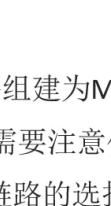


交换那些事儿M-LAG技术专题新系列
M-LAG典型案例及FAQ
每期推送2个案例+4个FAQ
让你在遇到类似问题时不迷茫不发愁
解决方案手到擒来~
点击上方话题可回顾往期内容



本期内容

典型案例：

- Case1 多级M-LAG组网中mac漂移案例
- Case2 M-LAG组网keepalive链路协议down案例

FAQ：

- Q1 更换故障M-LAG设备时先down keepalive还是先down peer-link链路？
- Q2 可以将两组IRF设备组建成M-LAG吗？
- Q3 M-LAG中镜像配置需要注意什么？
- Q4 M-LAG中keepalive链路的选择和配置需要注意什么？

DRNI更名为M-LAG

原DR口>>M-LAG接口
原DR group>>M-LAG group
原IPL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

多级M-LAG组网中mac漂移案例

组 / 网 / 说 / 明

设备及版本：S10508X R7596P10

组网：两台S10508X和两台S6800组成多级M-LAG组网，两台S6800作为接入层，纯二层转发，两台S10508X是网关设备。两级M-LAG间通过M-LAG group（BAGG301）互联，接入层S6800下行接入一台单挂设备。整体组网如下图所示：

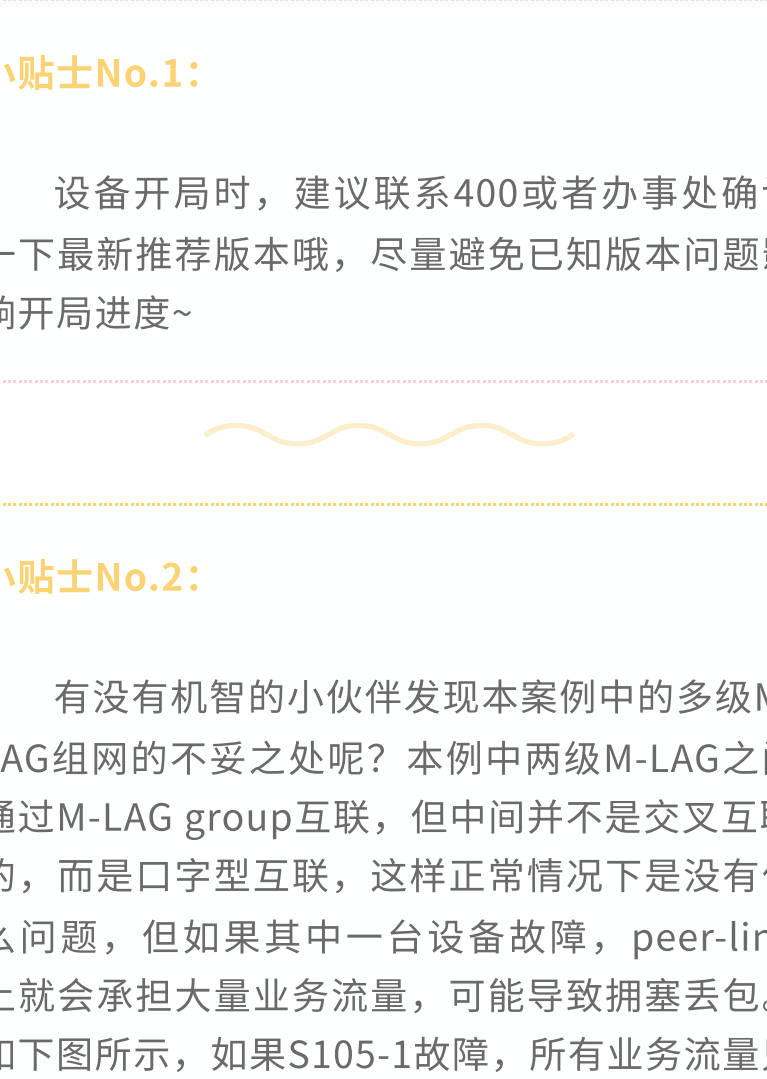


问 / 题 / 描 / 述

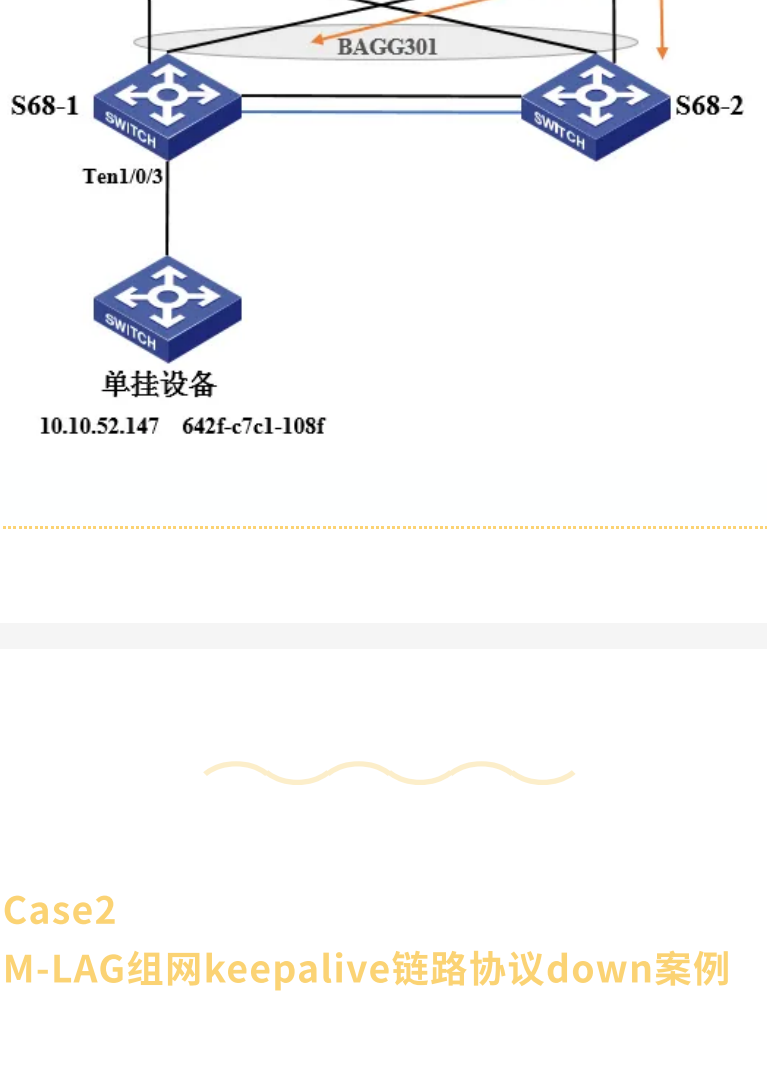
现场在对新上多级M-LAG组网业务进行测试时，发现单挂设备的mac地址在S6800-1的上行M-LAG接口学到了，进一步查看发现S6800-1和S6800-2上均存在该单挂设备的mac漂移现象，S6800-1上的漂移接口为上行M-LAG接口和单挂口Ten1/0/3，S6800-2上的漂移接口为上行M-LAG接口和peer-link接口。

过 / 程 / 分 / 析

正常情况下，在S68-1上单挂终端的mac地址就应该学在互联接口Ten1/0/3口；在S68-2上，单挂在S68-1下的终端mac，应该学习在S68-2的peer-link接口上，如下图绿色五角星所示。但现场S68-1和S68-2的上行M-LAG接口都学到了单挂终端的mac（如下图橙色五角星所示），怀疑两台S68设备在上行口收到了源mac为单挂终端mac的报文。



由此我们将目光转向了上行的两台S105X，在S105X上，单挂终端的mac和ARP都正常学习在M-LAG接口上。我们在两台S105X上做基于源mac的报文流统，发现在S105X-1上，只在M-LAG接口BAGG301上收到了源mac为单挂终端mac的报文；在S105-2上看入方向收到了从peer-link过来的报文，且还从M-LAG接口发送了源mac为单挂终端mac的报文给S68，如下图所示：



S68-2上从上行M-LAG接口收到了源mac为终端mac的报文，在M-LAG接口学到了这个mac地址，再通过rlink报文同步给S68-1，因此S68-1也在上行M-LAG接口学到了单挂终端的mac，现场两台S68上的mac漂移现象也就由此产生。

那么为什么S105X-2会将源mac为单挂终端mac的报文从M-LAG接口再发回去呢？经过底层排查确认，故障原因是现场开局调试时，peer-link接口偶尔会被手动震荡，震荡过程中两台M-LAG设备间通过peer-link同步M-LAG接口状态的过程出现异常误差，导致其中一台M-LAG设备上记录的M-LAG接口信息错误。简单来说就是，在S105X-2上，并没有正确记录S105X-1上的M-LAG接口信息，当收到S105X-1从BAGG301收到的报文后，又从本端的BAGG301转发了出去。

解 / 决 / 方 / 法

临时规避方法：震荡S105X-1上的BAGG301，重新触发出向S105X-2同步M-LAG接口。

解决方案：S10500X系列交换机的R762X版本解决了这个已知问题，升级到相应版本即可。

小贴士No.1：

设备开局时，建议联系400或者办事处确认一下最新推荐版本哦，尽量避免已知版本问题影响开局进度~

小贴士No.2：

有没有机智的小伙伴发现本案例中的多级M-LAG组网的不妥之处呢？本例中两级M-LAG之间通过M-LAG group互联，但中间并不是交叉互联的，而是口字型互联，这样正常情况下是没有什么问题，但如果其中一台设备故障，peer-link上就会承担大量业务流量，可能导致拥塞丢包。如下图所示，如果S105-1故障，所有业务流量只能走S105-2下行到S68-2上：



更合理的组网规划是在两级M-LAG之间交叉互联，这样即使单设备故障后，也能负载均衡：



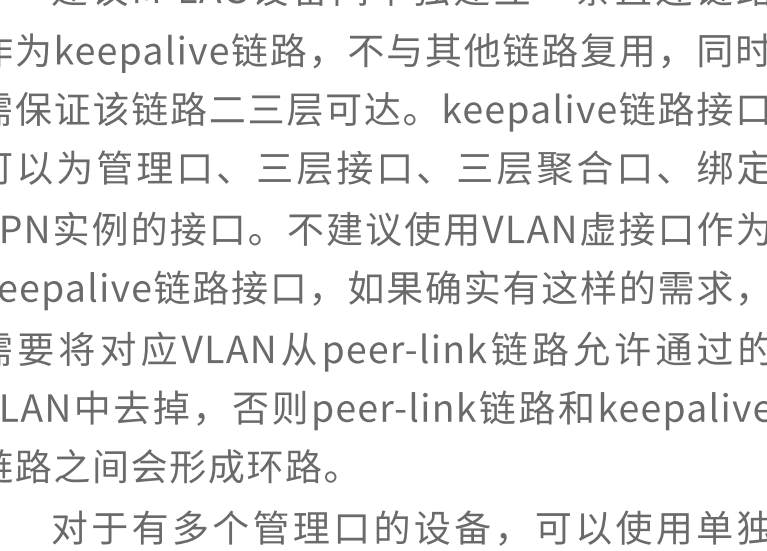
Case2

M-LAG组网keepalive链路协议down案例

组 / 网 / 说 / 明

设备及版本：S6900-F F2712

组网：现场新开局，TOR-1和TOR-2两台S6900-F设备组建M-LAG，局部组网如下图：



问 / 题 / 描 / 述

TOR-1和TOR-2两台S6900-F设备成功组建M-LAG，通过IPv6地址建立keepalive链路，但keepalive连接无法建立，链路一直处于协议down状态。

过 / 程 / 分 / 析

查看M-LAG配置没有问题，主、从设备上都有对端keepalive source地址的ND表项，互ping能通：



drni keepalive ipv6 destination 2409:8086:86F4:FF::1 source 2409:8086:86F4:FF::1 vpn-instance NFV-DAD

在两台设备上通过debugging drni keepalive命令查看keepalive报文交互情况，发现TOR-1设备上有收有发，但是TOR-2设备上则只有发、没有收。同时我们通过流统确认TOR-2设备keepalive链路的方向是收到了报文的：

从上述排查我们可以得出结论，TOR-2上收到了TOR-1发出的keepalive报文，但是没有处理，由此我们怀疑在TOR-2上keepalive使用的端口号被占用了，或者TOR-2收到的keepalive报文有问题，因此我们先看了TOR-2上的UDP端口号占用情况，发现keepalive缺省使用的UDP port 6400被其他应用给占用了：

keepalive缺省使用的UDP port 6400被占用，导致TOR-2无法正常接收并处理TOR-1发送过来的keepalive报文。

解 / 决 / 方 / 法

临时规避方法：配置keepalive时使用命令**udp-port udp-number**手工指定未被占用的端口号；如果设备发生重启，配置的UDP端口号存在被其他应用再次抢占的风险。

解决方案：F2712版本后续会评估发布补丁解决，将UDP port 6400预留给M-LAG keepalive使用，不允许其他应用抢占。

- FAQ -

Q1：更换故障M-LAG设备时先down keepalive还是先down peer-link链路？

A1：在替换故障M-LAG设备时，请注意先关闭keepalive物理端口，再关闭peer-link物理端口，否则会出现设备先被MAD down再被解除MAD down，M-LAG接口震荡的现象。

Q2：可以将两组IRF设备组建成M-LAG吗？

A2：不支持在IRF设备上（包括运行在IRF模式的单台设备）配置M-LAG，配置M-LAG前请确保设备出于独立运行模式。

Q3：M-LAG中镜像配置需要注意什么？

A3：在M-LAG设备上配置镜像时，请避免出现镜像源端口为聚合组A的成员端口，镜像目的端口、出端口或者反射端口为聚合组B的成员端口的情况，以免聚合组B成员端口接收到聚合组A成员端口的镜像LACP报文，引起聚合震荡。

Q4：M-LAG中keepalive链路的选择和配置需要注意什么？

A4：建议M-LAG设备间单独建立一条直连链路作为keepalive链路，不与其他链路复用，同时需保证该链路三层可达。keepalive链路接口可以为管理口、三层接口、三层聚合口、绑定VPN实例的接口。不建议使用VLAN虚接口作为keepalive链路接口，如果确实有这样的需求，需要将对应VLAN从peer-link链路允许通过的VLAN中去掉，否则peer-link链路和keepalive链路之间会形成环路。

对于有多个管理口的设备，可以使用单独的管理口作为keepalive链路接口，不能与管理网共用链路。

对于多单板/多子卡设备，建议和peer-link链路接口部署在不同的单板/子卡上。

keepalive链路接口（包括物理口和逻辑口）请务必配置为M-LAG保留接口。

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