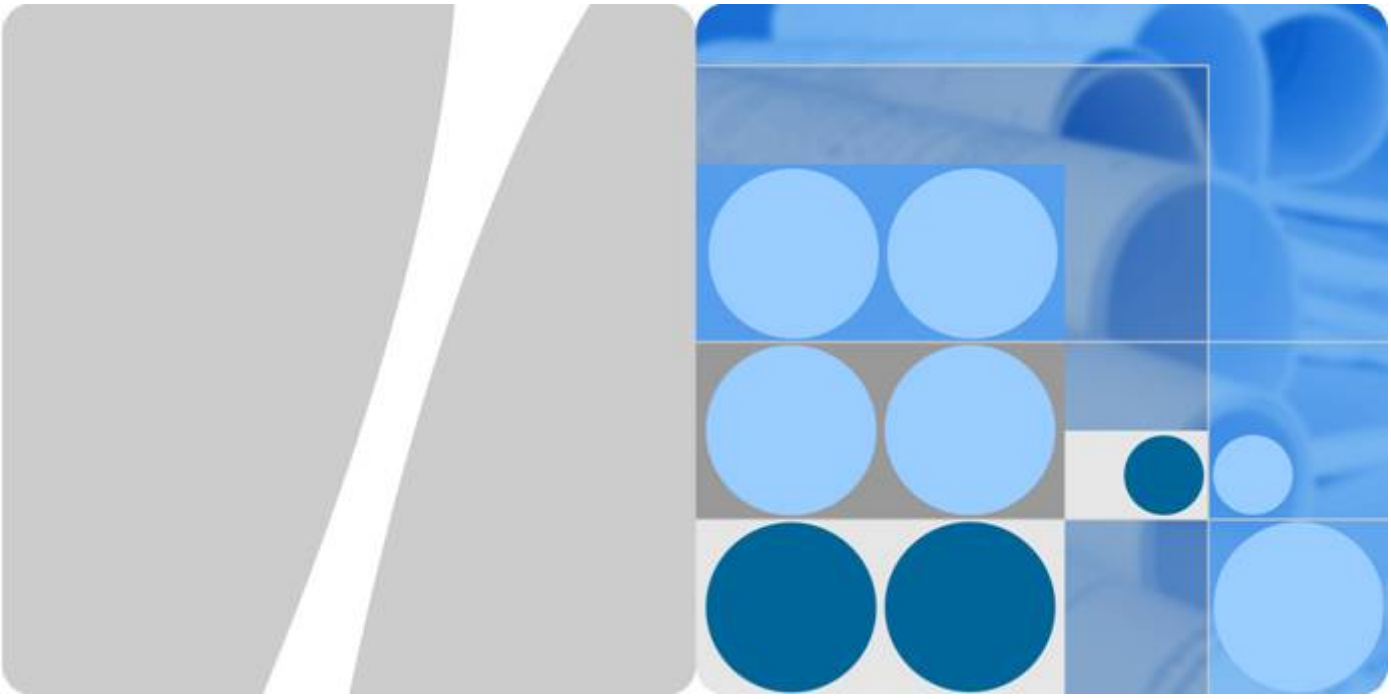


CloudEngine 8800, 7800, 6800, 5800 系列交换机

M-LAG 技术白皮书

文档版本　01

发布日期　2020-04-20

华为技术有限公司

HUAWEI

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： https://e.huawei.com

目 录

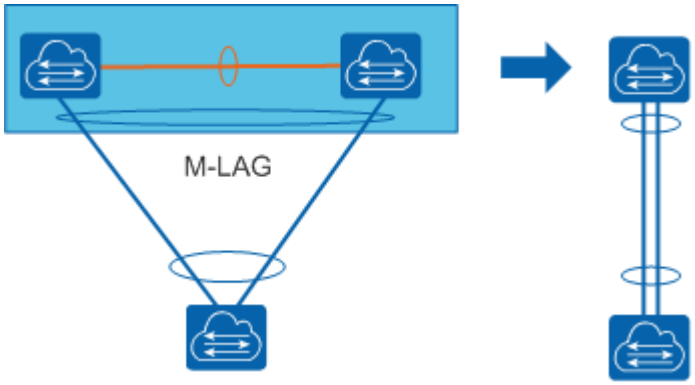
1 M-LAG（跨设备链路聚合）配置

1.1 M-LAG 简介

定义

M-LAG（Multichassis Link Aggregation Group）即跨设备链路聚合组，是一种实现跨设备链路聚合的机制，如图1-1所示，将两台接入交换机以同一个状态和被接入的设备进行链路聚合协商，从而把链路可靠性从单板级提高到了设备级，组成双活系统。

图 1-1 M-LAG 示意图



目的

M-LAG作为一种跨设备链路聚合的技术，除了具备增加带宽、提高链路可靠性、负载分担的优势外，还具备以下优势：

- 更高的可靠性
  把链路可靠性从单板级提高到了设备级。
- 简化组网及配置
  可以将M-LAG理解为一种横向虚拟化技术，将双归接入的两台设备在逻辑上虚拟成一台设备。M-LAG提供了一个没有环路的二层拓扑同时实现冗余备份，不再需要繁琐的生成树协议配置，极大的简化了组网及配置。
- 独立升级
  两台设备可以分别进行升级，保证有一台设备正常工作即可，对正在运行的业务几乎没有影响。

相关资料

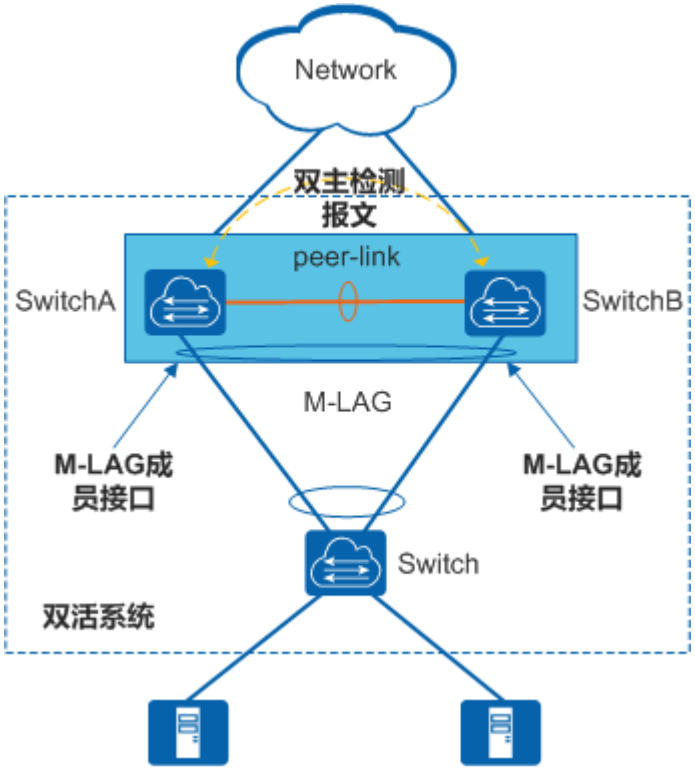
- M-LAG最佳实践：**CloudEngine系列交换机 M-LAG技术专题**
- M-LAG特性介绍视频：**M-LAG特性介绍**

1.2 M-LAG 原理描述

1.2.1 M-LAG 的基本概念

如图1-2所示，用户侧设备Switch（可以是交换机或主机）通过M-LAG机制与另外两台设备（SwitchA和SwitchB）进行跨设备链路聚合，共同组成一个双活系统。这样可以实现SwitchA和SwitchB共同进行流量转发的功能，保证网络的可靠性。

图 1-2 M-LAG 基本拓扑



下面介绍下M-LAG涉及的相关概念，如表1-1所示。

表 1-1 M-LAG 基本概念

概念	说明
DFS Group	动态交换服务组DFS Group（Dynamic Fabric Service Group），主要用于部署M-LAG设备之间的配对，M-LAG双归设备之间的接口状态，表项等信息同步需要依赖DFS Group协议进行同步。

概念	说明
DFS主设备	部署M-LAG且状态为主的设备，通常也称为M-LAG主设备。
DFS备设备	部署M-LAG且状态为备的设备，通常也称为M-LAG备设备。 **说明** DFS Group的角色区分为主和备，正常情况下，主设备和备设备同时进行业务流量的转发，转发行为没有区别，仅在故障场景下，主备设备的行为会有差别。
双主检测链路	双主检测链路，又称为心跳链路，是一条三层互通链路，用于M-LAG主备设备间发送双主检测报文。 **说明** 正常情况下，双主检测链路不会参与M-LAG的任何转发行为，只在故障场景下，用于检查是否出现双主的情况。双主检测链路可以通过外部网络承载（比如，如果M-LAG上行接入IP网络，那么两台双归设备通过IP网络可以互通，那么互通的链路就可以作为双主检测链路）。也可以单独配置一条三层可达的链路来作为双主检测链路（比如通过管理口）。
peer-link接口	peer-link链路两端直连的接口均为peer-link接口。
peer-link链路	peer-link链路是一条直连链路且必须做链路聚合，用于交换协商报文及传输部分流量。接口配置为peer-link接口后，该接口上不能再配置其它业务。 为了增加peer-link链路的可靠性，推荐采用多条链路做链路聚合。
HB DFS主设备	通过心跳链路来协商的状态为主的设备。 **说明** 通过心跳链路报文来协商的设备HB DFS主备状态在正常情况下，对M-LAG的转发行为不会产生影响，仅用于二次故障恢复场景下，在原DFS主设备或备设备故障恢复且peer-link链路仍然故障时，触发HB DFS状态为备的设备上相应端口Error-Down，避免M-LAG设备在双主情况下出现的流量异常。
HB DFS备设备	通过心跳链路来协商的状态为备的设备。 **说明** 通过心跳链路报文来协商的设备HB DFS主备状态在正常情况下，对M-LAG的转发行为不会产生影响，仅用于二次故障恢复场景下，在原DFS主设备或备设备故障恢复且peer-link链路仍然故障时，触发HB DFS状态为备的设备上相应端口Error-Down，避免M-LAG设备在双主情况下出现的流量异常。
M-LAG成员接口	M-LAG主备设备上连接用户侧主机（或交换设备）的Eth-Trunk接口。 为了增加可靠性，推荐链路聚合配置为LACP模式。 M-LAG成员接口角色也区分主和备，与对端同步成员口信息时，状态由Down先变为Up的M-LAG成员接口成为主M-LAG成员口，对端对应的M-LAG成员口为备。 **说明** 仅在M-LAG接入组播场景下，M-LAG成员接口的主备角色存在转发行为差异。

1.2.2 M-LAG 协议交互原理

基于M-LAG组成的双活系统提供了设备级的可靠性，那么M-LAG是如何建立的？如图1-3所示，M-LAG的建立过程有如下几个步骤：

1. DFS Group配对

   当M-LAG两台设备完成配置后，设备首先通过peer-link链路发送DFS Group的Hello报文。当设备收到对端的Hello报文后，会判断报文中携带的DFS Group编号是否和本端相同，如果两台设备的DFS Group编号相同，则两台设备DFS Group配对成功。

2. DFS Group协商主备

   配对成功后，两台设备会向对端发送DFS Group的设备信息报文，设备根据报文中携带的DFS Group优先级以及系统MAC地址确定出DFS Group的主备状态。

   以SwitchB为例，当SwitchB收到SwitchA发送的报文时，SwitchB会查看并记录对端信息，然后比较DFS Group的优先级，如果SwitchA的DFS Group优先级高于本端的DFS Group优先级，则确定SwitchA为DFS主设备，SwitchB为DFS备设备。如果SwitchA和SwitchB的DFS Group优先级相同，比较两台设备的MAC地址，确定MAC地址小的一端为DFS主设备。

   说明

   DFS Group的角色区分为主和备，正常情况下，主设备和备设备同时进行业务流量的转发，转发行为没有区别，仅在故障场景下，主备设备的行为会有差别。

3. M-LAG成员接口协商主备

   在DFS Group协商出主备状态后，M-LAG的两台设备会通过peer-link链路发送M-LAG设备信息报文，报文中携带了M-LAG成员接口的配置信息。在成员口信息同步完成后，确定M-LAG成员接口的主备状态。

   与对端同步成员口信息时，状态由Down先变为Up的M-LAG成员接口成为主M-LAG成员口，对端对应的M-LAG成员口为备，且主备状态默认不回切，即：当M-LAG成员接口状态为主的设备故障恢复后，先前由备状态升级为主状态的接口仍保持主状态，恢复故障的M-LAG成员接口状态为备，此处与DFS Group协商主备状态不一致。

   说明

   仅在M-LAG接入组播场景下，M-LAG成员接口的主备角色存在转发行为差异。

4. 双主检测

   协商出M-LAG主备后，两台设备之间会通过双主检测链路按照1s的周期发送M-LAG双主检测报文，一旦设备感知peer-link故障，会按照100ms的周期发送三个双主检测链路报文，加速检测。当两台设备均能够收到对端发送的报文时，双活系统即开始正常的工作。

   正常情况下，双主检测链路不会参与M-LAG的任何转发行为，只在DFS Group配对失败或者peer-link故障场景下，用于检查是否出现双主的情况，所以即便双主检测失败也不会影响M-LAG正常工作。双主检测链路可以通过外部网络承载（比如，如果M-LAG上行接入IP网络，那么两台双归设备通过IP网络可以互通，那么互通的链路就可以作为双主检测链路）。也可以单独配置一条三层可达的链路来作为双主检测链路（比如通过管理口）。

   - （推荐）双主检测链路通过管理网口互通，DFS Group绑定的管理网口IP地址要保证可以相互通信，管理网口下绑定VPN实例，保证双主检测报文与业务流量隔离。
   - 双主检测链路通过业务网络互通，DFS Group绑定的IP地址要保证可以三层互通。如果peer-link接口之间建立路由邻居关系，则业务网络双主检测报文

会直接通过最优路由经peer-link链路传输。一旦peer-link故障，路由收敛期间，双主检测报文通过次优路径传输到对端，双主检测时间会慢0.5秒或者1秒的时间。

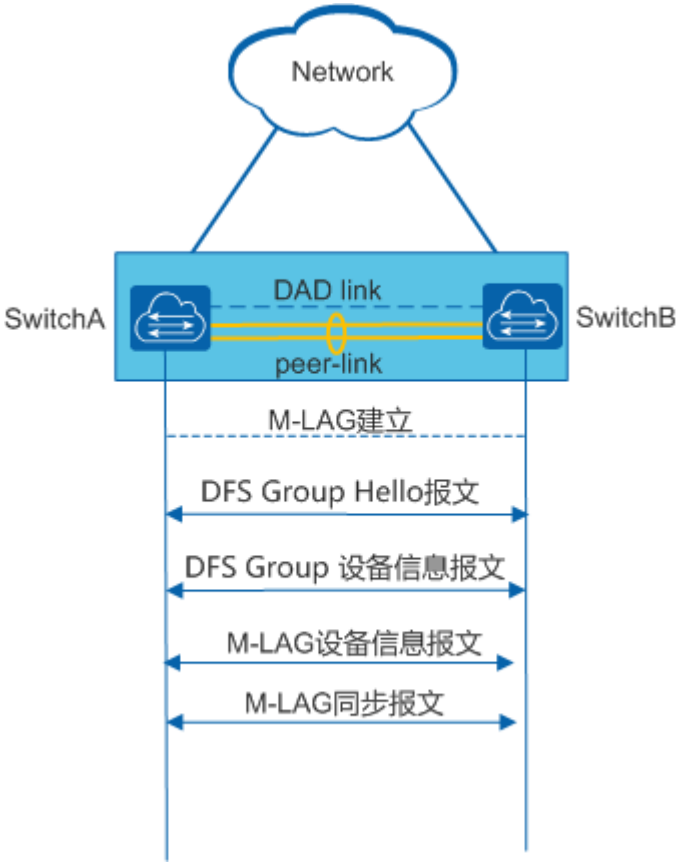
说明

在V200R005C10版本及之后版本，两台设备在心跳链路Up之后即会按照周期发送双主检测报文。若DFS Group绑定了本端和对端的IP地址，则在二次故障恢复场景下（设备已使能二次故障增强功能），即原DFS主设备或备设备故障恢复且peer-link链路仍然故障时，M-LAG设备根据双主检测报文中携带的DFS信息协商出HB DFS主备状态，触发HB DFS状态为备的设备相应端口Error-Down，从而避免双主场景下的流量异常。

5. M-LAG同步信息

正常工作后，两台设备之间会通过peer-link链路发送M-LAG同步报文实时同步对端的信息，M-LAG同步报文中包括MAC表项、ARP表项以及STP、VRRP协议报文信息等，并发送M-LAG成员端口的状态，这样任意一台设备故障都不会影响流量的转发，保证正常的业务不会中断。

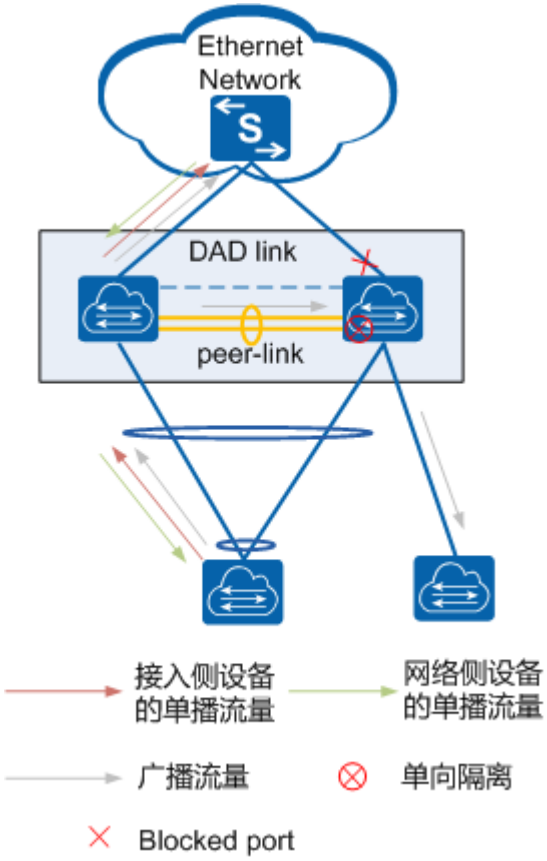
图 1-3 M-LAG 的建立



1.2.3 M-LAG 防环机制

M-LAG本身具有防环机制，可以构造出一个无环网络。那么M-LAG是如何构造无环网络的呢？如图1-4所示，从接入设备或网络侧到达M-LAG配对设备的单播流量，会优先从本地转发出去，peer-link链路一般情况下不用来转发数据流量。当流量通过peer-link链路广播到对端M-LAG设备，在peer-link链路与M-LAG成员口之间设置单方向的流量隔离，即从peer-link口进来的流量不会再从M-LAG口转发出去，所以不会形成环路，这就是M-LAG单向隔离机制。

图 1-4 M-LAG 接入二层网络流量转发示意图



单向隔离机制

机制生效前提

当M-LAG两台设备协商出M-LAG主备后，系统通过M-LAG同步报文判断接入设备是否双活接入：

- 若接入设备双活接入M-LAG系统，则M-LAG两台设备下发对应M-LAG成员口的单向隔离配置，来隔离由peer-link口发往M-LAG成员口的流量。

  说明

  M-LAG防环机制中的单向隔离仅对广播流量等泛洪流量生效。

- 若接入设备单归接入M-LAG系统，则M-LAG系统不会下发对应M-LAG成员口的单向隔离配置。

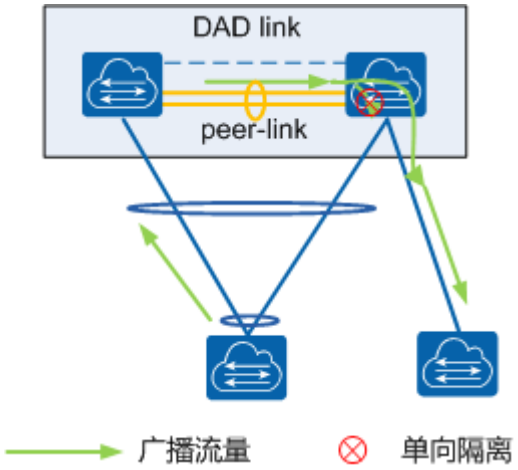
单向隔离机制实现原理

如图1-5所示，在设备双活接入M-LAG场景下，设备会默认按下列顺序下发全局ACL配置：

- Rule1：允许通过源端口为peer-link接口，目的端口为M-LAG成员口的三层单播报文；
- Rule2：拒绝通过源端口为peer-link接口，目的端口为M-LAG成员口的所有报文；

设备通过匹配ACL规则组来对实现peer-link接口与M-LAG成员口之间的单向隔离，隔离有peer-link接口发往M-LAG成员口的广播等泛洪流量。当M-LAG设备感知到本端的M-LAG成员口状态为Down时，会通过peer-link发送M-LAG同步报文，通知对端设备撤销自动下发的相应的M-LAG成员端口的单向隔离ACL规则组。

图 1-5 M-LAG 单向隔离示意图



1.2.4 M-LAG 配置一致性检查

M-LAG是由两台设备组成的一个双活系统，可将M-LAG理解为一种横向虚拟化技术，将M-LAG的两台设备在逻辑上虚拟成一台设备，形成一个统一的二层逻辑节点。这带来了逻辑拓扑的清晰高效，也决定了M-LAG两端设备的某些配置需要保持一致，否则可能会导致M-LAG无法正常工作或者成环等问题。

但M-LAG运用于企业网中时，却面临一个突出的问题：部署企业网数据中心时，通过手工配置、人工比对来保证每一个M-LAG系统两端设备的配置一致性，不仅处理效率低下，更多的是带来诸多潜在的误配置风险。

为了解决上述问题，华为公司提出了M-LAG配置一致性检查的解决方案。该解决方案中，通过M-LAG机制自带的配置一致性检查功能，去订阅M-LAG系统两端设备的各模块配置。我们可以通过检查功能返回的比对结果，及时地调整M-LAG两端设备的配置部署，防止组网成环或者数据丢包等问题发生。

M-LAG配置一致性检查将设备配置分为两类，如表1-2所示，分别为关键配置（Type 1）和一般配置（Type 2）。根据对关键配置检查不一致时的处理方式，M-LAG一致性又分为严格模式（strict）和松散模式（loose）。

- 关键配置（Type 1）：如果在M-LAG系统两端设备不一致，会导致成环、状态正常但长时间丢包等问题。

  严格模式下，如果M-LAG两端设备存在Type 1配置不一致，会导致M-LAG备设备上成员口处于ERROR DOWN状态，且触发设备对Type 1类型配置检查不一致的告警。

  松散模式下，如果M-LAG两端设备存在Type 1配置不一致，则会触发设备对两种类型配置检查不一致的告警。

- 一般配置（Type 2）：如果在M-LAG系统两端设备不一致，可能会导致M-LAG运行状态异常。与Type 1类型的配置相比较而言，Type 2类型的配置问题更容易被发现，对组网环境的影响也相对较小。

  无论处于何种模式，如果M-LAG两端设备存在以下Type 2配置不一致，则会触发设备对两种类型配置检查不一致的告警。

表 1-2 M-LAG 配置一致性检查配置分类列表

视图	配置	类型
全局	STP功能是否使能	Type 1
	STP工作模式配置	
	BPDU保护功能是否使能	
	STP多生成树实例与VLAN的映射关系配置 **说明** 设备默认仅检查ID为0的STP进程内多生成树实例与VLAN的映射关系。	
M-LAG成员口	STP功能是否使能	
	STP端口的Root保护功能是否使能	
	M-LAG成员接口的LACP模式配置	
全局	VLAN配置	Type 2
	静态MAC地址表项 ● 静态MAC地址表项指定接口为M-LAG成员口 ● VXLAN隧道的静态MAC地址表项	
	动态MAC的老化时间	

视图	配置	类型
	静态ARP表项 ● 短静态ARP表项 ● 长静态ARP表项 – 若静态ARP表项指定出接口，则仅检查出接口为M-LAG成员口的静态ARP。 – 若静态ARP表项指定所属VLAN，则直接比较VLAN ID。 – 若静态ARP表项指定出接口和所属VLAN，则直接比较出接口为M-LAG成员口的静态ARP表项和VLAN ID。 – VXLAN IPv4隧道的静态ARP表项 **说明** 设备不支持检查指定VPN实例的短静态ARP表项，若长静态ARP表项的出接口为M-LAG成员口且绑定了VPN实例或者所属VLAN对应的VLANIF接口绑定了VPN实例，设备同样不支持检查该静态ARP表项。	
	动态ARP的老化时间	
	广播域桥域BD（Bridge Domain）配置 ● BD ID ● BD关联VNI	

视图	配置	类型
	VBDIF接口配置 ● VBDIF接口的BD ID ● VBDIF接口IPv4地址 ● VBDIF接口IPv6地址 ● VBDIF接口VRRP4备份组 ● VBDIF接口MAC地址 ● VBDIF接口状态 **说明** 对于VBDIF接口MAC地址，设备默认仅检查虚拟MAC地址。 针对IPv6地址以及VRRP4备份组的配置检查，仅在VBDIF接口Up时才进行。若VBDIF接口状态为Down，则认为该接口下没有相关配置。	
	VLANIF接口配置 ● VLAN ID ● VLANIF接口IPv4地址 ● VLANIF接口IPv6地址 ● VLANIF接口VRRP4备份组 ● VLANIF接口MAC地址 ● VLANIF接口状态 **说明** 对于VLANIF接口MAC地址，设备默认仅检查虚拟MAC地址。 针对IPv6地址以及VRRP4备份组的配置检查，仅在VLANIF接口Up时才进行。若VLANIF接口状态为Down，则认为该接口下没有相关配置。	
M-LAG成员口	STP端口优先级配置	
	接口加入VLAN配置	
	M-LAG成员口参数配置	

视图	配置	类型
	M-LAG成员口所属Eth-Trunk接口成员口个数 **说明** 仅比较Eth-Trunk接口的成员口数量，对于成员口物理状态Up/Down或者成员口带宽不予检查。	

1.2.5 M-LAG 正常工作场景流量转发

M-LAG双活系统建立成功后即进入正常的工作，M-LAG主备设备负载分担共同进行流量的转发，转发行为没有区别。下面介绍M-LAG在正常工作情况下是如何进行流量转发的。

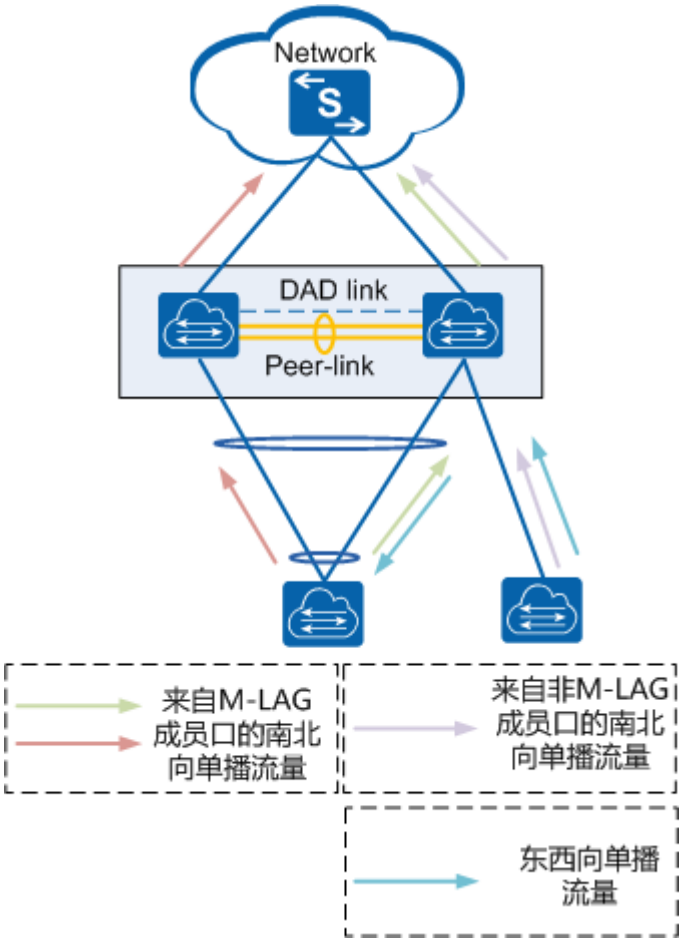
单播流量转发

如图1-6所示，M-LAG双活系统在接入设备双归接入场景下的已知单播流量转发：

对于南北向单播流量，在M-LAG接入侧，M-LAG的成员设备接收到接入设备通过链路捆绑负载分担发送的流量后，共同进行流量转发。到达M-LAG主备设备发往网络侧的流量则根据路由表转发流量。

对于东西向单播流量，在全部组建M-LAG，没有孤立端口的场景下，二层流量通过M-LAG本地优先转发，三层流量通过双活网关转发，都不经过peer-Link链路，直接由M-LAG主备设备转发至对应成员口。

图 1-6 M-LAG 已知单播流量转发示意图



组播流量转发

- **M-LAG接入二层网络**

  M-LAG上行接入二层网络，那么二层网络必须要保证发往M-LAG的流量只有一份，否则会有成环的风险。如图1-7所示，假设右侧M-LAG上行接口被STP协议阻塞：

  在ServerB作为组播源、ServerA作为组播组成员时，M-LAG主备都可以转发组播流量，在网络侧只引流一份流量的情况下，接收到流量的设备直接转发到本地的M-LAG成员口。如果本地M-LAG成员口故障，则组播流量如图1-8所示会从peer-link绕行，转发至M-LAG系统另一台设备的成员口进行转发。

  在ServerA作为组播源、ServerB作为组播组成员时，组播源的流量通过负载分担发送至M-LAG主备设备，由于右端M-LAG设备的上行接口被阻塞，那么右端设备的组播出接口指向peer-link链路。

图 1-7 M-LAG 接入二层网络组播流量转发示意图

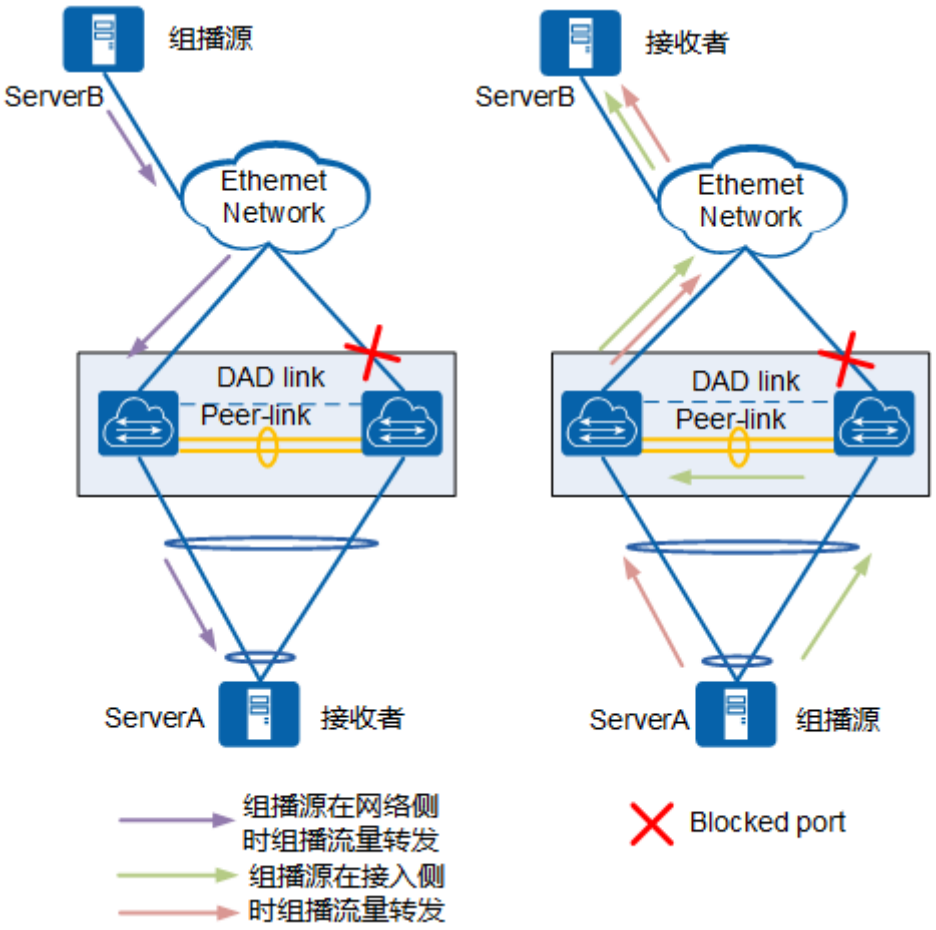
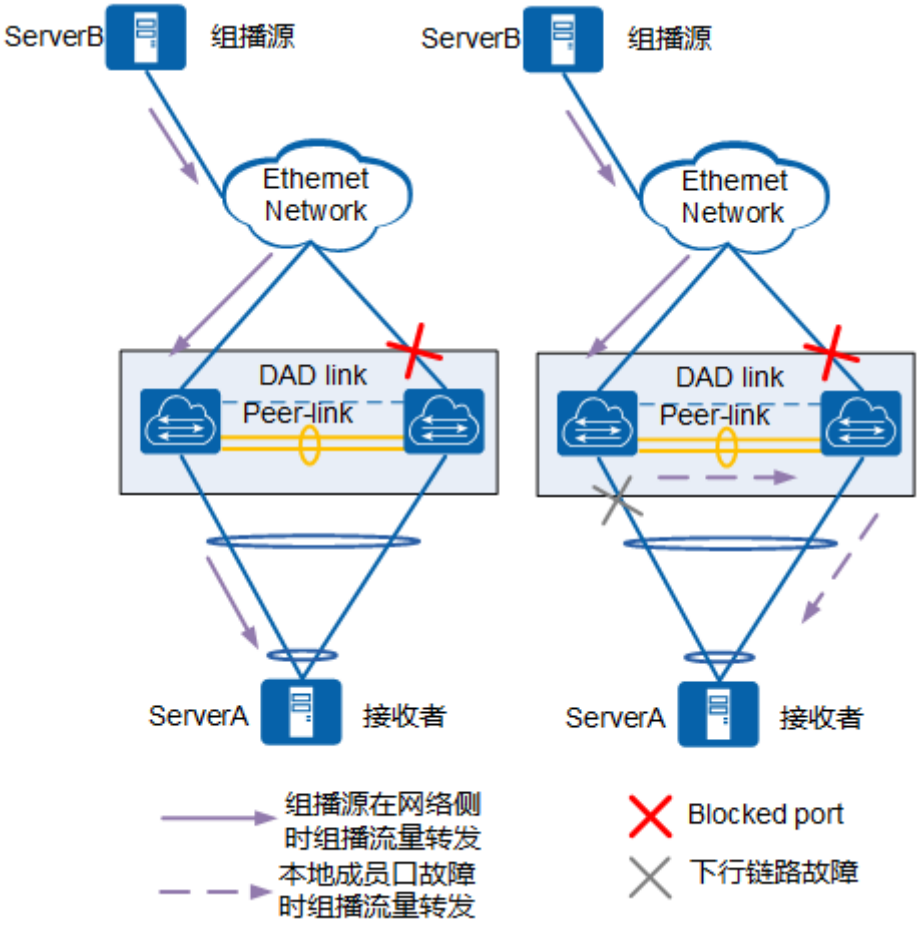


图 1-8 M-LAG 接入二层网络组播流量成员口故障转发示意图



- **M-LAG接入三层网络**

  M-LAG上行接入三层网络，M-LAG系统成员设备需要支持二三层组播混跑。如图1-9所示，M-LAG双活系统在接入设备双归接入场景下的组播流量转发：

  在ServerB作为组播源、ServerA作为组播组成员时，M-LAG主备设备都从组播源引流，且按照以下规则由M-LAG主备设备在本地查找组播表后将流量负载分担转发至组播组成员：

  - 若组播组地址最后一位为奇数（例如225.1.1.1或FF1E::1、FF1E::B），则由M-LAG成员口状态为主的设备转发至组播组成员；
  - 若组播组地址最后一位为偶数（例如225.1.1.2或FF1E::2、FF1E::A），则由M-LAG成员口状态为备的设备转发至组播组成员；

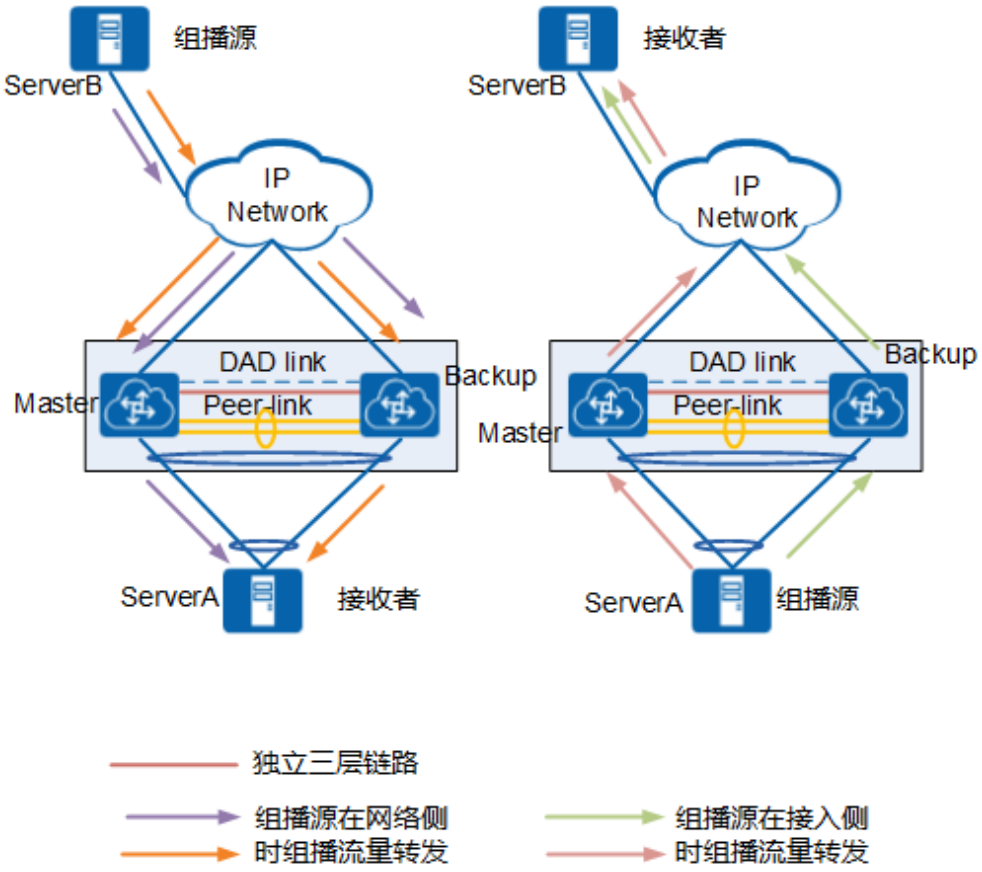
  说明

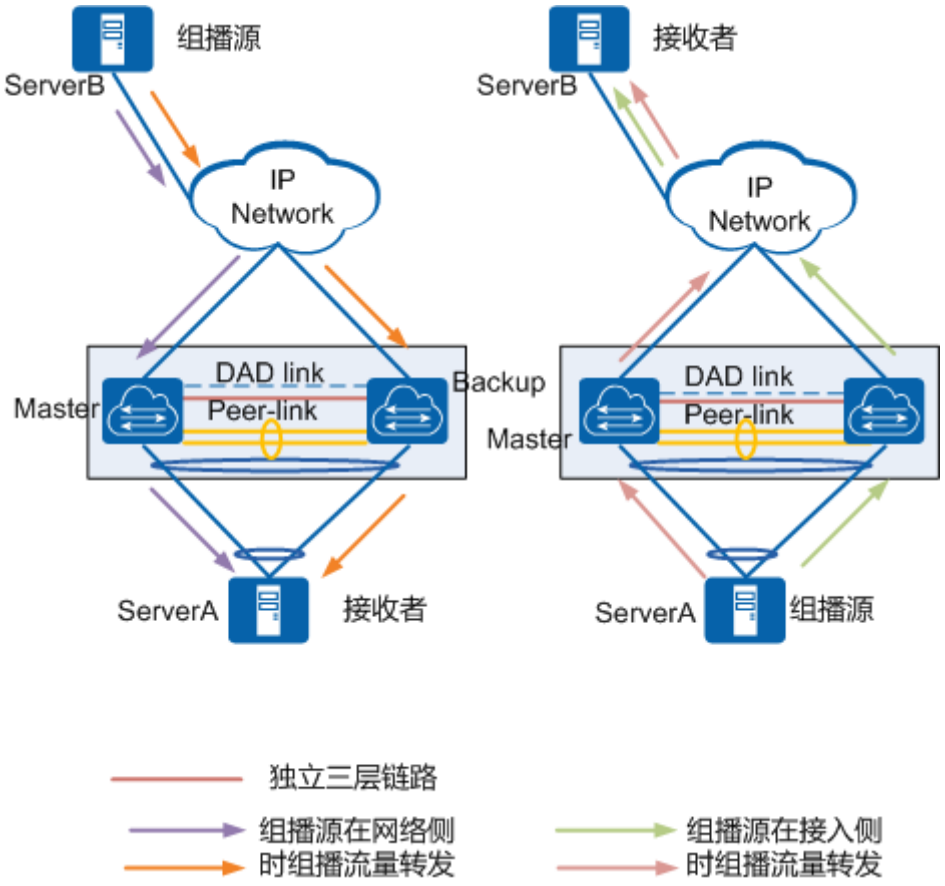
  对于V200R003C00之前的版本，仅M-LAG成员口状态为主的设备转发组播流量到接收者，从V200R003C00版本开始，M-LAG成员口状态为主备的设备均可以转发组播流量到接收者，实现负载分担。当M-LAG系统两台设备版本不一致时，组播流量转发规则以低版本为准。

  从V200R003C00版本开始，对于CE6870EI、CE6875EI，由单机或堆叠组成的M-LAG均支持IPv6三层组播，其余款型不支持。

  在ServerA作为组播源、ServerB作为组播组成员时，组播源发出的流量负载分担到M-LAG系统主备设备，主备设备收到流量后在本地查找组播表将报文发送出去。

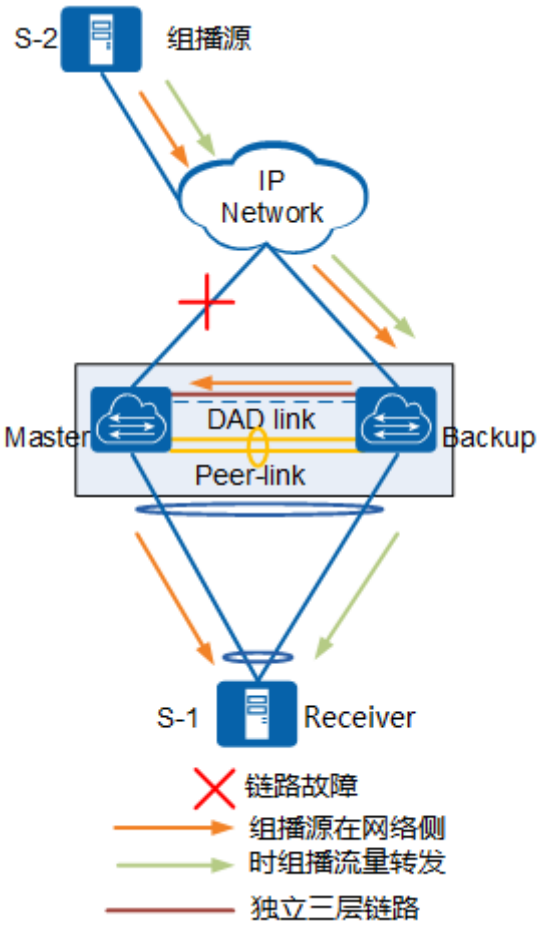
图 1-9 M-LAG 接入三层网络组播流量转发示意图





区别于单播流量，由组播流量转发示意图可以看出，M-LAG系统在转发组播流量时需要在M-LAG两台设备间配置一条独立三层链路。因为在故障场景下，可能出现网络侧只有单链路上行，此时M-LAG主备设备间部署一条独立的单独L3链路可以用来传输组播报文。如图1-10所示，在网络侧链路连接M-LAG备设备场景下，由peer-link接口转发的组播报文由于单向隔离无法转发至指定的M-LAG成员口，组播地址最后一位为奇数的组播报文是无法通过peer-link链路绕行至M-LAG成员口状态为主的设备，只能由独立三层链路转发至该设备。

图 1-10 M-LAG 单归接入三层网络组播流量转发示意图

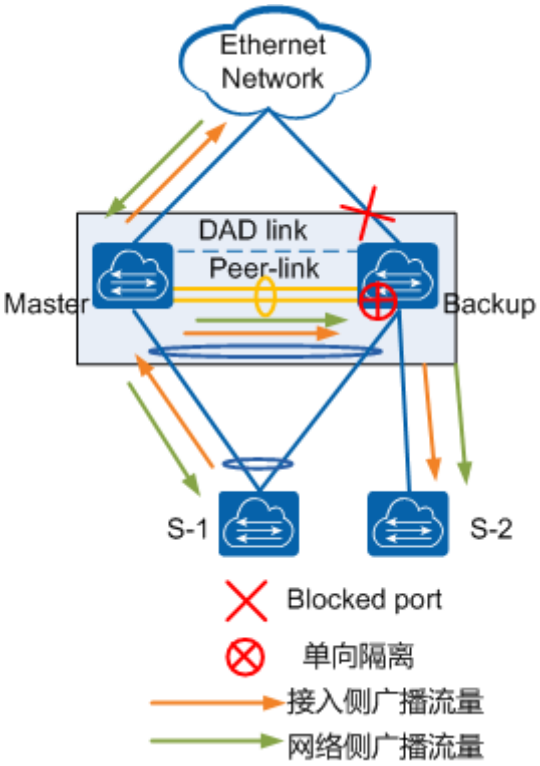


广播流量转发

- **M-LAG接入二层网络**

  M-LAG上行接入二层网络，那么二层网络必须要保证发往M-LAG的流量只有一份，否则会有成环的风险。此处以M-LAG主设备的转发为例，如图1-11所示，假设右侧M-LAG上行接口被STP协议阻塞，M-LAG主设备收到广播流量后向各个下一跳转发，当流量到达M-LAG备设备时，由于peer-link与M-LAG成员接口存在单向隔离机制，到达备设备的流量不会向S-1转发。

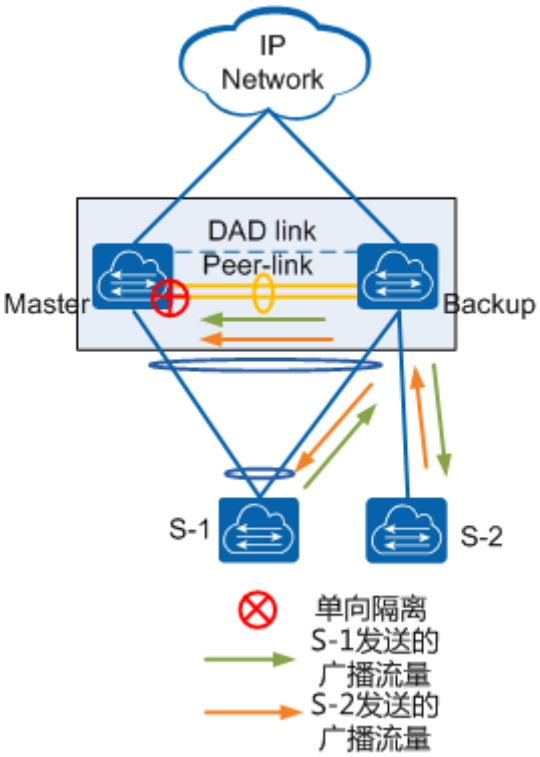
图 1-11 M-LAG 接入二层网络广播流量转发示意图



- **M-LAG接入三层网络**

  此处以M-LAG备设备的转发为例，如图1-12所示，M-LAG备设备收到广播流量后向各个下一跳转发，当流量到达M-LAG主设备时，由于peer-link与M-LAG成员接口存在单向隔离机制，到达主设备的流量不会向S-1转发。

图 1-12 M-LAG 接入三层网络广播流量转发示意图

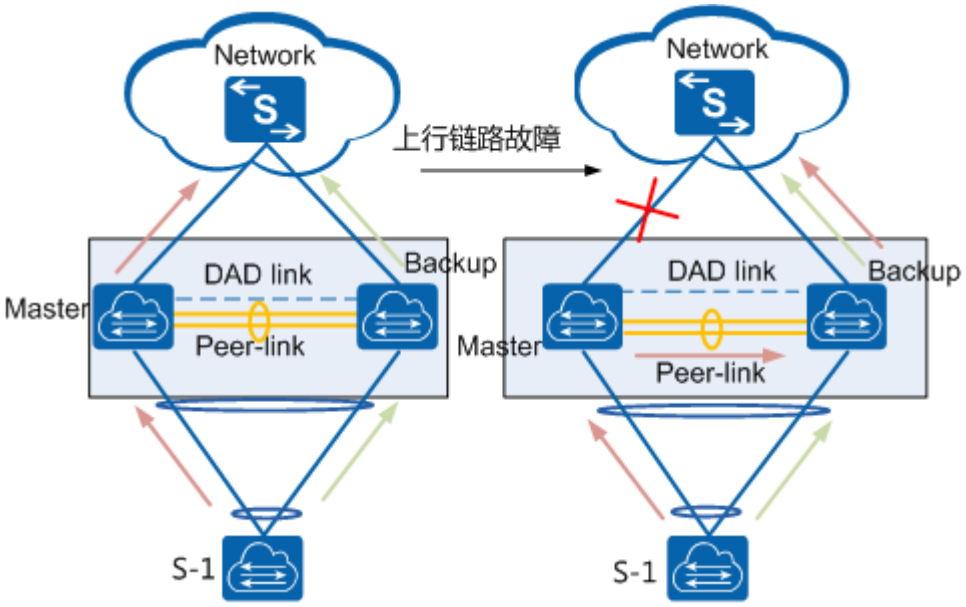


1.2.6 M-LAG 故障场景流量转发

M-LAG作为一种跨设备链路聚合的技术，把链路可靠性从单板级提高到了设备级。如果出现故障（不管是链路故障、设备故障还是peer-link故障），M-LAG都能够保证正常的业务不受影响，下面介绍M-LAG在故障情况下是如何保障业务的正常运行的。

上行链路故障

图 1-13 上行链路故障组网示意图

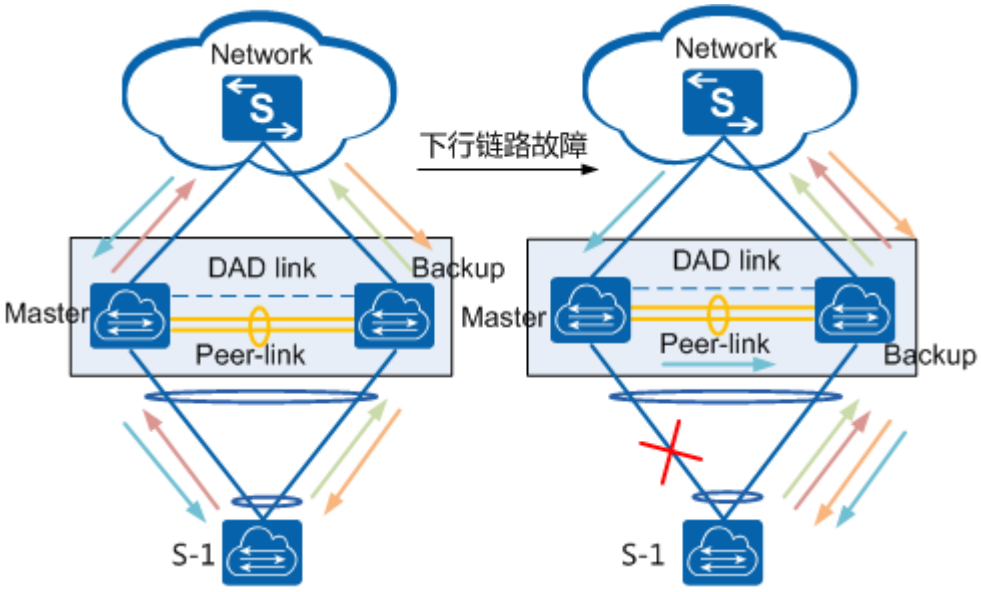


双主检测报文一般是通过管理网络传输，所以上行链路故障一般不影响M-LAG主备设备的双主检测，对于双活系统没有影响，M-LAG主备设备仍能够正常转发。如图1-13所示，M-LAG接入普通以太网场景，由于M-LAG主设备的上行链路故障，通过M-LAG主设备的流量均经过peer-link链路进行转发。

如果双主检测链路通过业务网络互通，且故障的上行链路恰好为双主检测链路，此时对于M-LAG正常工作没有影响。一旦peer-link也发生故障，双主检测无法进行，则会出现丢包现象。

下行链路故障

图 1-14 下行链路故障组网示意图

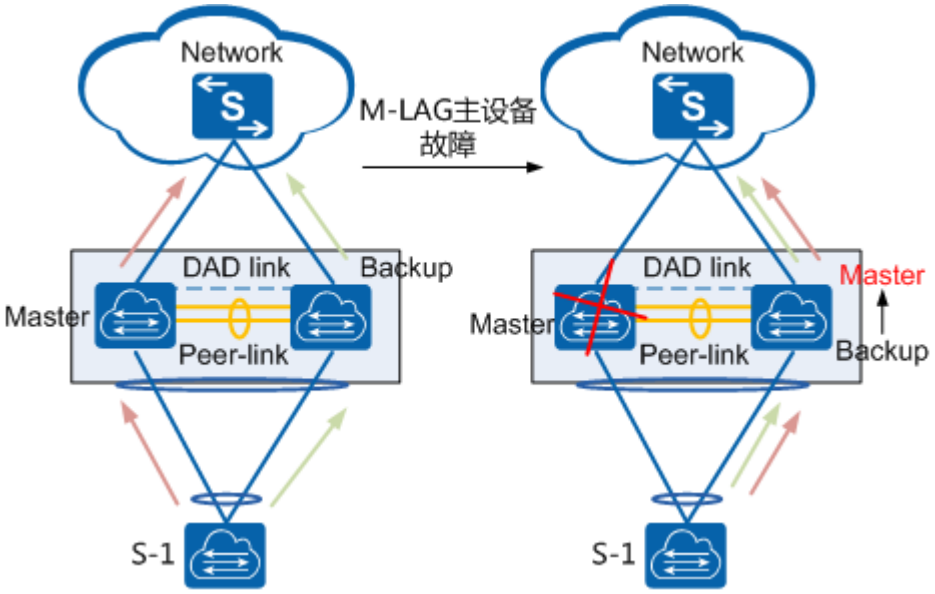


当下行M-LAG成员口故障时，DFS Group主备状态不会变化，但如果故障M-LAG成员口状态为主，则备M-LAG成员口状态由备升主，流量切换到该链路上进行转发。发生故障的M-LAG成员口所在的链路状态变为Down，双归场景变为单归场景。故障M-LAG成员口的MAC地址指向peer-link接口。在故障M-LAG成员口恢复后，M-LAG成员口状态不再回切，由备升主的M-LAG成员口状态仍为主，原主M-LAG成员口在故障恢复后状态为备。可以执行**display dfs-group** *dfs-group-id* **node** *node-id* **m-lag**命令来查看成员接口当前状态。

对于组播源在网络侧，组播成员在接入侧的组播流量，当M-LAG主设备的M-LAG成员口故障时，通过M-LAG同步报文通知对端设备进行组播表项刷新，M-LAG主备设备不再按照组播地址奇偶进行负载分担，而是所有组播流量都由端口状态Up的M-LAG备设备进行转发，反之亦然。

M-LAG 主设备故障

图 1-15 M-LAG 主设备故障组网示意图

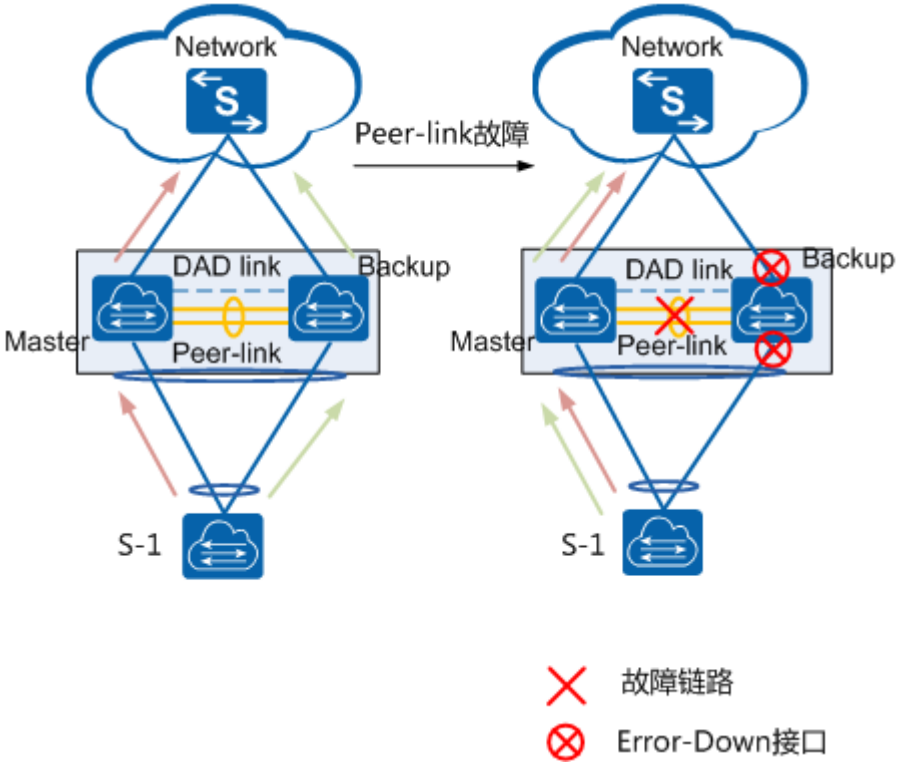


M-LAG主设备故障，M-LAG备设备将升级为主，其设备侧Eth-Trunk链路状态仍为Up，流量转发状态不变，继续转发流量。M-LAG主设备侧Eth-Trunk链路状态变为Down，双归场景变为单归场景。

如果是M-LAG备设备发生故障，M-LAG的主备状态不会发生变化，M-LAG备设备侧Eth-Trunk链路状态变为Down。M-LAG主设备侧Eth-Trunk链路状态仍为Up，流量转发状态不变，继续转发流量，双归场景变为单归场景。

peer-link 故障

图 1-16 peer-link 故障组网示意图



缺省情况下，M-LAG应用在普通以太网络、VXLAN网络或IP网络的双归接入，peer-link故障但双主检测心跳状态正常时，会触发M-LAG备设备上除逻辑端口、管理网口、peer-link接口和堆叠口以外的其他接口处于Error-Down状态。M-LAG应用在TRILL网络的双归接入，peer-link故障但双主检测心跳状态正常时，会触发M-LAG备设备上的M-LAG接口处于Error-Down状态。

peer-link故障恢复时，处于Error Down状态的M-LAG接口默认将在240s后自动恢复为Up状态，处于Error Down状态的其它接口将立即自动恢复为Up状态。

通过命令可以配置M-LAG场景下peer-link故障但双主检测心跳状态正常时，触发Error-Down的端口包括逻辑端口。如当M-LAG应用在VXLAN网络或IP网络的双归接入，peer-link故障但双主检测状态正常时，会触发M-LAG备设备上VLANIF接口、VBDIF接口、LoopBack接口以及M-LAG成员口处于Error-Down状态。

说明

在配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口后，若M-LAG系统peer-link接口故障恢复，为保证大规格VLANIF接口下的ARP同步正常，设备将在DFS Group配对成功后延迟6s恢复VLANIF接口、VBDIF接口、LoopBack接口为Up状态。此时，如果在接口下配置了接口三层协议状态延时Up时间，则VLANIF接口、VBDIF接口、LoopBack接口恢复Up状态的延迟时间为两者之和。

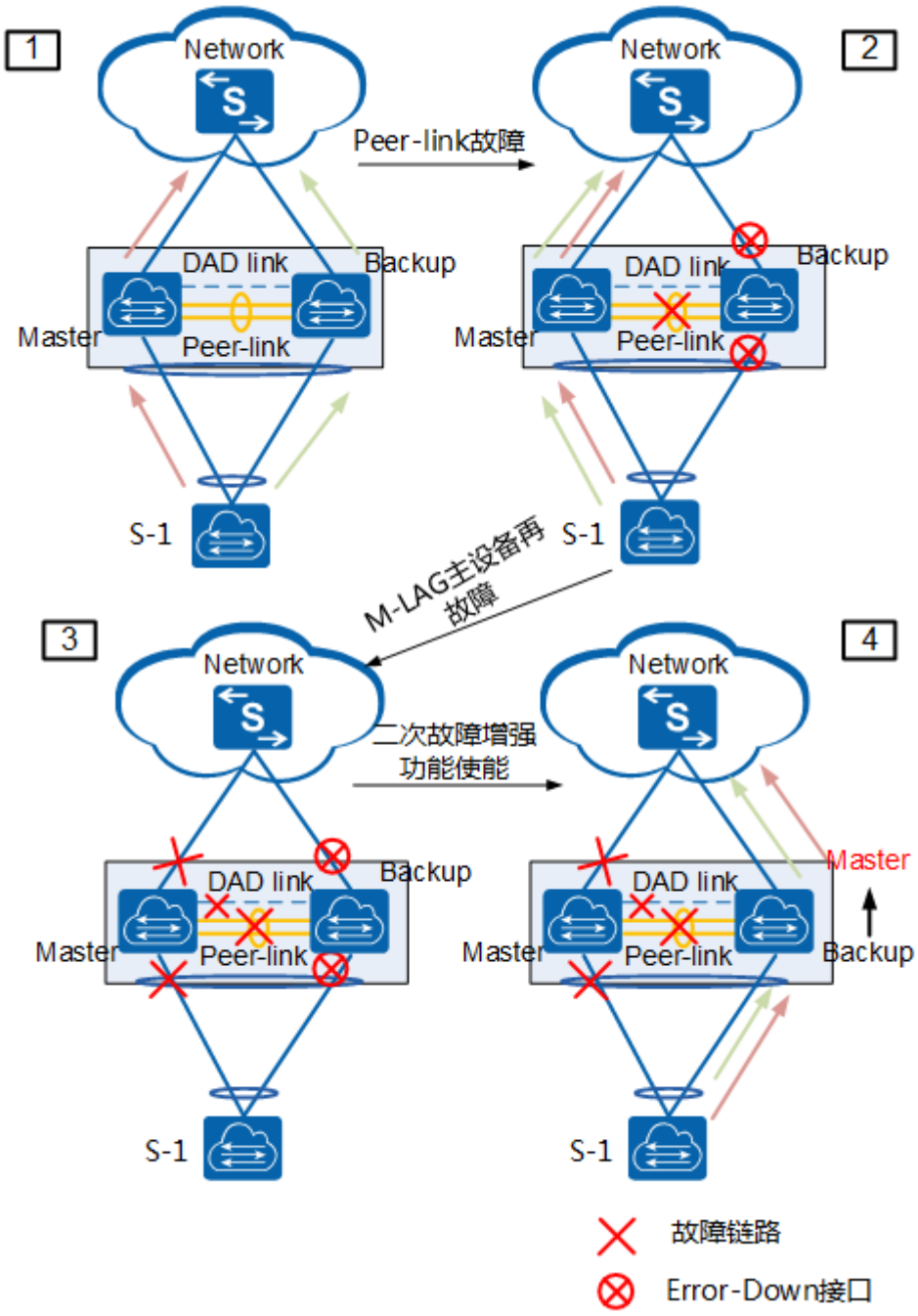
通过在端口下配置命令可以灵活配置某个端口在M-LAG场景下peer-link故障但双主检测心跳状态正常时是否将端口Error-Down。配置和设备端口Error-Down对应情况如表1-3所示。

表 1-3 设备在 peer-link 故障但双主检测正常时接口 Error-Down 情况

设备配置情况	M-LAG接入普通以太网络、VXLAN网络或IP网络
设备缺省情况	除逻辑端口、管理网口、peer-link接口和堆叠口以外的接口处于ERROR DOWN状态。
设备仅配置suspend功能	仅M-LAG成员口以及配置该功能的接口处于ERROR DOWN状态。
设备仅配置reserved功能	除配置该功能的接口、逻辑端口、管理网口、peer-link接口和堆叠口以外的接口处于ERROR DOWN状态。
设备同时配置suspend功能和reserved功能	仅M-LAG成员口以及配置suspend功能的接口处于ERROR DOWN状态。

M-LAG 二次故障（peer-link 故障+M-LAG 设备故障）

图 1-17 M-LAG 二次故障场景下使能二次故障增强功能组网示意图



如图1-17中2所示，在M-LAG应用于双归接入时，当peer-link故障但双主检测心跳状态正常会触发DFS备设备上某些端口处于Error-Down状态，此时DFS状态为主的设备继续工作。在该场景的基础上，若DFS状态为主的设备由于断电、整机故障重启等其他故障导致主设备不能工作时，由图1-17中3所示，此时M-LAG主备设备皆不能正常转发流量。

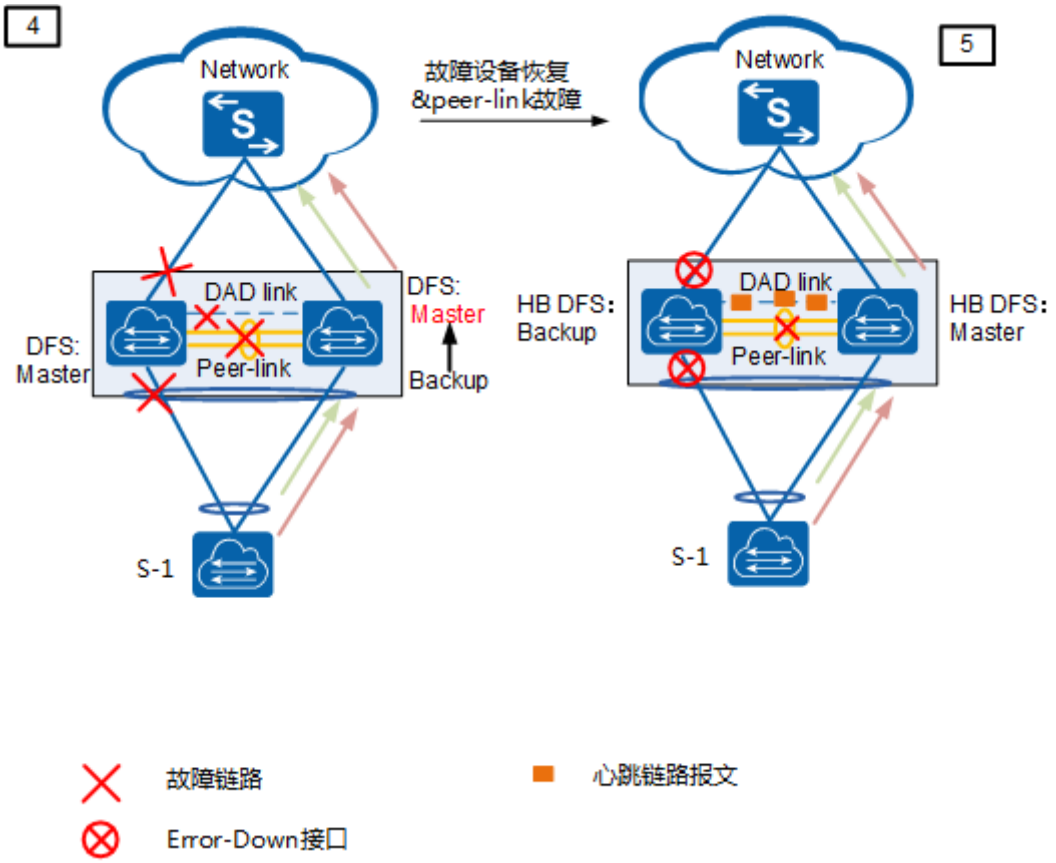
在该场景下，可以借助M-LAG二次故障增强功能来实现该故障场景下业务不中断的可靠性要求，如**图1-17**所示，通过M-LAG二次故障增强功能来说明不同的故障阶段和产生的行为：

1. Peer-link链路故障： 若Peer-link链路故障但双主检测心跳链路状态正常将会触发DFS状态为备的设备上某些端口处于ERROR DOWN（端口Error-Down范围可以参见**peer-link故障**）状态，DFS状态为主的设备继续工作。
2. DFS状态为主的设备故障：若DFS状态为主的设备在peer-link链路故障后由于断电、整机故障重启等其他故障导致不能工作时，此时M-LAG主备设备皆不能转发流量，业务中断。
3. 二次故障增强功能使能：在上述场景基础下，若M-LAG已使能二次故障增强功能，则DFS状态为备的设备会借助M-LAG双主检测机制感知到DFS主设备故障（在一定周期内接收不到任何的M-LAG双主检测心跳报文）后，将升级为DFS主设备并恢复设备上处于ERROR DOWN状态的端口为Up状态，继续转发流量。
4. 设备故障恢复：若原DFS状态为主的设备故障恢复后但peer-link故障仍故障
   - 若配置LACP M-LAG的系统ID在一定时间内切换为本设备的LACP系统ID，则在LACP协商时接入侧仅选择上行链路中的一条链路为活动链路，实际流量转发正常。
   - 若配置LACP M-LAG的系统ID为缺省情况，即系统ID不回切，M-lAG两台设备均使用同一系统ID来与接入侧设备协商，链路均能被选中成为活动链路。该场景下，由于peer-link链路仍然故障，M-LAG两端无法同步对端的优先级、系统MAC等信息，形成M-LAG两台设备双主的情况，可能导致组播流量异常。此时，如**图1-18**所示，可以借助心跳链路报文中携带必要的DFS Group协商主备的必要信息（如DFS Group优先级、系统MAC等）来协商M-LAG两台设备的HB DFS主备信息，触发HB DFS状态为备的设备上某些端口处于ERROR DOWN（端口Error-Down范围可以参见**peer-link故障**）状态，HB DFS状态为主的设备继续工作。

     说明

     若在peer-link故障后，二次故障的设备为DFS状态为备的设备，则此时不会对流量转发行为产生影响，仍由DFS状态为主的设备进行流量转发。

图 1-18 M-LAG 二次故障场景下设备故障恢复组网示意图



1.3 M-LAG 应用场景

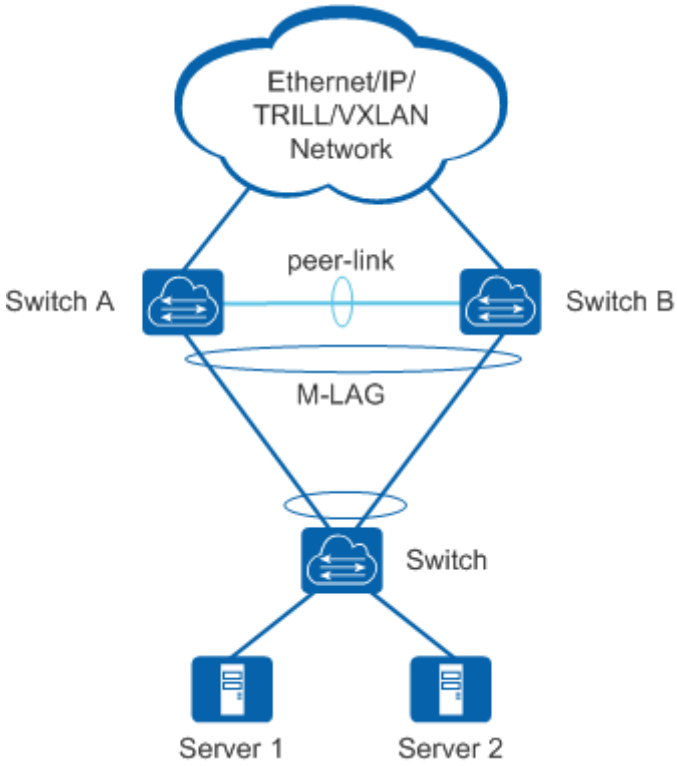
M-LAG特性主要应用于将服务器或交换机双归接入普通以太网络、TRILL（Transparent Interconnection of Lots of Links）、VXLAN（Virtual eXtensible Local Area Network）和IP网络。一方面可以起到负载分担流量的作用，另一方面可以起到备份保护的作用。由于M-LAG支持多级互联，M-LAG的组网可以分为单级M-LAG和多级M-LAG。

单级 M-LAG 场景

- **交换机的双归接入**

  如图1-19所示，为了保证可靠性，交换机在接入网络时需要考虑链路的冗余备份，采用部署MSTP等破环协议的方式可以实现，但是这种方式下链路的利用率很低，浪费大量的带宽资源。为了实现冗余备份同时提高链路的利用率，在SwitchA与SwitchB之间部署M-LAG，实现Switch的双归接入。这样SwitchA与SwitchB形成负载分担，共同进行流量转发，当其中一台设备发生故障时，流量可以快速切换到另一台设备，保证业务的正常运行。

图 1-19 交换机双归接入应用组网图



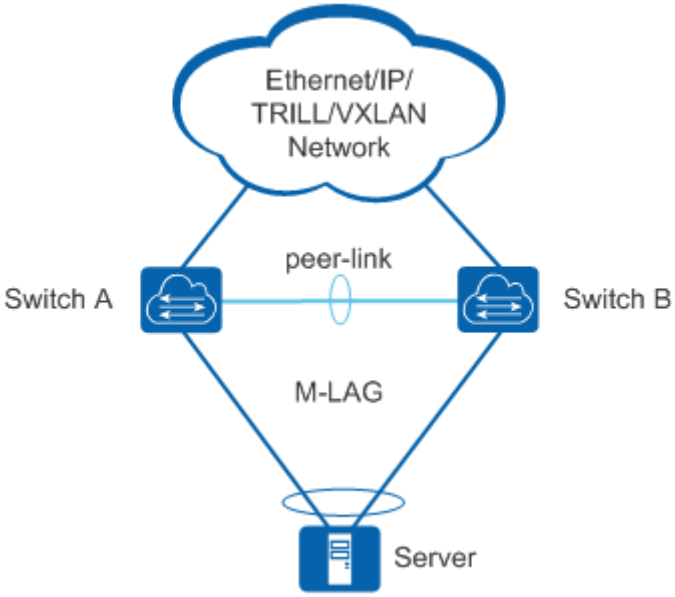
- **服务器的双归接入**

  如图1-20所示，为了保证可靠性，服务器一般采用链路聚合的方式接入网络，如果服务器接入的设备故障将导致业务的中断。为了避免这个问题的发生，服务器可以采用跨设备链路聚合的方式接入网络，在SwitchA与SwitchB之间部署M-LAG，实现服务器的双归接入。SwitchA与SwitchB形成负载分担，共同进行流量转发，当其中一台设备发生故障时，流量可以快速切换到另一台设备，保证业务的正常运行。

  说明

  服务器双归接入时的配置和一般的链路聚合配置没有差异，必须保证服务器侧和交换机侧的链路聚合模式一致，推荐两端均配置为LACP模式。

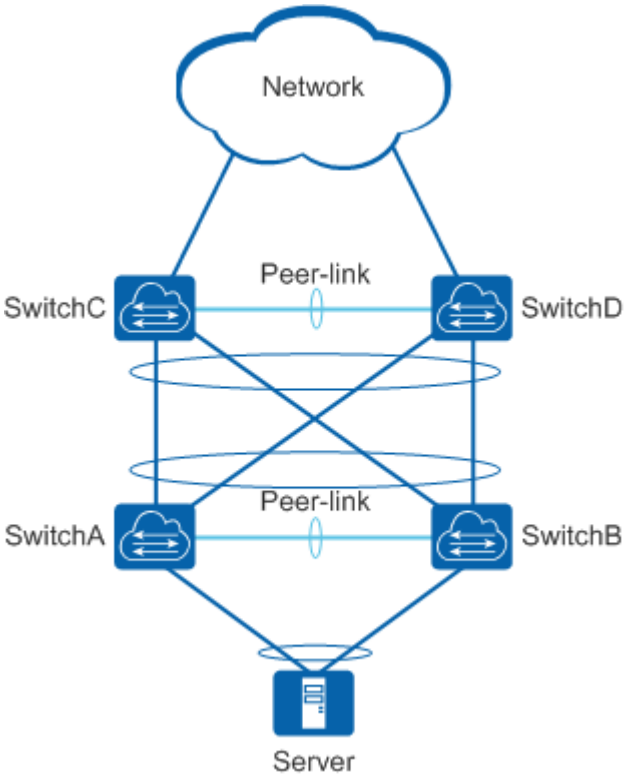
图 1-20 服务器双归接入应用组网图



多级 M-LAG 场景

如图1-21所示，SwitchA和SwitchB之间部署M-LAG后，在SwitchC和SwitchD之间部署M-LAG并与下层的M-LAG进行级联，这样不仅可以简化组网，而且在保证可靠性的同时可以扩展双归接入服务器的数量。多级M-LAG互联必须基于V-STP方式进行配置。

图 1-21 多级 M-LAG 互联应用组网图



1.4 M-LAG 配置任务概览

您可以根据实际组网采用以下方式进行M-LAG的配置：

表 1-4 M-LAG 配置任务概览

方式	描述	对应任务	配置注意点
根桥方式	M-LAG主设备和备设备均作为STP网络中的根桥且配置相同的桥ID，将两台设备模拟成同一个根桥，M-LAG主备设备在二层网络中不受其他组网变化的影响，保证正常的工作。	1. **配置根桥和桥ID** 2. **1.6.2 配置DFS Group** 3. **1.6.3 配置M-LAG一致性检查** 4. **1.6.4 配置peer-link** 5. **1.6.5 配置M-LAG成员接口** 6. **1.6.6 （可选）配置双活网关** 7. **1.6.7 （可选）配置peer-link故障场景下端口状态** 8. **1.6.8 （可选）使能M-LAG三层转发增强功能**	• 将M-LAG主设备和备设备均配置为根桥且配置相同的桥ID。 • 必须去使能peer-link接口的STP功能。
V-STP方式（推荐）	利用V-STP机制将M-LAG主设备和备设备的STP协议虚拟成一台设备的STP协议，对外呈现为一台设备进行STP协议计算。	1. **配置V-STP** 2. **1.7.2 配置DFS Group** 3. **1.7.4 配置M-LAG一致性检查** 4. **1.7.5 配置peer-link** 5. **1.7.6 配置M-LAG成员接口** 6. **1.7.7 （可选）配置双活网关** 7. **1.7.8 （可选）配置peer-link故障场景下端口状态** 8. **1.7.9 （可选）使能M-LAG三层转发增强功能**	必须使能peer-link接口的STP功能同时使能V-STP功能。

说明

- 根桥方式下，组成M-LAG的两台设备在二层网络中必须作为根桥，且不支持全根桥方式M-LAG的级联。
- V-STP方式下，组成M-LAG的设备在二层网络中可以不作为根桥，组网灵活且支持M-LAG的级联。V-STP功能还可以解决M-LAG中错误配置或错误连线导致的环路问题，所以推荐采用V-STP方式。

1.5 M-LAG 配置注意事项

涉及网元

无需其他网元配合。

License 支持

M-LAG特性是交换机的基本特性，无需获得License许可即可应用此功能。

版本支持

表 1-5 支持本特性的最低软件版本

产品	最低支持版本
CE8868EI	V200R005C10
CE8861EI	V200R005C10
CE8860EI	V100R006C00
CE8850-32CQ-EI	V200R002C50
CE8850-64CQ-EI	V200R005C00
CE7850EI	V100R005C10
CE7855EI	V200R001C00
CE6810EI	V100R005C10
CE6810LI	V100R005C10
CE6850EI	V100R005C10
CE6850HI/CE6850U-HI/ CE6851HI	V100R005C10
CE6855HI	V200R001C00
CE6856HI	V200R002C50
CE6857EI	V200R005C10
CE6860EI	V200R002C50
CE6865EI	V200R005C00
CE6870-24S6CQ-EI/ CE6870-48S6CQ-EI	V200R001C00
CE6870-48T6CQ-EI	V200R002C50
CE6875EI	V200R003C00

产品	最低支持版本
CE6880EI	V200R005C00
CE6881/CE6820/CE6863	V200R005C20
CE6881K	V200R019C10
CE6881E	V200R019C10
CE6863K	V200R019C10
CE5880EI	V200R005C10
CE5810EI	V100R005C10
CE5850EI	V100R005C10
CE5850HI	V100R005C10
CE5855EI	V100R005C10

说明

如果需要了解软件版本与交换机具体型号的配套信息，请查看**硬件查询工具**。

软件版本演进关系：

- 除CE6881、CE6881E、CE6881K、CE6863、CE6863K和CE6820

  V100R001C00 -> V100R00200 -> V100R003C00 -> V100R003C10 -> V100R005C00 -> V100R005C10 -> V100R006C00 -> V200R001C00 -> V200R002C50 -> V200R003C00 -> V200R005C00 -> V200R005C10 -> V200R019C00 -> V200R019C10
- 对于CE6881、CE6881E、CE6881K、CE6863、CE6863K和CE6820

  V200R005C20 -> V200R019C10 -> V200R020C00

特性依赖和限制

M-LAG配置的注意事项

- 组建M-LAG时，必须使用经过华为数据中心交换机认证的光模块或光电转换模块，若使用高速线缆或AOC光线缆，必须使用从华为原厂采购的线缆。非认证光模块或光电转换模块、非原厂采购的线缆的可靠性无法保证，可能导致业务不稳定。由非华为数据中心交换机认证光模块或光电转换模块、非原厂采购的线缆导致的问题，华为将不承担责任，并在原则上不予以解决。
- 组成M-LAG的两台设备类型要保持一致。如一端是SVF系统，另一端要求必须也是SVF系统；一端是CloudEngine 8800, 7800, 6800, 5800系列交换机，另一端也要是CloudEngine 8800, 7800, 6800, 5800系列交换机。推荐两台设备型号、版本均保持一致。
- 组建M-LAG的两台设备需要配置根桥和桥ID或者V-STP，对外呈现为一台设备进行STP协议计算，否则可能存在环路的风险。
- 基于根桥方式配置M-LAG时，组成M-LAG的两台设备的桥ID要配置相同，根优先级都配置为最高，保证M-LAG的两台设备为根节点。
- 在基于根桥方式配置M-LAG的场景中，不支持STP多进程；在基于V-STP配置M-LAG的场景中，V200R002C50的之前版本不支持MSTP模式，不支持STP多进程，V200R002C50及之后版本不支持MSTP模式，支持STP多进程。

- 在V-STP场景中，推荐按照如下顺序进行M-LAG的配置和物理连线。
  a. 配置V-STP。
  b. 配置DFS Group和peer-link接口。
  c. 用线缆连接M-LAG主备设备的peer-link接口。
  d. 配置M-LAG成员接口并用线缆连接M-LAG主备设备和用户侧主机（或交换设备）。
- 为了保证流量的正常转发，主备设备上绑定m-lag-id相同的M-LAG接口相关的配置必须保持一致。
- 如果配置组成M-LAG的两台设备作为网关时，无论服务器是双归接入还是单归接入，需要关注以下注意点：
  - 选择通过VRRP/VRRP6虚拟网关的方式接入，在两台设备上创建相同的VLAN，启用VLANIF接口，且相应的VLANIF接口不要配置ARP严格学习功能，否则会导致ARP表项无法同步。VBDIF接口下不支持通过VRRP/VRRP6虚拟网关的方式配置。M-LAG双归接入仅支持VRRP/VRRP6双活模式。
  - （推荐）选择通过配置VLANIF接口/VBDIF接口下相同的IP地址和MAC地址的方式接入（该方式从V100R006C00版本开始支持）。在V200R002C50版本及之前版本，VLANIF接口/VBDIF接口下配置相同的IP和MAC地址，会产生IP地址和MAC地址冲突告警hwEthernetARPMACIPConflict，该常见触发此告警属于正常现象，可以忽略。如果要屏蔽该告警，可以执行命令**undo snmp-agent trap enable feature-name arp trap-name hwethernetarpmacipconflict**来关闭告警开关。关闭该告警后，则无法通过告警检查网络中存在的环路，可能造成用户业务中断，请谨慎操作。在V200R003C00SPC810版本及之后版本，若VLANIF接口/VBDIF接口下配置相同的IP和MAC地址，则不会产生冲突告警。
- 在M-LAG故障场景下，三层流量的收敛性能与设备、端口下学习到的ARP表项成正比，对于ARP量较大的场景，收敛性能会较差。
- 为了防止设备重启和peer-link故障导致STP网络震荡、保证故障回切的性能，在M-LAG接口、peer-link接口和其他业务接口下配置延时Up时间为30s及以上。缺省情况下，M-LAG接口上报Up状态的延时时间在V200R005C00版本及其之前版本为120s，在V200R005C10版本及其之后版本为240s。
- 在设备重启或单板复位后，接口物理状态切换Up状态，但上层协议模块状态未满足转发要求，导致流量丢包。为了保证故障回切性能，缺省情况下，M-LAG接口上报Up状态的延时时间在V200R005C00版本及其之前版本为120s，在V200R005C10版本及其之后版本为240s。若此时同时在VLANIF接口上配置三层协议状态延时Up时间，则需要保证M-LAG成员口的延时Up时间长于VLANIF接口的延时时间，否则ND同步失败的表项依赖于ND Miss触发学习。
- 从V200R002C50版本开始，设备支持在M-LAG成员接口配置静态MAC地址。当M-LAG成员口故障时，故障成员口的MAC地址指向peer-link接口。
- 在V200R005C10及其之前版本，若M-LAG双归接入场景下配置指定M-LAG成员口为出接口的静态ARP，则对应M-LAG成员口故障时，设备不支持将出接口指向peer-link接口，从而导致流量无法正常转发。请勿在M-LAG双归接入时配置指定M-LAG成员口为出接口的静态ARP表项。在V200R019C00及其之后版本，除CE6810LI、CE6810EI、CE6850EI及CE5800系列（除CE5880EI）外的设备支持通过使能M-LAG三层转发增强功能，为报文出端口是M-LAG成员接口的动静态ARP表项申请备份的FRR资源，出接口指向peer-link接口，形成主备路径转发。但为静态ARP表项申请的FRR资源在M-LAG成员口状态为Down且对应的VLANIF接口仍为Up时将不会释放，导致FRR资源消耗增加。
- 在V200R005C00及其之前版本，M-LAG双归接入场景下，若配置指定M-LAG成员口为出接口的静态IPv6邻居表项，则对应M-LAG成员口故障时，设备不支持将出

接口指向peer-link接口，从而导致流量无法正常转发。请勿在M-LAG双归接入时配置指定M-LAG成员口为出接口的静态IPv6邻居。在V200R005C10版本，除CE6810LI、CE5880EI以及CE6880EI 的设备支持通过使能M-LAG三层转发增强功能，为报文出端口是M-LAG成员接口的所有ND表项申请备份的FRR资源，出接口指向peer-link接口，形成主备路径转发。但为静态IPv6邻居表项申请的FRR资源在M-LAG成员口状态为Down且对应的VLANIF接口仍为Up时，将不会释放对应的系统资源。从V200R019C00版本开始，仅CE6810LI不支持使能M-LAG三层转发增强功能。

- 在配置独立双主检测链路时，建议使用主接口作为双主检测链路接口。如果使用VLANIF接口，则需要保证peer-link接口不允许通过该VLAN，否则会有环路或MAC漂移的现象。
- 为了提高可靠性，将聚合链路的成员接口分布在不同的单板上，防止某一单板故障导致peer-link故障。
- 在V200R002C50及之前版本，当设备通过M-LAG双归接入到网关设备，在M-LAG设备上配置大量二层子接口的同时在设备光接口插入光电模块，该操作会导致重启M-LAG系统其中一台设备后，业务会有较长时间的丢包现象。此时先手动将流量切换到另外一台M-LAG设备，再进行该设备升级重启。
- M-LAG场景下，在和NAS(Network Attached Storage)设备或者负载均衡器对接时，由于一些NAS设备或者负载均衡器设备（例如，使能了Auto Last Hop功能的F5负载均衡器）不会按正常的ARP请求逻辑学习网关MAC地址，而是通过分析从网关来的数据流，使用第一次收到的数据流的源MAC作为网关MAC地址。此时，除CE6870EI和CE6875EI外的交换机需要在组建M-LAG的两台设备对应的VLANIF接口上配置相同的MAC地址，否则经由NAS设备或者负载均衡器的转发的流量会由于peer-link与M-LAG成员接口之间存在的单向隔离机制出现流量不通的问题。
- 在设备配置M-LAG一致性检查为严格模式时，若设备检测出M-LAG两台设备存在Type1类型的配置不一致情况（设备会将M-LAG备设备上成员口置于ERROR DOWN状态，且触发设备对Type 1类型配置检查不一致的告警），请设备管理员立即对相应配置进行调整，且管理员勿对设备进行重启操作。

  如果管理员未对相应配置进行调整且重启M-LAG主设备，则可能导致主设备在设备恢复时，由于M-LAG两端重新协商，Type1类型配置不一致导致Error-Down M-LAG备设备端口。 此时，M-LAG主设备的端口由于M-LAG成员口延时Up的机制存在，导致M-LAG主备设备皆不能正常转发流量，业务中断。

  若M-LAG设备未使能M-LAG配置一致性检查功能且M-LAG主备设备存在Type1和Type2类型的不一致配置，此时可能导致流量转发异常。请客户针对M-LAG主备设备进行手动调整，保持M-LAG两端Type1和Type2类型的配置保持一致，并使能M-LAG配置一致性检查功能。
- 在接入设备采用二层子接口方式双归接入M-LAG主备设备场景下，若一侧二层子接口状态为Down，会导致南北向流量由于M-LAG单向隔离机制（设备主接口双活接入M-LAG，则隔离由peer-link口发往M-LAG成员口的除三层已知单播报文外的所有流量）无法由peer-link链路绕行转发，出现流量丢包现象。
- 在配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口后，若M-LAG系统peer-link接口故障恢复，为保证大规格VLANIF接口下的ARP同步正常，设备将在DFS Group配对成功后延迟6s恢复VLANIF接口、VBDIF接口、LoopBack接口为Up状态。此时，如果在接口下配置了接口三层协议状态延时Up时间，则VLANIF接口、VBDIF接口、LoopBack接口恢复Up状态的延迟时间为两者之和。
- 仅CE6857EI、CE6865EI、CE6881、CE6881K、CE6863、CE6863K、CE6881E、CE6820、CE8861EI和CE8868EI设备支持通过M-LAG成员口与对接设备建立BFD会话。

- 在通过M-LAG成员口与对接设备建立BFD会话场景下，若M-LAG两端设备通过peer-link接口同步BFD协议报文时，BFD协议报文会进入优先级为6的端口队列进行转发。若此时存在业务报文进入优先级为7的端口队列进行转发且端口采用默认的PQ调度方式，当业务报文在一段时间内以100%的输出链路速率到达，则会造成由于BFD协议报文被丢弃而出现的BFD会话震荡。若存在其他协议报文通过peer-link接口转发，可能同样由于调度优先级问题而被丢弃。
- M-LAG成员，从V200R003C00之前版本升级到V200R019C10及之后版本，升级过程中M-LAG配置一致性检查会出现失败；从V200R003C00到V200R005C10之间任一版本升级到V200R019C10及之后版本，在版本升级过程中不支持M-LAG配置一致性检查，升级完成后再进行M-LAG配置一致性检查。

M-LAG与其他业务同时部署时的注意事项

当M-LAG业务和其他多种业务同时配置时，可能会因为ACL资源不足使部分业务下发失败。在M-LAG场景中，可以保证同时配置成功的业务可以参见**CloudEngine系列交换机 ACL技术专题**中的“CSS/M-LAG业务叠加场景”。

对于多数业务特性，M-LAG中支持的情况和普通物理设备支持情况相同，但是以下特性有区别，如**表1-6**所示。

表 1-6 M-LAG 与其他业务特性约束

特性	注意事项
堆叠	设备支持先组建堆叠，再用堆叠系统作为独立单元建立M-LAG。
SVF	设备支持先组建SVF，再用SVF系统作为独立单元建立M-LAG。SVF系统上M-LAG成员口只能同时属于Spine或者Leaf，不支持部分属于Spine，部分属于Leaf。
VBST	M-LAG特性与VBST特性互斥，不能同时配置。
CFM	M-LAG特性与CFM特性互斥，不能同时配置。
GVRP	Eth-Trunk接口视图下的M-LAG功能与GVRP功能互斥。
DHCP	• 组成M-LAG的两台设备不支持配置DHCP Snooping。 • DHCP Relay需要在组成M-LAG的两台设备同时配置。 • DHCP Server需要在组成M-LAG的两台设备同时配置，并且两台设备地址池的地址不能重叠。
ARP	针对CE5855EI、CE7850EI、CE7855EI、CE6850HI、CE6850U-HI、CE6851HI、CE6855HI、CE6856HI、CE6857EI、CE6865EI、CE8850EI、CE8861EI、CE8868EI、CE6860EI和CE8860EI设备，配置转发资源模式为large-arp模式或指定了ARP表项的UFT灵活资源模式和配置**m-lag forward layer-3 ipv4 enhance enable**命令互斥。

特性	注意事项
IP单播路由	● M-LAG设备和被接入设备无法通过M-LAG成员口建立路由邻居关系。 ● 两台M-LAG设备之间如果需要建立路由邻居关系，建议手工在两台M-LAG设备上配置Router ID，如果设备自行获取Router ID，可能由于Router ID冲突导致路由邻居建立不成功。
IPv4组播	对于V100R006C00之前的版本：M-LAG不支持IPv4三层组播。 从V100R006C00版本开始：由单机、堆叠或SVF组成的M-LAG均支持IPv4三层组播。配置时需要注意： ● M-LAG主备设备之间除了peer-link链路外，还需要有直连的三层链路，并且该链路两端的接口需要关闭STP功能。 ● M-LAG主备设备上的组播配置必须保持一致。 ● M-LAG主备设备上所有需要运行三层组播业务的VLANIF接口必须使能PIM-SM和IGMP，并且对应的VLAN内必须使能IGMP Snooping。 ● M-LAG主备设备上组播用户侧接口需要配置PIM Silent。 ● 如果M-LAG主备设备之间的三层链路通过VLANIF口互联，则该VLANIF口上需要使能PIM协议，且该VLANIF对应的VLAN在peer-link接口上必须配置为不允许通过。 ● 如果单播路由选路时peer-link链路被选为到达RP或组播源的最优链路，可能会导致出接口为peer-link接口的组播流量不通。为了避免此问题，需要保证M-LAG主备设备之间的三层链路的路由开销值小于或等于peer-link链路的路由开销值，使得单播路由选路时选择到M-LAG主备设备之间的三层链路。 在组播接收者（Receiver）双归接入M-LAG的场景中： ● 对于V200R003C00之前的版本，仅M-LAG主成员接口转发组播流量到Receiver。 ● 从V200R003C00版本开始，M-LAG主备成员接口均可以转发组播流量到Receiver，实现负载分担。负载分担的规则是：组播组地址最后一位十进制数字为奇数时（例如225.1.1.1），流量由M-LAG主成员接口转发；组播组地址最后一位十进制数字为偶数时（例如225.1.1.2），流量由M-LAG备成员接口转发。 ● 当M-LAG主备设备版本不一致时，组播流量转发规则以低版本为准。

特性	注意事项
IPv6组播	对于V200R003C00之前的版本：M-LAG不支持IPv6三层组播。 从V200R003C00版本开始：对于CE6870EI和CE6875EI，由单机或堆叠组成的M-LAG均支持IPv6三层组播，其余款型不支持。配置时需要注意： ● M-LAG主备设备之间除了peer-link链路外，还需要有直连的三层链路，并且该链路两端的接口需要关闭STP功能。 ● M-LAG主备设备上的组播配置必须保持一致。 ● M-LAG主备设备上所有需要运行三层组播业务的VLANIF接口必须使能PIM-SM（IPv6）和MLD，并且对应的VLAN内必须使能MLD Snooping。 ● M-LAG主备设备上组播用户侧接口需要配置PIM Silent（IPv6）。 ● 如果M-LAG主备设备之间的三层链路通过VLANIF口互联，则该VLANIF口上需要使能PIM（IPv6）协议，且该VLANIF对应的VLAN在peer-link接口上必须配置为不允许通过。 ● 如果单播路由选路时peer-link链路被选为到达RP或组播源的最优链路，可能会导致出接口为peer-link接口的组播流量不通。为了避免此问题，需要保证M-LAG主备设备之间的三层链路的路由开销值小于或等于peer-link链路的路由开销值，使得单播路由选路时选择到M-LAG主备设备之间的三层链路。 对于V200R003C00及之后版本，在组播接收者（Receiver）双归接入M-LAG的场景中： M-LAG主备成员接口均可以转发组播流量到Receiver，实现负载分担。负载分担的规则是：组播组地址最后一位十六进制数字为奇数时（例如FF1E::1、FF1E::B），流量由M-LAG主成员接口转发；组播组地址最后一位十六进制数字为偶数时（例如FF1E::2、FF1E::A），流量由M-LAG备成员接口转发。
FCoE	当设备上同时存在FSB和FCF或者FSB和NPV的配置时，M-LAG特性无法使用。
IPSG	组成M-LAG的两台设备不支持配置IPSG。
VPLS	组成M-LAG的两台设备不支持作为PE设备。

特性	注意事项
VXLAN	● 在V200R003C00版本之前，M-LAG与流封装类型为QinQ的二层子接口互斥，不能同时配置。在V200R003C00版本及之后版本，M-LAG支持与流封装类型为终结型QinQ的二层子接口同时配置，不支持与透传型QinQ二层子接口同时配置。 ● 对于V200R019C10之前的版本，部署了M-LAG的设备不支持再配置Segment VXLAN实现DCI二层互联。对于V200R019C10版本及之后版本，CE6881、CE6881K、CE6863、CE6863K、CE6881E组成的M-LAG支持配置映射VNI模式的Segment VXLAN实现DCI二层互联，其他款型设备组成的M-LAG仍然不支持配置Segment VXLAN实现DCI二层互联。 ● 对于CE5880EI、CE6881、CE6881K、CE6863、CE6863K、CE6881E和CE6880EI，IPv6 VXLAN与M-LAG功能互斥，不可以同时部署。 ● 在分布式网关的场景下，当部署VXLAN双活接入且网关处于环回模式时，网络中的不同M-LAG系统的NVE接口必须配置成不同的MAC地址。例如设备A和设备B组成M-LAG 1系统，设备C和设备D组成M-LAG 2系统，那么M-LAG 1和M-LAG 2的NVE接口必须配置成不同的MAC地址。 ● 在设备接入VXLAN网络中，若在VBDIF接口下配置双活网关，不支持通过VRRP/VRRP6方式来配置。
风暴控制	不建议在peer-link的物理成员端口上配置组播报文风暴控制功能，否则可能导致M-LAG同步报文被抑制，引起M-LAG系统的数据报文转发异常。
端口安全	● 配置端口安全后，接口学习到的动态MAC地址会转换为安全动态MAC地址或Sticky MAC地址。安全动态MAC地址和Sticky MAC地址为静态MAC地址类型，无法在M-LAG两边设备上通过peer-link接口同步。 ● 双归M-LAG口上配置了端口安全时，可能存在M-LAG两边设备上安全动态MAC地址或Sticky MAC地址不一致的情况。

1.6 基于根桥方式配置 M-LAG

M-LAG主设备和备设备均作为STP网络中的根桥且配置相同的桥ID，将两台设备模拟成同一个根桥，M-LAG主备设备在二层网络中不受其他组网变化的影响，保证正常的工作。

1.6.1 配置根桥和桥 ID

背景信息

采用根桥方式配置M-LAG时，必须将M-LAG主设备和备设备均作为STP网络中的根桥且配置相同的桥ID，将两台设备模拟成同一个根桥。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**stp** [**instance** *instance-id*] **root primary**，配置当前设备为根桥设备。

缺省情况下，交换设备不作为任何生成树的根桥。配置后该设备优先级值自动为0，将不能更改设备优先级。

如果不指定**instance**，则配置设备在实例0上为根桥设备。

步骤3 执行命令**stp bridge-address** *mac-address*，配置设备参与生成树计算的桥MAC。

缺省情况下，当前设备参与生成树计算的桥MAC是设备的MAC地址。在配置M-LAG主备设备桥MAC时，建议选取两台设备中MAC地址较小的作为参与生成树计算的桥MAC。

步骤4 执行命令**commit**，提交配置。

----结束

1.6.2 配置 DFS Group

背景信息

动态交换服务组DFS Group，主要用于设备之间的配对。为了实现双主检测报文的交互，DFS Group需要绑定IP地址，绑定的IP地址用于和对端进行通信。

设备双归接入普通以太网络、VXLAN或IP网络时，需要配置DFS Group绑定IP地址，前提是已经配置两台双归设备对应三层接口的IP地址且保证互通。当接入的是VPN网络时，还需要配置DFS Group绑定VPN实例，前提是设备上已经创建VPN实例。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**dfs-group** *dfs-group-id*，创建DFS Group并进入DFS-Group视图。

步骤3 根据实际场景选择配置DFS Group绑定IP地址。

设备双归接入普通以太网络、VXLAN或IP网络时，配置DFS Group绑定IP地址。以下任选一种即可，不支持同时配置。

- 执行命令**source ip** *ip-address* [**vpn-instance** *vpn-instance-name*] [**peer** *peer-ip-address* [**udp-port** *port-number*]]，配置DFS Group绑定IPv4地址和VPN实例。
- 执行命令**source ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] [**peer** *peer-ipv6-address* [**udp-port** *port-number*]]，配置DFS Group绑定IPv6地址和VPN实例。

若在配置DFS Group绑定设备互通的心跳IP时同时指定对端设备的心跳IP和UDP通信端口号，则在该配置生效时，M-LAG两端设备即开始收发心跳报文并协商HB DFS主备状态。在二次故障增强场景下，若原DFS状态为主的设备故障恢复且peer-link链路仍然故障，则M-LAG两台设备根据HB DFS的主备状态来触发状态为备的设备相应接口Error-Down，避免M-LAG两台设备形成双主状态而导致的流量异常，提高设备可靠性。

步骤4 （可选）执行命令**priority** *priority*，配置DFS Group的优先级。

优先级用于两台设备间进行主备协商，值越大优先级越高，优先级高的为主用设备。

如果优先级相同，那么比较两台设备的系统MAC地址，MAC地址较小的为主用设备。

缺省情况下，DFS Group的优先级为100。

步骤5 （可选）执行命令**m-lag up-delay** *value* [**auto-recovery interval** *interval-time*]，配置M-LAG接口上报Up状态的延时时间。

为了保证故障回切的性能，针对设备重启、子卡复位或Peer-link故障恢复的场景，M-LAG接口默认配置了240秒的延迟Up时间，未配置延迟自动恢复时间。

步骤6 （可选）执行命令**set lacp system-id switch-delay** { *switch-delay-time* | **immediately** }，配置LACP M-LAG的系统ID延迟切换时间。

缺省情况下，LACP M-LAG的系统ID不切换。**immediately**表示LACP M-LAG的系统ID立即切换；*switch-delay-time*为0时表示LACP M-LAG的系统ID不做切换。

步骤7 （可选）执行命令**authentication-mode hmac-sha256 password** *password*，指定DFS Group协议报文所使用的验证模式及验证口令。

缺省情况下，没有配置DFS Group协议报文的验证模式。

步骤8 （可选）执行命令**dfs-master led enable**，使能堆叠状态指示灯显示DFS Group主备状态功能。

缺省情况下，堆叠状态指示灯显示DFS Group主备状态功能未使能。

在使能堆叠状态指示灯显示DFS Group主备状态功能后，切换堆叠指示灯显示为DFS主备状态，DFS状态为主的设备堆叠状态显示灯常亮，而DFS状态为备的设备堆叠状态显示灯常灭。

步骤9 （可选）执行命令**dual-active detection error-down** { **delay** *delay-time* | **disable** }，配置关闭或者延迟执行在peer-link故障但双主检测心跳状态正常时Error-Down备设备上除管理网口、peer-link接口和堆叠口以外接口的动作。

缺省情况下，在peer-link故障但双主检测心跳状态正常会触发备设备上除管理网口、peer-link接口和堆叠口以外的接口处于Error-Down。

当接入设备以单归方式三层接入M-LAG主备设备时，在peer-link故障但双主检测心跳状态正常即M-LAG双主场景下，并不影响设备流量的转发。此时，可以执行**dual-active detection error-down**命令用来配置关闭或者延迟执行在peer-link故障但双主检测心跳状态正常时Error-Down备设备上除管理网口、peer-link接口和堆叠口以外接口的动作，避免流量丢包。

在接入设备以M-LAG双归接入或者二层接入方式接入M-LAG主备时，不可以关闭或者延迟Error-Down动作。

步骤10 （可选）执行命令**dual-active detection enhanced enable**，使能M-LAG场景下二次故障增强功能。

在M-LAG应用于双归接入时，当peer-link故障但双主检测心跳状态正常会触发DFS备设备上某些端口处于Error-Down状态，此时DFS状态为主的设备继续工作。在该场景的基础上，若DFS状态为主的设备由于断电、整机故障重启等其他故障导致主设备不能工作时，M-LAG主备设备皆不能正常转发流量。

此时，可以使能M-LAG二次故障增强功能，在Peer-Link故障且DFS设备二次故障后借助双主检测机制，DFS状态为备的设备在感知DFS状态为主的设备故障后，恢复备设备上处于Error-Down状态的端口，继续转发流量，从而提升二次故障场景下业务不中断的可靠性。

在M-LAG二次故障场景下，若设备故障恢复后peer-link链路故障仍然存在，则有可能导致M-LAG形成双主，建议在配置DFS Group绑定IP地址时同时指定对端设备的IP地址。如此，在设备故障恢复后，若peer-link故障仍然存在，则触发HB DFS状态为备的设备上相应端口Error-Down，避免M-LAG设备在双主情况下出现的流量异常。

步骤11 （可选）执行命令**dual-active detection error-down mode routing-switch**，配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口。

缺省情况下，M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围不包括逻辑端口。M-LAG应用于TRILL网络的双归接入时，peer-link故障但双主检测状态正常时会触发备设备上M-LAG接口处于ERROR DOWN状态。当M-LAG应用于普通以太网络或IP网络的双归接入时，peer-link故障但双主检测状态时正常会触发备设备上除逻辑端口、配置reserved功能的接口、管理网口、peer-link接口和堆叠口以外的物理接口处于ERROR DOWN状态。

在配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口后，当M-LAG应用于IP网络或者VXLAN网络时，仅会触发备设备上VLANIF接口、VBDIF接口、LoopBack接口以及M-LAG成员接口处于ERROR DOWN状态。

说明

在配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口后，若M-LAG系统peer-link接口故障恢复，为保证大规格VLANIF接口下的ARP同步正常，设备将在DFS Group配对成功后延迟6s恢复VLANIF接口、VBDIF接口、LoopBack接口为Up状态。此时，如果在接口下配置了接口三层协议状态延时Up时间，则VLANIF接口、VBDIF接口、LoopBack接口恢复Up状态的延迟时间为两者之和。

步骤12 （可选）执行命令**peer-link mac-address remain enable**，使能特定条件下不再触发本端或对端设备删除peer-link接口上对应的MAC地址。

缺省情况下，未配置特定条件下不再触发本端或对端设备删除peer-link接口上对应的MAC地址。

步骤13 （可选）执行命令**pim synchronize enable**，使能DR优先级修改时M-LAG备设备自动同步DR优先级。

缺省情况下，未使能DR优先级修改时同步DR优先级。在两组M-LAG双活系统场景下，M-LAG的成员均接入PIM网络，一组M-LAG的DR优先级高，一组M-LAG的DR优先级低。当DR优先级高的一方出现故障，通过新一轮的竞选另外一方的M-LAG主设备上的DR优先级变高，为保证整个M-LAG系统组播数据不丢包，需要配置DR优先级同步功能，将M-LAG备设备上的DR优先级同步变高。

步骤14 （可选）执行命令**vrrp synchronize enable**，使能VRRP优先级修改时M-LAG备设备自动同步VRRP优先级。

缺省情况下，未使能VRRP优先级修改时同步VRRP优先级。两组M-LAG双活系统组成VRRP主备备份。当VRRP主用设备发生异常导致报文不能转发时，VRRP备用设备的M-LAG主设备上的优先级变高，为保证整个M-LAG系统不丢包，需要配置VRRP优先级同步功能，将M-LAG备设备上的VRRP优先级同步变高。

步骤15 执行命令**commit**，提交配置。

----结束

1.6.3 配置 M-LAG 一致性检查

前提条件

部署M-LAG系统的两端设备之间的DFS Group已配对成功并协商出主备关系。

背景信息

M-LAG配置一致性检查将设备配置分为两类，如表1-7所示，分别为关键配置（Type 1）和一般配置（Type 2）。根据对关键配置检查不一致时的处理方式，M-LAG一致性又分为严格模式（strict）和松散模式（loose）。

- 关键配置（Type 1）：如果在M-LAG系统两端设备不一致，会导致成环、状态正常但长时间丢包等问题。

  严格模式下，如果M-LAG两端设备存在Type 1配置不一致，会导致M-LAG备设备上成员口处于ERROR DOWN状态，且触发设备对Type 1类型配置检查不一致的告警。

  松散模式下，如果M-LAG两端设备存在Type 1配置不一致，则会触发设备对两种类型配置检查不一致的告警。
- 一般配置（Type 2）：如果在M-LAG系统两端设备不一致，可能会导致M-LAG运行状态异常。与Type 1类型的配置相比较而言，Type 2类型的配置问题更容易被发现，对组网环境的影响也相对较小。

  无论处于何种模式，如果M-LAG两端设备存在以下Type 2配置不一致，则会触发设备对两种类型配置检查不一致的告警。

表 1-7 M-LAG 配置一致性检查配置分类列表

视图	配置	类型
全局	STP功能是否使能	Type 1
	STP工作模式配置	
	BPDU保护功能是否使能	
	STP多生成树实例与VLAN的映射关系配置 **说明** 设备默认仅检查ID为0的STP进程内多生成树实例与VLAN的映射关系。	
M-LAG成员口	STP功能是否使能	
	STP端口的Root保护功能是否使能	
	M-LAG成员接口的LACP模式配置	
全局	VLAN配置	Type 2

视图	配置	类型
	静态MAC地址表项 ● 静态MAC地址表项指定接口为M-LAG成员口 ● VXLAN隧道的静态MAC地址表项	
	动态MAC的老化时间	
	静态ARP表项 ● 短静态ARP表项 ● 长静态ARP表项 – 若静态ARP表项指定出接口，则仅检查出接口为M-LAG成员口的静态ARP。 – 若静态ARP表项指定所属VLAN，则直接比较VLAN ID。 – 若静态ARP表项指定出接口和所属VLAN，则直接比较出接口为M-LAG成员口的静态ARP表项和VLAN ID。 – VXLAN IPv4隧道的静态ARP表项 **说明** 设备不支持检查指定VPN实例的短静态ARP表项，若长静态ARP表项的出接口为M-LAG成员口且绑定了VPN实例或者所属VLAN对应的VLANIF接口绑定了VPN实例，设备同样不支持检查该静态ARP表项。	
	动态ARP的老化时间	
	广播域桥域BD（Bridge Domain）配置 ● BD ID ● BD关联VNI	

视图	配置	类型
	VBDIF接口配置 ● VBDIF接口的BD ID ● VBDIF接口IPv4地址 ● VBDIF接口IPv6地址 ● VBDIF接口VRRP4备份组 ● VBDIF接口MAC地址 ● VBDIF接口状态 **说明** 对于VBDIF接口MAC地址，设备默认仅检查虚拟MAC地址。 针对IPv6地址以及VRRP4备份组的配置检查，仅在VBDIF接口Up时才进行。若VBDIF接口状态为Down，则认为该接口下没有相关配置。	
	VLANIF接口配置 ● VLAN ID ● VLANIF接口IPv4地址 ● VLANIF接口IPv6地址 ● VLANIF接口VRRP4备份组 ● VLANIF接口MAC地址 ● VLANIF接口状态 **说明** 对于VLANIF接口MAC地址，设备默认仅检查虚拟MAC地址。 针对IPv6地址以及VRRP4备份组的配置检查，仅在VLANIF接口Up时才进行。若VLANIF接口状态为Down，则认为该接口下没有相关配置。	
M-LAG成员口	STP端口优先级配置	
	接口加入VLAN配置	
	M-LAG成员口参数配置	

视图	配置	类型
	M-LAG成员口所属Eth-Trunk接口成员口个数 **说明** 仅比较Eth-Trunk接口的成员口数量，对于成员口物理状态Up/Down或者成员口带宽不予检查。	

操作步骤

- 配置M-LAG一致性检查功能：
  a. 执行命令**system-view**，进入系统视图。
  b. 执行命令**dfs-group** *dfs-group-id*，创建DFS Group并进入DFS-Group视图。
  c. 执行命令**consistency-check enable mode** { **strict** | **loose** }，使能M-LAG配置一致性检查并指定检查模式。

     缺省情况下，M-LAG配置一致性检查功能处于去使能状态。
  d. 执行命令**commit**，提交配置。
- 检查一致性检查运行状态及M-LAG系统主备设备的相关配置信息：
  - 执行命令**display dfs-group**，查看M-LAG一致性检查结果。
  - 执行命令**display dfs-group consistency-check** { **global** | **interface m-lag** *m-lag-id* | **static-arp** | **static-mac** }，用来查看M-LAG系统主备设备的相关配置信息。
  - 执行命令**display dfs-group consistency-check status**，用来查看M-LAG配置一致性检查功能的运行状态。

----结束

异常处理

- 配置M-LAG一致性检查模式为松散模式后，一旦M-LAG两端设备存在Type 1或者Type 2类型配置不一致，则会触发设备对两种类型配置检查不一致的告警：“ETRUNK_1.3.6.1.4.1.2011.5.25.178.8.2.1 hwMLagConsistencyCheckType1”、“ETRUNK_1.3.6.1.4.1.2011.5.25.178.8.2.3 hwMLagConsistencyCheckType2”。

  当调整M-LAG两端设备的配置部署一致时，M-LAG一致性检查成功后，则告警恢复。
- 配置M-LAG一致性检查模式为严格模式后，一旦M-LAG两端设备存在Type 1类型配置不一致，则会导致M-LAG备设备上的M-LAG成员接口处于ERROR DOWN状态，且触发设备对Type 1类型配置检查不一致的告警“ETRUNK_1.3.6.1.4.1.2011.5.25.178.8.2.1 hwMLagConsistencyCheckType1”。

  Error-Down是指设备检测到故障后将接口状态设置为ERROR DOWN状态，此时接口不能收发报文，接口指示灯为常灭。可以通过**display error-down recovery**命令查看设备上所有被Error-Down的接口信息。

  接口被Error-Down时，建议先调整M-LAG两端设备的配置部署，不建议直接手动恢复或在系统视图下执行命令**error-down auto-recovery cause m-lag interval**

*interval-value*使能接口状态自动恢复为Up的功能，否则可能会导致业务多包、丢包或不通等故障，请谨慎操作。

在设备配置M-LAG一致性检查为严格模式时，若设备检测出M-LAG两台设备存在Type1类型的配置不一致情况，建议设备管理员立即对相应配置进行调整，且不建议管理员对设备进行重启操作。如果管理员未对相应配置进行调整且重启M-LAG主设备，则可能导致主设备在设备恢复时，由于M-LAG两端重新协商，Type1类型配置不一致导致Error-Down M-LAG备设备端口。 此时，M-LAG主设备的端口由于M-LAG成员口延时Up的机制存在，导致M-LAG主备设备皆不能正常转发流量，业务中断。

----结束

1.6.4 配置 peer-link

背景信息

peer-link链路是位于部署M-LAG的两台设备之间的一条直连聚合链路，用于交互协议报文和传输部分流量，保证M-LAG的正常工作。

前提条件

部署M-LAG的两台设备之间的直连链路已经配置为聚合链路。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**interface eth-trunk** *trunk-id*，进入Eth-Trunk接口视图。

步骤3 执行命令**trunkport** *interface-type* { *interface-number1* [**to** *interface-number2*] } &<1-n>，增加成员接口。

批量增加成员接口时，若其中某个接口加入失败，则全部回退，此接口之前的接口也不会加入到Eth-trunk接口中。

说明

对于CE5810EI，“n”的取值为8，对于CE5880EI、CE6881、CE6881K、CE6820、CE6863、CE6863K、CE6881E和CE6880EI，“n”的取值为128，其他款型交换机“n”的取值由命令**assign forward eth-trunk mode**确定。

步骤4 执行命令**mode lacp-static**，配置Eth-Trunk的工作模式为静态LACP模式。

缺省情况下，Eth-Trunk的工作模式为手工负载分担模式。为了提高M-LAG的可靠性，必须配置为静态LACP模式。

步骤5 执行命令**undo stp enable**，去使能接口的STP功能。

缺省情况下，接口的STP功能处于使能状态。

说明

由于两端设备需要模拟成同一个STP根桥，保证设备直连接口不会被阻塞掉，需要将接口STP功能去使能。

步骤6 执行命令**peer-link** *peer-link-id*，配置接口为peer-link接口。

缺省情况下，接口不是peer-link接口。

- 接口配置为peer-link接口后，缺省加入所有VLAN。
- 接口配置为peer-link接口后，该接口上不能再配置其他业务。
- 如果后续需要配置ERPS的控制VLAN、TRILL的Carrier VLAN或FCoE VLAN，需要执行**步骤7**将peer-link接口退出控制VLAN、Carrier VLAN或FCoE VLAN，否则无法配置。
- 如果后续配置了网络侧的VLANIF且该VLANIF接口为双主检测链路，建议执行**步骤7**将peer-link接口退出相应的vlan，否则有可能会造成心跳检测失效等问题。

步骤7 （可选）执行命令**port vlan exclude** { { *vlan-id1* [**to** *vlan-id2*] } &<1-10> }，配置peer-link接口不允许通过的VLAN。

缺省情况下，未配置peer-link接口不允许通过的VLAN。

步骤8 执行命令**commit**，提交配置。

----结束

1.6.5 配置 M-LAG 成员接口

前提条件

部署M-LAG的两台设备与用户侧设备之间的链路已经分别配置为聚合链路。为了提高可靠性，避免M-LAG在配置过程中的误接或者成环，建议将链路聚合模式配置为LACP模式。

操作步骤

- 当聚合链路工作模式采用手工负载分担模式时，执行如下操作：
  a. 执行命令**system-view**，进入系统视图。
  b. 执行命令**interface eth-trunk** *trunk-id*，进入Eth-Trunk接口视图。
  c. 执行命令**trunkport** *interface-type* { *interface-number1* [**to** *interface-number2*] } &<1-n>，增加成员接口。

     批量增加成员接口时，若其中某个接口加入失败，则全部回退，此接口之前的接口也不会加入到Eth-trunk接口中。

     说明

     对于CE5810EI，“n”的取值为8，对于CE5880EI、CE6881、CE6881K、CE6820、CE6863、CE6863K、CE6881E和CE6880EI，“n”的取值为128，其他款型交换机“n”的取值由命令**assign forward eth-trunk mode**确定。
  d. 执行命令**dfs-group** *dfs-group-id* **m-lag** *m-lag-id*，配置绑定DFS Group和用户侧Eth-Trunk接口，即配置为M-LAG成员接口。

     说明

     部署M-LAG的两台设备配置绑定M-LAG的ID必须保持一致。
  e. 执行命令**commit**，提交配置。
- （推荐）当聚合链路工作模式采用LACP模式时，执行如下操作：
  a. 执行命令**system-view**，进入系统视图。

b. 执行命令**interface eth-trunk** *trunk-id*，进入Eth-Trunk接口视图。

c. 执行命令**trunkport** *interface-type* { *interface-number1* [**to** *interface-number2*] } &<1-n>，增加成员接口。

批量增加成员接口时，若其中某个接口加入失败，则全部回退，此接口之前的接口也不会加入到Eth-trunk接口中。

说明

对于CE5810EI，“n”的取值为8，对于CE5880EI、CE6881、CE6881K、CE6820、CE6863、CE6863K、CE6881E和CE6880EI，“n”的取值为128，其他款型交换机“n”的取值由命令**assign forward eth-trunk mode**确定。

d. 执行命令**mode** { **lacp-static** | **lacp-dynamic** }，配置Eth-Trunk的工作模式为LACP模式。

e. 执行命令**dfs-group** *dfs-group-id* **m-lag** *m-lag-id*，配置绑定DFS Group和用户侧Eth-Trunk接口，即配置为M-LAG成员接口。

说明

部署M-LAG的两台设备配置绑定M-LAG的ID必须保持一致。

f. （可选）配置LACP M-LAG的系统优先级和系统ID：

- 执行命令**quit**，退出Eth-Trunk接口视图。

  说明

  从V200R001C00版本开始，当DFS配对成功时，状态为主的设备会将本身的LACP M-LAG的系统优先级、系统ID自动同步给状态为备的设备，状态为备的设备M-LAG成员接口使用同步过来的LACP M-LAG的系统优先级、系统ID进行LACP协商，无须再手动配置设备的LACP M-LAG的系统优先级和系统ID。

- 执行命令**lacp m-lag priority** *priority*，配置LACP M-LAG的系统优先级。

  LACP M-LAG的系统优先级的缺省值是32768。

  - LACP M-LAG的系统优先级适用于LACP模式的Eth-Trunk组成的M-LAG，而LACP的系统优先级适用于LACP模式的Eth-Trunk接口，可通过**lacp priority**命令配置。
  - 部署M-LAG的两台设备上成员口Eth-Trunk接口的LACP M-LAG的系统优先级必须保持一致。

- 执行命令**lacp m-lag system-id** *mac-address*，配置LACP M-LAG的系统ID。

  LACP M-LAG的系统ID在系统视图下的缺省值为主控板的以太口MAC地址。

  - LACP M-LAG的系统ID仅适用于LACP模式的Eth-Trunk组成的M-LAG。
  - 部署M-LAG的两台设备上成员口Eth-Trunk接口的LACP M-LAG的系统ID必须保持一致。
  - 在配置LACP M-LAG的系统ID时，建议选取两台设备中MAC地址较小的作为系统ID。

g. 执行命令**commit**，提交配置。

----结束

1.6.6 （可选）配置双活网关

前提条件

M-LAG成员接口已经加入对应的VLAN或者M-LAG成员口所属Eth-Trunk接口的二层子接口已加入对应的BD。

背景信息

在M-LAG双归接入IP网络或VXLAN网络的场景中，M-LAG主备设备需要同时作为三层网关，必须保证M-LAG成员接口对应的VBDIF接口具有相同的IP地址和MAC地址。您可以通过在VBDIF接口上配置相同的IP地址并使用**mac-address**命令配置相同的虚拟MAC地址。

操作步骤

- 通过配置VLANIF/VBDIF接口IP地址和MAC地址实现双活网关功能。
  a. 执行命令**system-view**，进入系统视图。
  b. 执行命令**interface** { **vlanif** *vlan-id* | **vbdif** *bd-id* }，进入VLANIF接口或者VBDIF接口视图。
  c. 配置接口IP地址：
     - 当网络是IPv4网络时，执行命令**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]，配置IPv4地址。
     - 当网络是IPv6网络时，执行如下步骤：
       1) 执行命令**ipv6 enable**，使能接口的IPv6功能。
       2) 执行命令**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*|*prefix-length* } 或 **ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*|*prefix-length* } **eui-64**，配置接口的全球单播地址。

     缺省情况下，接口上没有配置IP地址。

     主备设备上M-LAG成员接口对应的VLANIF接口的IP地址必须配置相同。
  d. 执行命令**mac-address**（VLANIF接口视图） *mac-address*或者**mac-address**（VBDIF接口视图） *mac-address*，配置VLANIF/VBDIF接口的虚拟MAC地址。

     缺省情况下，VLANIF接口的MAC地址和系统的MAC地址保持一致。

     主备设备上M-LAG成员接口对应的VLANIF接口的虚拟MAC地址必须配置相同。
  e. （可选）执行命令**protocol up-delay-time** *time-value*，配置接口三层协议状态延时Up时间。

     缺省情况下，VLANIF接口延时Up时间为1秒。当ARP表项数量较多的时候，配置延时UP的时间也相应要加长。

     设备故障恢复或者peer-link故障恢复时，存在大批量的ARP表项等待批量同步。配置Up延迟时间可以使得接口协议层的状态在ARP表项同步完成后再从Down变成Up，以避免协议报文被丢弃，减少链路故障恢复时的丢包时间，提高了收敛性能。

说明

在设备重启或单板复位后，接口物理状态切换Up状态，但上层协议模块状态未满足转发要求，导致流量丢包。为了保证故障回切性能，缺省情况下，M-LAG成员端口上报Up状态的延时时间为240秒。若此时同时在VLANIF接口上配置三层协议状态延时Up时间，则需要保证M-LAG成员口的延时Up时间长于VLANIF接口的延时时间，否则ND同步失败的表项依赖于ND Miss触发学习。

f. 执行命令**commit**，提交配置。

----结束

1.6.7 （可选）配置 peer-link 故障场景下端口状态

背景信息

当M-LAG应用于普通以太网络、VXLAN网络或IP网络的双归接入时，peer-link故障但双主检测心跳状态正常会触发备设备上除管理网口、peer-link接口和堆叠口以外的接口处于Error-Down状态。一旦peer-link故障恢复，处于ERROR DOWN状态的M-LAG接口默认将在240s后自动恢复为Up状态，处于ERROR DOWN状态的其它接口将立即自动恢复为Up状态。

但在实际组网应用中，当某些上行端口运行路由协议或者是双主检测心跳口时是不希望被Error-Down的。此时，可以根据实际情况选择配置下列功能。

在peer-link故障但双主检测正常时，配置下列功能，设备接口Error-Down情况如表1-8所示。

表 1-8 设备在 peer-link 故障但双主检测正常时接口 Error-Down 情况

设备配置情况	M-LAG接入普通以太网络、VXLAN网络或IP网络
设备缺省情况	除管理网口、peer-link接口和堆叠口以外的接口处于ERROR DOWN状态。
设备仅配置suspend功能	仅M-LAG成员口以及配置该功能的接口处于ERROR DOWN状态。
设备仅配置reserved功能	除配置该功能的接口、管理网口、peer-link接口和堆叠口以外的接口处于ERROR DOWN状态。
设备同时配置suspend功能和reserved功能	仅M-LAG成员口以及配置suspend功能的接口处于ERROR DOWN状态。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**interface** *interface-type interface-number*，进入接口视图。

步骤3 执行命令**m-lag unpaired-port reserved**，配置备设备接口在peer-link故障但双主检测正常时不被Error-Down。

- 可以通过在备设备的其他接口上配置**m-lag unpaired-port suspend**，指定特定的接口在peer-link故障但双主检测心跳状态正常时自动被Error-Down。
- 建议在M-LAG主备设备对应接口同时配置该命令，保证M-LAG主备倒换后设备接口Error-Down情况一致。
- 该命令不支持在peer-link接口、M-LAG成员口下配置。

步骤4 执行命令**commit**，提交配置。

----结束

1.6.8 （可选）使能 M-LAG 三层转发增强功能

背景信息

为提升M-LAG成员口故障时的收敛速度，可以使能M-LAG三层转发增强，为M-LAG成员接口的所有ARP/ND表项申请备份的FRR资源，形成下行出接口的主备备份路径，在M-LAG成员口故障时能够快速切换出接口为peer-link接口。

说明

CE6810LI不支持IPv6场景下的M-LAG三层转发增强功能。

CE5800系列交换机（除CE5880EI外）、CE6810EI、CE6810LI、CE6850EI 不支持IPv4场景下的M-LAG三层转发增强功能。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**m-lag forward layer-3 enhanced enable**（CE6881、CE6881K、CE6863、CE6863K、CE6881E和CE6820）或者**m-lag forward layer-3 { ipv4 | ipv6 } enhanced enable**（除CE6881、CE6881K、CE6863、CE6863K、CE6881E和CE6820），使能M-LAG三层转发增强功能。

缺省情况下，M-LAG三层转发增强功能未使能。

在使能M-LAG三层转发增强功能后，对于报文出端口为M-LAG成员接口的所有ARP/ND表项会申请备份的FRR资源，出接口指向peer-link接口，形成主备路径转发。一旦FEI侧感知到M-LAG成员接口故障，设备双归接入变成单归，则将对应ARP/ND表项的下一跳由M-LAG成员接口切换为Peer-link接口，提升故障场景下的切换性能。

说明

- 在使能M-LAG三层转发增强功能后，由于下一跳资源消耗的增加可能导致主备路径下发不成功，引起流量丢包。
- 在使能M-LAG三层转发增强功能后，若配置Eth-Trunk接口加退M-LAG成员口时需要指定该Eth-Trunk口清除学习到的所有ARP/ND表项，避免由于上层协议模块不感知M-LAG成员口变化导致的FRR资源浪费。
- 在使能M-LAG三层转发增强功能后，需要间隔300s才可以配置去使能M-LAG三层转发增强功能；在去使能了M-LAG三层转发增强功能后，需要间隔300s才可以配置使能M-LAG三层转发增强功能。
- 针对CE6881、CE6881K、CE6863K、CE6881E和CE6863设备如果需要使用IPv4场景下M-LAG三层转发增强功能，需要同时配置**arp broadcast-detect enable**命令。

步骤3 执行命令**commit**，提交配置。

----结束

1.6.9 检查基于根桥方式配置 M-LAG 的配置结果

操作步骤

- 执行命令**display dfs-group** *dfs-group-id* [**node** *node-id* **m-lag** [**brief**] | **peer-link**]，查看M-LAG的信息。

----结束

后续处理

完成M-LAG配置后，如果peer-link故障但心跳状态正常会导致状态为备的设备上部分接口处于ERROR DOWN状态。Error-Down是指设备检测到故障后将接口状态设置为ERROR DOWN状态，此时接口不能收发报文，接口指示灯为常灭。可以通过**display error-down recovery**命令可以查看设备上所有被Error-Down的接口信息。

当M-LAG应用于普通以太网络、VXLAN网络或IP网络的双归接入时，peer-link故障但双主检测心跳状态正常会触发状态为备的设备上除管理网口、peer-link接口和堆叠口以外的接口处于Error-Down状态。一旦peer-link故障恢复，处于Error-Down状态的M-LAG接口默认将在240s后自动恢复为Up状态，处于Error-Down状态的其它物理接口将自动恢复为Up状态。

接口被Error-Down时，建议先排除引起peer-link故障的原因，不建议直接手动恢复或在系统视图下执行命令**error-down auto-recovery cause m-lag interval** *interval-value*使能接口状态自动恢复为Up的功能，否则可能会导致业务多包、丢包或不通等故障，请谨慎操作。

1.7 基于 V-STP 方式配置 M-LAG（推荐）

利用V-STP机制将M-LAG主设备和备设备的STP协议虚拟成一台设备的STP协议，对外呈现为一台设备进行STP协议计算。

1.7.1 配置 V-STP

背景信息

V-STP（Virtual Spanning Tree Protocol）是二层拓扑管理特性，其核心思想是将两台设备的STP协议虚拟成一台设备的STP协议，对外呈现为一台设备进行STP协议计算。

STP可以感知M-LAG主备协商状态，M-LAG主备设备配置了V-STP使能之后，在M-LAG主备协商成功后，两台设备被虚拟化成一台设备进行端口角色计算和快速收敛计算。STP需要同步M-LAG主备的桥MAC信息和实例优先级信息。M-LAG主备协商成功后，M-LAG备设备使用M-LAG主设备同步过来的桥MAC信息和实例优先级信息进行STP计算和收发报文，保证虚拟化成一台设备后的STP计算参数。

当前，V-STP只能用于M-LAG组网，可以解决多级M-LAG互联场景和组成M-LAG的设备作为非根桥场景的需求。

配置V-STP功能时，需要保证组成M-LAG的两台设备上STP/RSTP定时器配置一致，否则可能导致网络拓扑震荡。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**stp mode** { **stp** | **rstp** }，配置交换设备的生成树协议为STP或RSTP模式。

缺省情况下，交换设备运行MSTP模式。

V-STP场景中，不支持MSTP模式，但支持STP多进程（MSTP进程默认是MSTP模式，目前V-STP场景仅支持STP及RSTP模式，需要将MSTP进程配置为STP或RSTP模式，作为STP进程）。

步骤3 （可选）执行命令**stp bridge-address** *mac-address*，配置设备参与生成树计算的桥MAC。

缺省情况下，当前设备参与生成树计算的桥MAC是设备的MAC地址。

为防止设备重启、DFS主备倒换时导致STP网络结构震荡、保证故障回切的性能，在主备设备优先级相同的情况下建议配置DFS状态为备的设备较大的桥MAC地址。

步骤4 执行命令**stp v-stp enable**，使能STP进入跨设备组合工作模式。

缺省情况下，STP进入跨设备组合工作模式处于去使能状态。

步骤5 执行命令**commit**，提交配置。

----结束

1.7.2 配置 DFS Group

背景信息

动态交换服务组DFS Group，主要用于设备之间的配对。为了实现双主检测报文的交互，DFS Group需要绑定IP地址，绑定的IP地址用于和对端进行通信。

设备双归接入普通以太网络、VXLAN或IP网络时，需要配置DFS Group绑定IP地址，前提是已经配置两台双归设备对应三层接口的IP地址且保证互通。当接入的是VPN网络时，还需要配置DFS Group绑定VPN实例，前提是设备上已经创建VPN实例。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**dfs-group** *dfs-group-id*，创建DFS Group并进入DFS-Group视图。

步骤3 根据实际场景选择配置DFS Group绑定IP地址。

设备双归接入普通以太网络、VXLAN或IP网络时，配置DFS Group绑定IP地址。以下任选一种即可，不支持同时配置。

- 执行命令**source ip** *ip-address* [**vpn-instance** *vpn-instance-name*] [**peer** *peer-ip-address* [**udp-port** *port-number*]]，配置DFS Group绑定IPv4地址和VPN实例。
- 执行命令**source ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] [**peer** *peer-ipv6-address* [**udp-port** *port-number*]]，配置DFS Group绑定IPv6地址和VPN实例。

若在配置DFS Group绑定设备互通的心跳IP时同时指定对端设备的心跳IP和UDP通信端口号，则在该配置生效时，M-LAG两端设备即开始收发心跳报文并协商HB DFS主备状

态。在二次故障增强场景下，若原DFS状态为主的设备故障恢复且peer-link链路仍然故障，则M-LAG两台设备根据HB DFS的主备状态来触发状态为备的设备相应接口Error-Down，避免M-LAG两台设备形成双主状态而导致的流量异常，提高设备可靠性。

步骤4 （可选）执行命令**priority** *priority*，配置DFS Group的优先级。

优先级用于两台设备间进行主备协商，值越大优先级越高，优先级高的为主用设备。

如果优先级相同，那么比较两台设备的系统MAC地址，MAC地址较小的为主用设备。

缺省情况下，DFS Group的优先级为100。

步骤5 （可选）执行命令**m-lag up-delay** *value* [**auto-recovery interval** *interval-time*]，配置M-LAG接口上报Up状态的延时时间。

为了保证故障回切的性能，针对设备重启、子卡复位或Peer-link故障恢复的场景，M-LAG接口默认配置了240秒的延迟Up时间，未配置延迟自动恢复时间。

步骤6 （可选）执行命令**set lacp system-id switch-delay** { *switch-delay-time* | **immediately** }，配置LACP M-LAG的系统ID延迟切换时间。

缺省情况下，LACP M-LAG的系统ID不切换。**immediately**表示LACP M-LAG的系统ID立即切换；*switch-delay-time*为0时表示LACP M-LAG的系统ID不做切换。

步骤7 （可选）执行命令**authentication-mode hmac-sha256 password** *password*，指定DFS Group协议报文所使用的验证模式及验证口令。

缺省情况下，没有配置DFS Group协议报文的验证模式。

步骤8 （可选）执行命令**dfs-master led enable**，使能堆叠状态指示灯显示DFS Group主备状态功能。

缺省情况下，堆叠状态指示灯显示DFS Group主备状态功能未使能。

在使能堆叠状态指示灯显示DFS Group主备状态功能后，切换堆叠指示灯显示为DFS主备状态，DFS状态为主的设备堆叠状态显示灯常亮，而DFS状态为备的设备堆叠状态显示灯常灭。

步骤9 （可选）执行命令**dual-active detection error-down** { **delay** *delay-time* | **disable** }，配置关闭或者延迟执行在peer-link故障但双主检测心跳状态正常时Error-Down备设备上除管理网口、peer-link接口和堆叠口以外接口的动作。

缺省情况下，在peer-link故障但双主检测心跳状态正常会触发备设备上除管理网口、peer-link接口和堆叠口以外的接口处于Error-Down。

当接入设备以单归方式三层接入M-LAG主备设备时，在peer-link故障但双主检测心跳状态正常即M-LAG双主场景下，并不影响设备流量的转发。此时，可以执行**dual-active detection error-down**命令用来配置关闭或者延迟执行在peer-link故障但双主检测心跳状态正常时Error-Down备设备上除管理网口、peer-link接口和堆叠口以外接口的动作，避免流量丢包。

在接入设备以M-LAG双归接入或者二层接入方式接入M-LAG主备时，不可以关闭或者延迟Error-Down动作。

步骤10 （可选）执行命令**dual-active detection enhanced enable**，使能M-LAG场景下二次故障增强功能。

在M-LAG应用于双归接入时，当peer-link故障但双主检测心跳状态正常会触发DFS备设备上某些端口处于Error-Down状态，此时DFS状态为主的设备继续工作。在该场景的基础上，若DFS状态为主的设备由于断电、整机故障重启等其他故障导致主设备不能工作时，M-LAG主备设备皆不能正常转发流量。

此时，可以使能M-LAG二次故障增强功能，在Peer-Link故障且DFS设备二次故障后借助双主检测机制，DFS状态为备的设备在感知DFS状态为主的设备故障后，恢复备设备上处于Error-Down状态的端口，继续转发流量，从而提升二次故障场景下业务不中断的可靠性。

在M-LAG二次故障场景下，若设备故障恢复后peer-link链路故障仍然存在，则有可能导致M-LAG形成双主，建议在配置DFS Group绑定IP地址时同时指定对端设备的IP地址。如此，在设备故障恢复后，若peer-link故障仍然存在，则触发HB DFS状态为备的设备上相应端口Error-Down，避免M-LAG设备在双主情况下出现的流量异常。

步骤11 （可选）执行命令**dual-active detection error-down mode routing-switch**，配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口。

缺省情况下，M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围不包括逻辑端口。M-LAG应用于TRILL网络的双归接入时，peer-link故障但双主检测状态正常时会触发备设备上M-LAG接口处于ERROR DOWN状态。当M-LAG应用于普通以太网络或IP网络的双归接入时，peer-link故障但双主检测状态时正常会触发备设备上除逻辑端口、配置reserved功能的接口、管理网口、peer-link接口和堆叠口以外的物理接口处于ERROR DOWN状态。

在配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口后，当M-LAG应用于IP网络或者VXLAN网络时，仅会触发备设备上VLANIF接口、VBDIF接口、LoopBack接口以及M-LAG成员接口处于ERROR DOWN状态。

说明

在配置M-LAG场景下peer-link故障但双主检测心跳状态正常时触发端口Error-Down的范围包括逻辑端口后，若M-LAG系统peer-link接口故障恢复，为保证大规格VLANIF接口下的ARP同步正常，设备将在DFS Group配对成功后延迟6s恢复VLANIF接口、VBDIF接口、LoopBack接口为Up状态。此时，如果在接口下配置了接口三层协议状态延时Up时间，则VLANIF接口、VBDIF接口、LoopBack接口恢复Up状态的延迟时间为两者之和。

步骤12 （可选）执行命令**peer-link mac-address remain enable**，使能特定条件下不再触发本端或对端设备删除peer-link接口上对应的MAC地址。

缺省情况下，未配置特定条件下不再触发本端或对端设备删除peer-link接口上对应的MAC地址。

步骤13 （可选）执行命令**pim synchronize enable**，使能DR优先级修改时M-LAG备设备自动同步DR优先级。

缺省情况下，未使能DR优先级修改时同步DR优先级。在两组M-LAG双活系统场景下，M-LAG的成员均接入PIM网络，一组M-LAG的DR优先级高，一组M-LAG的DR优先级低。当DR优先级高的一方出现故障，通过新一轮的竞选另外一方的M-LAG主设备上的DR优先级变高，为保证整个M-LAG系统组播数据不丢包，需要配置DR优先级同步功能，将M-LAG备设备上的DR优先级同步变高。

步骤14 （可选）执行命令**vrrp synchronize enable**，使能VRRP优先级修改时M-LAG备设备自动同步VRRP优先级。

缺省情况下，未使能VRRP优先级修改时同步VRRP优先级。两组M-LAG双活系统组成VRRP主备备份。当VRRP主用设备发生异常导致报文不能转发时，VRRP备用设备的M-LAG主设备上的优先级变高，为保证整个M-LAG系统不丢包，需要配置VRRP优先级同步功能，将M-LAG备设备上的VRRP优先级同步变高。

步骤15 执行命令**commit**，提交配置。

----结束

1.7.3 （可选）配置 STP 多进程

背景信息

STP多进程，主要用于不同M-LAG接入设备间的生成树独立计算。使能STP多进程功能后，每个进程可以管理M-LAG设备上的部分M-LAG成员端口，设备将以进程为单位进行STP协议计算，不在此进程内的端口将不参与此进程的协议计算，加快了STP生成树协议的收敛速度。

说明

M-LAG主备设备需要保持STP进程配置一致（进程数目、进程ID、STP使能），否则V-STP功能不可用。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**stp process** *process-id*，创建一个指定ID的STP进程并进入该STP进程视图。

步骤3 执行命令**stp mode** { **stp** | **rstp** }，配置STP进程的工作模式。

缺省情况下，当前STP进程的工作模式为MSTP。V-STP不支持MSTP模式，需要将进程的工作模式切换为STP/RSTP。正常启动后，设备默认存在ID为0的STP进程，系统视图和接口视图中的STP相关配置都属于此进程。

步骤4 执行命令**stp enable**，使能交换设备STP进程的MSTP功能。

缺省情况下，进程下的STP功能默认不使能。

步骤5 执行命令**commit**，提交配置。

----结束

1.7.4 配置 M-LAG 一致性检查

前提条件

部署M-LAG系统的两端设备之间的DFS Group已配对成功并协商出主备关系。

背景信息

M-LAG配置一致性检查将设备配置分为两类，如表1-9所示，分别为关键配置（Type 1）和一般配置（Type 2）。根据对关键配置检查不一致时的处理方式，M-LAG一致性又分为严格模式（strict）和松散模式（loose）。

- 关键配置（Type 1）：如果在M-LAG系统两端设备不一致，会导致成环、状态正常但长时间丢包等问题。

  严格模式下，如果M-LAG两端设备存在Type 1配置不一致，会导致M-LAG备设备上成员口处于ERROR DOWN状态，且触发设备对Type 1类型配置检查不一致的告警。

  松散模式下，如果M-LAG两端设备存在Type 1配置不一致，则会触发设备对两种类型配置检查不一致的告警。

- 一般配置（Type 2）：如果在M-LAG系统两端设备不一致，可能会导致M-LAG运行状态异常。与Type 1类型的配置相比较而言，Type 2类型的配置问题更容易被发现，对组网环境的影响也相对较小。

  无论处于何种模式，如果M-LAG两端设备存在以下Type 2配置不一致，则会触发设备对两种类型配置检查不一致的告警。

表 1-9 M-LAG 配置一致性检查配置分类列表

视图	配置	类型
全局	STP功能是否使能	Type 1
	STP工作模式配置	
	BPDU保护功能是否使能	
	STP多生成树实例与VLAN的映射关系配置 **说明** 设备默认仅检查ID为0的STP进程内多生成树实例与VLAN的映射关系。	
M-LAG成员口	STP功能是否使能	
	STP端口的Root保护功能是否使能	
	M-LAG成员接口的LACP模式配置	
全局	VLAN配置	Type 2
	静态MAC地址表项 ● 静态MAC地址表项指定接口为M-LAG成员口 ● VXLAN隧道的静态MAC地址表项	
	动态MAC的老化时间	

视图	配置	类型
	静态ARP表项 ● 短静态ARP表项 ● 长静态ARP表项 – 若静态ARP表项指定出接口，则仅检查出接口为M-LAG成员口的静态ARP。 – 若静态ARP表项指定所属VLAN，则直接比较VLAN ID。 – 若静态ARP表项指定出接口和所属VLAN，则直接比较出接口为M-LAG成员口的静态ARP表项和VLAN ID。 – VXLAN IPv4隧道的静态ARP表项 **说明** 设备不支持检查指定VPN实例的短静态ARP表项，若长静态ARP表项的出接口为M-LAG成员口且绑定了VPN实例或者所属VLAN对应的VLANIF接口绑定了VPN实例，设备同样不支持检查该静态ARP表项。	
	动态ARP的老化时间	
	广播域桥域BD（Bridge Domain）配置 ● BD ID ● BD关联VNI	

视图	配置	类型
	VBDIF接口配置 ● VBDIF接口的BD ID ● VBDIF接口IPv4地址 ● VBDIF接口IPv6地址 ● VBDIF接口VRRP4备份组 ● VBDIF接口MAC地址 ● VBDIF接口状态 **说明** 对于VBDIF接口MAC地址，设备默认仅检查虚拟MAC地址。 针对IPv6地址以及VRRP4备份组的配置检查，仅在VBDIF接口Up时才进行。若VBDIF接口状态为Down，则认为该接口下没有相关配置。	
	VLANIF接口配置 ● VLAN ID ● VLANIF接口IPv4地址 ● VLANIF接口IPv6地址 ● VLANIF接口VRRP4备份组 ● VLANIF接口MAC地址 ● VLANIF接口状态 **说明** 对于VLANIF接口MAC地址，设备默认仅检查虚拟MAC地址。 针对IPv6地址以及VRRP4备份组的配置检查，仅在VLANIF接口Up时才进行。若VLANIF接口状态为Down，则认为该接口下没有相关配置。	
M-LAG成员口	STP端口优先级配置	
	接口加入VLAN配置	
	M-LAG成员口参数配置	

视图	配置	类型
	M-LAG成员口所属Eth-Trunk接口成员口个数 **说明** 仅比较Eth-Trunk接口的成员口数量，对于成员口物理状态Up/Down或者成员口带宽不予检查。	

操作步骤

- 配置M-LAG一致性检查功能：
  a. 执行命令**system-view**，进入系统视图。
  b. 执行命令**dfs-group** *dfs-group-id*，创建DFS Group并进入DFS-Group视图。
  c. 执行命令**consistency-check enable mode** { **strict** | **loose** }，使能M-LAG配置一致性检查并指定检查模式。

     缺省情况下，M-LAG配置一致性检查功能处于去使能状态。
  d. 执行命令**commit**，提交配置。
- 检查一致性检查运行状态及M-LAG系统主备设备的相关配置信息：
  - 执行命令**display dfs-group**，查看M-LAG一致性检查结果。
  - 执行命令**display dfs-group consistency-check** { **global** | **interface m-lag** *m-lag-id* | **static-arp** | **static-mac** }，用来查看M-LAG系统主备设备的相关配置信息。
  - 执行命令**display dfs-group consistency-check status**，用来查看M-LAG配置一致性检查功能的运行状态。

----结束

异常处理

- 配置M-LAG一致性检查模式为松散模式后，一旦M-LAG两端设备存在Type 1或者Type 2类型配置不一致，则会触发设备对两种类型配置检查不一致的告警：“ETRUNK_1.3.6.1.4.1.2011.5.25.178.8.2.1 hwMLagConsistencyCheckType1”、“ETRUNK_1.3.6.1.4.1.2011.5.25.178.8.2.3 hwMLagConsistencyCheckType2”。

  当调整M-LAG两端设备的配置部署一致时，M-LAG一致性检查成功后，则告警恢复。
- 配置M-LAG一致性检查模式为严格模式后，一旦M-LAG两端设备存在Type 1类型配置不一致，则会导致M-LAG备设备上的M-LAG成员接口处于ERROR DOWN状态，且触发设备对Type 1类型配置检查不一致的告警“ETRUNK_1.3.6.1.4.1.2011.5.25.178.8.2.1 hwMLagConsistencyCheckType1”。

  Error-Down是指设备检测到故障后将接口状态设置为ERROR DOWN状态，此时接口不能收发报文，接口指示灯为常灭。可以通过**display error-down recovery**命令查看设备上所有被Error-Down的接口信息。

  接口被Error-Down时，建议先调整M-LAG两端设备的配置部署，不建议直接手动恢复或在系统视图下执行命令**error-down auto-recovery cause m-lag interval**

*interval-value*使能接口状态自动恢复为Up的功能，否则可能会导致业务多包、丢包或不通等故障，请谨慎操作。

在设备配置M-LAG一致性检查为严格模式时，若设备检测出M-LAG两台设备存在Type1类型的配置不一致情况，建议设备管理员立即对相应配置进行调整，且不建议管理员对设备进行重启操作。如果管理员未对相应配置进行调整且重启M-LAG主设备，则可能导致主设备在设备恢复时，由于M-LAG两端重新协商，Type1类型配置不一致导致Error-Down M-LAG备设备端口。 此时，M-LAG主设备的端口由于M-LAG成员口延时Up的机制存在，导致M-LAG主备设备皆不能正常转发流量，业务中断。

----结束

1.7.5 配置 peer-link

背景信息

peer-link链路是位于部署M-LAG的两台设备之间的一条直连聚合链路，用于交互协议报文和传输部分流量，保证M-LAG的正常工作。

前提条件

部署M-LAG的两台设备之间的直连链路已经配置为聚合链路。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**interface eth-trunk** *trunk-id*，进入Eth-Trunk接口视图。

步骤3 执行命令**trunkport** *interface-type* { *interface-number1* [**to** *interface-number2*] } &<1-n>，增加成员接口。

批量增加成员接口时，若其中某个接口加入失败，则全部回退，此接口之前的接口也不会加入到Eth-trunk接口中。

说明

对于CE5810EI，“n”的取值为8，对于CE5880EI、CE6881、CE6881K、CE6820、CE6863、CE6863K、CE6881E和CE6880EI，“n”的取值为128，其他款型交换机“n”的取值由命令**assign forward eth-trunk mode**确定。

步骤4 执行命令**mode lacp-static**，配置Eth-Trunk的工作模式为静态LACP模式。

缺省情况下，Eth-Trunk的工作模式为手工负载分担模式。为了提高M-LAG的可靠性，必须配置为静态LACP模式。

步骤5 执行命令**stp enable**，使能接口的STP功能。

缺省情况下，接口的STP功能处于使能状态。

步骤6 （可选）执行命令**stp binding process** *process-id1* [**to** *process-id2*] **link-share**，配置peer-link作为STP多进程的共享链路，指定peer-link端口参与多个STP进程的状态计算。

步骤7 执行命令**peer-link** *peer-link-id*，配置接口为peer-link接口。

缺省情况下，接口不是peer-link接口。

- 接口配置为peer-link接口后，缺省加入所有VLAN。
- 接口配置为peer-link接口后，该接口上不能再配置其他业务。
- 如果后续需要配置ERPS的控制VLAN、TRILL的Carrier VLAN或FCoE VLAN，需要执行**步骤8**将peer-link接口退出控制VLAN、Carrier VLAN或FCoE VLAN，否则无法配置。
- 如果后续配置了网络侧的VLANIF且该VLANIF接口为双主检测链路，建议执行**步骤8**将peer-link接口退出相应的vlan，否则有可能会造成心跳检测失效等问题。

步骤8 （可选）执行命令**port vlan exclude** { { *vlan-id1* [**to** *vlan-id2*] } &<1-10> }，配置peer-link接口不允许通过的VLAN。

缺省情况下，未配置peer-link接口不允许通过的VLAN。

步骤9 执行命令**commit**，提交配置。

----结束

1.7.6 配置 M-LAG 成员接口

前提条件

部署M-LAG的两台设备与用户侧设备之间的链路已经分别配置为聚合链路。为了提高可靠性，避免M-LAG在配置过程中的误接或者成环，建议将链路聚合模式配置为LACP模式。

操作步骤

- 当聚合链路工作模式采用手工负载分担模式时，执行如下操作：
  a. 执行命令**system-view**，进入系统视图。
  b. 执行命令**interface eth-trunk** *trunk-id*，进入Eth-Trunk接口视图。
  c. 执行命令**trunkport** *interface-type* { *interface-number1* [**to** *interface-number2*] } &<1-n>，增加成员接口。

     批量增加成员接口时，若其中某个接口加入失败，则全部回退，此接口之前的接口也不会加入到Eth-trunk接口中。

     说明

     对于CE5810EI，“n”的取值为8，对于CE5880EI、CE6881、CE6881K、CE6820、CE6863、CE6863K、CE6881E和CE6880EI，“n”的取值为128，其他款型交换机“n”的取值由命令**assign forward eth-trunk mode**确定。
  d. 执行命令**dfs-group** *dfs-group-id* **m-lag** *m-lag-id*，配置绑定DFS Group和用户侧Eth-Trunk接口，即配置为M-LAG成员接口。

     说明

     部署M-LAG的两台设备配置绑定M-LAG的ID必须保持一致。
  e. （可选）执行命令**stp binding process** *process-id*，把当前端口加入指定ID的STP进程中。

     在使能STP多进程功能后，每个进程可以管理M-LAG设备上的部分M-LAG成员端口，设备将以进程为单位进行STP协议计算，不在此进程内的端口将不参与此进程的协议计算，需要将指定M-LAG成员口加入到指定ID的STP进程中。

- 如果属于不同进程的M-LAG成员端口同属于一个广播域，在peer-link接口被阻塞的情况下可能会造成环路，所以需要划分不同进程的M-LAG成员端口不同的广播域。
- M-LAG成员端口切换进程之前，先执行命令**shutdown**关闭当前接口并不再配置业务。在M-LAG主备设备上的相关成员端口进程切换成功后，执行命令**undo shutdown**开启当前接口再进行业务配置。

f. 执行命令**commit**，提交配置。

- （推荐）当聚合链路工作模式采用LACP模式时，执行如下操作：

a. 执行命令**system-view**，进入系统视图。

b. 执行命令**interface eth-trunk** *trunk-id*，进入Eth-Trunk接口视图。

c. 执行命令**trunkport** *interface-type* { *interface-number1* [**to** *interface-number2*] } &<1-n>，增加成员接口。

批量增加成员接口时，若其中某个接口加入失败，则全部回退，此接口之前的接口也不会加入到Eth-trunk接口中。

说明

对于CE5810EI，“n”的取值为8，对于CE5880EI、CE6881、CE6881K、CE6820、CE6863、CE6863K、CE6881E和CE6880EI，“n”的取值为128，其他款型交换机“n”的取值由命令**assign forward eth-trunk mode**确定。

d. 执行命令**mode** { **lacp-static** | **lacp-dynamic** }，配置Eth-Trunk的工作模式为LACP模式。

e. 执行命令**dfs-group** *dfs-group-id* **m-lag** *m-lag-id*，配置绑定DFS Group和用户侧Eth-Trunk接口，即配置为M-LAG成员接口。

说明

部署M-LAG的两台设备配置绑定M-LAG的ID必须保持一致。

f. （可选）执行命令**stp binding process** *process-id*，把当前与接入链路相连的端口加入指定ID的STP进程中。

在使能STP多进程功能后，每个进程可以管理M-LAG设备上的部分M-LAG成员端口，设备将以进程为单位进行STP协议计算，不在此进程内的端口将不参与此进程的协议计算，需要将指定M-LAG成员口加入到指定ID的STP进程中。

- 如果属于不同进程的M-LAG成员端口同属于一个广播域，在peer-link接口被阻塞的情况下可能会造成环路，所以需要划分不同进程的M-LAG成员端口不同的广播域。
- M-LAG成员端口切换进程之前，先执行命令**shutdown**关闭当前接口并不再配置业务。在M-LAG主备设备上的相关成员端口进程切换成功后，执行命令**undo shutdown**开启当前接口再进行业务配置。

g. （可选）配置LACP M-LAG的系统优先级和系统ID：

- 执行命令**quit**，退出Eth-Trunk接口视图。

说明

从V200R001C00版本开始，当DFS配对成功时，状态为主的设备会将本身的LACP M-LAG的系统优先级、系统ID自动同步给状态为备的设备，状态为备的设备M-LAG成员接口使用同步过来的LACP M-LAG的系统优先级、系统ID进行LACP协商，无须再手动配置设备的LACP M-LAG的系统优先级和系统ID。

- 执行命令**lacp m-lag priority** *priority*，配置LACP M-LAG的系统优先级。

  LACP M-LAG的系统优先级的缺省值是32768。

  - LACP M-LAG的系统优先级适用于LACP模式的Eth-Trunk组成的M-LAG，而LACP的系统优先级适用于LACP模式的Eth-Trunk接口，可通过**lacp priority**命令配置。
  - 部署M-LAG的两台设备上成员口Eth-Trunk接口的LACP M-LAG的系统优先级必须保持一致。

- 执行命令**lacp m-lag system-id** *mac-address*，配置LACP M-LAG的系统ID。

  LACP M-LAG的系统ID在系统视图下的缺省值为主控板的以太口MAC地址。

  - LACP M-LAG的系统ID适用于LACP模式的Eth-Trunk组成的M-LAG。
  - 部署M-LAG的两台设备上成员口Eth-Trunk接口的LACP M-LAG的系统ID必须保持一致。
  - 在配置LACP M-LAG的系统ID时，建议选取两台设备中MAC地址较小的作为系统ID。

h. 执行命令**commit**，提交配置。

----结束

1.7.7 （可选）配置双活网关

前提条件

M-LAG成员接口已经加入对应的VLAN或者M-LAG成员口所属Eth-Trunk接口的二层子接口已加入对应的BD。

背景信息

在M-LAG双归接入IP网络或VXLAN网络的场景中，M-LAG主备设备需要同时作为三层网关，必须保证M-LAG成员接口对应的VBDIF接口具有相同的IP地址和MAC地址。您可以通过在VBDIF接口上配置相同的IP地址并使用**mac-address**命令配置相同的虚拟MAC地址。

操作步骤

- 通过配置VLANIF/VBDIF接口IP地址和MAC地址实现双活网关功能。
  a. 执行命令**system-view**，进入系统视图。
  b. 执行命令**interface** { **vlanif** *vlan-id* | **vbdif** *bd-id* }，进入VLANIF接口或者VBDIF接口视图。

c. 配置接口IP地址：

- 当网络是IPv4网络时，执行命令**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]，配置IPv4地址。
- 当网络是IPv6网络时，执行如下步骤：
  1) 执行命令**ipv6 enable**，使能接口的IPv6功能。
  2) 执行命令**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*|*prefix-length* } 或 **ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*|*prefix-length* } **eui-64**，配置接口的全球单播地址。

缺省情况下，接口上没有配置IP地址。

主备设备上M-LAG成员接口对应的VLANIF接口的IP地址必须配置相同。

d. 执行命令**mac-address**（VLANIF接口视图） *mac-address*或者**mac-address**（VBDIF接口视图） *mac-address*，配置VLANIF/VBDIF接口的虚拟MAC地址。

缺省情况下，VLANIF接口的MAC地址和系统的MAC地址保持一致。

主备设备上M-LAG成员接口对应的VLANIF接口的虚拟MAC地址必须配置相同。

e. （可选）执行命令**protocol up-delay-time** *time-value*，配置接口三层协议状态延时Up时间。

缺省情况下，VLANIF接口延时Up时间为1秒。当ARP表项数量较多的时候，配置延时UP的时间也相应要加长。

设备故障恢复或者peer-link故障恢复时，存在大批量的ARP表项等待批量同步。配置Up延迟时间可以使得接口协议层的状态在ARP表项同步完成后再从Down变成Up，以避免协议报文被丢弃，减少链路故障恢复时的丢包时间，提高了收敛性能。

说明

在设备重启或单板复位后，接口物理状态切换Up状态，但上层协议模块状态未满足转发要求，导致流量丢包。为了保证故障回切性能，缺省情况下，M-LAG成员端口上报Up状态的延时时间为240秒。若此时同时在VLANIF接口上配置三层协议状态延时Up时间，则需要保证M-LAG成员口的延时Up时间长于VLANIF接口的延时时间，否则ND同步失败的表项依赖于ND Miss触发学习。

f. 执行命令**commit**，提交配置。

----结束

1.7.8 （可选）配置 peer-link 故障场景下端口状态

背景信息

当M-LAG应用于普通以太网络、VXLAN网络或IP网络的双归接入时，peer-link故障但双主检测心跳状态正常会触发备设备上除管理网口、peer-link接口和堆叠口以外的接口处于Error-Down状态。一旦peer-link故障恢复，处于ERROR DOWN状态的M-LAG接口默认将在240s后自动恢复为Up状态，处于ERROR DOWN状态的其它接口将立即自动恢复为Up状态。

但在实际组网应用中，当某些上行端口运行路由协议或者是双主检测心跳口时是不希望被Error-Down的。此时，可以根据实际情况选择配置下列功能。

在peer-link故障但双主检测正常时，配置下列功能，设备接口Error-Down情况如表1-10所示。

表 1-10 设备在 peer-link 故障但双主检测正常时接口 Error-Down 情况

设备配置情况	M-LAG接入普通以太网络、VXLAN网络或IP网络
设备缺省情况	除管理网口、peer-link接口和堆叠口以外的接口处于ERROR DOWN状态。
设备仅配置suspend功能	仅M-LAG成员口以及配置该功能的接口处于ERROR DOWN状态。
设备仅配置reserved功能	除配置该功能的接口、管理网口、peer-link接口和堆叠口以外的接口处于ERROR DOWN状态。
设备同时配置suspend功能和reserved功能	仅M-LAG成员口以及配置suspend功能的接口处于ERROR DOWN状态。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**interface** *interface-type interface-number*，进入接口视图。

步骤3 执行命令**m-lag unpaired-port reserved**，配置备设备接口在peer-link故障但双主检测正常时不被Error-Down。

- 可以通过在备设备的其他接口上配置**m-lag unpaired-port suspend**，指定特定的接口在peer-link故障但双主检测心跳状态正常时自动被Error-Down。
- 建议在M-LAG主备设备对应接口同时配置该命令，保证M-LAG主备倒换后设备接口Error-Down情况一致。
- 该命令不支持在peer-link接口、M-LAG成员口下配置。

步骤4 执行命令**commit**，提交配置。

----结束

1.7.9 （可选）使能 M-LAG 三层转发增强功能

背景信息

为提升M-LAG成员口故障时的收敛速度，可以使能M-LAG三层转发增强，为M-LAG成员接口的所有ARP/ND表项申请备份的FRR资源，形成下行出接口的主备备份路径，在M-LAG成员口故障时能够快速切换出接口为peer-link接口。

说明

CE6810LI不支持IPv6场景下的M-LAG三层转发增强功能。

CE5800系列交换机（除CE5880EI外）、CE6810EI、CE6810LI、CE6850EI 不支持IPv4场景下的M-LAG三层转发增强功能。

操作步骤

步骤1 执行命令**system-view**，进入系统视图。

步骤2 执行命令**m-lag forward layer-3 enhanced enable**（CE6881、CE6881K、CE6863、CE6863K、CE6881E和CE6820）或者**m-lag forward layer-3 { ipv4 | ipv6 } enhanced enable**（除CE6881、CE6881K、CE6863、CE6863K、CE6881E和CE6820），使能M-LAG三层转发增强功能。

缺省情况下，M-LAG三层转发增强功能未使能。

在使能M-LAG三层转发增强功能后，对于报文出端口为M-LAG成员接口的所有ARP/ND表项会申请备份的FRR资源，出接口指向peer-link接口，形成主备路径转发。一旦FEI侧感知到M-LAG成员接口故障，设备双归接入变成单归，则将对应ARP/ND表项的下一跳由M-LAG成员接口切换为Peer-link接口，提升故障场景下的切换性能。

说明

- 在使能M-LAG三层转发增强功能后，由于下一跳资源消耗的增加可能导致主备路径下发不成功，引起流量丢包。
- 在使能M-LAG三层转发增强功能后，若配置Eth-Trunk接口加退M-LAG成员口时需要指定该Eth-Trunk口清除学习到的所有ARP/ND表项，避免由于上层协议模块不感知M-LAG成员口变化导致的FRR资源浪费。
- 在使能M-LAG三层转发增强功能后，需要间隔300s才可以配置去使能M-LAG三层转发增强功能；在去使能了M-LAG三层转发增强功能后，需要间隔300s才可以配置使能M-LAG三层转发增强功能。
- 针对CE6881、CE6881K、CE6863K、CE6881E和CE6863设备如果需要使用IPv4场景下M-LAG三层转发增强功能，需要同时配置**arp broadcast-detect enable**命令。

步骤3 执行命令**commit**，提交配置。

----结束

1.7.10 检查基于 V-STP 方式配置 M-LAG 的配置结果

操作步骤

- 执行命令**display dfs-group** *dfs-group-id* [**node** *node-id* **m-lag** [**brief**] | **peer-link**]，查看M-LAG的信息。
- 执行命令**display stp** [**process** *process-id*] **v-stp**，查看虚拟生成树实例（V-STP）的状态信息和统计信息。

----结束

后续处理

完成M-LAG配置后，如果peer-link故障但心跳状态正常会导致状态为备的设备上部分接口处于ERROR DOWN状态。Error-Down是指设备检测到故障后将接口状态设置为ERROR DOWN状态，此时接口不能收发报文，接口指示灯为常灭。可以通过**display error-down recovery**命令可以查看设备上所有被Error-Down的接口信息。

当M-LAG应用于普通以太网络、VXLAN网络或IP网络的双归接入时，peer-link故障但双主检测心跳状态正常会触发状态为备的设备上除管理网口、peer-link接口和堆叠口以外的接口处于Error-Down状态。一旦peer-link故障恢复，处于Error-Down状态的M-LAG接口默认将在240s后自动恢复为Up状态，处于Error-Down状态的其它物理接口将自动恢复为Up状态。

接口被Error-Down时，建议先排除引起peer-link故障的原因，不建议直接手动恢复或在系统视图下执行命令**error-down auto-recovery cause m-lag interval** *interval-value*使能接口状态自动恢复为Up的功能，否则可能会导致业务多包、丢包或不通等故障，请谨慎操作。

1.8 维护 M-LAG

通过维护M-LAG，通过实时查看M-LAG的故障原因，可以实时监控M-LAG的运行状况。当历史故障信息较多时，可以将这些故障信息清除。

1.8.1 监控 M-LAG 运行状况

背景信息

通过监控M-LAG运行状况，当M-LAG故障时，可以查看故障原因，通过获取的信息进行故障定位。

操作步骤

步骤1 执行命令**display m-lag troubleshooting** { **history** | **current** }，查看M-LAG发生故障的原因。

该命令最多显示最近100次故障的原因。

----结束

1.8.2 清除 M-LAG 历史故障原因信息

背景信息

当需要查看一定时间内的M-LAG历史故障原因信息时，可以先清除设备上已有的M-LAG历史故障原因信息，对故障原因重新进行记录。

说明

清除M-LAG历史故障原因信息后，之前的历史故障信息将无法恢复，请务必仔细确认。

操作步骤

- 在用户视图下，执行命令**reset m-lag troubleshooting history** ，清除设备上已有的M-LAG历史故障原因信息。

----结束

1.9 M-LAG 配置举例

介绍M-LAG的配置举例。配置示例中包括组网需求、配置思路、操作步骤等。

本节仅列举单特性的配置示例。如果您想了解更多综合场景配置案例、特性典型配置案例、对接案例、替换案例及行业案例，请参考典型配置案例。

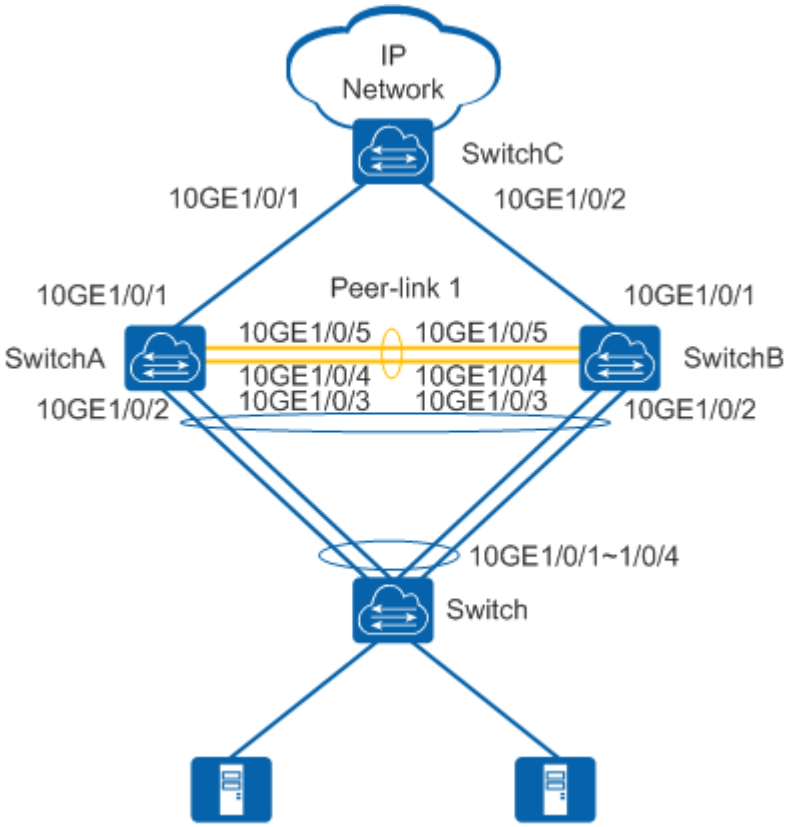
1.9.1 配置交换机双归接入 IP 网络示例（V-STP 方式）

组网需求

如图1-22所示，通过配置M-LAG双归接入IP网络可以满足以下要求：

- 当一条接入链路发生故障时，流量可以快速切换到另一条链路，保证可靠性。
- 为了高效利用带宽，两条链路同时处于active状态，可实现使用负载分担的方式转发流量。

图 1-22 配置 M-LAG 双归接入 IP 组网图



配置思路

采用如下的思路配置M-LAG双归接入IP网络：

1. 在Switch上配置上行接口绑定在一个Eth-Trunk中。
2. 分别在SwitchA和SwitchB上配置V-STP、DFS Group、peer-link和M-LAG接口。
3. 分别在SwitchA和SwitchB上配置VLANIF接口IP地址和MAC地址，作为接入设备的双活网关。
4. 分别在SwitchA、SwitchB和SwitchC上配置OSPF功能，保证三层互通。

说明

在V-STP场景下，为防止接口因生成树协议计算结果被阻塞，可以通过配置主接口实现三层互通或者去使能IP网络侧的生成树协议。

5. 分别在SwitchA和SwitchB上配置Monitor Link关联上行接口和下行接口，避免因上行链路故障导致用户侧流量无法转发而丢弃。

操作步骤

步骤1 在Switch上配置上行接口绑定在一个Eth-Trunk中

配置Switch。

```
<HUAWEI> system-view
[~HUAWEI] sysname Switch
[*HUAWEI] commit
[~Switch] vlan batch 11
[*Switch] interface eth-trunk 20
[*Switch-Eth-Trunk20] mode lacp-static
[*Switch-Eth-Trunk20] port link-type trunk
[*Switch-Eth-Trunk20] port trunk allow-pass vlan 11
[*Switch-Eth-Trunk20] trunkport 10ge 1/0/1 to 1/0/4
[*Switch-Eth-Trunk20] quit
[*Switch] commit
```

步骤2 分别在SwitchA和SwitchB上配置V-STP、DFS Group、peer-link和M-LAG接口

配置SwitchA。

```
<HUAWEI> system-view
[~HUAWEI] sysname SwitchA
[*HUAWEI] commit
[~SwitchA] stp mode rstp
[*SwitchA] stp v-stp enable
[*SwitchA] interface loopback 0
[*SwitchA-LoopBack0] ip address 10.1.1.1 32
[*SwitchA-LoopBack0] quit
[*SwitchA] dfs-group 1
[*SwitchA-dfs-group-1] source ip 10.1.1.1
[*SwitchA-dfs-group-1] priority 150
[*SwitchA-dfs-group-1] quit
[*SwitchA] interface eth-trunk 1
[*SwitchA-Eth-Trunk1] trunkport 10ge 1/0/4
[*SwitchA-Eth-Trunk1] trunkport 10ge 1/0/5
[*SwitchA-Eth-Trunk1] mode lacp-static
[*SwitchA-Eth-Trunk1] peer-link 1
[*SwitchA-Eth-Trunk1] quit
[*SwitchA] vlan batch 11
[*SwitchA] interface eth-trunk 10
[*SwitchA-Eth-Trunk10] mode lacp-static
[*SwitchA-Eth-Trunk10] port link-type trunk
[*SwitchA-Eth-Trunk10] port trunk allow-pass vlan 11
[*SwitchA-Eth-Trunk10] trunkport 10ge 1/0/2
[*SwitchA-Eth-Trunk10] trunkport 10ge 1/0/3
[*SwitchA-Eth-Trunk10] dfs-group 1 m-lag 1
[*SwitchA-Eth-Trunk10] quit
[*SwitchA] commit
```

配置SwitchB。

```
<HUAWEI> system-view
[~HUAWEI] sysname SwitchB
[*HUAWEI] commit
[~SwitchB] stp mode rstp
[*SwitchB] stp v-stp enable
[*SwitchB] interface loopback 0
```



```
[*SwitchB-LoopBack0] ip address 10.1.1.2 32
[*SwitchB-LoopBack0] quit
[*SwitchB] dfs-group 1
[*SwitchB-dfs-group-1] source ip 10.1.1.2
[*SwitchB-dfs-group-1] priority 120
[*SwitchB-dfs-group-1] quit
[*SwitchB] interface eth-trunk 1
[*SwitchB-Eth-Trunk1] trunkport 10ge 1/0/4
[*SwitchB-Eth-Trunk1] trunkport 10ge 1/0/5
[*SwitchB-Eth-Trunk1] mode lacp-static
[*SwitchB-Eth-Trunk1] peer-link 1
[*SwitchB-Eth-Trunk1] quit
[*SwitchB] vlan batch 11
[*SwitchB] interface eth-trunk 10
[*SwitchB-Eth-Trunk10] mode lacp-static
[*SwitchB-Eth-Trunk10] port link-type trunk
[*SwitchB-Eth-Trunk10] port trunk allow-pass vlan 11
[*SwitchB-Eth-Trunk10] trunkport 10ge 1/0/2
[*SwitchB-Eth-Trunk10] trunkport 10ge 1/0/3
[*SwitchB-Eth-Trunk10] dfs-group 1 m-lag 1
[*SwitchB-Eth-Trunk10] quit
[*SwitchB] commit
```

步骤3 分别在SwitchA和SwitchB上配置VLANIF接口IP地址和MAC地址，作为接入设备的双活网关

两端的虚拟IP和虚拟MAC配置要求完全一致，目的是为M-LAG提供相同的虚拟IP和虚拟MAC。

配置SwitchA。

```
[~SwitchA] interface vlanif 11
[*SwitchA-Vlanif11] ip address 10.2.1.1 24
[*SwitchA-Vlanif11] mac-address 0000-5e00-0101
[*SwitchA-Vlanif11] quit
[*SwitchA] commit
```

配置SwitchB。

```
[~SwitchB] interface vlanif 11
[*SwitchB-Vlanif11] ip address 10.2.1.1 24
[*SwitchB-Vlanif11] mac-address 0000-5e00-0101
[*SwitchB-Vlanif11] quit
[*SwitchB] commit
```

步骤4 分别在SwitchA、SwitchB和SwitchC上配置OSPF功能，保证三层互通

配置SwitchA。

```
[~SwitchA] interface 10ge 1/0/1
[~SwitchA-10GE1/0/1] undo portswitch
[*SwitchA-10GE1/0/1] ip address 10.3.1.1 24
[*SwitchA-10GE1/0/1] quit
[*SwitchA] ospf 1
[*SwitchA-ospf-1] area 0
[*SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.1 0.0.0.0
[*SwitchA-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[*SwitchA-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[*SwitchA-ospf-1-area-0.0.0.0] quit
[*SwitchA-ospf-1] quit
[*SwitchA] commit
```

配置SwitchB。

```
[~SwitchB] interface 10ge 1/0/1
[~SwitchB-10GE1/0/1] undo portswitch
[*SwitchB-10GE1/0/1] ip address 10.4.1.1 24
[*SwitchB-10GE1/0/1] quit
[*SwitchB] ospf 1
[*SwitchB-ospf-1] area 0
[*SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.2 0.0.0.0
[*SwitchB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
```



```
[*SwitchB-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[*SwitchB-ospf-1-area-0.0.0.0] quit
[*SwitchB-ospf-1] quit
[*SwitchB] commit
```

配置SwitchC。

```
<HUAWEI> system-view
[~HUAWEI] sysname SwitchC
[*HUAWEI] commit
[~SwitchC] interface 10ge 1/0/1
[~SwitchC-10GE1/0/1] undo portswitch
[*SwitchC-10GE1/0/1] ip address 10.3.1.2 24
[*SwitchC-10GE1/0/1] quit
[*SwitchC] interface 10ge 1/0/2
[*SwitchC-10GE1/0/2] undo portswitch
[*SwitchC-10GE1/0/2] ip address 10.4.1.2 24
[*SwitchC-10GE1/0/2] quit
[*SwitchC] ospf 1
[*SwitchC-ospf-1] area 0
[*SwitchC-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[*SwitchC-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[*SwitchC-ospf-1-area-0.0.0.0] quit
[*SwitchC-ospf-1] quit
```

步骤5 分别在SwitchA和SwitchB上配置Monitor Link关联上行接口和下行接口

配置SwitchA。

```
[~SwitchA] monitor-link group 1
[*SwitchA-mtlk-group1] port 10ge 1/0/1 uplink
[*SwitchA-mtlk-group1] port eth-trunk 10 downlink 1
[*SwitchA-mtlk-group1] quit
[*SwitchA] commit
```

配置SwitchB。

```
[~SwitchB] monitor-link group 1
[*SwitchB-mtlk-group1] port 10ge 1/0/1 uplink
[*SwitchB-mtlk-group1] port eth-trunk 10 downlink 1
[*SwitchB-mtlk-group1] quit
[*SwitchB] commit
```

步骤6 验证配置结果

执行命令**display dfs-group**，查看M-LAG的相关信息。

查看DFS Group编号为1的M-LAG信息。

```
[~SwitchA] display dfs-group 1 m-lag
*             : Local node
Heart beat state : OK
Node 1 *
  Dfs-Group ID   : 1
  Priority       : 150
  Address        : ip address 10.1.1.1
  State          : Master
  Causation      : -
  System ID      : 0025-9e95-7c31
  SysName        : SwitchA
  Version        : V100R006C00
  Device Type    : CE6850EI
Node 2
  Dfs-Group ID   : 1
  Priority       : 120
  Address        : ip address 10.1.1.2
  State          : Backup
  Causation      : -
  System ID      : 0025-9e95-7c11
  SysName        : SwitchB
  Version        : V100R006C00
  Device Type    : CE6850EI
```


查看SwitchA上的M-LAG信息。

```
[~SwitchA] display dfs-group 1 node 1 m-lag brief
* - Local node

M-Lag ID     Interface      Port State    Status              Consistency-chec
      1      Eth-Trunk 10   Up            active(*)-active    --


Failed reason:
   1 -- Relationship between vlan and port is inconsistent
   2 -- STP configuration under the port is inconsistent
   3 -- STP port priority configuration is inconsistent
   4 -- LACP mode of M-LAG is inconsistent
   5 -- M-LAG configuration is inconsistent
   6 -- The number of M-LAG members is inconsistent
```

查看SwitchB上的M-LAG信息。

```
[~SwitchB] display dfs-group 1 node 2 m-lag brief
* - Local node

M-Lag ID     Interface      Port State    Status              Consistency-chec
      1      Eth-Trunk 10   Up            active-active(*)    --


Failed reason:
   1 -- Relationship between vlan and port is inconsistent
   2 -- STP configuration under the port is inconsistent
   3 -- STP port priority configuration is inconsistent
   4 -- LACP mode of M-LAG is inconsistent
   5 -- M-LAG configuration is inconsistent
   6 -- The number of M-LAG members is inconsistent
```

通过以上显示信息可以看到，“Heart beat state”的状态是“OK”，表明心跳状态正常；SwitchA作为Node 1，优先级为150，“State”的状态是“Master”；SwitchB作为Node 2，优先级为120，“State”的状态是“Backup”。同时“Causation”的状态是“-”，Node 1的“Port State”状态为“Up”，Node 2的“Port State”状态为“Up”，且Node 1和Node 2的M-LAG状态均为“active”，表明M-LAG的配置正确。

----结束

配置文件

- SwitchA的配置文件

```
#
sysname SwitchA
#
dfs-group 1
 priority 150
 source ip 10.1.1.1
#
vlan batch 11
#
stp mode rstp
stp v-stp enable
#
interface Vlanif11
 ip address 10.2.1.1 255.255.255.0
 mac-address 0000-5e00-0101
#
interface Eth-Trunk1
 mode lacp-static
 peer-link 1
#
interface Eth-Trunk10
```



```
 port link-type trunk
 port trunk allow-pass vlan 11
 mode lacp-static
 dfs-group 1 m-lag 1
#
interface 10GE1/0/1
 undo portswitch
 ip address 10.3.1.1 255.255.255.0
#
interface 10GE1/0/2
 eth-trunk 10
#
interface 10GE1/0/3
 eth-trunk 10
#
interface 10GE1/0/4
 eth-trunk 1
#
interface 10GE1/0/5
 eth-trunk 1
#
interface LoopBack0
 ip address 10.1.1.1 255.255.255.255
#
monitor-link group 1
 port 10GE1/0/1 uplink
 port Eth-Trunk10 downlink 1
#
ospf 1
 area 0.0.0.0
  network 10.1.1.1 0.0.0.0
  network 10.2.1.0 0.0.0.255
  network 10.3.1.0 0.0.0.255
#
return
```

- SwitchB的配置文件

```
#
sysname SwitchB
#
dfs-group 1
 priority 120
 source ip 10.1.1.2
#
vlan batch 11
#
stp mode rstp
stp v-stp enable
#
interface Vlanif11
 ip address 10.2.1.1 255.255.255.0
 mac-address 0000-5e00-0101
#
interface Eth-Trunk1
 mode lacp-static
 peer-link 1
#
interface Eth-Trunk10
 port link-type trunk
 port trunk allow-pass vlan 11
 mode lacp-static
 dfs-group 1 m-lag 1
#
interface 10GE1/0/1
 undo portswitch
 ip address 10.4.1.1 255.255.255.0
#
interface 10GE1/0/2
 eth-trunk 10
#
```



```
interface 10GE1/0/3
 eth-trunk 10
#
interface 10GE1/0/4
 eth-trunk 1
#
interface 10GE1/0/5
 eth-trunk 1
#
interface LoopBack0
 ip address 10.1.1.2 255.255.255.255
#
monitor-link group 1
 port 10GE1/0/1 uplink
 port Eth-Trunk10 downlink 1
#
ospf 1
 area 0.0.0.0
  network 10.1.1.2 0.0.0.0
  network 10.2.1.0 0.0.0.255
  network 10.4.1.0 0.0.0.255
#
return
```

- SwitchC的配置文件

```
#
sysname SwitchC
#
interface 10GE1/0/1
 undo portswitch
 ip address 10.3.1.2 255.255.255.0
#
interface 10GE1/0/2
 undo portswitch
 ip address 10.4.1.2 255.255.255.0
#
ospf 1
 area 0.0.0.0
  network 10.3.1.0 0.0.0.255
  network 10.4.1.0 0.0.0.255
#
return
```

- Switch的配置文件

```
#
sysname Switch
#
vlan batch 11
#
interface Eth-Trunk20
 port link-type trunk
 port trunk allow-pass vlan 11
 mode lacp-static
#
interface 10GE1/0/1
 eth-trunk 20
#
interface 10GE1/0/2
 eth-trunk 20
#
interface 10GE1/0/3
 eth-trunk 20
#
interface 10GE1/0/4
 eth-trunk 20
#
return
```


1.10 M-LAG 技术专题

介绍M-LAG在实际配置过程中的推荐部署模型以及配置建议。

以上章节仅列举单特性的配置过程以及示例。如果您想了解更多综合场景下M-LAG的推荐部署模型以及配置建议，请参考**M-LAG技术专题**。