

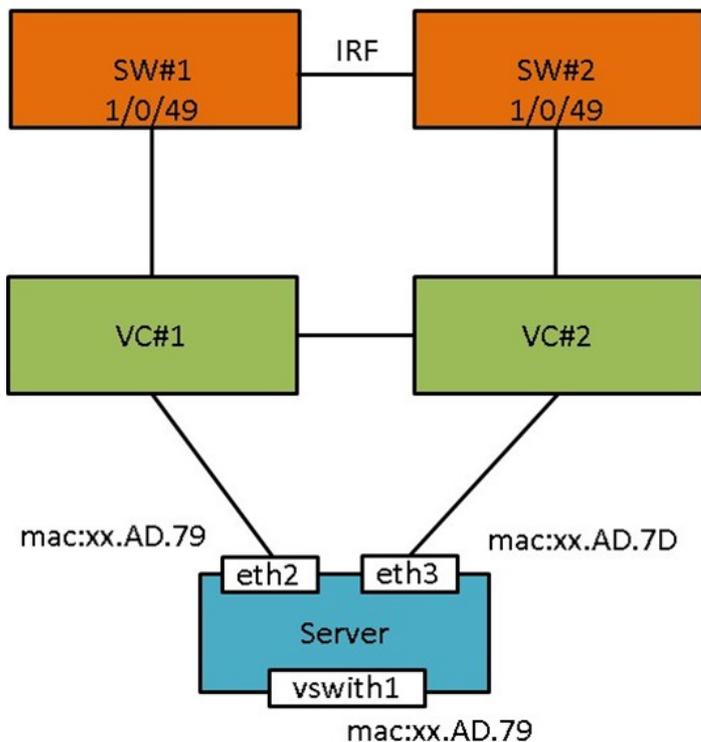
主机网卡开启LLDP功能导致无法正常ping通虚拟交换机地址问题

陈明槐 2016-04-20 发表

某局点的CAS平台使用了UIS刀片服务器，并且使用了VC设备作为网络互联模板，现场组网如下，服务器的eth2和eth3分别连接VC#1和VC#2，VC#1和VC#2分别连接SW#1和SW#2，并且VC#1和VC#2通过内部口互联，SW#1和SW#2做IRF堆叠。

虚拟交换机vswitch1的IP地址为192.168.12.19，它的网关为192.168.12.254，该网关配置在交换机上。

现场发现从服务器（192.168.12.19）ping交换机（192.168.12.254）可以正常ping通，但是从交换机（192.168.12.254）ping服务器（192.168.12.19）出现时通时不通的现象。



步骤1: 登录CAS云计算管理平台，查看主机的虚拟交换机配置，确认虚拟交换机vswitch1的物理接口使用了eth2和eth3，链路聚合模式为静态链路聚合，负载分担模式为主备负载分担。



步骤2: 登录CVK主机后台执行ifconfig命令，查看网卡eth2的mac地址为“FC:15:B4:1C:AD:79”，网卡eth3的mac地址为“FC:15:B4:1C:AD:7D”。内核端口vswitch1的mac地址为“FC:15:B4:1C:AD:79”，说明CVK对外发送的报文会被封装mac地址“FC:15:B4:1C:AD:79”。

```
root@HZ-CAS02-CV19:~# ifconfig
```

```

eth3 Link encap:Ethernet HWaddr FC:15:B4:1C:AD:7D

inet6 addr: fe80::fe15:b4ff:fe1c:ad7d/64 Scope:Link

UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1

RX packets:11079145 errors:0 dropped:0 overruns:0 frame:0

TX packets:441 errors:0 dropped:0 overruns:0 carrier:0

collisions:0 txqueuelen:1000

RX bytes:885930719 (844.8 Mb) TX bytes:28000 (27.3 Kb)

eth2 Link encap:Ethernet HWaddr FC:15:B4:1C:AD:79

inet6 addr: fe80::fe15:b4ff:fe1c:ad79/64 Scope:Link

UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1

RX packets:11107927 errors:0 dropped:0 overruns:0 frame:0

TX packets:438 errors:0 dropped:0 overruns:0 carrier:0

collisions:0 txqueuelen:1000

RX bytes:888236637 (847.0 Mb) TX bytes:27752 (27.1 Kb)

vswitch1 Link encap:Ethernet HWaddr FC:15:B4:1C:AD:79

inet addr:192.168.12.19 Bcast:192.168.12.255 Mask:255.255.255.0

inet6 addr: fe80::fe15:b4ff:fe1c:ad79/64 Scope:Link

UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1

RX packets:28322 errors:0 dropped:0 overruns:0 frame:0

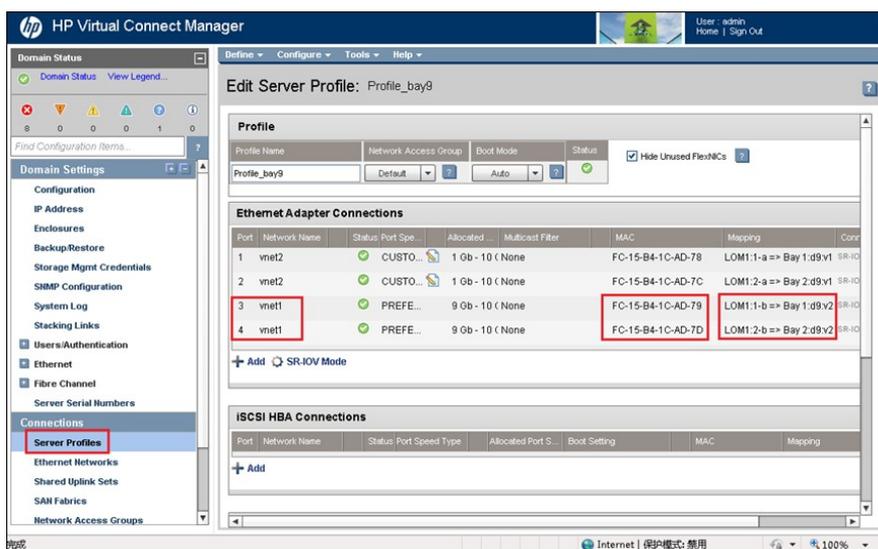
TX packets:45 errors:0 dropped:0 overruns:0 carrier:0

collisions:0 txqueuelen:0

RX bytes:1363398 (1.3 Mb) TX bytes:3994 (3.9 Kb)

```

步骤3: 登录VC的Server Profile配置页面，查看到mac地址“FC:15:B4:1C:AD:79”映射的VC为Bay1（即VC#1），mac地址“FC:15:B4:1C:AD:7D”映射的VC为Bay2（即VC#2），并且都使用了vnet1



步骤4: 结合上述的配置和现场组网，当CVK的虚拟交换机vswitch1的报文分为如下两种情况:

网卡eth2为主时:

CVK主机 (192.168.12.19) 到网关 (192.168.12.254) 的报文的流程为: “vswitch1->eth2->VC#1->SW#1”。

vswitch1的mac地址“FC:15:B4:1C:AD:79”会记录在VC#1的下行口和SW#1的下行口。

网关 (192.168.12.254) 到CVK主机 (192.168.12.19) 的报文会根据SW#1和VC#1的mac地址表项转发, 流程为: “SW#1->VC#1->eth2->vswitch1”。

网卡eth3为主时:

CVK主机 (192.168.12.19) 到网关 (192.168.12.254) 的报文的流程为: “vswitch1->eth3->VC#2->VC#1->SW#1”。

vswitch1的mac地址“FC:15:B4:1C:AD:79”会记录在VC#2的下行口、VC#1的互联口和SW#1的下行口。

网关 (192.168.12.254) 到CVK主机 (192.168.12.19) 的报文会根据SW#1、VC#1和VC#2的mac地址表项转发, 流程为: “SW#1->VC#1->VC#2->eth3->vswitch1”。

步骤5: 经过现场的进一步测试发现: 当eth2网卡为主时CVK主机 (192.168.12.19) ping网关 (192.168.12.254) 和网关 (192.168.12.254) ping CVK主机 (192.168.12.19) 都是通的。

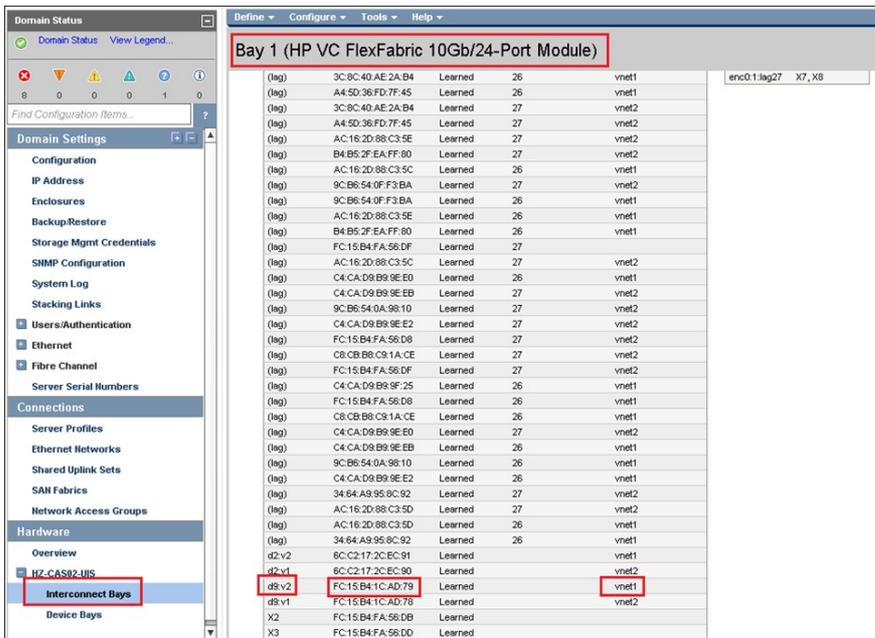
而当eth3网卡为主时CVK主机 (192.168.12.19) ping网关 (192.168.12.254) 是通的, 网关 (192.168.12.254) ping CVK主机 (192.168.12.19) 时通时不通。

下图表示网卡eth3为主。

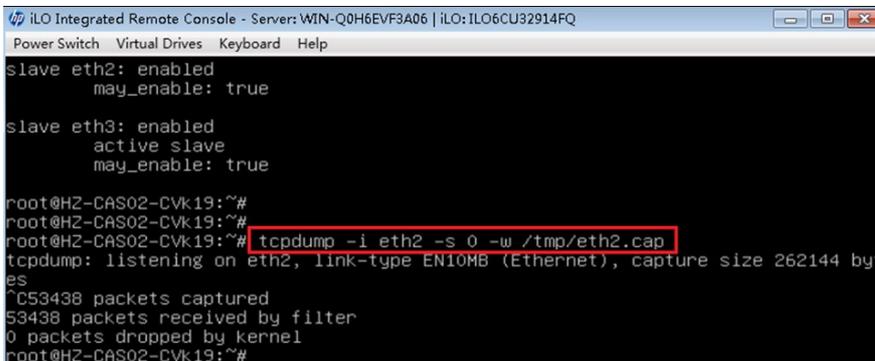
```
root@HZ-CAS02-CVK19:~# ovs-appctl bond/show
---- vswitch1_bond ----
bond_mode: active-backup
bond may use recirculation: no, Recirc-ID : -1
bond-hash-basis: 0
updelay: 0 ms
downdelay: 0 ms
lacp_status: off
active slave mac: fc:15:b4:1c:ad:7d(eth3)
slave eth2: enabled
may_enable: true
slave eth3: enabled
active slave
may_enable: true
```

步骤6: 登录VC的配置页面, 查看Bay1 (VC#1) 的mac地址表项, 发现当网卡eth3为主时VC#1记录了从d9:v2 (即第9槽位的刀片服务器) 获取的mac地址“FC:15:B4:1C:AD:79”信息

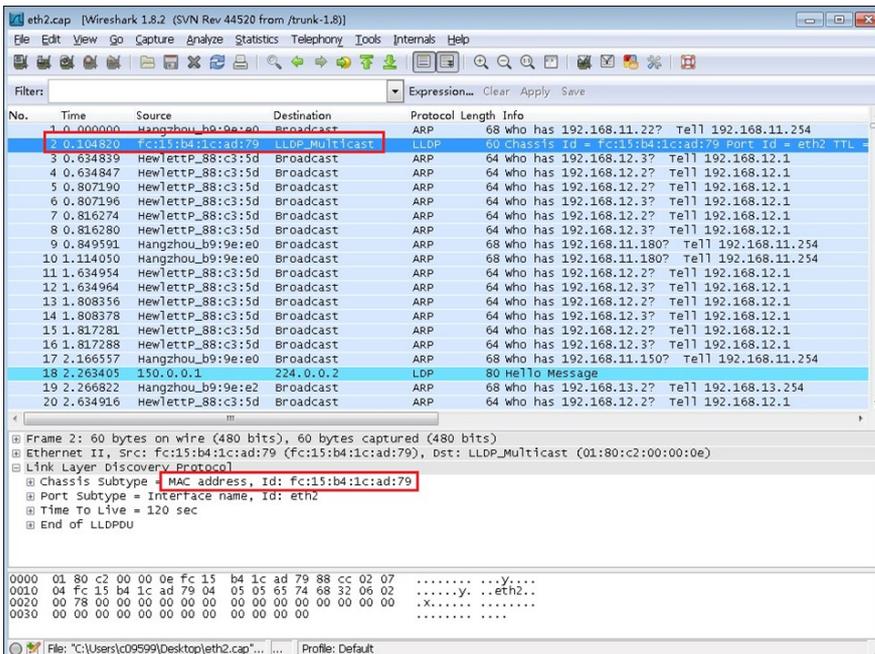
而按照步骤4中的分析, 当网卡eth3为主时VC#1中仅在互联口有 “FC:15:B4:1C:AD:79”的mac地址信息。



步骤7: 执行tcpdump命令对网卡eth2进行抓包, 如下图所示。



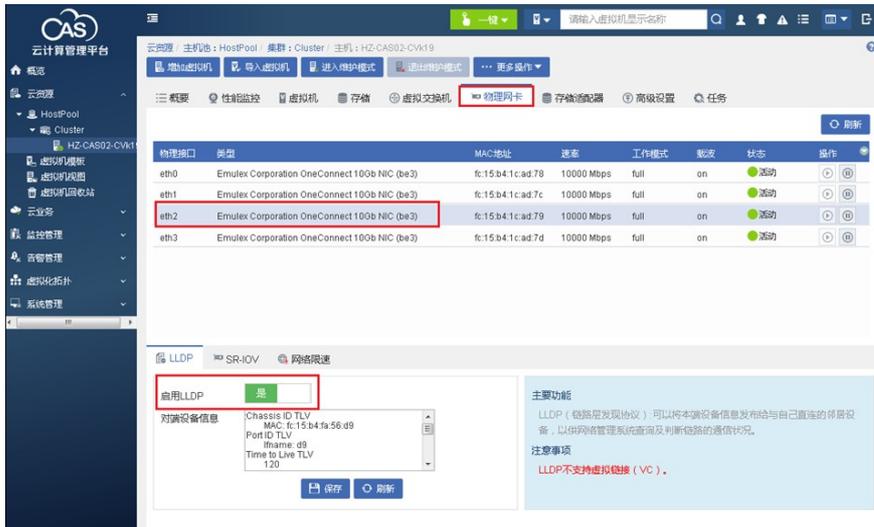
使用Wireshark工具分析报文发现, 虽然网卡eth2状态为备, 但是它在不不停的发生LLDP_multicast报文。



因此, 问题原因可以基本确定: 虚拟交换机vswitch (mac: FC:15:B4:1C:AD:79) 的负载分担模式为主备负载分担时, 网卡eth2 (mac也是 FC:15:B4:1C:AD:79) 会定时发送LLDP_multicast, 从而影响VC#1上的mac地址表项信息。导致VC#1收到SW#1发送给vswitch1的报文时, 没有通过VC#1的互联口转发给VC#2, 而是直接下发给eth2网卡, 此时eth2的状态为备用状态, 不处理报文, 直接丢弃VC#1转发的报文, 导致ping不通。

那么，接下来的问题是：网卡eth2为什么会定时发送LLDP_multicast报文呢？

步骤8：登录CAS云计算管理平台，发现网卡eth2开启了LLDP功能。LLDP（链路层发现协议）的功能可以将本端设备信息发布给与自己直连的邻居设备，以供网络管理系统查询及判断链路的通信状况。



在CAS云计算管理平台关闭网卡的LLDP功能，问题解决。