

交换那些事儿|M-LAG技术专题新系列

M-LAG典型案例及FAQ

每期推送2个案例+4个FAQ

让你在遇到类似问题时不迷茫不发愁

解决方案手到擒来~

往期内容可点击上方话题复习

本期内容

典型案例:

- **Case1 M-LAG+EVPN组网leaf单挂终端无法访问外网**
- **Case2 M-LAG组网下挂终端ping设备上地址不通**

FAQ:

- **Q1 M-LAG+VRRP场景下流量负载分担?**
- **Q2 M-LAG单挂设备有没有特殊配置要求?**
- **Q3 M-LAG成员接口物理down后是否会影响M-LAG group的状态?**
- **Q4 M-LAG+EVPN组网中，border设备配置group ip后为何不以虚地址发布路由?**

DRNI更名为M-LAG

原DR口>>M-LAG接口

原DR group>>M-LAG group

原IP链路>>peer-link链路

原IPP口>>peer-link接口

本文涉及命令行暂不做修改

Case1

M-LAG+EVPN组网leaf单挂终端无法访问外网

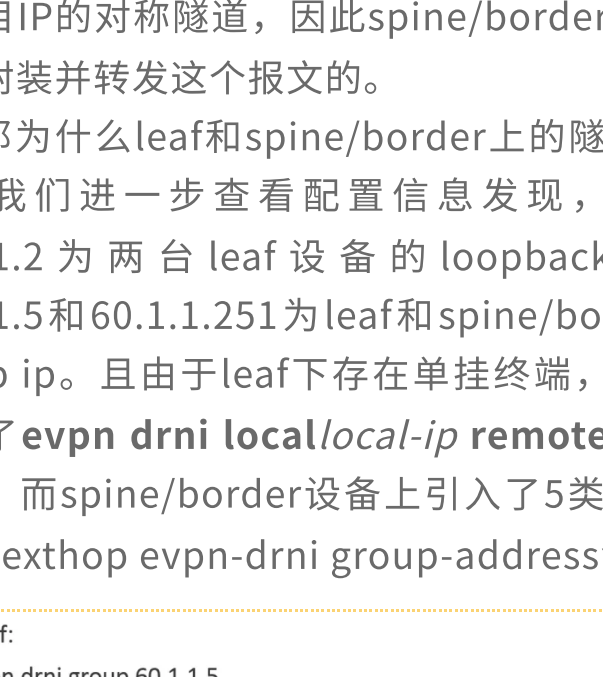
组 / 网 / 说 / 明

设备及版本:

S12508X-AF R2820

S6860-54HF F2707

组网: 两台S6860起M-LAG作为一组leaf设备，上行两台S12508X-AF同样起M-LAG作为spine/border设备，leaf与spine/border之间建立隧道，leaf设备下的终端全单挂。下图为局部组网图。



问 / 题 / 描 / 述

现场M-LAG+EVPN组网，两台S6860起M-LAG作为leaf，远端两台12508X-AF起M-LAG作为Spine/border设备，测试发现leaf下单挂的终端无法访问外网，leaf设备下单挂的arp、mac表项学习正常。在终端ping外网时抓包，ICMP报文从leaf设备发出，spine/border设备也收到了ICMP报文，但并未上行继续向外发出，可以确认丢包发生在spine/border上。

过 / 程 / 分 / 析

spine/border未将报文继续转发出去，我们首先怀疑从leaf发送过来的报文是否封装正确，从抓包来看，报文是正常的，VXLAN外层源目IP为隧道源目IP，内层为原始报文源目IP:



既然报文没有问题，那为什么spine/border没有正常解封封装报文并转发呢？在“[交换那些事儿 | 技术专题篇 - DRNI+EVPN组网进阶介绍](#)”中我们讲解过，设备收到带封装的报文后，如果其封装的源目IP在设备上可以找到对称隧道，就可以解封封装，反之则无法正常解封封装并转发。我们在leaf上查看到spine/border的隧道只有如下一条，抓到的报文也是封装的这个隧道的源目IP:

source 60.1.1.5, destination 60.1.1.251

抓到的报文也是封装的这个隧道的源目IP;

在spine/border上查看到leaf的隧道有两条:

source 60.1.1.251, destination 60.1.1.1

source 60.1.1.251, destination 60.1.1.2

很明显spine/border上这两条隧道都不是报文封装源目IP的对称隧道，因此spine/border就是无法正常解封封装并转发这个报文的。

那为什么leaf和spine/border上的隧道会不对称呢？我们进一步查看配置信息发现，60.1.1.1和60.1.1.2为两台leaf设备的loopback 0地址，60.1.1.5和60.1.1.251为leaf和spine/border设备的group ip。且由于leaf下存在单挂终端，所以leaf上配置了evpn drni local/local-ip remote remote-ip命令；而spine/border设备上引入了5类路由，也配置了nexthop evpn-drni group-address命令:

```
Leaf:
evpn drni group 60.1.1.5
evpn drni local 60.1.1.2 remote 60.1.1.1

spine/border:
evpn drni group 60.1.1.251
nexthop evpn-drni group-address
```

在这种配置下，由于leaf全单挂，leaf设备会以group ip发布3类路由，以local ip发布2类路由，spine/border只会通过group ip发布引入的5类路由。leaf收到spine/border发布的5类路由后，根据路由下一跳建立虚地址隧道source 60.1.1.5, destination 60.1.1.251; spine/border上会收到leaf发布的3类和2类路由，但由于spine/border上没有vsi配置，无法接收3类路由，因此只能通过2类路由的下一跳建立到leaf实地址的隧道。



从上面的分析可以看出，在现场这种组网配置下，leaf和spine/border之间的隧道无法对称，spine/border也就无法正常解封封装报文，导致了现场的故障现象。那么应该如何解决呢？

解 / 决 / 方 / 法

在spine/border上配置vxlan default-decapsulation source interface命令强制解封封装即可解决问题。配置该命令后，spine/border收到带VXLAN封装的报文后会直接强制解封封装，再走转发流程正常转发。

M-LAG+EVPN隧道建立机制可参考:

[交换那些事儿 | 技术专题篇 - DRNI+EVPN组网及配置介绍](#)

[交换那些事儿 | 技术专题篇 - DRNI+EVPN组网进阶介绍](#)

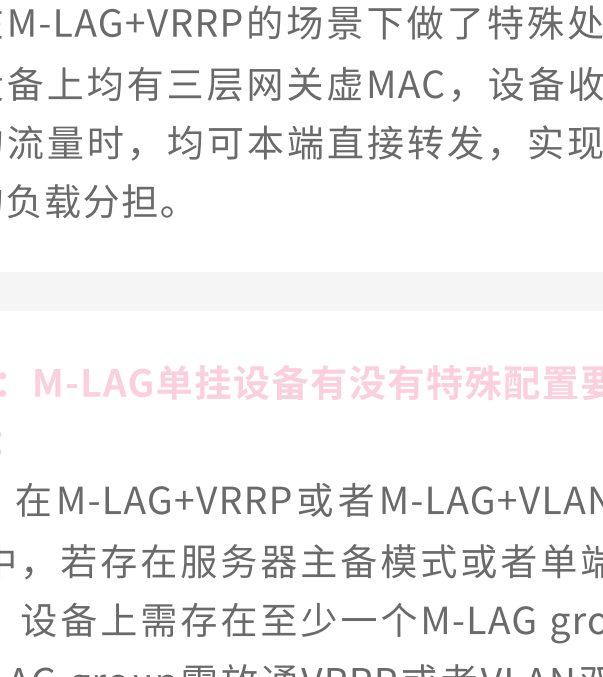
Case2

M-LAG组网下挂终端ping设备上地址不通

组 / 网 / 说 / 明

设备及版本: S10508 R7596P09

组网: 两台S10508起M-LAG，没有配置VLAN双活或者VRRP网关，单台M-LAG设备上有终端对应的VLAN，M-LAG接口对接下行设备。下图为局部组网图。



问 / 题 / 描 / 述

现场M-LAG组网，M-LAG设备上未起VLAN双活或者VRRP网关，设备下行通过M-LAG接口BAGG6对接一台接入设备，BAGG6放通VLAN4019。其中只有M-LAG-1设备上有interface vlan-interface 4019地址，地址为10.0.251.5，M-LAG-2设备上未起VLAN4019的三层虚接口，接入设备上int vlan 4019的地址为10.0.251.6，现场测试从接入设备上ping 10.0.251.5不通。

过 / 程 / 分 / 析

根据往期内容我们知道，现场这种配置是不规范的，那为什么这种不规范的配置会导致不通呢？

对于直连不通的情况，我们首先要确认ARP/MAC的学习情况，查看表项发现，在M-LAG-1设备上，接入设备的ARP学习在了peer-link接口，MAC学习在M-LAG接口:

```
[M-LAG-1]display arp 10.0.251.6
Type: S-Static D-Dynamic O-Openflow R-Rule M-Multiport I-Invalid
IP address MAC address VLAN/VSI name Interface Aging Type
10.0.251.6 1451-7ea4-b589 4019 BAGG1 1200 D

[M-LAG-1]display mac 1451-7ea4-b589
MAC Address VLAN ID State Port/Nickname Aging
1451-7ea4-b589 4019 Learned BAGG6 Y
```

正常情况下，ARP和MAC都应该学习在M-LAG接口，才能实现M-LAG的正常转发。我们先来复习一下M-LAG组网中，两台M-LAG设备间ARP和MAC的同步过程:

M-LAG设备的M-LAG接口在收到ARP请求报文后，会正常二层泛洪广播一份到对端M-LAG设备，对端M-LAG设备上的arp学习在Peer-link接口；同时本端将收到的ARP请求报文中送平台后，还会通过rlink报文将ARP报文再同步到对端M-LAG设备，此时对端M-LAG设备的ARP会学在M-LAG接口，并将上一步学习在peer-link接口的ARP覆盖，如下图:

与ARP同步报文实现不同，M-LAG中MAC同步的是表项，M-LAG接口学习的MAC表项直接同步到另一台设备的M-LAG接口。

在现场故障环境中，我们在设备上做ARP流统，发现ARP报文从M-LAG-2的M-LAG接口进入设备，由于M-LAG-2设备上没有int vlan 4019，ARP报文不会上送CPU，也就不会触发rlink封装同步，M-LAG-1上只会收到通过二层广播从peer-link接口泛洪过来的ARP报文，因此M-LAG-1设备上ARP的出接口为peer-link接口。而M-LAG中MAC是同步的表项，因此M-LAG-1和M-LAG-2上接入设备的MAC地址都是学在M-LAG接口的。

接下来我们在设备上做ICMP流统，发现接入设备的ICMP request报文正常可以到M-LAG-1上，但M-LAG-1回包时，ICMP reply报文会从peer-link接口发给M-LAG-2设备，但是M-LAG-2设备未将报文转发给下行Access设备。这是因为ICMP报文在M-LAG-2设备上做二层转发，出端口为M-LAG接口，M-LAG-2设备判断两端都有被选中的M-LAG接口，因此，从peer-link链路接收的流量会直接被丢弃，不会再从M-LAG接口发出。

解 / 决 / 方 / 法

按照规范配置，在两台M-LAG设备上都正常配置VLAN4019的三层虚接口即可。

M-LAG三层网关配置及表项同步机制可参考:

[交换那些事儿 | 技术专题篇 - DRNI+VRRP及DRNI+VLAN双活组网配置介绍](#)

[交换那些事儿 | 技术专题篇 - DRNI组网ARP及MAC表项同步机制介绍](#)

- FAQ -

Q1: M-LAG+VRRP场景下流量负载分担?

A1:

传统的VRRP标准模式，只有主设备能上行转发，VRRP设备设备收到下行进来的流量后，会转发到VRRP主设备上，再从VRRP主转出。而在M-LAG+VRRP的场景下做了特殊处理，两台设备上均有三层网关虚MAC，设备收到下行来的流量时，均可本端直接转发，实现上行链路的负载分担。

Q2: M-LAG单挂设备有没有特殊配置要求?

A2:

在M-LAG+VRRP或者M-LAG+VLAN双活组网中，若存在服务器主备模式或者单端口接入时，设备上需存在至少一个M-LAG group，该M-LAG group需放通VRRP或者VLAN双活网关所在的VLAN。如果没有放通相应VLAN的M-LAG group存在，可能会导致该VLAN内的单挂业务转发不通。

Q3: M-LAG成员接口物理down后是否会影响M-LAG group的状态?

A3:

不会。两台M-LAG设备只要有一个设备上存在选中的成员接口，两台设备的M-LAG group都会处于UP状态，即使本设备上的成员端口down掉也不会影响M-LAG聚合状态，该实现用于两设备M-LAG接口表项的同步。

Q4: M-LAG+EVPN组网中，border设备配置group ip后为何不以虚地址发布路由?

A4:

border设备一般只引入5类外部路由，BGP l2vpn地址族下引入的5类路由不受evpn drni group ip命令配置的group ip影响，需要同时在l2vpn地址族下配置nexthop虚地址后，border才能以虚地址（group ip）发布5类路由。

每多处理一个问题

对M-LAG的理解也更多一层

经验案例多看看

典型配置多学习

大家都可以成为大佬~

有问题欢迎留言哦~

END

扫码关注我们哦