L2VPN Workgroup
INTERNET-DRAFT
Intended Status: Standards Track

Ali Sajassi
Samer Salam
Cisco

Yakov Rekhter
John Drake
Juniper

Expires: August 18, 2013

February 18, 2013

IP Inter-Subnet Forwarding in E-VPN
draft-sajassi-l2vpn-evpn-inter-subnet-forwarding-00

Abstract

E-VPN provides an extensible and flexible multi-homing VPN solution for intra-subnet connectivity among hosts/VMs over an MPLS/IP network. However, there are scenarios in which inter-subnet forwarding among hosts/VMs across different IP subnets is required, while maintaining the multi-homing capabilities of E-VPN. This document describes an IRB solution based on E-VPN to address such requirements.

Table of Contents

Terminology

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 RFC 2119 [RFC2119].

1 Introduction

E-VPN provides an extensible and flexible multi-homing VPN solution for intra-subnet connectivity among hosts/VMs over an MPLS/IP network. However, there are scenarios where, in addition to intra-subnet forwarding, inter-subnet forwarding is required among hosts/VMs across different IP subnets, while maintaining the multi-homing capabilities of E-VPN. This document describes an IRB solution based on E-VPN to address such requirements.

2 Inter-Subnet Forwarding Scenarios

The inter-subnet forwarding scenarios for E-VPN can be divided into six categories. The first two scenarios, along with their corresponding solutions, are described in [EVPN-IPVPN-INTEROP]. The solutions for scenarios 3 through 6 are the focus of this document.

1. Connecting IP-VPN sites and E-VPN NVEs without route aggregation
2. Connecting IP-VPN NVEs and E-VPN NVEs without route aggregation
3. Connecting E-VPN NVEs within a DC
4. Connecting E-VPN NVEs in different DCs without route aggregation
5. Connecting E-VPN NVEs in different DCs with route aggregation
6. Connecting IP-VPN sites and E-VPN NVEs with route aggregation

In the above scenarios, the term "route aggregation" refers to the case where a node situated at the edge of the data center network behaves as a default gateway for all VM addresses that are unknown to the data center switches. Effectively, this WAN edge switch implements a gateway functionality. The absence of route aggregation refers to the scenario where all data center switches are aware of all VM addresses (in a given EVI/VRF context), for both VMs in the local as well as remote data centers.

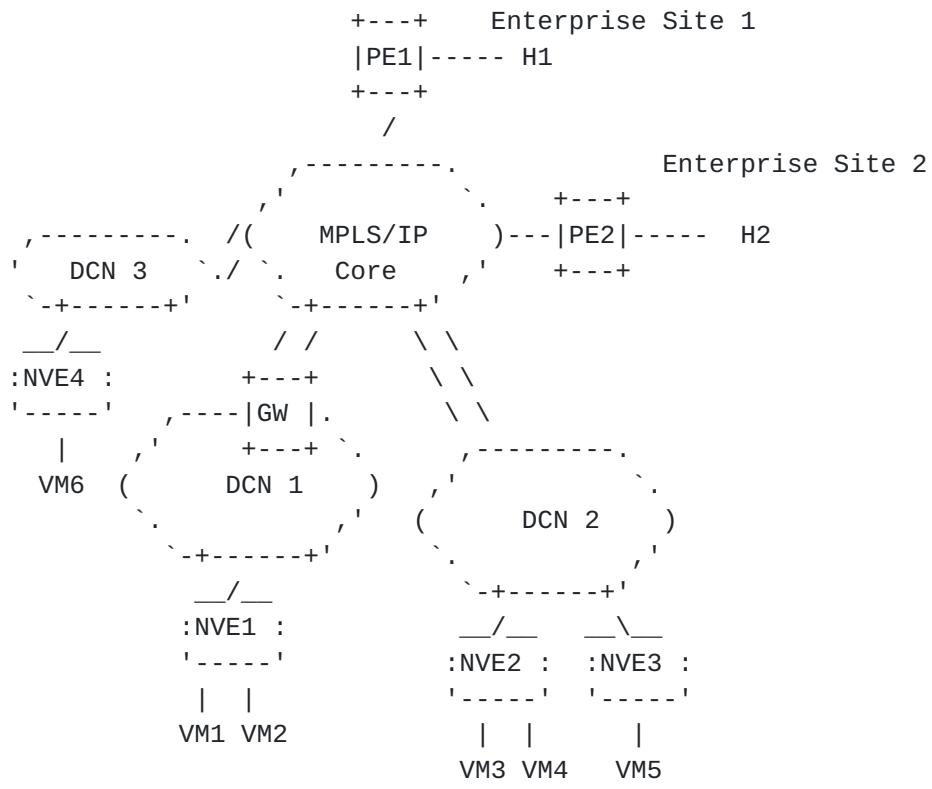

Figure 2: Interoperability Use-Cases

In what follows, we will describe scenarios 3 through 6 in more detail.

2.1 Connecting E-VPN NVEs within a DC

In this scenario, connectivity is required between hosts (e.g. VMs) in the same data center, and those hosts belong to different IP subnets. Each subnet is associated with a single EVI on the NVEs. Furthermore, all the EVIs in question belong to the same VRF.

As an example, consider VM3 and VM5 of Figure 2 above. Assume that connectivity is required between these two VMs where VM3 belongs to the IP3 subnet whereas VM5 belongs to the IP5 subnet. NVE2 has an EVI3 associated with IP3 subnet and NVE3 has an EVI5 associated with the IP5 subnet. Both EVI3 and EVI5 are associated with the same VRFa.

2.2 Connecting E-VPN NVEs in different DCs without route aggregation

This case is similar to that of section 2.1 above albeit for the fact that the hosts belong to different data centers that are interconnected over a WAN (e.g. MPLS/IP PSN). The data centers in question here are seamlessly interconnected to the WAN, i.e. no gateways are used on the data center WAN edge.

As an example, consider VM3 and VM6 of Figure 2 above. Assume that connectivity is required between these two VMs where VM3 belongs to the IP3 subnet whereas VM6 belongs to the IP6 subnet. NVE2 has an EVI3 associated with IP3 subnet and NVE4 has an EVI6 associated with the IP6 subnet. Both EVI3 and EVI6 are associated with the same VRFa.

2.3 Connecting E-VPN NVEs in different DCs with route aggregation

In this scenario, connectivity is required between hosts (e.g. VMs) in different data centers, and those hosts belong to different IP subnets. What makes this case different from that of Section 2.2 is that at least one of the data centers in question has a gateway as the WAN edge switch. Because of that, the NVEs in the data centers with gateways do not have the addresses of the hosts situated in remote data centers.

As an example, consider VM1 and VM5 of Figure 2 above. Assume that connectivity is required between these two VMs where VM1 belongs to the IP1 subnet whereas VM5 belongs to the IP5 subnet. NVE3 has an EVI5 associated with the IP5 subnet and NVE1 has an EVI1 associated with the IP1 subnet. Both EVI1 and EVI5 are associated with the same VRFa. Due to the gateway at the edge of DCN 1, NVE1 does not have the address of VM5 in its VRFa table.

2.4 Connecting IP-VPN sites and E-VPN NVEs with route aggregation

In this scenario, connectivity is required between hosts (e.g. VMs) in a data center and hosts in an enterprise site connected through IP-VPN. The NVE within the data center is an E-VPN NVE, whereas the NVE in the enterprise site is an IP-VPN NVE. Furthermore, the data center in question has a gateway as the WAN edge switch. Because of that, the NVE in the data center does not have the addresses of the hosts situated in the enterprise site.

As an example, consider end-station H1 and VM2 of Figure 2. Assume that connectivity is required between the end-station and the VM, where VM2 belongs to the IP2 subnet whereas H1 belongs to the IP1 subnet. NVE1 has and EVI2 associated with the IP2 subnet. EVI2 is associated with VRFa. On IP-VPN PE1, the IP1 subnet is in VRFa as well. Due to the gateway at the edge of DCN 1, NVE1 does not have the address of H1 in its VRFa table.

3 Concepts needed before solution description

3.1 Default GW & MAC address aliasing versus single MAC/IP

3.2 VM Mobility we can go from scenarios 2.2 to scenario 2.4 (describe how E-VPN provides capability

4 Operational Models for Inter-Subnet Forwarding

4.1 Among E-VPN NVEs within a DC

When an E-VPN MAC advertisement route is received by the NVE, the IP address associated with the route is used to populate the VRF, whereas the MAC address associated with the route is used to populate both the bridge-domain MAC table, as well as the adjacency associated with the IP route in the VRF.

When an Ethernet frame is received by an ingress NVE, it performs a lookup on the destination MAC address in the associated EVI. If the MAC address corresponds to its IRB Interface MAC address, the ingress NVE deduces that the packet must be inter-subnet routed. Hence, the ingress NVE performs an IP lookup in the associated VRF table. The lookup identifies both the next-hop (i.e. egress) NVE to which the packet must be forwarded, in addition to an adjacency that contains a MAC rewrite and an MPLS label stack. The MAC rewrite holds the MAC address associated with the destination host (as populated by the E-VPN MAC route), instead of the MAC address of the next-hop NVE. The ingress NVE then rewrites the destination MAC address in the packet with the address specified in the adjacency. It also rewrites the source MAC address with its IRB Interface MAC address. The ingress NVE, then, forwards the frame to the next-hop (i.e. egress) NVE after encapsulating it with the MPLS label stack. Note that this label stack includes the LSP label as well as the EVI label that was advertised by the egress NVE. When the MPLS encapsulated packet is received by the egress NVE, it uses the EVI label to identify the bridge-domain table. It then performs a MAC lookup in that table, which yields the outbound interface to which the Ethernet frame must be forwarded. Figure 2 below depicts the packet flow, where NVE1 and NVE2 are the ingress and egress NVEs, respectively.

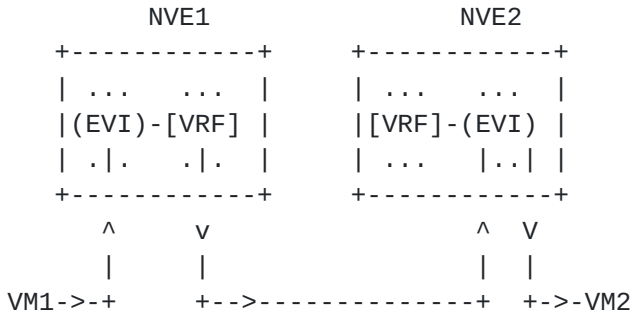

Figure 2: Inter-Subnet Forwarding Among E-VPN NVEs within a DC

Note that the forwarding behavior on the egress NVE is similar to E-

VPN intra-subnet forwarding. In other words, all the packet processing associated with the inter-subnet forwarding semantics is confined to the ingress NVE.

It should also be noted that [E-VPN] provides different level of granularity for the EVI label. Besides identifying bridge domain table, it can be used to identify the egress interface or a destination MAC address on that interface. If EVI label is used for egress interface or destination MAC address identification, then no MAC lookup is needed in the egress EVI and the packet can be directly forwarded to the egress interface just based on EVI label lookup.

4.2 Among E-VPN NVEs in Different DCs Without Route Aggregation

[This section will be expanded in the future revision].

4.3 Among E-VPN NVEs in Different DCs with Route Aggregation

In this scenario, the NVEs within a given data center do not have entries for the MAC/IP addresses of hosts in remote data centers. Rather, the NVEs have a default IP route pointing to the WAN gateway for each VRF. This is accomplished by the WAN gateway advertising for a given E-VPN that spans multiple DC a default VPN-IP route that is imported by the NVEs of that E-VPN that are in the gateway's own DC.

When an Ethernet frame is received by an ingress NVE, it performs a lookup on the destination MAC address in the associated EVI. If the MAC address corresponds to the IRB Interface MAC address, the ingress NVE deduces that the packet must be inter-subnet routed. Hence, the ingress NVE performs an IP lookup in the associated VRF table. The lookup, in this case, matches the default route which points to the local WAN gateway. The ingress NVE then rewrites the destination MAC address in the packet with the IRB Interface MAC address of the local WAN gateway. It also rewrites the source MAC address with its own IRB Interface MAC address. The ingress NVE, then, forwards the frame to the WAN gateway after encapsulating it with the MPLS label stack. Note that this label stack includes the LSP label as well as the IP-VPN label that was advertised by the local WAN gateway. When the MPLS encapsulated packet is received by the local WAN gateway, it uses the IP-VPN label to identify the VRF table. It then performs an IP lookup in that table. The lookup identifies both the remote WAN gateway (of the remote data center) to which the packet must be forwarded, in addition to an adjacency that contains a MAC rewrite and an MPLS label stack. The MAC rewrite holds the MAC address associated with the ultimate destination host (as populated by the E-VPN MAC route). The local WAN gateway then rewrites the destination MAC address in the packet with the address specified in the adjacency. It also rewrites the source MAC address with its IRB Interface MAC address.

The local WAN gateway, then, forwards the frame to the remote WAN gateway after encapsulating it with the MPLS label stack. Note that this label stack includes the LSP label as well as a VPN label that was advertised by the remote WAN gateway. When the MPLS encapsulated packet is received by the remote WAN gateway, it simply swaps the VPN label with the EVI label advertised by the egress NVE. This implies that the remote WAN gateway must allocate the VPN label at least at the granularity of a (VRF, egress NVE) tuple. The remote WAN gateway then forward the packet to the egress NVE. The egress NVE then performs a MAC lookup in the EVI (identified by the received EVI label) to determine the outbound port to send the traffic on.

Figure 4 below depicts the forwarding model.

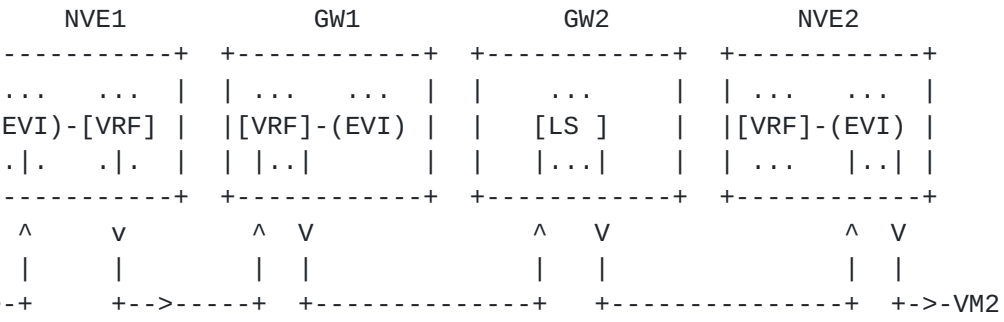

Figure 4: Inter-Subnet Forwarding Among E-VPN NVEs in Different DCs with Route Aggregation

4.4 Among IP-VPN Sites and E-VPN NVEs with Route Aggregation

In this scenario, the NVEs within a given data center do not have entries for the IP addresses of hosts in remote enterprise sites. Rather, the NVEs have a default IP route pointing to the WAN gateway for each VRF.

When an Ethernet frame is received by an ingress NVE, it performs a lookup on the destination MAC address in the associated EVI. If the MAC address corresponds to the IRB Interface MAC address, the ingress NVE deduces that the packet must be inter-subnet routed. Hence, the ingress NVE performs an IP lookup in the associated VRF table. The lookup, in this case, matches the default route which points to the local WAN gateway. The ingress NVE then rewrites the destination MAC address in the packet with the IRB Interface MAC address of the local WAN gateway. It also rewrites the source MAC address with its own IRB Interface MAC address. The ingress NVE, then, forwards the frame to the WAN gateway after encapsulating it with the MPLS label stack. Note that this label stack includes the LSP label as well as the IP-

VPN label that was advertised by the local WAN gateway. When the MPLS encapsulated packet is received by the local WAN gateway, it uses the IP-VPN label to identify the VRF table. It then performs an IP lookup in that table. The lookup identifies the next hop ASBR to which the packet must be forwarded. The local gateway in this case strips the Ethernet encapsulation and forwards the IP packet to the ASBR using a label stack comprising of an LSP label and a VPN label that was advertised by the ASBR. When the MPLS encapsulated packet is received by the ASBR, it simply swaps the VPN label with the IP-VPN label advertised by the egress PE. This implies that the remote WAN gateway must allocate the VPN label at least at the granularity of a (VRF, egress PE) tuple. The ASBR then forwards the packet to the egress PE. The egress PE then performs an IP lookup in the VRF (identified by the received IP-VPN label) to determine where to forward the traffic.

Figure 5 below depicts the forwarding model.

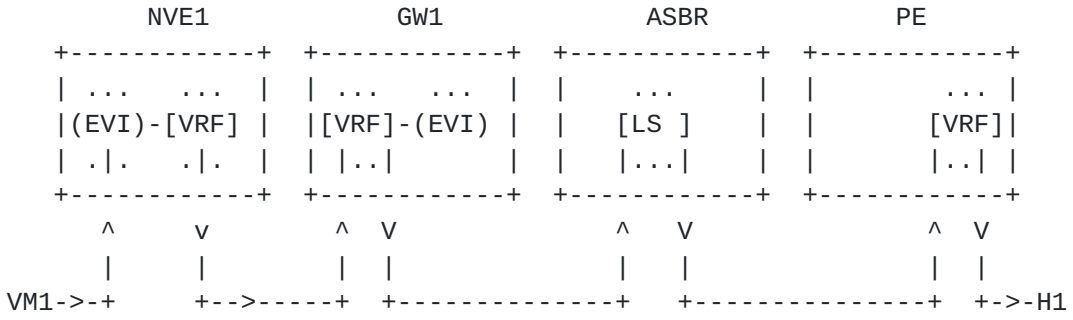

Figure 5: Inter-Subnet Forwarding Among IP-VPN Sites and E-VPN NVEs with Route Aggregation

5 VM Mobility

describe how mobility works

5 Acknowledgement

6 Security Considerations

7 IANA Considerations

8 References

8.1 Normative References

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

8.2 Informative References

[EVPN] Sajassi et al., "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-00.txt, work in progress, February, 2012.

[EVPN-IPVPN-INTEROP] Sajassi et al., "E-VPN Seamless Interoperability with IP-VPN", draft-sajassi-l2vpn-evpn-ipvpn-interop-01, work in progress, October, 2012.

[DC-MOBILITY] Aggarwal et al., "Data Center Mobility based on BGP/MPLS, IP Routing and NHRP", draft-raggarwa-data-center-mobility-03.txt, work in progress, June, 2012.

Authors' Addresses

Ali Sajassi
Cisco
Email: sajassi@cisco.com

Samer Salam
Cisco
Email: ssalam@cisco.com

Yakov Rekhter
Juniper Networks
Email: yakov@juniper.net

John E. Drake
Juniper Networks
Email: jdrake@juniper.net