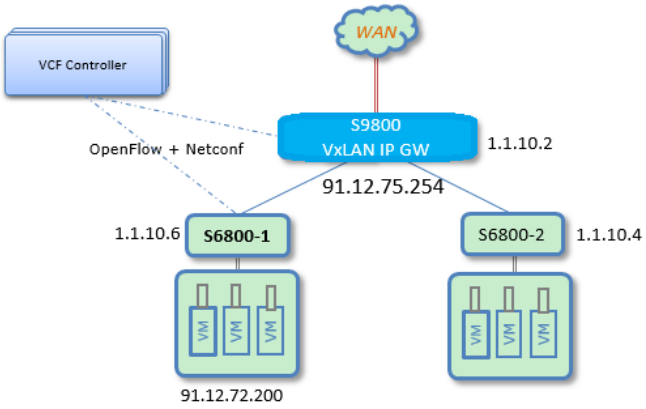
SDN网络VCFC强控方案下S9800部分流量三层转发不通问题分析

EVPN 转发不通 程飞 2017-04-21 发表

现场SDN网络，使用的是VCFC强控方案，即VCFC下发流表指导交换机转发，S6800为L2VTEP，S9800为L3VTEP网关。现场突然发现VLAN 72里面的流量访问网关不通，之前该VLAN的流量能正常访问。
拓扑图如下：



S6800-1下一台虚机91.12.72.200不能ping通网关S9800(91.12.75.254)，它的MAC是3c8c-404e-dd46。将虚机迁移到另外一台S6800后，可以ping通这台S9800。

1．由于现场反馈业务是突然不通，因此进行流量统计，首先在S6800-1上进行统计，发现在S6800-1的AC口收到了报文。由于S6800出口不支持对封装为VXLAN的报文进行流量统计，因此在S9800入方向进行流量统计：
[S6800-1]dis acl 3000
Advanced ACL 3000, named -none-, 4 rules,
ACL&＃39;s step is 5
rule 3 permit ip source 91.12.72.200 0 destination 91.12.75.254 0
rule 4 permit ip source 91.12.75.254 0 destination 91.12.72.200 0
S6800-1入方向收到报文：
<S6800-1>dis qos policy interface
Interface: Ten-GigabitEthernet1/1/24
Direction: Inbound
Policy: test
Classifier: test
Operator: AND
Rule(s) :
If-match acl 3000
Behavior: test
Accounting enable:
9 (Packets)
Interface: Ten-GigabitEthernet2/1/24
Direction: Inbound
Policy: test
Classifier: test
Operator: AND
Rule(s) :
If-match acl 3000
Behavior: test
Accounting enable:
9 (Packets)

S9800上通过匹配VXLAN内层IP：
[S9810]dis acl 3001
Advanced ACL 3001, named -none-, 2 rules,
ACL&＃39;s step is 5

rule 0 permit vxlan source 91.12.75.254 0 destination 91.12.72.200 0 inner-protocol ip inner-source 9
1.12.75.254 0 inner-destination 91.12.72.200 0
rule 1 permit vxlan source 91.12.72.200 0 destination 91.12.75.254 0 inner-protocol ip inner-source 9
1.12.72.200 0 inner-destination 91.12.75.254 0

S9800上入方向没有收到：
[S9810]dis qos policy interface
Interface: FortyGigE1/1/0/4
Direction: Inbound
Policy: test
Classifier: test
Operator: AND
Rule(s) :
If-match acl 3001
Behavior: test
Accounting enable:
0 (Packets)
Interface: FortyGigE2/1/0/4
Direction: Inbound
Policy: test
Classifier: test
Operator: AND
Rule(s) :
If-match acl 3001
Behavior: test
Accounting enable:
0 (Packets)

2．根据流量统计结果，怀疑是S6800-1的问题，因此查看S6800-1到网关的mac表项：

==========display l2vpn mac-address==========
MAC Address State VSI Name Link ID/Name Aging
0050-5691-b7b1 Dynamic SDN_VSI_168 XGE1/1/3 Aging
0050-5691-c0f0 Dynamic SDN_VSI_168 XGE1/1/2 Aging
3c8c-4003-c4b2 Dynamic SDN_VSI_168 Tunnel257 Aging
3c8c-404e-dd46 Openflow SDN_VSI_168 Tunnel257 NotAging
3c8c-404e-dd46 Openflow SDN_VSI_72 Tunnel258 NotAging

如上l2vpn mac表，可以看到在VSI SDN_VSI_72里（这个VSI对应VLAN72），S6800-1到网关S9800（3c8c-404e-dd46）的出接口是tunnel258，该隧道的目的IP地址是1.1.10.4，这个目的地址并不是S9800的VTEP地址，而是另外一台S6800-2的VTEP地址。而其他正常vlan的业务到网关的隧道出口是tunnel257，是正常的S9800的VTEP的IP（1.1.10.2）
Tunnel258
Current state: UP
Line protocol state: UP
Description: Tunnel258 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: 10:46:16 Wed 04/05/2017
Tunnel source 1.1.10.6, destination 1.1.10.4 //这个地址是S6800-2的VTEP IP
Tunnel protocol/transport UDP_VXLAN/IP

Tunnel257
Current state: UP
Line protocol state: UP
Description: Tunnel257 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: 10:46:16 Wed 04/05/2017
Tunnel source 1.1.10.6, destination 1.1.10.2 //这个地址是S9800的VTEP IP
Tunnel protocol/transport UDP_VXLAN/IP

3．l2vpn mac地址表是控制器下发生成的，因此怀疑是控制器下发错误导致，查看控制器下发的流表来确定，发现确实是控制器下发的到网关的出接口隧道为tunnel258.

```
<S6800-1>dis open ins 1 flow-table
Instance 1 flow table information:
Table 0 information:
 Table type: MAC-IP, flow entry count: 139, total flow entry count: 139
……
Flow entry 162 information:
 COOKIE: 0x4c324757415057, priority: 29999, hard time: 0, idle time: 0, flags:
 check_overlap, byte count: --, packet count: --
Match information:
 Ethernet destination MAC address: 3c8c-404e-dd46
 Ethernet destination MAC address mask: ffff-ffff-ffff
 Experimenter:
  Address ID: 72
Instruction information:
 Write actions:
  Output interface: Tun258
  Set field:
   Tunnel ID: 72
……
```

4．查看VCFC的日志，发现环境中出现了一台虚机MAC地址为网关MAC地址3c:8c:40:4e:dd:46，这个虚机也位于VLAN72，并且这个虚机从1.1.10.4这台S6800-2上线，看日志打印为91.12.73.39这台虚机的MAC从原来正常MAC修改成了网关保留MAC地址：

```
[2017-04-03 13:47:17.321] INFO  e8-335b7551ae44-212-thread-1  [NEM][FwdDevice][processMessa
ge][1.1.10.5] MSG_MODIFY_VPORT, old:
  Vm IP Address is:[91.12.73.39]
  Vm Mac Address is:[00:50:56:bc:f3:3b]
  Vm Vni:[72]
  Vlan Id is:[72]
  Host IP Address is:[1.1.10.3]
  DataPath Id is:[00:01:60:0b:03:8a:b3:b6]
  Openflow Port is:[0xf]
  Network Uuid is:[3369dd20-4c88-4b2b-8c57-9a3b47141bed]
  Subnet Uuid is:[2a2aefa4-80bc-485b-83d9-9199aa79c45d]
  VRoute Uuid is:[cf3fa3ce-42d5-40d3-95a7-dbb79fbc9df6]
  bExternal is:[false]
  vtepIPAddr is:[1.1.10.4]
  tenantid is:[b931d808-cedd-4e0d-ad47-5a0b94db218d]
  Hash Value is:[925378128]
  Port Uuid is:[ffffff00-0048-5b0c-4927-ffffffffffff]
  virIpMacAddrList:[
    virIpMacAddr: 91.12.73.3900:50:56:bc:f3:3b
  ], new:
  Vm IP Address is:[91.12.73.39]
  Vm Mac Address is:[3c:8c:40:4e:dd:46] //mac被修改成了和S9800网关mac相同
  Vm Vni:[72]
  Vlan Id is:[72]
  Host IP Address is:[1.1.10.3]
  DataPath Id is:[00:01:60:0b:03:8a:b3:b6]
  Openflow Port is:[0xf]
  Network Uuid is:[3369dd20-4c88-4b2b-8c57-9a3b47141bed]
  Subnet Uuid is:[2a2aefa4-80bc-485b-83d9-9199aa79c45d]
  VRoute Uuid is:[cf3fa3ce-42d5-40d3-95a7-dbb79fbc9df6]
  bExternal is:[false]
  vtepIPAddr is:[1.1.10.4]
  tenantid is:[b931d808-cedd-4e0d-ad47-5a0b94db218d]
  Hash Value is:[925378128]
  Port Uuid is:[ffffff00-0048-5b0c-4927-ffffffffffff]
  virIpMacAddrList:[
    virIpMacAddr: 91.12.73.393c:8c:40:4e:dd:46 ].
```

所以VCFC给1.1.10.6这台S6800-1下发了MAC流表，出接口为tunnel258，而tunnel258的目的IP为1.1.10.4的这台S6800-2交换机。

5．查看VCFC诊断里异常的ARP从S6800-2的 XGE2/1/1接口上来的，在这个接口下找到这个异常虚机。

UuidFromDBPort: ffffff00-0048-c0a8-4903-ffffffffffff
VNI: 72
startMigrateTimeStamp: 2108882011995198
migrateTimes: 0
DataPath-ID: 00:01:60:0b:03:8a:b3:b6
VLAN-ID: 72
IP: 91.12.73.39
MAC: 3c:8c:40:4e:dd:46
PortNumber: ArpPort [port=0xce, groupId=null, type=TYPE_PORT]
PortName: XGE2/1/1
TimeToLive: 1

6、在S9800上流量统计不到的原因是，该报文的目的IP是S6800-2的地址，因此在S9800上还是VXLAN报文，没有解封装，因此流量统计不到。

虚机MAC不能使用网关MAC，需要将异常虚机的MAC修改为原有MAC。

对于OpenFlow交换机，服务链、安全策略、将报文上送控制器的流表、带内网管这些功能的OpenFlow流表都需要底层ACL表来实现，而其他的普通转发流表，是通过MAC-IP表实现，即三层转发采用ARP表，二层转发采用MAC表。平台下发OpenFlow流表后，根据流表类型下发成MAC-IP表或者ACL表。