



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

June 14, 2012

CCDE, CCENT, CCSI, Cisco Eos, Cisco Explorer, Cisco HealthPresence, Cisco IronPort, the Cisco logo, Cisco Nurse Connect, Cisco Pulse, Cisco SensorBase, Cisco StackPower, Cisco StadiumVision, Cisco TelePresence, Cisco TrustSec, Cisco Unified Computing System, Cisco WebEx, DCE, Flip Channels, Flip for Good, Flip Mino, Flipshare (Design), Flip Ultra, Flip Video, Flip Video (Design), Instant Broadband, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn, Cisco Capital, Cisco Capital (Design), Cisco:Financed (Stylized), Cisco Store, Flip Gift Card, and One Million Acts of Green are service marks; and Access Registrar, Aironet, AITouch, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Lumin, Cisco Nexus, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, Continuum, EtherFast, EtherSwitch, Event Center, Explorer, Follow Me Browsing, GainMaker, iLYNX, IOS, iPhone, IronPort, the IronPort logo, Laser Link, LightStream, Linksys, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, PCNow, PIX, PowerKEY, PowerPanels, PowerTV, PowerTV (Design), PowerVu, Prisma, ProConnect, ROSA, SenderBase, SMARTnet, Spectrum Expert, StackWise, WebEx, and the WebEx logo are registered trademarks of Cisco and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1002R)

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Overlay Transport Virtualization (OTV) Inter-DC Multicast Traffic over Unicast Transport
© 2012 Cisco Systems, Inc. All rights reserved.



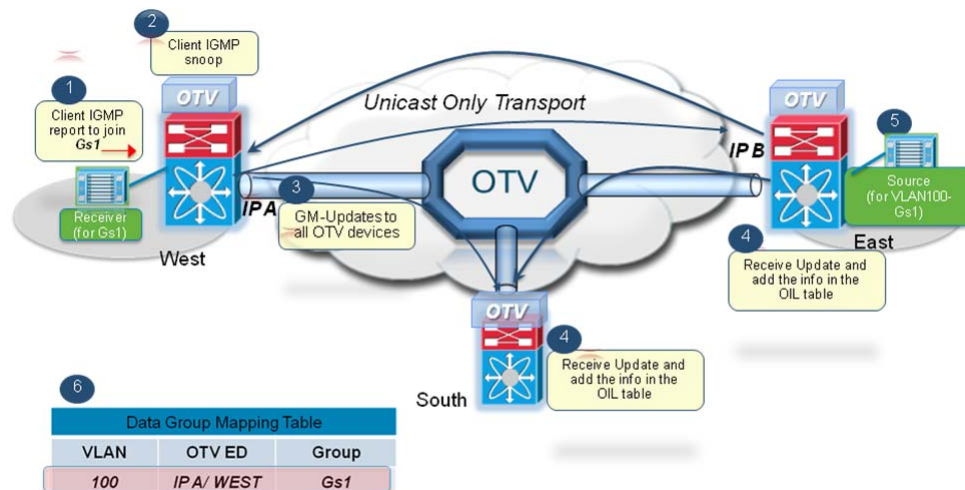
Inter-DC Multicast Traffic over Unicast Transport

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.

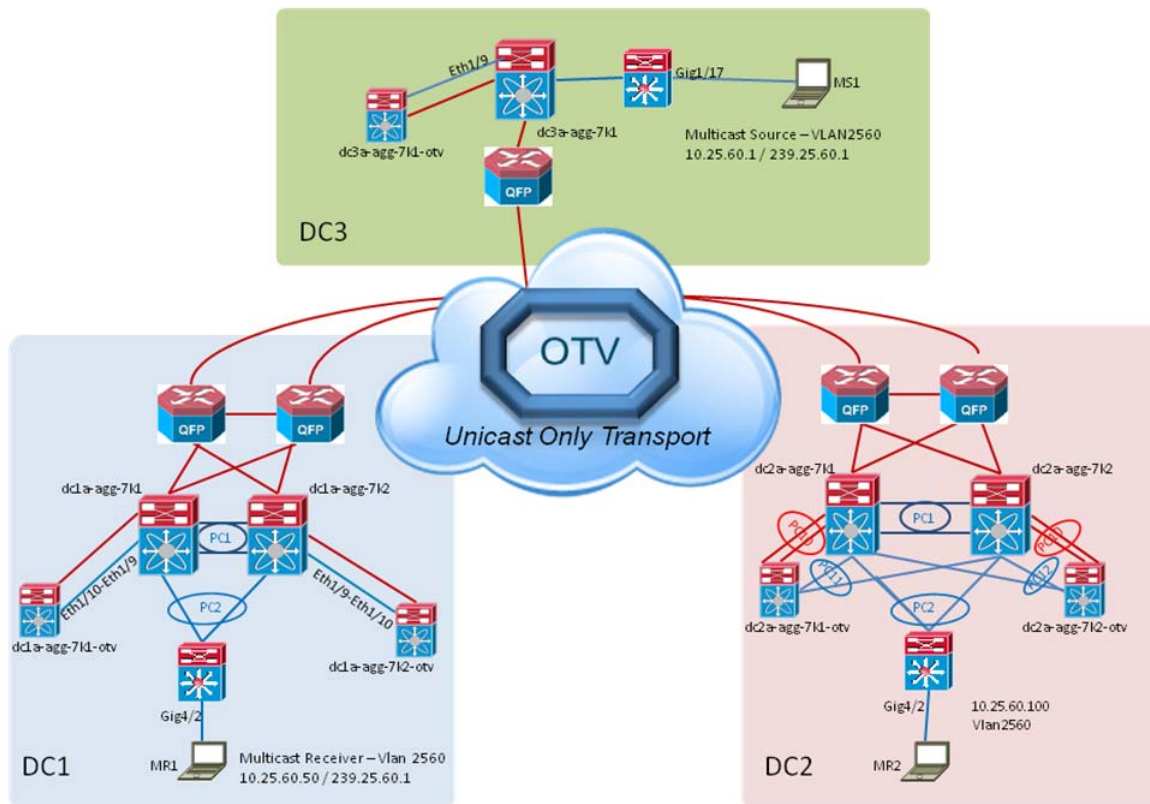
- Step 3** Upon snooping IGMP reports, the Edge Device announces the receivers in a Group-Membership Update (which is an OTV Control Packet) to all EDs that belong to the same logical overlay.
- Step 4** On reception of the GM Update, an ED will add the edge device to the appropriate multicast outbound 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: 224.0.0.13 (224.0.0.13)" identifies the Source IP address 0.0.0.0.

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

PIM_Hello_FROM_OTV.enc [Wireshark 1.6.4 (SVN Rev 39941 from /trunk-1.6)]

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help

Filter: ip.src == 0.0.0.0 Expression... Clear Apply

No.	Time	Source	Destination	Protocol	Length	Info
52	2.00015381	0.0.0.0	224.0.0.13	PIMv2	72	Hello

- Frame 52: 72 bytes on wire (576 bits), 72 bytes captured (576 bits)
 - Arrival Time: May 24, 2012 16:12:26.000548818 Eastern Daylight Time
 - Epoch Time: 1337890346.000548818 seconds
 - [Time delta from previous captured frame: 0.000000000 seconds]
 - [Time delta from previous displayed frame: 0.000000000 seconds]
 - [Time since reference or first frame: 2.000153819 seconds]
 - Frame Number: 52
 - Frame Length: 72 bytes (576 bits)
 - Capture Length: 72 bytes (576 bits)
 - [Frame is marked: False]
 - [Frame is ignored: False]
 - [Protocols in frame: eth:vlan:ip:pim]
 - [Coloring Rule Name: Broadcast]
 - [Coloring Rule String: eth[0] & 1]
- Ethernet II, Src: Cisco_c2:ef:c2 (00:1b:54:c2:ef:c2), Dst: IPv4mcast_00:00:0d (01:00:5e:00:00:0d)
 - Destination: IPv4mcast_00:00:0d (01:00:5e:00:00:0d)
 - Source: Cisco_c2:ef:c2 (00:1b:54:c2:ef:c2)
 - Type: 802.1Q Virtual LAN (0x8100)
- 802.1Q Virtual LAN, PRI: 6, CFI: 0, ID: 2560
 - 110 = Priority: Voice, < 10ms latency and jitter (6)
 - ...0 = CFI: Canonical (0)
 - ... 1010 0000 0000 = ID: 2560
 - Type: IP (0x0800)
 - Trailer: 09bda257
- Internet Protocol Version 4, Src: 0.0.0.0 (0.0.0.0), Dst: 224.0.0.13 (224.0.0.13)
 - Version: 4
 - Header length: 20 bytes
 - Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00: Not-ECT (Not ECN-Capable Transport))
 - Total Length: 50
 - Identification: 0x84c2 (33986)
 - Flags: 0x00
 - Fragment offset: 0
 - Time to live: 1
 - Protocol: PIM (103)
 - Header checksum: 0x53d6 [correct]
 - Source: 0.0.0.0 (0.0.0.0)
 - Destination: 224.0.0.13 (224.0.0.13)
- Protocol Independent Multicast
 - 0010 = Version: 2
 - 0000 = Type: Hello (0)
 - Reserved byte(s): 00
 - Checksum: 0x572f [correct]
 - PIM options: 4

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.

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

```
dcla-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)

dcla-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 aggregation 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.



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

CCDE, CCENT, Cisco Eos, Cisco Lumin, Cisco Nexus, Cisco StadiumVision, the Cisco logo, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn is a service mark; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, iQuick Study, IronPort, the IronPort logo, LightStream, Linksys, MediaTone, MeetingPlace, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0805R)