Configuring EVPN Type 5 for QFX10000 Series Switches

Configuration Example

August 2016

Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, California 94089
USA
408-745-2000
www.juniper.net

Contents

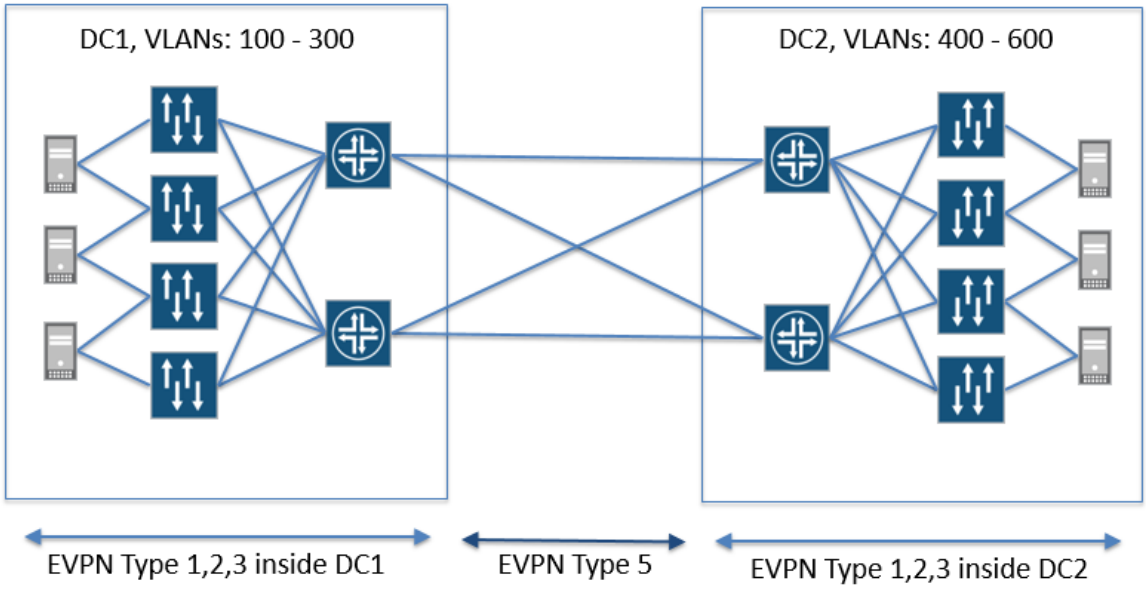
Introduction

This document describes how to configure EVPN Type 5 on QFX10000 Series switches. Currently, only EVPN/VXLAN and symmetric Type 5 is supported.

Customer Use Case

Advantages of using EVPN Type 5 in DCI

When a single Layer 2 domain is not extended across multiple data center (DC) locations, the IP subnet belonging to this Layer 2 domain is confined within a single DC. If all Layer 2 domains within each DC network satisfy this requirement, there is no longer a need to advertise separate MAC+IP routes for each tenant between DCs, as host routes for the tenants can be aggregated. Thus, the Layer 2 inter-DC connectivity issue can be simply transformed to an inter-DC Layer 3 IP prefix reachability issue. A unified EVPN control plane can be used to accomplish Layer 3 route advertisement between multiple DC locations, avoiding the need for an additional L3 VPN protocol family.



Summary of advantages

- No need to exchange all hosts routes between DC locations, resulting in lower RIB/FIB requirements on DCI equipment
- No need to use multiple protocol families (EVPN and L3 VPN) to advertise Layer 2 and Layer 3 reachability information

Technical Overview

Understanding EVPN Type 5 Implementation on QFX 10000 Series Switches

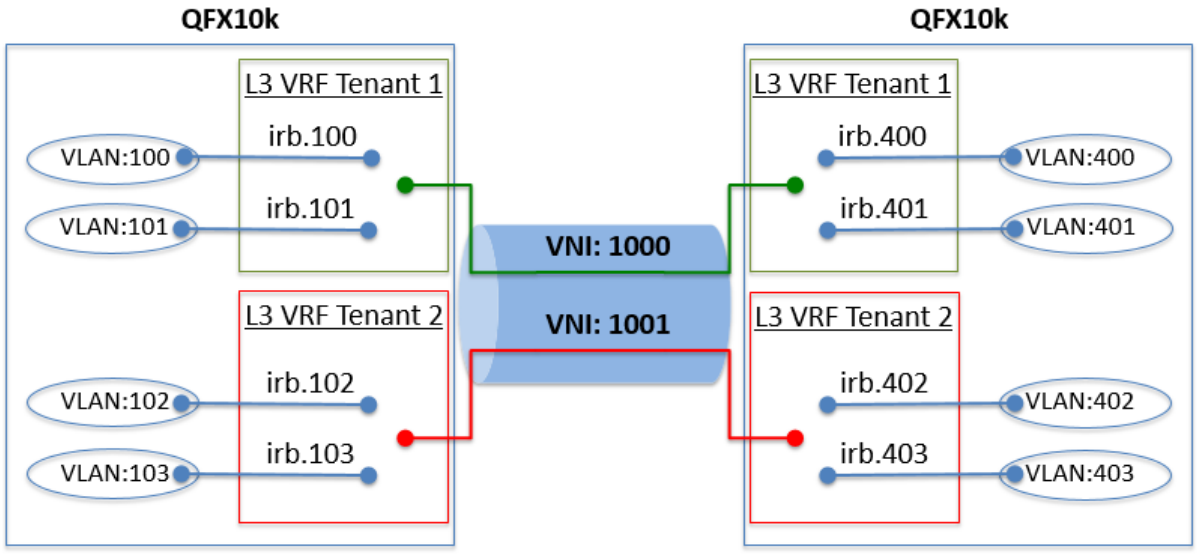
IETF draft “IP Prefix Advertisement in EVPN” (https://tools.ietf.org/html/draft-ietf-bess-evpn-prefix-advertisement) defines IP Prefix Route encoding:

```
+---------------------------------------+
|      RD   (8 octets)                  |
+---------------------------------------+
|Ethernet Segment Identifier (10 octets)|
+---------------------------------------+
|  Ethernet Tag ID (4 octets)           |
+---------------------------------------+
|  IP Prefix Length (1 octet)           |
+---------------------------------------+
|  IP Prefix (4 or 16 octets)           |
+---------------------------------------+
|  GW IP Address (4 or 16 octets)       |
+---------------------------------------+
|  MPLS Label (3 octets)                |
+---------------------------------------+
```

IETF draft “Integrated Routing and Bridging in EVPN” (https://tools.ietf.org/html/draft-ietf-bess-evpn-inter-subnet-forwarding) defines two operational modes for inter-subnet forwarding in EVPN: symmetric and asymmetric.

In asymmetric mode, all packet processing associated with inter-subnet forwarding is done on the ingress network virtualization endpoint (NVE). Forwarding behavior at the egress NVE is similar to EVPN intra-subnet forwarding, i.e. it performs only a MAC address lookup. This explains why the mode is called “asymmetric”—the number of lookups is different on the ingress and egress NVEs. There are multiple ways of implementing asymmetric inter-subnet forwarding, however as it is not supported on QFX10000 Series switches there is no further detail about this mode in this document.

By contrast, “symmetric” mode performs an equal number of lookups—IP and MAC—on both the ingress and egress NVEs. Symmetric mode is also called the VRF-to-VRF model. In this scenario, for a given tenant (e.g., an IP-VPN service), an NVE has one MAC-VRF, which consists of multiple bridge domains (one bridge domain per VLAN). The MAC-VRFs on an NVE for a given tenant are associated with an IP-VRF corresponding to that tenant (or IP-VPN service) via their IRB interfaces. A globally unique VNI will be provisioned for each customer L3 VRF. The VNI will be used to identify the customer on each DC.



With symmetric mode, all received EVPN Type 5 routes will resolve over the BGP next hop. The overlay gateway address inside a Type 5 route will always be set to 0.0.0.0. Since VXLAN encapsulation requires an inner Ethernet header (inner MAC SA/DA), and since for inter-subnet traffic the source/destination host MAC addresses cannot be used, the ingress

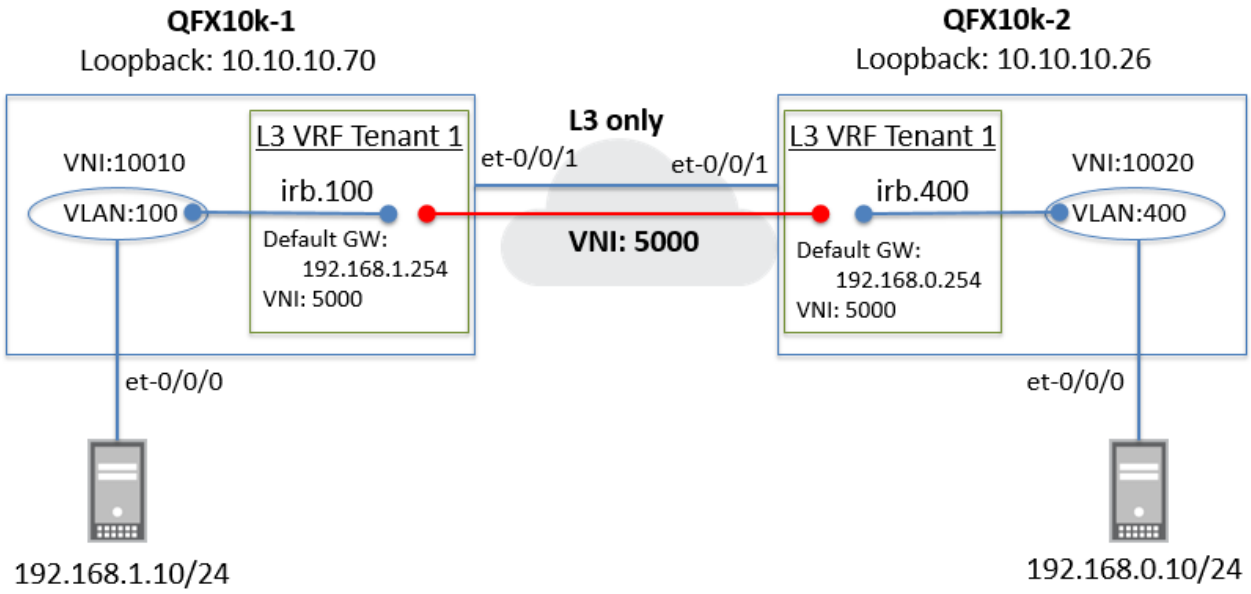
NVE's MAC address (the “chassis MAC”) is used as the inner MAC SA/DA. The NVE's MAC address is the device MAC address and it is common across all MAC-VRFs and IP-VRFs. This MAC address is advertised using a new EVPN BGP extended community called Router's MAC.

Configuration Example: Configuring EVPN Type 5 on QFX10000 Series Switches

Requirements

- Two QFX10002 switches
- Junos OS release 15.1X53.D30

Topology



Packet Flow

1. 192.168.1.10 sends an ARP request to the local IRB address (192.168.1.254) on QFX10k-1.
2. QFX10k-1 replies with an ARP reply. The target MAC address is set to the local chassis MAC address (IRB-MAC1).
3. 192.168.1.10 sends an IP packet with Src IP: 192.168.1.10, Dst IP: 192.168.0.10, Dst MAC: IRB-MAC1
4. QFX10k-1 receives the packet on its IRB interface and does a lookup for 192.168.0/24. The next hop is QFX10k-2’s IP address 10.10.10.26, which resolves over VNI 5000.
5. QFX10k-1 sets the inner L2 header Dst MAC address to **00:00:5E:00:53:48** (the chassis MAC address of QFX10k-2), and adds a VXLAN header with the VNI set to 5000.
6. QFX10k-2 receives the VXLAN-encapsulated packet. The VNI value 5000 tells QFX10k-2 to do an IP lookup in the L3 VRF.
7. QFX10k-2 sends an ARP request over the EVPN BUM tree (broadcast), or forwards the traffic according to the EVPN or local MAC database, if the end host’s IP address already resolved.

Configuration

Step 1: Configure core, L2 access and IRB interfaces

QFX10k-1

```
interfaces {
    et-0/0/1 {
        mtu 9192;
        unit 0 {
            family inet {
                address 10.0.0.1/24;
            }
        }
    }
    et-0/0/0 {
        unit 0 {
            family ethernet-switching {
                interface-mode trunk;
                vlan {
                    members all;
                }
            }
        }
    }
    irb {
        unit 100 {
            family inet {
                address 192.168.1.254/24;
            }
        }
    }
    lo0 {
        unit 0 {
            family inet {
                address 10.10.10.70/32 {
                    primary;
                }
            }
        }
        unit 1 {
            family inet {
                address 10.10.10.170/32 {
                    primary;
                }
            }
        }
    }
}
```

QFX10k-2

```
interfaces {
    et-0/0/1 {
        mtu 9192;
        unit 0 {
            family inet {
                address 10.0.0.2/24;
            }
        }
    }
    et-0/0/0 {
        unit 0 {
            family ethernet-switching {
                interface-mode trunk;
                vlan {
                    members all;
                }
            }
        }
    }
    irb {
        unit 400 {
            family inet {
                address 192.168.0.254/24;
            }
        }
    }
    lo0 {
        unit 0 {
            family inet {
                address 10.10.10.26/32 {
                    primary;
                }
            }
        }
        unit 1 {
            family inet {
                address 10.10.10.126/32 {
                    primary;
                }
            }
        }
    }
}
```

Step 2: Configure OSPF, BGP, EVPN and AS number

QFX10k-1

```
protocols {
    bgp {
        group interop {
            type internal;
            local-address 10.10.10.70;
            family evpn {
                signaling;
            }
            neighbor 10.10.10.26;
        }
    }
    ospf {
        area 0.0.0.0 {
            interface all {
                interface-type p2p;
            }
            interface em0.0 {
                disable;
            }
        }
    }
    evpn {
        encapsulation vxlan;
        multicast-mode ingress-replication;
    }
}

routing-options {
    router-id 10.10.10.70;
    autonomous-system 300;
}
```

QFX10k-2

```
protocols {
    bgp {
        group interop {
            type internal;
            local-address 10.10.10.26;
            family evpn {
                signaling;
            }
            neighbor 10.10.10.70;
        }
    }
    ospf {
        area 0.0.0.0 {
            interface all {
                interface-type p2p;
            }
            interface em0.0 {
                disable;
            }
        }
    }
    evpn {
        encapsulation vxlan;
        multicast-mode ingress-replication;
    }
}

routing-options {
    router-id 10.10.10.26;
    autonomous-system 300;
}
```

Step 3: Configure switch-options

On QFX10000 Series switches, there is only one virtual-switch instance. This means that all 4096 VLANs belong to this virtual-switch. The VTEP source interface must be configured to use lo0.0. The vrf-target under switch-options is used only for EVPN Type 1 routes (multi-homed segments).

QFX10k-1

```
switch-options {
    vtep-source-interface lo0.0;
    route-distinguisher 10.10.10.70:1;
    vrf-target target:300:1001;
}
```

QFX10k-2

```
switch-options {
    vtep-source-interface lo0.0;
    route-distinguisher 10.10.10.26:1;
    vrf-target target:300:1001;
}
```

Step 4: Configure L2 VLANs/VNIs mapping

QFX10k-1

```
vlans {
    bd100 {
        vlan-id 100;
        l3-interface irb.100;
        vxlan {
            vni 10010;
            ingress-node-replication;
        }
    }
}
```

QFX10k-2

```
vlans {
    bd400 {
        vlan-id 400;
        l3-interface irb.400;
        vxlan {
            vni 10020;
            ingress-node-replication;
        }
    }
}
```

Step 5: Configure VRF-to-VRF for EVPN Type 5

QFX10k-1

```
routing-instances {
    TYPE5 {
        instance-type vrf;
        interface lo0.1;
        interface irb.100;
        route-distinguisher 10.10.10.70:5000;
        vrf-target target:300:5000;
        protocols {
            evpn {
                ip-prefix-support {
                    forwarding-mode symmetric;
                    encapsulation vxlan;
                    vni 5000;
                }
            }
        }
    }
}
```

QFX10k-2

```
routing-instances {
    TYPE5 {
        instance-type vrf;
        interface lo0.1;
        interface irb.400;
        route-distinguisher 10.10.10.26:5000;
        vrf-target target:300:5000;
        protocols {
            evpn {
                ip-prefix-support {
                    forwarding-mode symmetric;
                    encapsulation vxlan;
                    vni 5000;
                }
            }
        }
    }
}
```

Verification

Verify the state of the BGP session

```
user@qfx10k-1> show bgp summary
Groups: 1 Peers: 1 Down peers: 0
Table          Tot Paths  Act Paths Suppressed    History Damp State    Pending
bgp.evpn.0
                   1          1          0          0          0          0
Peer                     AS      InPkt     OutPkt    OutQ   Flaps Last Up/Dwn
State|#Active/Received/Accepted/Damped...
10.10.10.26             300        138         33       0      45       12:55 Establ
  bgp.evpn.0: 1/1/1/0
  default-switch.evpn.0: 0/0/0/0
  __default_evpn__.evpn.0: 0/0/0/0
  TYPE5.evpn.0: 1/1/1/0
```

Verify that 192.168.1/24 was exchanged as a Type 5 route

The overlay gateway address should be set to 0. The Router' s MAC extended community must be set to the chassis MAC address of QFX10k-2.

```
user@qfx10k-1> show route receive-protocol bgp 10.10.10.26 extensive
inet.0: 23 destinations, 24 routes (23 active, 0 holddown, 0 hidden)
:vxlan.inet.0: 7 destinations, 7 routes (6 active, 0 holddown, 1 hidden)
TYPE5.inet.0: 3 destinations, 3 routes (3 active, 0 holddown, 0 hidden)
inet6.0: 1 destinations, 1 routes (1 active, 0 holddown, 0 hidden)
bgp.evpn.0: 1 destinations, 1 routes (1 active, 0 holddown, 0 hidden)
* 5:10.10.10.26:5000::0::192.168.0.0::24/304 (1 entry, 0 announced)
     Import Accepted
     Route Distinguisher: 10.10.10.26:5000
     Route Label: 5000
     Overlay gateway address: 0.0.0.0
     Nexthop: 10.10.10.26
     MED: 0
     Localpref: 100
     AS path: ? (Originator)
     Communities: target:300:5000 encapsulation0:0:0:0:vxlan router-mac 00:00:5E:00:53:48
__default_evpn__.evpn.0: 1 destinations, 1 routes (1 active, 0 holddown, 0 hidden)


TYPE5.evpn.0: 2 destinations, 2 routes (2 active, 0 holddown, 0 hidden)
* 5:10.10.10.26:5000::0::192.168.0.0::24/304 (1 entry, 1 announced)
     Import Accepted
     Route Distinguisher: 10.10.10.26:5000
     Route Label: 5000
     Overlay gateway address: 0.0.0.0
     Nexthop: 10.10.10.26
     MED: 0
     Localpref: 100
     AS path: ? (Originator)
     Communities: target:300:5000 encapsulation0:0:0:0:vxlan router-mac 00:00:5E:00:53:48
```