Overlay Transport Virtualization (OTV) Inter-DC Multicast Traffic over Unicast Transport

June 14, 2012

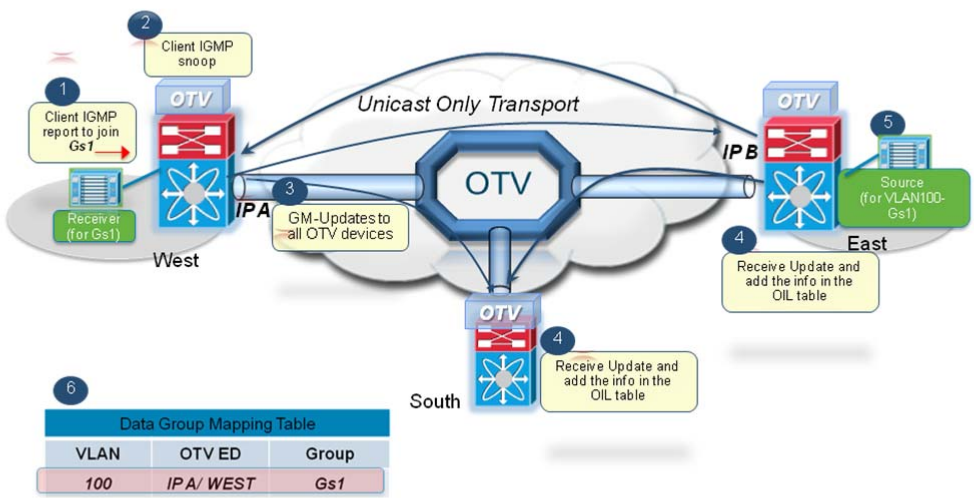
Inter-DC Multicast Traffic over Unicast Transport

The purpose of this paper is to discuss how Layer 2 Multicast packets with IP headers communicate across an OTV Unicast core. If non-IP Layer 2 Multicast packets are introduced into this environment, OTV will simply broadcast those packets to all data centers.

In certain scenarios there may be the requirement to establish Layer 2 multicast communication between remote sites which can be accomplished simply by adding a one line configuration on the Nexus 7000 in each data center. This is the case when a multicast source sending traffic to a specific group is deployed in a given VLAN in an East data center, whereas multicast receivers belonging to the same VLAN are placed in a West and South data center.

Figure 1 shows an IGMP Overview and Unicast OTV scenario.

Figure 1 IGMP Overview and Unicast OTV



Transport Process

The following simplified process defines the steps shown in Figure 1 necessary to establish a typical OTV inter-DC multicast over unicast transport configuration.

Step 1 Receiver (West) sends IGMP reports to join a multicast group.

Step 2 An Edge Device (ED) snoops these reports, but but does NOT forward them on the overlay network.

Step 3 Upon snooping IGMP reports, the Edge Device announces the receivers in a Group-Membership Update (which is an OTV Control Packet) to all EDs that belong to the same logical overlay.

Step 4 On reception of the GM Update, an ED will add the edge device to the appropriate multicast outbound interface list (OIL), (East and South).

Step 5 When the source begins sending traffic, the an Edge Device sees the overlay interface in the OIL and replicates multicast traffic to specific Edge Devices where an interested receiver is in that multicast group.

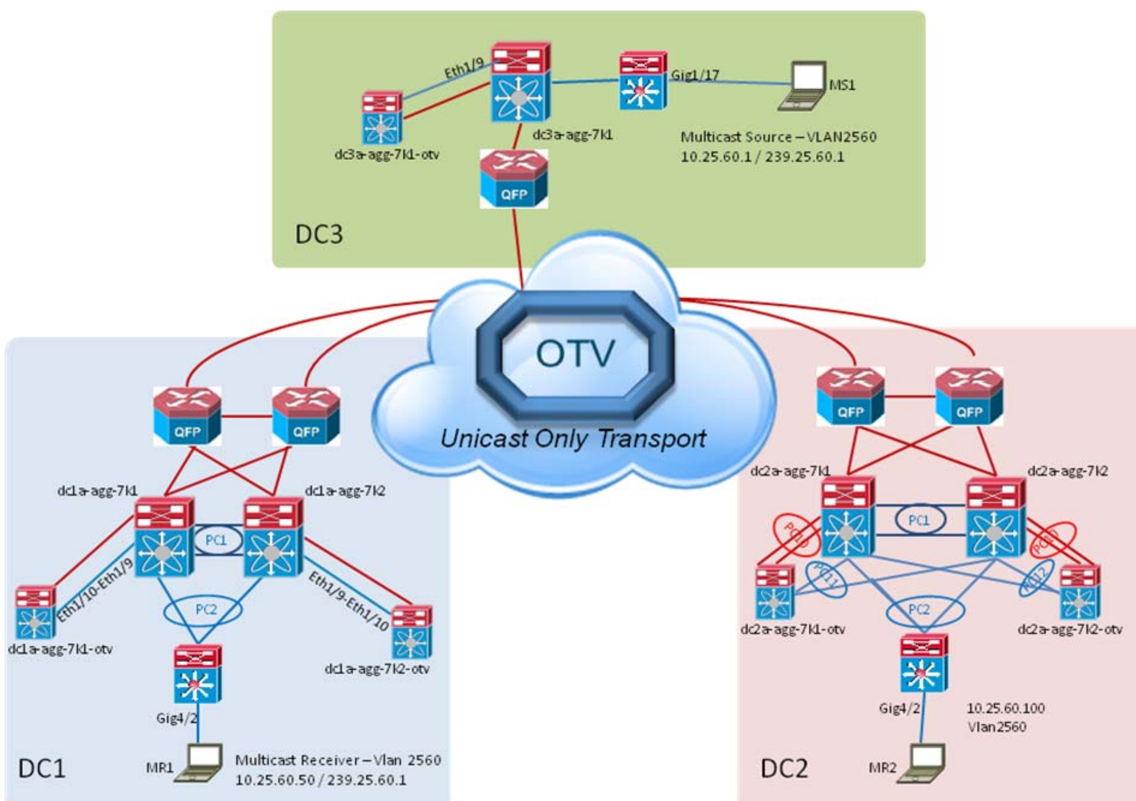
Step 6 Replication is optimized since only EDs with receivers will join the specific multicast group. South will not receive the multicast traffic, since there are no receivers.

OTV is configured on a separate Virtual Device Context (VDC) which enforces the separation between SVI routing and OTV encapsulation for a given VLAN.

Testbed Configuration

Figure 2 test bed was used to verify that L2 multicast traffic flows were established to only those data centers that had clients requesting the specific multicast groups.

Figure 2 Test Bed Configuration



As shown in Figure 2, OTV is configured in Unicast mode within each data center. This document assumes that the reader already has an understanding of OTV. For more information on OTV please visit on the following URL:

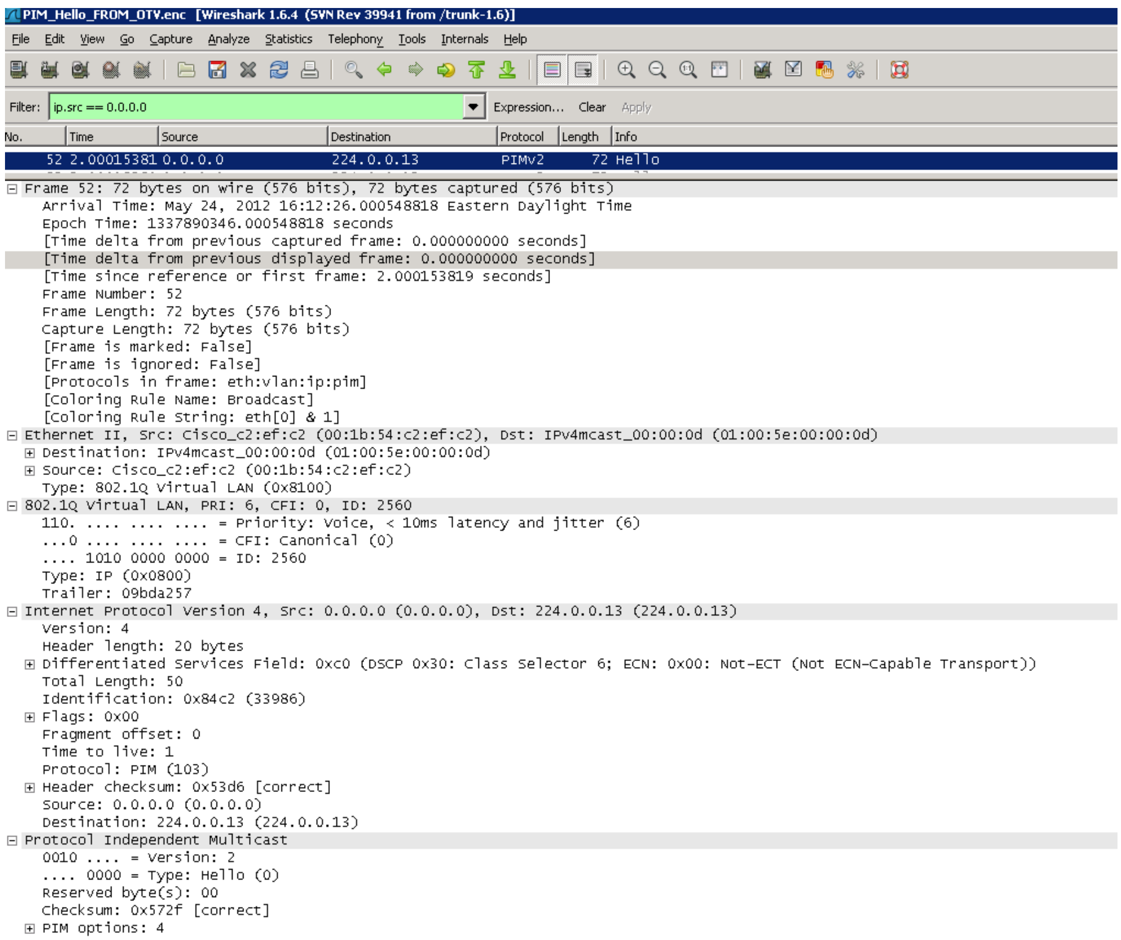
http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DCI/whitepaper/DCI_1.html

As a result of OTV being configured in Unicast mode, "dummy" IP PIM hello packets are generated by the OTV AED (Authoritative Edge Device) for that particular vlan, and sent to the aggregation VDC via the OTV internal interface. This packet allows for a multicast mrouter port to be created for that VLAN on the aggregation VDC, so that multicast data packets can be forwarded across the OTV transport. Figure 3 shows a packet capture of the PIM Hello packet that gets generated. Notice the source IP address of the packet 0.0.0.0. In the event of any special ACL filtering or Firewall appliance these packets should be permitted into the aggregation VDC on the Nexus 7000.

Figure 3 shows a "dummy" PIM Hello packet generated from the OTV AED.

Note Highlighted line "Internet Protocol Version 4, src: 0.0.0.0 (0.0.0.0), Dst: 224.0.01.3 (224.0.0.13)" identifies the Source IP address 0.0.0.0.

Figure 3 "Dummy" PIM Hello packet generated from the OTV AED



Data Center Configuration Procedure

The current configuration recommendation to allow L2 multicast traffic to function correctly is to configure a specific IGMP snooping querier for each VLAN that will need to receive multicast traffic. When an IGMP snooping querier is enabled, it sends out periodic IGMP queries that trigger IGMP report messages from hosts that want to receive IP multicast traffic. IGMP snooping listens to these IGMP reports to establish appropriate forwarding. Configuration of an IGMP snooping querier also creates an mrouter port on the access-layer switch as well

To configure an IGMP querier specific to a VLAN on a Nexus 7000 switch, the configuration is performed specifically on the VLAN and not on the SVI. All configurations are performed in the aggregation VDC of each Nexus 7000.

Note For L2 multicast, PIM is not configured on these SVIs. However, if PIM were configured on the SVI in each data center it would imply that there is a receiver for all multicast groups. Consequently, multicast traffic would get forwarded to all data centers, independent of an active multicast receiver in that data center. IGMP querier packets do not traverse the OTV core and will remain local to each data center.

The following procedure, using the appropriate corresponding NX-OS commands, allows you to configure an IGMP querier. Configuration examples are included.

Step 1 Enter global configuration mode.

configure terminal

```
switch# configure terminal
switch(config)#
```

Step 2 Enable IGMP snooping for the current VDC. The default is enabled.

ip igmp snooping

```
switch(config)# ip igmp snooping
```

Note If the global setting is disabled with the no form of this command, IGMP snooping on all VLANs is disabled, whether IGMP snooping is enabled on a VLAN or not. If you disable IGMP snooping, Layer 2 multicast frames flood to all modules.

Step 3 Beginning with Cisco Release 5.1(1), use this command to configure the IGMP snooping parameters you want for the VLAN. These configurations do not apply until you specifically create the specified VLAN.

vlan configuration *vlan-id*

```
switch(config)# vlan configuration 2560
switch(config-vlan-config)#
```

Step 4 Configure a snooping querier when you do not enable PIM because multicast traffic does not need to be routed. The IP address is used as the source in messages. The ip address configured needs to be within the subnet range for the specific VLAN.

ip igmp snooping querier *ip-address*

```
switch(config-vlan-config)# ip igmp snooping querier 10.25.56.253
```

Step 5 (Optional) Configures a snooping MRT for query messages when you do not enable PIM because multicast traffic does not need to be routed. The default value is 10 seconds. You may want to change this option depending upon how often the receiver sends out periodic IP IGMP join messages.

ip igmp snooping query-max-response-time *seconds*

```
switch(config-vlan-config)# ip igmp snooping query-max-response-time 10
```

Step 6 (Optional) Configures a snooping query interval when you do not enable PIM because multicast traffic does not need to be routed. The default value is 125 seconds.

ip igmp snooping query-interval *seconds*

```
switch(config-vlan-config)# ip igmp snooping query-interval 5
```

Step 7 (Optional) Exit from configuration mode.

exit

```
switch(config-vlan-config)# exit
```

Step 8 (Optional) Copies the running configuration to the startup configuration.

copy running-config startup-config

```
switch# copy running-config startup-config
```

OTV Multicast Enable Transport Configuration Example

The following sample configuration is provided.

```
dc1a-agg-7k1# show running-config vlan 2560
version 5.2(5)
vlan configuration 2560
  ip igmp snooping querier 10.25.60.253
dc1a-agg-7k1# show running-config interface vlan2560

version 5.2(5)
interface Vlan2560
  no shutdown
  mtu 9216
  no ip redirects
  ip address 10.25.60.253/24
  ip ospf passive-interface
  hsrp 1
    preempt delay minimum 180 reload 300
    priority 253
    timers  1  3
    ip 10.25.60.254
```

The following commands confirm that IP IGMP mrouter ports are created on the aggregation VDC.

```
dc1a-agg-7k1# show ip igmp snooping mrouter vlan 2560
Type: S - Static, D - Dynamic, V - vPC Peer Link
      I - Internal, F - Fabricpath core port
      U - User Configured
Vlan  Router-port   Type      Uptime      Expires
2560  Po1           SV        4d03h       never (See Figure 2 - DC1)
2560  Eth1/9        D         4d03h       00:04:57(See Figure 2 - DC1)

dc1a-agg-7k2# show ip igmp snooping mrouter vlan 2560
Type: S - Static, D - Dynamic, V - vPC Peer Link
I - Internal, F - Fabricpath core port
U - User Configured
Vlan  Router-port   Type      Uptime      Expires
2560  Po1           SV        3d04h       never (See Figure 2 - DC1)
2560  Eth1/9        D         3d04h       00:04:44 (See Figure 2 - DC1)
```

Once the mrouter port is created on the aggregration VDC, any IGMP report messages received will be forwarded to the OTV VDC, creating a (VLAN,*,G) for the specific VLAN and multicast group on all OTV edge devices.

1. Client MR1, multicast receiver in DC1, sends out a request to join the multicast group 239.25.60.1

```
dc1a-agg-7k2-otv# show otv mroute group 239.25.60.1
OTV Multicast Routing Table For Overlay200
(2560, *, 239.25.60.1), metric: 0, uptime: 00:00:54, igmp
  Outgoing interface list: (count: 1)
    Eth1/10, uptime: 00:00:54, igmp
```

2. OTV Control Plane messages are received by each data center and install the (Vlan,*,G)

```
dc2a-agg-7k2-otv# show otv mroute group 239.25.60.1
OTV Multicast Routing Table For Overlay200
(2560, *, 239.25.60.1), metric: 0, uptime: 00:00:58, overlay(r)
  Outgoing interface list: (count: 1)
    Overlay200, dc1a-agg-7k2-otv, uptime: 00:00:58, isis_otv-default

dc3a-agg-7k-otv# show otv mroute group 239.25.60.1
OTV Multicast Routing Table For Overlay200
(2560, *, 239.25.60.1), metric: 0, uptime: 00:01:01, overlay(r)
  Outgoing interface list: (count: 1)
    Overlay200, dc1a-agg-7k2-otv, uptime: 00:01:01, isis_otv-default
```

3. The Multicast source (10.25.60.1 -> 239.25.60.1) in DC2 begins sending traffic.

```
dc3a-agg-7k-otv# show otv mroute group 239.25.60.1
OTV Multicast Routing Table For Overlay200
(2560, *, 239.25.60.1), metric: 0, uptime: 00:00:26, overlay(r)
  Outgoing interface list: (count: 1)
    Overlay200, dc1a-agg-7k2-otv, uptime: 00:00:26, isis_otv-default
(2560, 10.25.60.1, 239.25.60.1), metric: 0, uptime: 00:00:16, site
  Outgoing interface list: (count: 1)
    Overlay200, dc1a-agg-7k2-otv, uptime: 00:00:16, otv
```

Brian Howard

Test Lead, Systems Development Unit, Cisco Systems

Brian Howard is a Test Lead Software Engineer in the Systems Development Unit focusing on data center interconnect (DCI) technologies. Recent DCI design and test efforts include Cisco Overlay Transport Virtualization, Advanced Virtual Private LAN Services (A-VPLS), Cisco Nexus 1000V Series Switches, Virtual Security Gateway, and LISP. He has provided quality initiatives and testing in Cisco Advanced Services and the Cisco Corporate Development Office for 12 years, focusing primarily on routing and switching, and most recently in data center virtualization using DCI.