

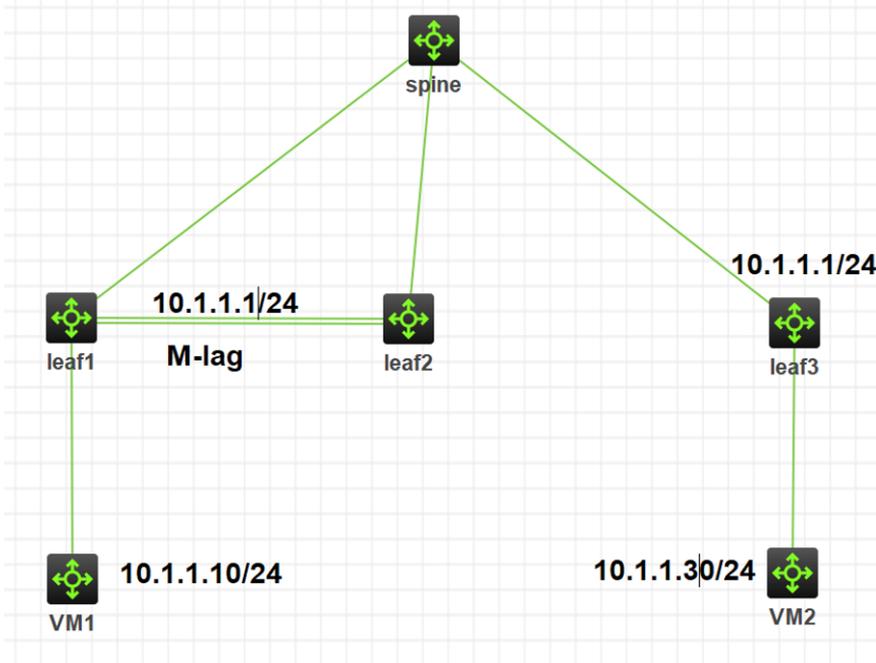
某局点S7800XPG设备EVPN+M-LAG组网二层报文转发丢包

ADDC方案 EVPN 刘倩 2023-03-06 发表

组网及说明

组网大致如下，leaf设备采用了M-LAG组网，但是下挂服务器采用了单挂接入，而非DR口接入，leaf上采用arp代答方式，同网段互访为二层互访

M-LAG配置了evpn m-lag local remote，远端leaf3上查看存在两条实隧道和1条虚隧道



告警信息

不sheji

问题描述

现场开局设备上无业务流量，会出现本地leaf3下挂终端偶发性无法ping通远端leaf1或者leaf2下挂终端。

访问不通时，在远端leaf1下查看l2vpn mac-address发现没有下挂终端的mac表项。但是有arp代答表项。

```
<leaf1>dis arp suppression vsi
```

IP address	MAC address	VSI name	Link ID	Aging(min)
10.1.1.30	28c2-b416-0902	vpna	0x5000000	N/A
10.1.1.10	28ca-d77c-0702	vpna	0x0	25

```
<leaf1>dis l2vpn mac-address
```

* - The output interface is issued to another VSI

MAC Address	State	VSI Name	Link ID/Name	Aging
28c2-b416-0902	EVPN	vpna	Tunnel0	NotAging

过程分析

现场开局设备上无业务流量，会出现本地leaf3下挂终端偶发性无法ping通远端leaf1或者leaf2下单挂终端。

访问不通时，在远端leaf1下查看l2vpn mac-address发现没有下挂终端的mac表项。此时有arp代答表项。

```
<leaf1>dis arp suppression vsi
IP address  MAC address  VSI name          Link ID  Aging(min)
10.1.1.30   28c2-b416-0902  vpna              0x5000000  N/A
10.1.1.10   28ca-d77c-0702  vpna              0x0        25
```

```
<leaf1>dis l2vpn mac-address
```

* - The output interface is issued to another VSI

```
MAC Address  State  VSI Name          Link ID/Name  Aging
28c2-b416-0902  EVPN   vpna              Tunnel0        NotAging
```

leaf3设备上从对端同步过来的MAC表项是不老化的，所以icmp报文到达leaf3后查询l2vpn mac-address，因为配置了evpn drni local remote，leaf1下终端是单挂接入的，所以icmp会选择实隧道单播到对端leaf1。

```
[leaf3]dis l2vpn mac-address
```

* - The output interface is issued to another VSI

```
MAC Address  State  VSI Name          Link ID/Name  Aging
28c2-b416-0902  Dynamic  vpna              GE1/0/2       Aging
28ca-d77c-0702  EVPN    vpna              Tunnel1        NotAging
```

在EVPN——M-LAG组网中，有一个特性，流量到达本地设备需要广播时，会从各个隧道都泛洪一份，实隧道和虚隧道都会泛洪一份，当流量泛洪到远端leaf时，实隧道的流量会丢弃，虚隧道的流量正常转发，此时远端leaf下挂终端可以正常接收一份流量。

而现场流量到达远端leaf3时，再次查询l2vpn mac-address时，因为leaf3上本地学习到的MAC表项由于无流量老化后，没有对应mac表项，从实隧道过来的流量会被丢弃，导致leaf3下挂的目的终端一份流量也收不到，如果leaf3侧主动触发一下学习MAC地址，流量就能正常转发了，所以最终导致现场偶发性不通的情况。

解决方法

可以采取两种解决办法：

(1) 学习到本地终端的arp代答表项老化时间是25分钟，时间达到20分钟后，设备会主动发起一次探测触发重学arp表项，相应的也可以学习mac地址表项。可以配置mac的老化时间为20分钟。

(2) vsi视图下配置arp mac-learning disable，这个功能是禁止EVPN从ARP/ND信息中学习MAC地址表项，EVPN仅通过MAC地址信息学习远端MAC地址表项。这样远端leaf上学习到的本地mac表项老化后，本地leaf同步的表项也会消失。所以流量在本地查询表项查询不到后就会泛洪，从各个实隧道和虚隧道都会泛洪一份，实隧道会丢弃但是虚隧道的流量可以正常转发

