcard S-PMSI A-D route the procedures of [rfc6625]) to determine whether to accept the data. If the data is accepted, then the V-hub further forwards it over the P-tunnel bound to the S-PMSI A-D route that has been re-advertised by the V-hub.

If multiple S-PMSIs received by a V-hub have been aggregated into the same P-tunnel, then the V-hub, prior to forwarding to the V-spokes associated with that V-hub the data received on this P-tunnel MAY de-aggregate and then re-aggregate (in a different way) this data using the state present in its MVPN-TIB that has been created as a result of receiving C-multicast routes from the V-spokes. Even if S-PMSIs received by the V-hub each have their own P-tunnel, the V-hub, prior to forwarding to the V-spokes the data received on these P-tunnels MAY aggregate these S-PMSIs using the state present in its MVPN-TIB that has been created as a result of receiving C-multicast routes from the V-spokes.

7.9. Use of Ingress Replication with I-PMSI A-D routes

The following modifications to the procedures specified in [rfc6514] for originating/receiving I-PMSI A-D routes enable to discard packets coming from a "wrong" PE when Ingress Replication is used for I-PMSI P-tunnels (for other types of P-tunnels the procedures specified in [rfc6513], [rfc6514] are sufficient).

The modifications to the procedures are required to be implemented (by all the PEs of a given MVPN) only under the following conditions:

+ At least one of those PEs is a V-hub or V-spoke PE for the given MVPN.

+ The given MVPN is configured to use the optional procedure of using Ingress Replication to instantiate an I-PMSI.

If Ingress Replication is used with I-PMSI A-D routes, then when a PE advertises such routes, the Tunnel Type in the PMSI Tunnel attribute MUST be set to Ingress Replication; the Leaf Information Required flag MUST be set to 1; the the attribute MUST carry no MPLS labels.

A PE that receives such I-PMSI A-D route MUST respond with a Leaf A-D route. The PMSI Tunnel attribute of that Leaf A-D route is constructed as follows. The Tunnel Type is set to Ingress Replication. The Tunnel Identifier MUST carry a routable address of the PE that originates the Leaf A-D route. The PMSI Tunnel attribute MUST carry a downstream assigned MPLS label that is used to demultiplex the traffic received over a unicast tunnel by the PE.

This document assumes that for a given MVPN all the PEs that have sites of that MVPN connected to them implement the procedures specified in this section.

8. An example of RT provisioning

Consider a VPN A that consists of 9 sites - site-1 through site-9. Each site is connected to its own PE - PE-1 through PE-9.

We designate PE-3, PE-6, and PE-9 as V-hubs.

To simplify the presentation the following example assumes that each V-spoke is associated with just one V-hub. However, as mentioned earlier, in practice each V-spoke should be associated with two or more V-hubs.

PE-1 and PE-2 are V-spokes associated with PE-3. PE-4 and PE-5 are V-spokes associated with PE-6. PE-7 and PE-8 are V-spokes associated with PE-9.

8.1. Unicast routing

All the PEs (both V-hubs and V-spokes) are provisioned to export routes using RT-A (just as it would be with "vanilla" any-to-any VPN).

All the V-hubs (PE-3, PE-6, and PE-9) are provisioned to import routes with RT-A (just as it would be with "vanilla" any-to-any VPN).

In addition, PE-3 is provisioned to originate a VPN-IP default route with RT-A-VH-1 (but not with RT-A), while PE-1 and PE-2 are provisioned to import routes with RT-A-VH-1.

Likewise, PE-6 is provisioned to originate a VPN-IP default route with RT-A-VH-2 (but not with RT-A), while PE-4 and PE-5 are provisioned to import routes with RT-A-VH-2.

Finally, PE-9 is provisioned to originate a VPN-IP default route with RT-A-VH-3 (but not with RT-A), while PE-7 and PE-8 are provisioned to import routes with RT-A-VH-3.

Now let's modify the above a bit by assuming that site-3 has Internet connectivity. Thus site-3 advertises an IP default route to PE-3. PE-3, in turn originates a VPN-IP default route. In this case, the VPN-IP default route carries RT-A and RT-A-VH-1 (rather than just RT-A-VH-1, as before), which results in importing this route to PE-6 and PE-9, as well as to PE-1 and PE-2.

If PE-7 and PE-8, in addition to importing VPN-IP default route from PE-9, also want to import each other VPN-IP routes, then PE-7 and PE-8 export their VPN-IP routes with two RTs: RT-A and RT-A-VH-3.

8.2. Multicast routing

All the PEs designated as V-spokes (PE-1, PE-2, PE-4, PE-5, PE-7, and PE-8) are provisioned to export their I-PMSI/S-PMSI/SA A-D routes using RT-A (just as it would be with "vanilla" any-to-any MVPN). Thus these routes could be imported by all the V-hubs (PE-3, PE-6, and PE-9).

The V-hub on PE-3 is provisioned to export its I-PMSI/S-PMSI/SA A-D routes with two RTs: RT-A and RT-A-VH-1. Thus these routes could be

imported by all the other V-hubs (PE-6 and PE-9), and also by the V-spokes, but only by the V-spokes associated with the V-hub on PE-3 (PE-1 and PE-2). In addition, the V-hub on PE-3 originates the Associated-V-spoke-only I-PMSI A-D route with RT-A-VH-1. This route could be imported only by the V-spokes associated with the V-hub on PE-3 (PE-1 and PE-2).

The V-hub on PE-6 is provisioned to export its I-PMSI/S-PMSI/SA A-D routes with two RTs: RT-A and RT-A-VH-2. Thus these routes could be imported by all the other V-hubs (PE-3 and PE-9), and also by the V-spokes, but only by the V-spokes associated with the V-hub on PE-6 (PE-4 and PE-5). In addition, the V-hub on PE-6 originates the Associated-V-spoke-only I-PMSI A-D route with RT-A-VH-2. This route could be imported only by the V-spokes associated with the V-hub on PE-6 (PE-4 and PE-5).

The V-hub on PE-9 is provisioned to export its I-PMSI/S-PMSI/SA A-D routes with two RTs: RT-A and RT-A-VH-3. Thus these routes could be imported by all the other V-hubs (PE-3 and PE-6), and also by the V-spokes, but only by the V-spokes associated with the V-hub on PE-9 (PE-7 and PE-8). In addition, the V-hub on PE-9 originates the Associated-V-spoke-only I-PMSI A-D route with RT-A-VH-3. This route could be imported only by the V-spokes associated with the V-hub on PE-9 (PE-7 and PE-8).

If PE-7 and PE-8, in addition to importing VPN-IP default route from PE-9, also want to import each other VPN-IP routes, then PE-7 and PE-8 export their I-PMSI/S-PMSI/SA A-D routes with two RTs: RT-A and RT-A-VH-3.

If the V-hub on PE-9 receives an S-PMSI/SA A-D route originated by either some other V-hub (PE-3 or PE-6), or by a V-spoke that is not associated with this V-hub (PE-1, or PE-2, or PE-4, or PE-5), the V-hub, before re-advertising this route replaces the RT(s) carried in this route with just one RT - RT-A-VH-3. Thus, the re-advertised route could be imported only by the V-spokes associated with the V-hub on PE-9 (PE-7 and PE-8).

9. Further Refinements

In some cases a VPN customer may not want to rely solely on an (IP) default route being advertised from a V-spoke to a CE, but may want CEs to receive all the VPN routes (e.g., for the purpose of faster detection of VPN connectivity failures, and activating some backup connectivity).

In this case an OPTIONAL approach would be to install in the V-

spoke's data plane only the VPN-IP default route advertised by the V-hub associated with the V-spoke, even if the V-spoke receives an IP default route from the CE, and keep all the VPN-IP routes in the V-spoke's control plane (thus being able to advertise these routes as IP routes from the V-spoke to the CEs). Granted, this would not change control plane resource consumption, but would reduce forwarding state on the data plane.

10. IANA Considerations

This document introduces no new IANA Considerations.

11. Security Considerations

This document introduces no new Security Considerations, above and beyond what is already specified in [rfc4364].

12. Acknowledgements

We would like to acknowledge Han Nguyen (AT&T) for his contributions to this document. We would like to thank Eric Rosen(Cisco) for his review and comments. We would also like to thank Samir Saad (AT&T) and Jeffrey (Zhaohui) Zhang (Juniper) for their review and comments.

13. Normative References

[rfc1918] Rekhter, Y., et.al., "Address Allocation for Private Internets", RFC 1918, February 1996.

[rfc3031] Rosen, E., et. al., "MPLS Architecture", RFC3031, January 2001.

[rfc4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[rfc4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[rfc4684] Pedro Marques, et al., "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC4684, November 2006

[rfc6513] Eric C. Rosen, Rahul Aggarwal, et. al., "Multicast in

MPLS/BGP IP VPNs", RFC6513, February 2012

[rfc6514] R. Aggarwal, E. Rosen, T. Morin, Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC6514, February 2012

[rfc6625] Eric Rosen, et. al., "Wildcards in Multicast VPN Auto-Discovery Routes", RFC6625, May 2012

14. Non-normative References

[seamless-MPLS-multicast] Yakov Rekhter, et. al., "Inter-Area P2MP Segmented LSPs", draft-ietf-mpls-seamless-mcast, work in progress.

15. Authors' Addresses

Huajin Jeng
AT&T
e-mail: hj2387@att.com

James Uttaro
AT&T
e-mail: ju1738@att.com

Luay Jalil
Verizon
e-mail: luay.jalil@verizon.com

Bruno Decraene
France Telecom
e-mail: bruno.decraene@orange.com

Rahul Aggarwal
e-mail: raggarwa_1@yahoo.com

Yakov Rekhter
Juniper Networks, Inc.
e-mail: yakov@juniper.net