L2MP based forwarding across vPC peer–link in Carmel ASIC based switches(Nexus 5548/5596)



Document ID: 115900

Contributed by Prashanth Krishnappa, Cisco TAC Engineer. Jul 03, 2014

Introduction Prerequisites Requirements Components Used Conventions Loop avoidance Related Information

In vPC topologies user traffic will be seen on peer–link only for orphan port traffic or flooded traffic (unknown unicast, broadcast, multicast). For this flood traffic, there is a requirement that switches make sure flood traffic received on one leg of the vPC is not sent back on the other vPC leg so that packets are not sent back towards source or duplicated to other vPCs.

In Carmel based switches (Nexus 55xx), vPC loop avoidance implementation is different compared to Gatos (Nexus 5010/5020) based implementation which uses a separate internal MCT VLAN for flooded traffic across peer–link.

Because Carmel based switches support L2MP or fabricpath, engineering decided to use L2MP based forwarding across the peer–link. With this model, vPC primary switch will have a switch–id of 2748(0xabc) while the vPC secondary will have a switch–id of 2749(0xabd). The Emulated switch–id of 2750(0xabe) will be used as source switch–id for frames which ingress a vPC but sent across the peer–link. All ports on the vPC primary will be members of FTAG 256 while that on the vPC secondary will be members of FTAG 257. In vPC primary switch, only orphan ports will be members of FTAG 256.

Prerequisites

Requirements

There are no specific requirements for this document.

Components Used

This document is not restricted to specific software and hardware versions.

Conventions

Refer to Cisco Technical Tips Conventions for more information on document conventions.

Loop avoidance

For broadcast/unknown unicast/multicast frames coming into vPC primary switch, they will be sent out with a FTAG of 256 across the peer–link. When the vPC secondary switch gets this frame across the vPC peer–link, it inspects the FTAG and since its 256, the vPC secondary switch will only send it out to FTAG 256 members which will be orphan ports only. For flood traffic from vPC secondary, it will be sent with FTAG of 257 and when the vPC primary switch gets this frame, it sends the received flood frame only to members of FTAG 257 which will be orphan ports only. This is how Carmel based switches implement vPC loop avoidance.

In order to deep dive L2MP/FTAG based forwarding of flood frames across peer-link, this topology is used:



switch. Port–channel 1 is the vPC peer–link. The IP addresses shown belong to interface VLAN 1 of the switches. Host 1 and Host 2 are Cisco switches connected via vPC in VLAN 1. These are called host 1 and host 2 in this document. There is orphan port in VLAN 1 connected to Eth1/32 on both switches.

Here is some command output from the switches:

```
N5K-C5596UP-109# show vpc
Legend:
          (*) - local vPC is down, forwarding via vPC peer-link
vPC domain id
                        : 2
Peer status
                       : peer adjacency formed ok
vPC keep-alive status
                       : peer is alive
Configuration consistency status : success
Per-vlan consistency status : success
Type-2 consistency status : success
vPC role
                      : primary
Number of vPCs configured
Peer Gateway
                      : 2
vPC Peer-link status
_____
id Port Status Active vlans
-- ---- -----
1
  Pol up 1
vPC status
_____
id Port Status Consistency Reason
                                          Active vlans
_____ _____
111Poll1upsuccesssuccess200Po200upsuccesssuccess
                                           1
                                            1
N5K-C5596UP-109# show platform fwm info l2mp myswid
switch id
_____
switch id manager
 _____
           _____
  vpc role: 0
  my primary switch id: 2748 (0xabc)
  emu switch id: 2750 (0xabe)
  peer switch id: 2749 (0xabd)
N5K-C5596UP-109# show vpc orphan-ports
Note:
-----::Going through port database. Please be patient.::------
VLAN
         Orphan Ports
_____
          _____
1
          Eth1/32
```

: 2

Peer status vPC keep-alive status Configuration consistency status Per-vlan consistency status Type-2 consistency status vPC role				peer adjacency peer is alive success success success secondary	formed	ok
Number Peer (Peer 9 Dual-a Gracer Auto-r	r of vPCs con Gateway gateway exclu active exclud ful Consisten recovery stat eer-link stat	figured ded VLAN ed VLANs cy Chec us us	: 15 : 5 : C :	2 Enabled - Enabled Disabled		
id 1	Port Status	Active	e vlans			
1 1	 Pol up	1				
vPC st	tatus					
id	Port	Status	Consistency	r Reason		Active vlans
111 200	Po111 Po200	up up	success	success success		1 1

N5K-C5596UP-110# show platform fwm info l2mp myswid

```
switch id
```

vPC domain id

switch id manager

vpc role: 1 my primary switch id: 2749 (0xabd) emu switch id: 2750 (0xabe)

peer switch id: 2748 (0xabc)
N5K-C5596UP-110# show vpc orphan-ports
Note:
-----::Going through port database. Please be patient.::-----

VLAN	Orphan Ports
1	Eth1/32

Now lets check on default FTAGs used and its members.

N5K-C5596UP-109# show platform fwm info l2mp ftag all L2MP FTAG

```
ftag[0x9565b1c] id: 256 (0x100)
Topology ID: 0x111
Ftag flags: 0 (invalid ftag-flags)
Is stale: FALSE
ftag_mask[0x973eca4]
ifindex array:
0x160000c7 0x1600006e 0x1a01f000
0x15010000 0x15020000 0x1600007e
0x16000000
ifmap[0x88400fc]
```

```
ifmap idx 6: ref 1, lu_mcq_alloced 0, lu_mcq 15 (orig 15) 'not pruned'
ifmap idx 6: prune_ifmap 0, prune ref count 0, prune_unvisited 0
ifmap_idx 6: oifls_macg_ref_cnt 0, num_oifls 0
ifmap idx 6: ifs - sup-eth1 sup-eth2 Po200 Po1 Po111 Eth1/32 Po127
rpf: (0x0)
alternate: 0
intf:
Pol (0x1600000)
ftag_ucast_index: 1
ftag_flood_index: 1
ftag_mcast_index: 32
ftag_alt_mcast_index: 48
_____
ftag[0x9565e3c] id: 257 (0x101)
Topology ID: 0x111
Ftag flags: 0 (invalid ftag-flags)
Is stale: FALSE
ftag_mask[0x95612b4]
ifindex array:
0x1a01f000 0x15010000 0x15020000
0x16000000
ifmap[0x883b81c]
ifmap idx 11: ref 1, lu_mcq_alloced 0, lu_mcq 14 (orig 14) 'not pruned'
ifmap idx 11: prune_ifmap 0, prune ref count 0, prune_unvisited 0
ifmap_idx 11: oifls_macg_ref_cnt 0, num_oifls 0
ifmap idx 11: ifs - sup-eth1 sup-eth2 Po1 Eth1/32
rpf: (0x0)
alternate: 1
intf:
Pol (0x1600000)
ftag_ucast_index: 0
ftag_flood_index: -1
ftag_mcast_index: 0
ftag_alt_mcast_index: 0
_____
N5K-C5596UP-109#
N5K-C5596UP-110# show platform fwm info l2mp ftag all
L2MP FTAG
_____
ftag[0x956a99c] id: 256 (0x100)
Topology ID: 0x111
Ftag flags: 0 (invalid ftag-flags)
Is stale: FALSE
ftag_mask[0x98b4764]
ifindex array:
0x16000066 0x1a01f000 0x15010000
0x15020000 0x16000000
ifmap[0x9635adc]
ifmap idx 4: ref 1, lu_mcq_alloced 0, lu_mcq 15 (orig 15) 'not pruned'
ifmap idx 4: prune_ifmap 0, prune ref count 0, prune_unvisited 0
ifmap_idx 4: oifls_macg_ref_cnt 0, num_oifls 0
ifmap idx 4: ifs - sup-eth1 sup-eth2 Po103 Po1 Eth1/32
rpf: (0x0)
alternate: 1
intf:
Po1 (0x1600000)
ftaq_ucast_index: 1
ftag_flood_index: -1
ftag_mcast_index: 32
ftag_alt_mcast_index: 48
_____
ftag[0x956acbc] id: 257 (0x101)
Topology ID: 0x111
Ftag flags: 0 (invalid ftag-flags)
Is stale: FALSE
```

```
ftag_mask[0x97359bc]
ifindex array:
0x160000c7 0x16000066 0x1600006e
0x1a01f000 0x15010000 0x15020000
0x1600007e 0x16000000
ifmap[0x95c624c]
ifmap idx 7: ref 1, lu_mcq_alloced 0, lu_mcq 16 (orig 16) 'not pruned'
ifmap idx 7: prune_ifmap 0, prune ref count 0, prune_unvisited 0
ifmap_idx 7: oifls_macg_ref_cnt 0, num_oifls 0
ifmap idx 7: ifs - sup-eth1 sup-eth2 Po200 Po103 Po1 Po111 Eth1/32 Po127
rpf: (0x0)
alternate: 0
intf:
Pol (0x1600000)
ftag_ucast_index: 0
ftag_flood_index: 1
ftag_mcast_index: 32
ftag_alt_mcast_index: 48
```

Test 1: Broadcast ARP traffic coming into vPC secondary

A non-existent IP 192.168.1.199 is pinged from host 1(192.168.1.101). Due to this, host 1 keeps sending out a broadcast ARP request asking "who is 192.168.1.199". Host 1 happens to hash this broadcast traffic to vPC secondary switch N5K-C5596UP-110, which in turn floods it to all ports in VLAN 1 including Po1 which is the vPC peer-link.

A TX SPAN of Port-channel 1 is captured to look at the fabric path headers of this ARP broadcast which is a multi-destination frame in FP terminology. Look at the fabric path header of this multi-destination frame.

e o o							
Elle Ed	Wew Go Cepture Analyze Statistics	Telephony Tools Internals (telp				
₩₩₩₩₩ E================================							
Filter Expression Cloar Apply Save							
No.	Time	Source	Destination	Protocol	Longth	Identification	into
22 8	1 2012-10-31 15:26:29.574603360	Ciaco_Ofib3:01	Broadcast	ARP	94		Who has 192.160.1.1997 Tell 192.168.1.101
	2 2012-10-31 15+26+46-578376630	Cisco_Of (b3:01	Broadcast	18.2	84		Who has 192.168.1.1997 Tell 192.168.1.101
	3 2012-10-31 15:26:48.577569140	Cisco_Of:b3:01	Broadcast	A8.7	04		Who has 192.160.1.1997 Tell 192.168.1.101
	4 2012-10-31 15+26+52.577605320	Cisco_0f(b3:01	Broadcast	A8.P	84		Who has 192.168.1.1997 Tell 192.168.1.101
	5 2012-10-31 15:27:00.577670840	Cisco_Of:b3:01	Broadcast	13.7	84		Who has 192.168.1.199? Tell 192.168.1.101



- Because the frame ingresses via a vPC(vPC 111), source switch-id is abe.00.0000.
- Destination is a broadcast MAC FF:FF:FF:FF:FF:FF
- FTAG is 257.

When this frame comes into the vPC primary switch, it will inspect the FTAG 257. Because only orphan ports are members of FTAG 257, this broadcast ARP frame will only be sent to Eth 1/32.

Test 2: Unknown unicast frame coming into vPC secondary

In order to introduce unknown unicast traffic, on host 1, I set up a static ARP for 192.168.1.99 with a static MAC of 0001.0002.0003 and do a ping to 192.168.1.99. The ICMP echo request arrives at N5K-C5596UP-110 and because it does not know where MAC 0001.0002.0003 is, it floods this frame in the VLAN including peer-link.

A TX SPAN of Port–channel 1 is captured to look at the fabric path headers of this unknown unicast flood frame, which is a multi–destination frame in FP terminology. Look at the fabric path header of this multi–destination frame.

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help								
	8	(🖻 🖾 🗶 💐 🗐	6 S. 🔅 e	🖗 🤏 ዥ 🚽		ର୍ ଖ୍ 🖭 🎆 🖾		
Filter:			-	Expression	Clear Apply Save			
No.	Time		Source		Destination	Protocol Le		
	1 2012-10-31	16:18:20.000000000	192.168.1.10)1	192.168.1.99	ICMP		
	2 2012-10-31	16:18:21.000396870	192.168.1.10	11	192.168.1.99	ICMP		
	3 2012-10-31	16:18:22.000788810	192.168.1.10)1	192.168.1.99	ICMP		
	4 2012-10-31	16:18:23.001732900	192.168.1.10	11	192.168.1.99	ICMP		
4								
▶ Frame	1: 122 bytes	on wire (976 bits), 1	22 bytes capt	ured (976 bits)			
❤ Cisco	FabricPath,	Src: abe.00.0000 Dst (01:bb:cc:dd:	01:01 01:bb:c	c:dd:01:01)			
MC	Destination:	01:bb:cc:dd:01:01 (01:	bb:cc:dd:01:0	1)				
⊽ Sou	rce: abe.00.0	000						
0	080 08 08	= End	Node ID: 0 (0	x800088)				
· ·	1	= U/L	bit: Locally	administered a	ddress (this is N	NOT the factory default)		
·	0	= I/G	bit: Individu	al address (un	icast)			
· ·	6) = 000/	DL Bit: Deliv	er in order (I	T DA) or Learn (1	T SA)		
	ande an and de mine de sinder	1010 1011 1110 = swite	ch-1d: 2/50 (uxuuuabe)				
	ourse LTD: D	(0x00)						
010	ource LID: 0	ETAC: 257						
010	10 0000 01	1000 = TTL 32						
 ⊽ Ether	net II Spo-	Cisco Of:b3:01 (54:7f:	ee.of.b3.01)	Det · FauinTra	02-00-03 (00-01-	00.02.00.03)		
✓ Des	tination: Equ	(ipTra_02:00:03_(00:01:	00:02:00:03)	Dati Edothilo	_02.00.00 (00.01.	00.02.00.000		
A	ddress: Equip	Tra 02:00:03 (00:01:00	:02:00:03)					
1 .		= LG b	it: Globally	unique address	(factory default			
	0	= IG b	it: Individua	l address (uni	cast)			
▼ Sou	▼ Source: Cisco Df:b3:01 (54:7f:ee:Df:b3:01)							
A	Address: Cisco 0f:b3:01 (54:7f:ee:0f:b3:01)							
.								
.								
Тур	e: 802.10 Vir	tual LAN (0x8100)						
▼ 802.1	▼ 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 1							
000 = Priority: Best Effort (default) (0)								
0 = CFI: Canonical (0)								
	. 0000 0000 0	001 = ID: 1						
Тур	e: IP (0x0800	<i>n</i>						
Totor	iter: DI36004	Wareign 4 Eng. 163-16	8 1 101 (100	168 1 1011 D-	+. 107 169 1 00 /	107 169 1 001		
Ver Ver	sion: 4	version 4, Src: 192.16	8.1.101 [192.	166.I.10IJ, DS	t: 192.168.1.99 (192,168,1,99]		
0000	1 bb cc dd 01	01 02 Ba ba 00 00 00	89 83 48 68		@`	******		
0010 00	0 01 00 02 00	03 54 7f ee 0f b3 01	81 00 00 01	T				
0020 08	8 00 45 00 00	54 93 71 00 00 ff 01	a4 1e c0 a8	ET.q				
0030 000	1 65 cD a8 D1	63 08 00 ee 5a b3 1a	71 01 6d 87	.ecZ.	.q.m.			
⊖ M Cl	sco FabricPath	(cfp), 16 bytes Packe	ts: 4 Dis	ofile: Default				

- Since the frame ingresses via a vPC(vPC 111), source switch-id is abe.00.0000
- Destination is a multicast MAC 01:bb:cc:dd:01:01
- FTAG is 257.

When this frame comes into the vPC primary switch, it will inspect the FTAG 257. Because only orphan ports are members of FTAG 257, this vPC primary will flood this frame only to orphan port Eth 1/32.

Due to the above mechanism, the following is the flow for the flooded traffic coming into the vPC secondary



Test 3: Broadcast ARP traffic coming into vPC Primary

A non-existent IP 192.168.1.200 is pinged from host 2(192.168.1.69). Due to this, host 2 keeps sending out a broadcast ARP request asking "who is 192.168.1.200". Host 2 happens to hash this broadcast traffic to vPC Primary switch N5K-C5596UP-109, which in turn floods it to all ports in VLAN 1 including Po1 which is the vPC peer-link.

A TX SPAN of Port-channel 1 is captured to look at the fabric path headers of this ARP broadcast which is a multi-destination frame in FP terminology. Look at the fabric path header of this multi-destination frame.

Elle Edit View Go Capture Analyze Statistics Telephony Tools Internals Help							
🗟 🖬 🗟 🎯 🚳 🗆 🖂 🗶 🍣	🖴 🔍 🗇 🌳 😜	春 🛃 🗐 📑 I 🔍 C	\ @, ≌				
Filter:	▼ Express	ion Clear Apply Save					
No. Time	Source	Destination	Protocol L				
1 2012-10-31 13:53:20.000000000	Cisco_48:4c:00	Broadcast	ARP				
2 2012-10-31 13:53:22.000140560	Cisco_48:4c:00	Broadcast	ARP				
3 2012-10-31 13:53:23.999955470	Cisco_48:4c:00	Broadcast	ARP				
4 2012-10-31 13:53:25.999978340	Cisco_48:4c:00	Broadcast	ARP				
5 2012-10-31 13:53:28.000098460	Cisco_48:4c:00	Broadcast	ARP				
6 2012-10-31 13:53:29.999967990	Cisco_48:4c:00	Broadcast	ARP				
7 2012-10-31 13:53:32.000172270	Cisco_48:4c:00	Broadcast	ARP				
8 2012-10-31 13:53:34.000140460	Cisco_48:4c:00	Broadcast	ARP				
9 2012-10-31 13:53:36.000116550	Cisco_48:4c:00	Broadcast	ARP				
10 2012-10-31 13:53:38.000081040	Cisco_48:4c:00	Broadcast	ARP				
11 2012-10-31 13:53:40.000048330	Cisco_48:4c:00	Broadcast	ARP				
 Frame 1: 84 bytes on wire (672 bits), Cieco EsbricRath, Src. abs 00 0000 per 	84 bytes captured (672	bits)					
<pre>MC Destination: Broadcast (ff:ff:ff:ff:ff:ff) * Source: abe.00.0000 0000 00 00 = End Node ID: 0 (0x000000)1</pre>							
0000 ff ff ff ff ff ff ff c2 0a be 00 00 08 90 340 20							

- Since the frame ingresses via a vPC(vPC 200), source switch-id is abe.00.0000
- Destination is a broadcast MAC FF:FF:FF:FF:FF
- FTAG is 256.

When this frame comes into the vPC secondary switch, it will inspect the FTAG 256. Because only orphan ports are members of FTAG 256, this broadcast ARP frame will only be sent to Eth 1/32.

Test 4: Unknown unicast frame coming into vPC Primary

In order to introduce unknown unicast traffic, on host 2, a static ARP for 192.168.1.200 is set up with a static MAC of 0003.0004.0005 and 192.168.1.200 is pinged. The ICMP echo request hashes to vPC primary N5K–C5596UP–109 and because it does not know where MAC 0003.0004.0005 is, it floods this frame in the VLAN including peer–link. A TX SPAN of Port–channel 1 is captured to look at the fabric path headers of this unknown unicast flood frame which is a multi–destination frame in FP terminology. Look at the fabric path header of this multi–destination frame.



- Since the frame ingresses via a vPC(vPC 200), source switch-id is abe.00.0000
- Destination is a multicast MAC 01:bb:cc:dd:01:01 which is used for unknown unicast flooding
 FTAG is 256.

When this frame comes into the vPC secondary switch, it will inspect the FTAG 257. Because only orphan ports are members of FTAG 256, this vPC primary will flood this frame only to orphan port Eth 1/32.

Due to the above mechanism, the following is the flow for the flooded traffic coming into the vPC Primary switch.



Related Information

• Technical Support & Documentation – Cisco Systems

Contacts & Feedback | Help | Site Map

© 2014 – 2015 Cisco Systems, Inc. All rights reserved. Terms & Conditions | Privacy Statement | Cookie Policy | Trademarks of Cisco Systems, Inc.

Updated: Jul 03, 2014

Document ID: 115900