M-LAG基础原理及组网应用介绍

合作伙伴支持部

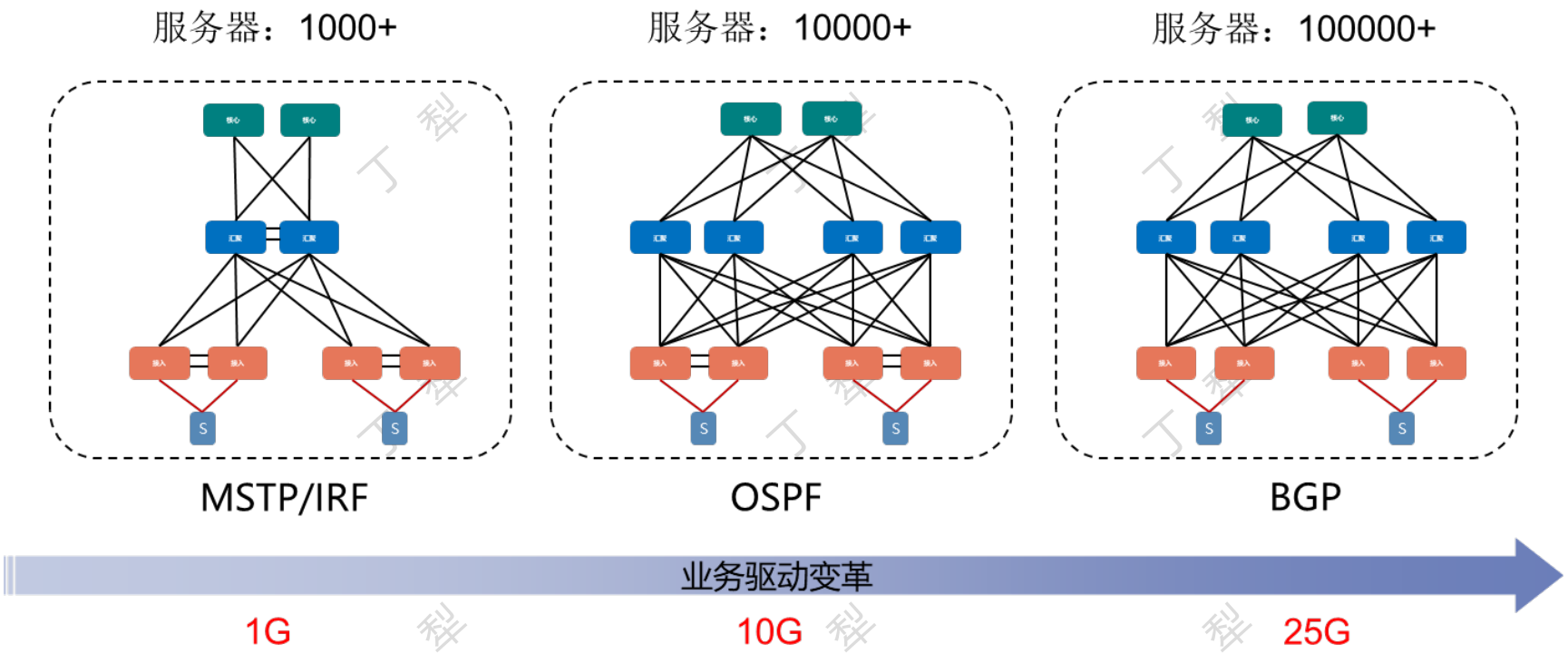
1

PART 01

去堆叠技术介绍

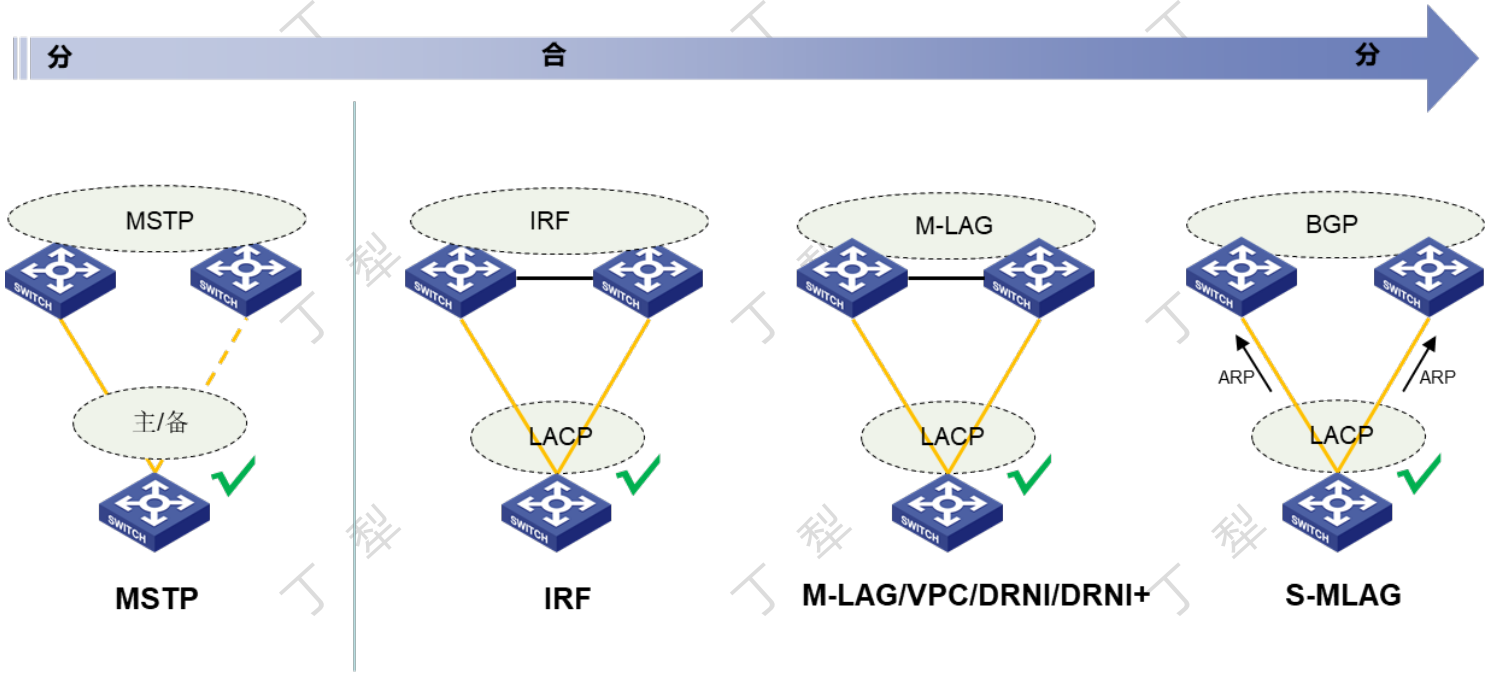
服务器接入发展

- 数据中心网络架构由传统的二层架构，过度成为了OSPF和BGP的全三层架构，理论上BGP三层架构组网中可以承载100000+的服务器。



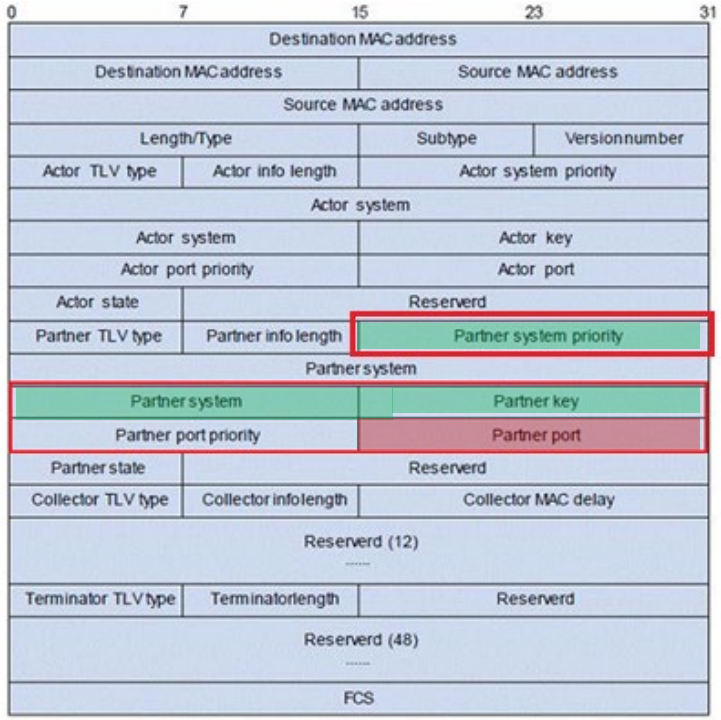
服务器接入方式发展

- 第一阶段：接入层的交换机不支持虚拟化，接入交换机独立运行，服务器网卡工作在主备模式，交换机表项依靠数据流量刷新。
- 第二阶段：接入层交换机支持IRF等虚拟化技术，将多台设备虚拟化为一台，支持与服务器进行链路聚合，从而实现链路双活提高链路利用率。
- 第三阶段：在M-LAG和S-MLAG的出现后，实现了在接入层交换机控制层面独立的情况下实现了接入层链路双活接入，S-MLAG实现已经被互联网用户所接受。



去堆叠技术

- 动态聚合——在去堆叠方案中实现跨设备链路聚合，需要解决两个问题：
  - 问题1：如何让服务器认为连接对端的接入交换机是同一台设备；
  - 问题2：两台接入设备上服务器表项的同步；
- 静态聚合——天然可以实现跨设备链路聚合，但静态聚合缺乏LACP报文对链路的监控和协商机制，同样需要解决上述问题2。



思考：动态聚合（IEEE802.3ad）能聚合成功的要素（针对LACPDU报文）

- 当Partner_System_Priority和Partner_System一致时，则认为对端设备为同一个设备。
- 本端的不同端口，接收LACPDU报文中要求Pantner_Port不一致，同时Partner_key一致时则可以聚合成功。

问题解决方法：

- 手工配置参数，确保相关Partner参数一致；
- 通过M-LAG协议报文（建立RLINK通道）同步转发表项，解决服务器之间的表项同步问题。

去堆叠技术与堆叠对比（IRF与M-LAG对比）

- 下表为IRF和M-LAG对比，组网可靠性要求高，升级过程要求业务中断时间短的场景推荐使用M-LAG。
- 需要注意的是，在同一组网环境中，不能同时部署IRF和M-LAG。

项目	IRF	M-LAG
控制面	• 所有成员设备控制面统一，集中管理 • 所有成员设备需要同步所有表项	• 两台独立设备，控制平面解耦 • 主要同步MAC表项/ARP表项/ND表项
设备面	紧耦合 • 硬件要求：芯片架构相同，要求同系列 • 软件要求：必须相同版本	松耦合 • 硬件要求：支持不同型号 • 软件要求：支持不同版本 （由于M-LAG的特性支持情况还在快速发展阶段，现阶段部分产品要求相同型号、版本）
版本升级	• 需要成员设备同步升级，或者主设备、从设备分开升级但操作较复杂 • 升级时业务最小中断时间2s左右	可独立升级，升级时业务中断时间小于1s 对于支持GIR（Graceful Insertion and Removal，平滑插入和移除）的版本，可以做到不中断。
配置管理	统一配置，统一管理，操作简单 耦合度高，和控制器配合存在单点故障可能	独立配置，M-LAG系统会进行配置一致性检查，具体业务配置需要手工保证 独立管理，耦合度低，和控制器配合使用不存在单点故障，可靠性更高
三层转发	无需特殊配置	双活接入，依赖三层接口特殊配置或VRRP

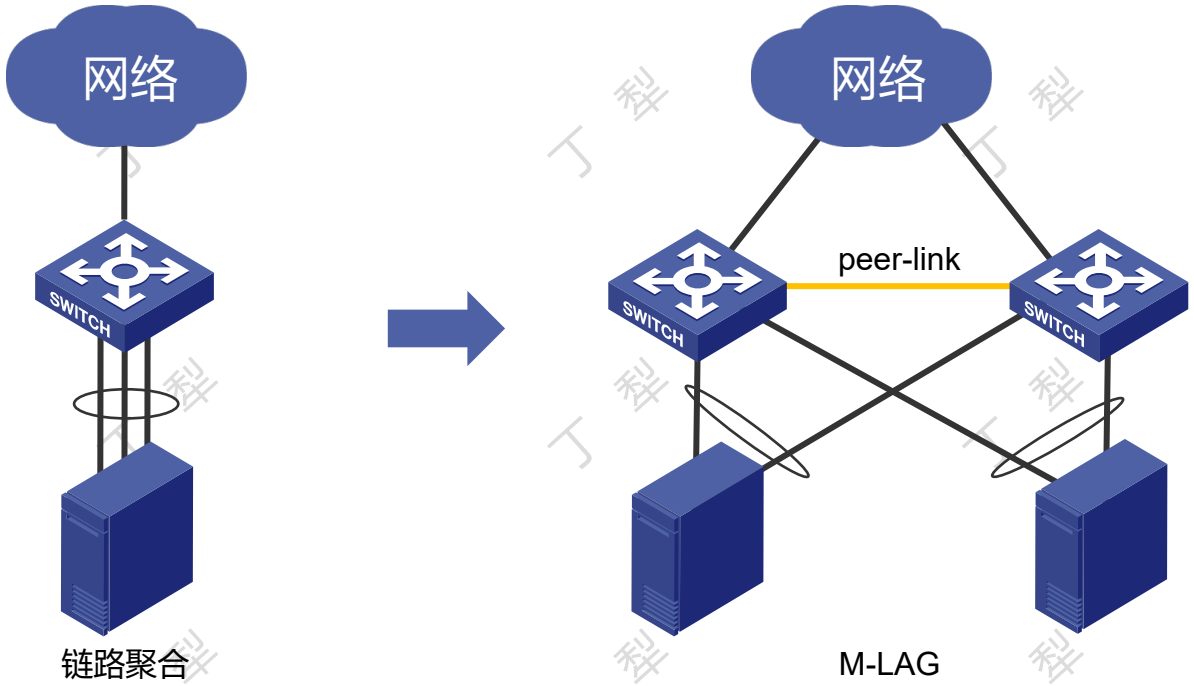
2

PART 02

M-LAG协议及相关机制介绍

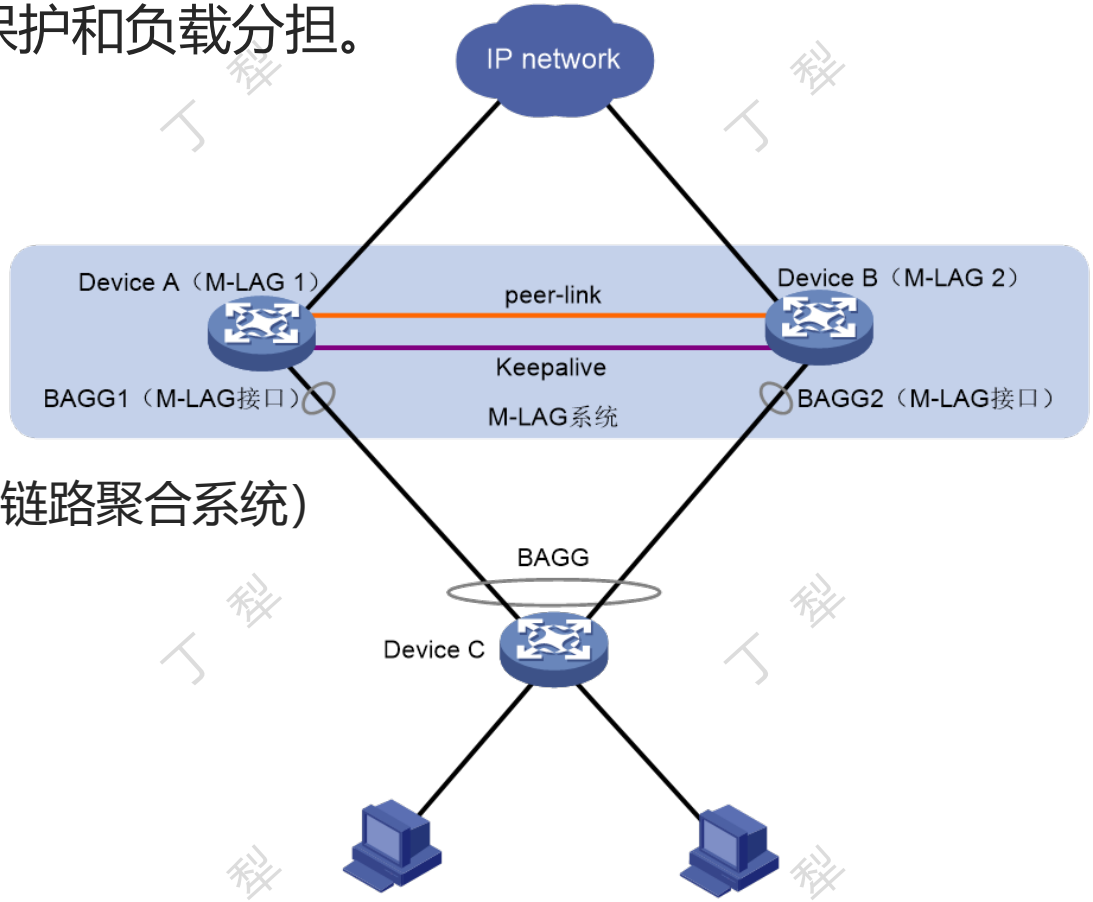
M-LAG 简介

- M-LAG（Multichassis link aggregation，跨设备链路聚合）是将两台物理设备虚拟成一台设备来实现跨设备链路聚合，从而提供设备级冗余保护和流量负载分担。
- M-LAG主要应用于双归接入组网，将可靠性从链路级提高到设备级。



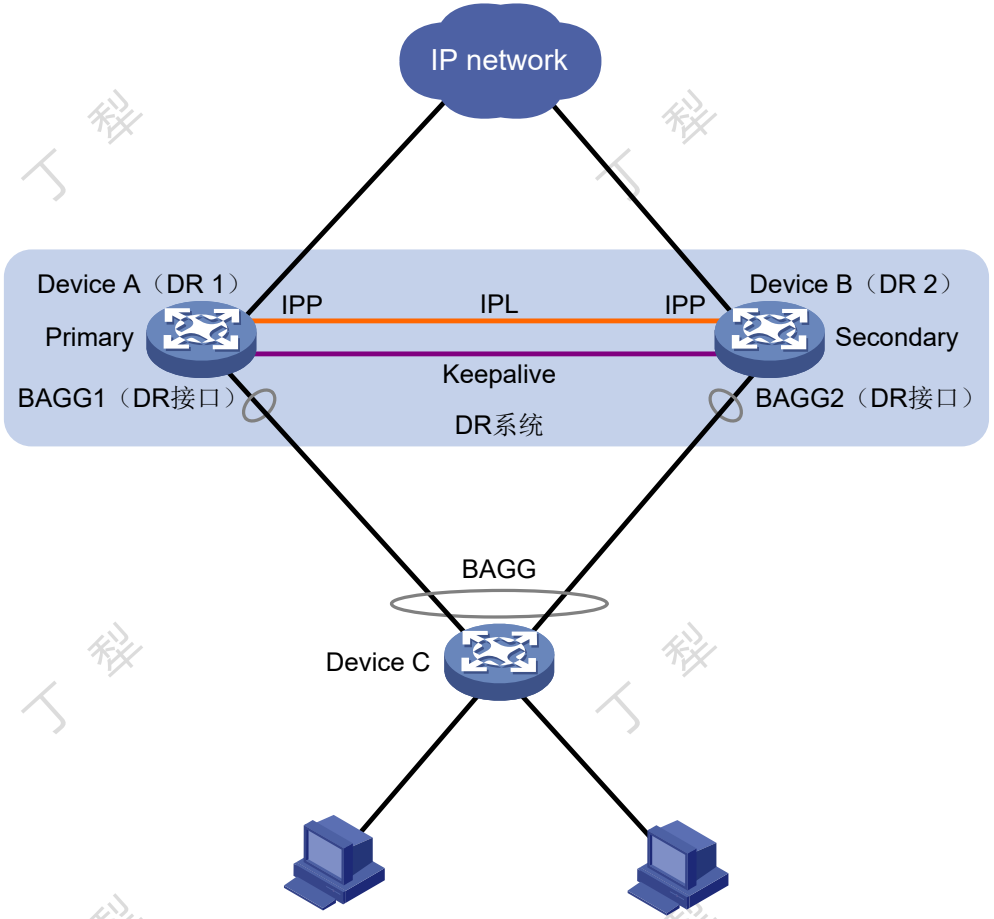
M-LAG网络模型及基本概念

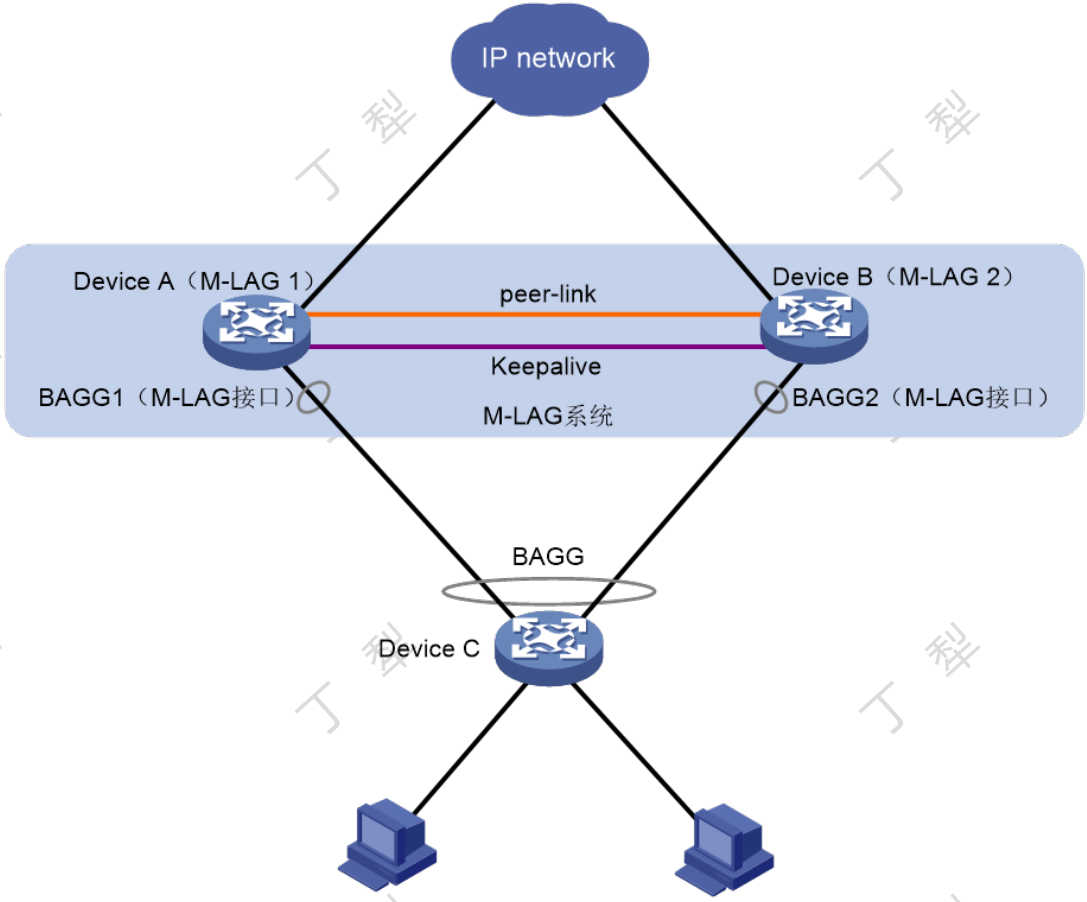
- M-LAG是一种设备二层链路聚合技术，将两台物理设备在二层聚合层面虚拟成一台设备来实现跨设备链路聚合，从而实现二层流量转发设备级冗余保护和负载分担。
- 协议标准：IEEE Std 802.1AX（各厂商都是私有实现）



- M-LAG相关基本概念：
  - M-LAG系统（ Multichassis link aggregation，跨设备链路聚合系统）
  - M-LAG接口（跨设备链路聚合接口）
  - peer-link接口（内部控制链路端口）
  - peer-link链路（ 内部控制链路）
  - Keepalive链路

DRNI→M-LAG


IP network
Device A（DR 1）
IPP
IPL
IPP
Device B（DR 2）
Primary
Secondary
Keepalive
BAGG1（DR接口）
BAGG2（DR接口）
DR系统
BAGG
Device C


IP network
Device A（M-LAG 1）
peer-link
Device B（M-LAG 2）
Keepalive
BAGG1（M-LAG接口）
BAGG2（M-LAG接口）
M-LAG系统
BAGG
Device C

M-LAG设备角色

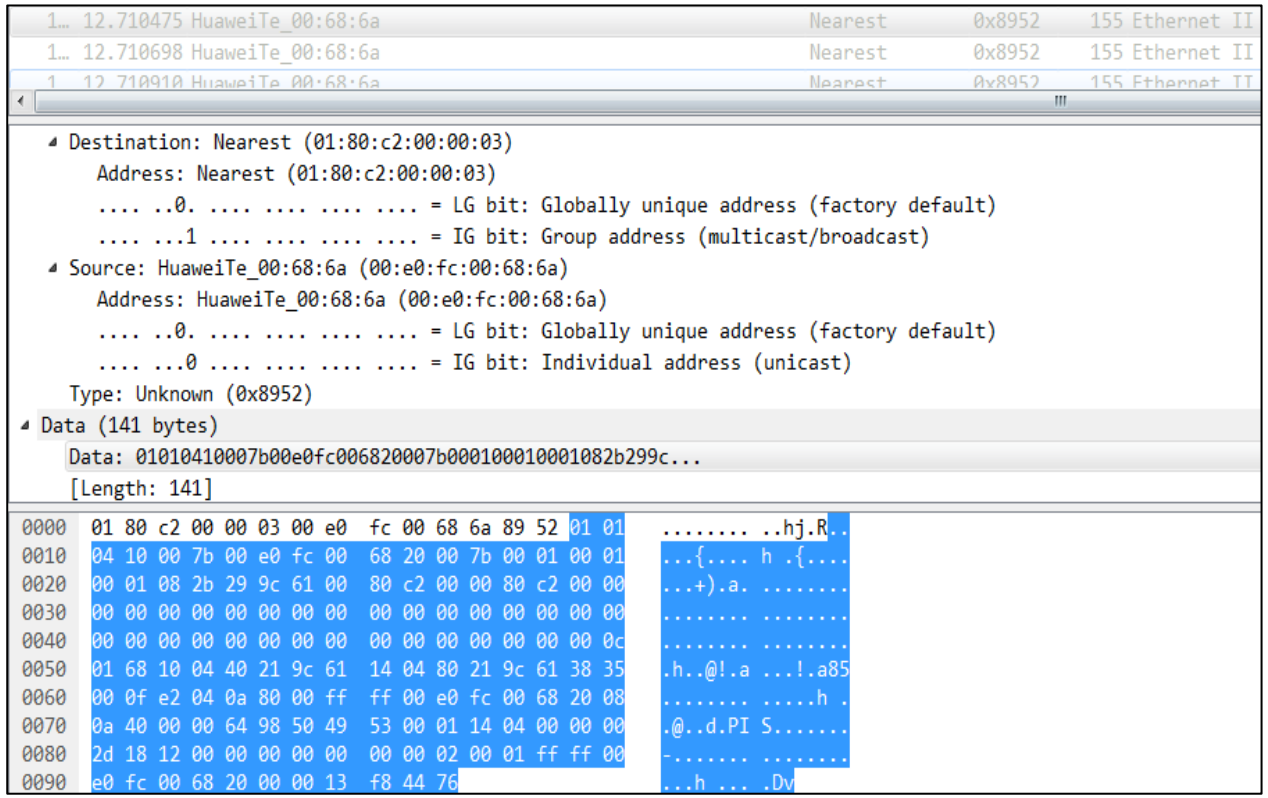
- 角色划分：
  - None：设备启动时的状态，无设备角色；当设备无可工作的M-LAG或者无peer-link接口时也会进入此状态；
  - Primary：主设备
  - Secondary：从设备
- 生效角色：实际工作角色
- 配置角色：根据配置值计算出的角色
- 角色保持：后加入的设备不会抢占已经存在的Primary角色
- 角色的应用场景：
  - M-LAG MAD
  - 集中式计算（STP等）

显示M-LAG设备角色信息。

```
<Sysname> display m-lag role
                    Effective role information
Factors                   Local                   Peer
Effective role            Primary                 Secondary
Initial role              None                    None
MAD DOWN state            Yes                     Yes
Health level              0                       0
Role priority             32768                   32768
Bridge MAC                3cd4-3ce1-0200          3cd4-437d-0300
Effective role trigger: Peer link calculation
Effective role reason: Bridge MAC

                    Configured role information
Factors                   Local                   Peer
Configured role           Primary                 Secondary
Role priority             32768                   32768
Bridge MAC                3cd4-3ce1-0200          3cd4-437d-0300
```

M-LAG系统的建立的协议和通道-DRCP

- M-LAG通过在peer-link链路上运行DRCP协议来交互分布式聚合的相关信息，以确定两台设备是否可以组成M-LAG系统。
- 运行该协议的设备之间通过互发DRCPDU（Distributed Relay Control Protocol Data Unit，分布式聚合控制协议数据单元）来交互分布式聚合的相关信息

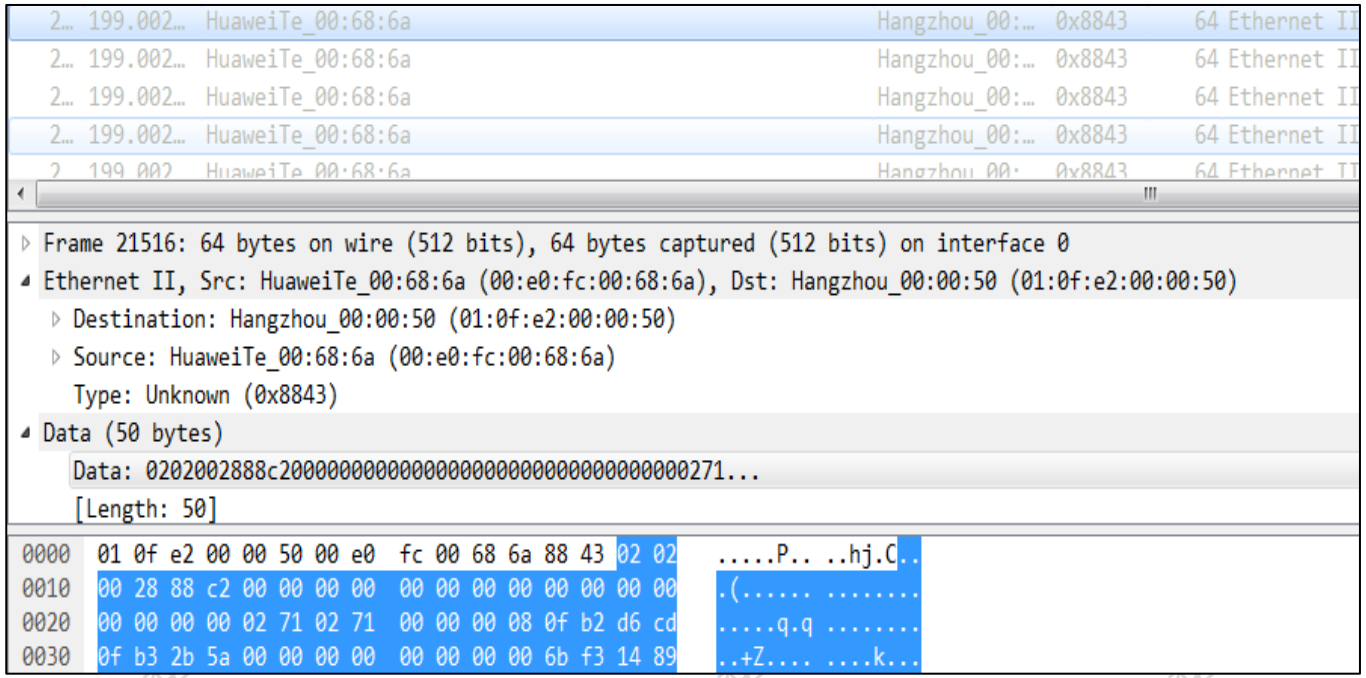


- 在peer-link链路上运行，所有的DRCP报文都是通过peer-link接口收发
- 基于数据链路层传输报文，无VLAN tag
- Eth协议号为8952
- 点对点传输，只有一跳，不能泛洪
- 不可靠传输，没有重传
- 交互信息包括：
  - 系统信息（Mac，Priority，Number）
  - 端口信息（M-LAGID，LocalPorts，PeerPorts，State）
  - 角色信息（Role，RolePriority， BridgeMAC）
- DRCP超时时间：长超时（90秒），短超时（3秒）

  m-lag drcp period short

M-LAG系统的建立的协议和通道-RLINK

- M-LAG通过在peer-link链路上运行RLINK通道来进行协议控制报文的交互和表项的同步等,eth.type协议号为8843



- RLINK的作用：
  - 同步协议控制信息（比如STP根信息）
  - 同步配置一致性检查数据
  - 同步协议报文（比如ARP/STP等报文）
  - 同步表项信息（比如MAC/ARP等）
  - M-LAG设备间的数据报文不通过该协议透传
- RLINK的实现：
  - 基于数据链路层传输报文，无VLAN tag
  - 无连接，通信前不用建立虚拟通道
  - 在peer-link链路上点对点传输，只有一跳，不能泛洪
  - 可靠传输，支持重传，有seq
  - 支持分片和重组
  - 支持保序

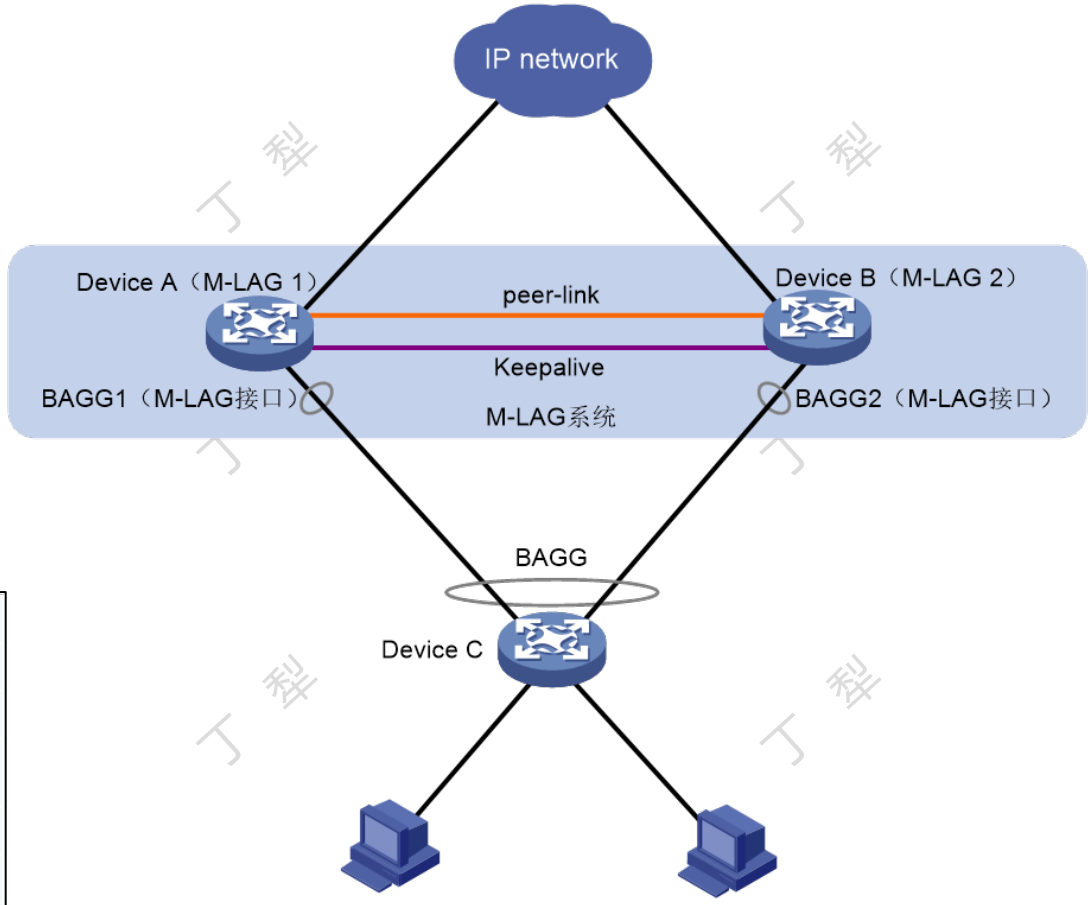
……

M-LAG系统的建立的基本配置-M-LAG聚合

2个M-LAG系统中相同的M-LAG聚合需要：

- 具有相同的System MAC，可单独配置，默认取全局
- 具有相同的System Priority，可单独配置，默认取全局
- 具有相同的操作KEY
- 具有不同的PortNumber
- 具有相同的聚合模式
- 具有相同的选择逻辑

Keepalive		
	支持三层业务口	通过业务网络互通
	支持管理口	通过管理网络互通
	支持vrf	在管理口通过vrf将keepalive报文与业务报文隔离
peer-link链路		
	支持二层聚合口	有peer-link二层物理链路
	支持VxLAN TUNNEL	支持无peer-link的overlay转发
M-LAG接口		
	支持二层聚合口	目前仅支持二层聚合的MLAG，动态、静态均可



M-LAG系统的建立的基本配置-系统参数

```
M-LAG system-mac 0001-0001-0001
M-LAG system-number 1
M-LAG system-priority 123
M-LAG role priority 65535
M-LAG mad exclude logical-interfaces
```

- M-LAG系统中相互配对的M-LAG接口的系统MAC地址必须相同；
- M-LAG系统中不同M-LAG设备的系统编号必须不同；
- M-LAG系统中相互配对的的M-LAG接口的系统优先级必须相同；
- M-LAG系统形成后，修改角色优先级、系统MAC、系统编号、系统优先级都会导致M-LAG系统分裂，不建议直接修改；
- 对于一个M-LAG聚合接口来说，system-mac和system-priority优先采用该聚合接口下的配置，只有该聚合接口下未进行配置时，才采用系统视图下的配置；
- 设备角色优先级用于两台设备间进行主从协商，值越小优先级越高，优先级高的为主设备。如果优先级相同，那么比较两台设备的桥MAC地址，桥MAC地址较小的为主设备。

M-LAG系统的建立的基本配置-peer-link接口/M-LAG接口

```
interface Bridge-Aggregation100  //peer-link接口
 port link-type trunk
 port trunk permit vlan all
 link-aggregation mode dynamic
 port m-lag peer-link 1
```

```
interface Bridge-Aggregation10  //m-lag接口
 port link-type trunk
 port trunk permit vlan 1 10
 link-aggregation mode dynamic
 port m-lag group 1
```

- 配置二层聚合接口或Tunnel接口为peer-link接口口
  - 该聚合接口不能是M-LAG接口；一台M-LAG设备上只能配置一个peer-link接口口；peer-link接口聚合口需要设置为动态聚合；目前peer-link接口口创建时默认permit vlan all；peer-link接口口协议类型的保持默认，例如不需要关闭stp等；两端M-LAG设备的peer-link接口口上允许通过的超长帧需要相同，否则会导致M-LAG设备间信息同步失败；
- 配置二层聚合接口加入分布式聚合组，即M-LAG聚合，一台设备上可以配置多个M-LAG接口，一个二层聚合接口只能加入一个分布式聚合组；

M-LAG系统的建立的基本配置-keepalive

```
M-LAG keepalive ip destination 1.1.1.2 source 1.1.1.1

%Sep 16 03:30:43:600 2019 DUT3 M-LAG/6/M-LAG_KEEPALIVELINK_DOWN: -MDC=1; Keepalive link went down.
%Sep 16 03:31:34:178 2019 DUT3 M-LAG/6/M-LAG_KEEPALIVELINK_UP: -MDC=1; Keepalive link came up.
```

- 需要配置M-LAG设备间通过Keepalive链路检测邻居状态，Keepalive报文周期发送用于检测对端设备是否故障，peer-link接口 DOWN时也可用来计算角色；
- Keepalive链路要独立于peer-link链路链路，否则会误检测；
- Keepalive报文是UDP报文，三层可达即可，可以使用网管口；
- 发包间隔：取值为100～10000，单位为毫秒，默认1000毫秒；
- 超时间隔：取值为3～20，单位为秒，默认5秒；

M-LAG的配置一致性检查

- M-LAG系统建立过程中会进行配置一致性检查（DRCP），以确保两端M-LAG设备配置一致，不影响M-LAG设备转发报文；
- M-LAG设备通过交换各自的配置信息，检查配置是否一致；
- 目前M-LAG支持对两种类型的配置一致性检查：
  - Type 1类型配置：影响M-LAG系统转发的配置。如果Type 1类型配置不一致，则将从设备上M-LAG接口置为down状态。
  - Type 2类型配置：仅影响业务模块的配置。如果Type 2类型配置不一致，从设备上M-LAG接口依然为up状态，不影响M-LAG系统正常工作。由Type 2类型配置对应的业务模块决定是否关闭该业务功能，其他业务模块不受影响。

    %Sep 18 16:16:02:323 2023 DeviceA M-LAG/6/MLAG_IFCHECK_INCONSISTENCY: Detected M-LAG interface Bridge-Aggregation3 type 2 configuration inconsistency.
  - Primary设备只打印LOG，没有动作。Secondary设备：
    - *全局Type1类型配置不一致，M-LAG DOWN掉所有M-LAG接口，打印LOG；*
    - *M-LAG接口Type1类型配置不一致，M-LAG DOWN掉本M-LAG接口，打印LOG；*
    - *Type2类型配置不一致，不DOWN端口，只打印LOG。*
  - 一致性配置检查显示中只显示两侧不一样的部分，相同的部分不做显示。

M-LAG的配置一致性检查

- Primary设备只打印LOG，没有动作。Secondary设备：
  - *全局Type1类型配置不一致，M-LAG DOWN掉所有M-LAG接口，打印LOG；*
  - *M-LAG接口Type1类型配置不一致，M-LAG DOWN掉本M-LAG接口，打印LOG；*
  - *Type2类型配置不一致，不DOWN端口，只打印LOG。*

配置项	配置命令	Type
LAGG	[intf]link-aggregation mode dynamic	Type1
VLAN	[intf] Local link type：Trunk（端口类型） [intf] Local PVID：10（端口PVID）	Type1
VLAN	[intf] VLAN permitted on local M-LAG interface：1-10（M-LAG接口实际允许通过的VLAN） [sys] Local VLAN interface：2-10，15，20-30，40，50	Type2
STP	[intf] stp enable （仅对M-LAG接口进行配置一致性检查） [sys] stp global enable / stp mode / stp vlan enable / MSTP VLAN-to-instance mappings等	Type1

M-LAG MAD机制

M-LAG MAD DOWN时机：

- peer-link链路故障后，为了防止从设备继续转发流量，M-LAG提供MAD机制，将Secondary设备上除M-LAG保留接口以外的接口置于M-LAG MAD DOWN状态（S105 R759X版本后通过健康度检查Down设备）。
- peer-link链路故障恢复后，为了防止丢包，完成角色选举后，Primary角色的设备M-LAG接口立即转发流量，Secondary设备要等待M-LAG MAD延迟恢复时间后（m-lag restore-delay 30）才能转发流量。

M-LAG 保留口：

- 系统保留口：管理口，peer-link口及其成员。
- 配置保留口：Keepalive相关口，M-LAG转发相关的VLAN虚接口，VSI虚接口等。

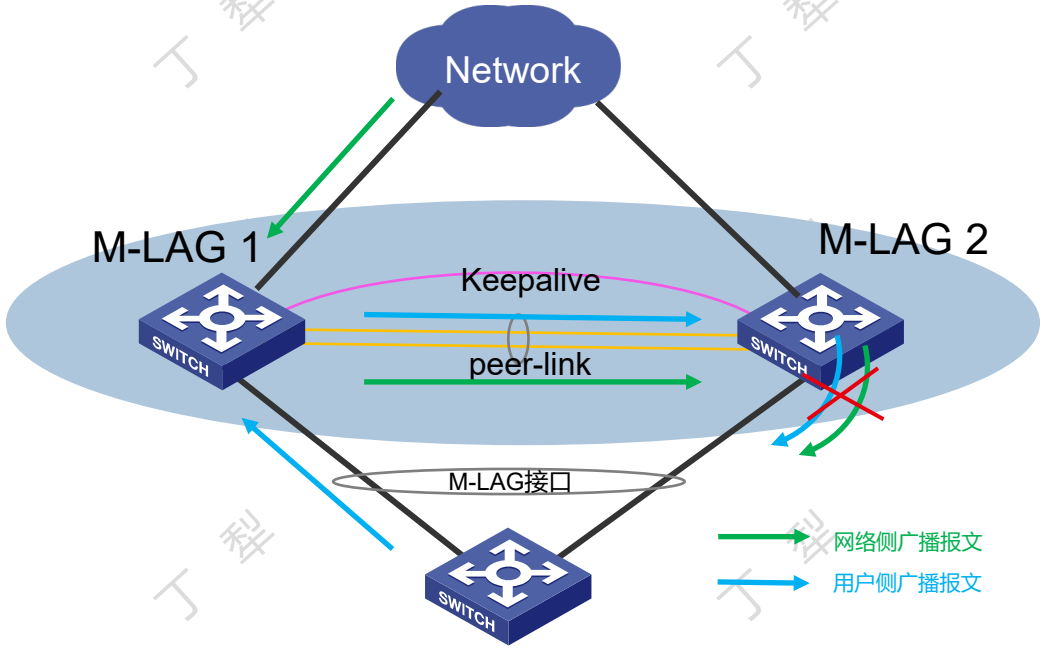
M-LAG设备重启后的自动恢复时间设定

- M-LAG系统中主从设备由于故障重启，仅一台M-LAG设备恢复启动后，缺省情况下，该设备处于None角色，所有M-LAG接口处于M-LAG DOWN状态。此时用户流量无法通过M-LAG接口转发。

- 为了避免上述情况出现，可以配置**m-lag auto-recovery reload-delay** *delay-value*功能，在设备重启后启动自动恢复定时器。
- 当自动恢复定时器超时后，该设备上M-LAG接口被置为非M-LAG DOWN状态：
  - 如果存在处于up状态的M-LAG接口，则用户流量可以正常转发；
  - 如果不存在处于up状态的M-LAG接口，则设备保持None角色，用户流量无法转发。

M-LAG防环机制—内部自动消环

- 不依赖STP这种环网协议；
- 单边绑定的M-LAG接口（即仅一台M-LAG设备配置M-LAG接口），缺省会自动将该M-LAG接口shutdown；
- 单边绑定的peer-link接口会自动link down；
- peer-link接口接收的流量不能发往，两边都有选中接口的，M-LAG接口。但M-LAG接口接收的流量可以发往peer-link接口，实现单向隔离；
- 本地M-LAG接口无选中接口后，Peer设备会解除单向隔离。



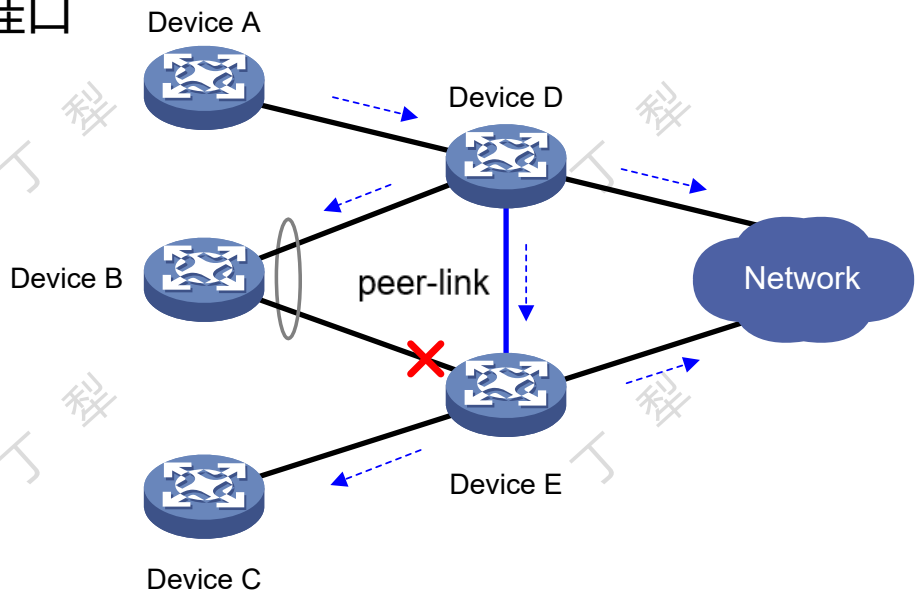
3

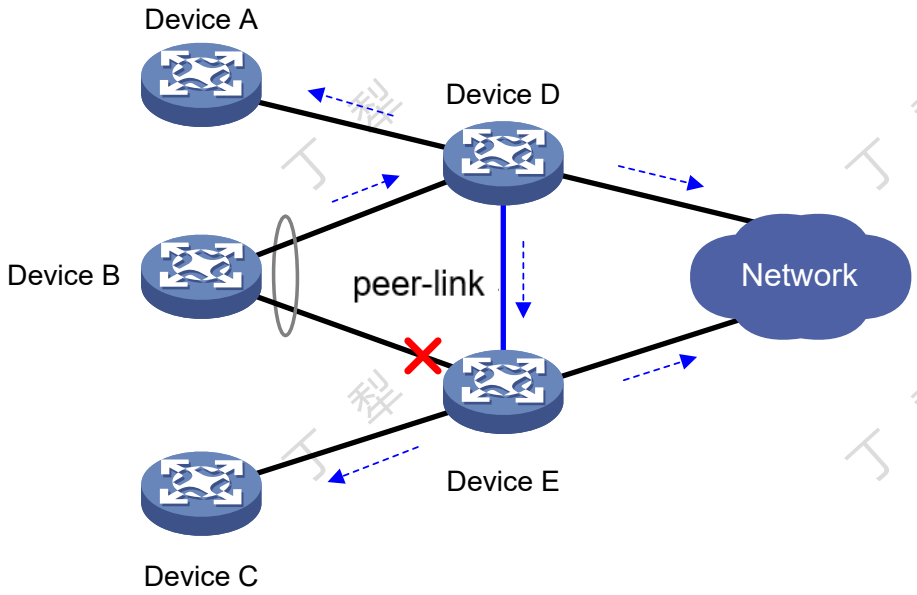
PART 03

M-LAG系统流量转发及故障处理

M-LAG的转发-用户侧广播流量

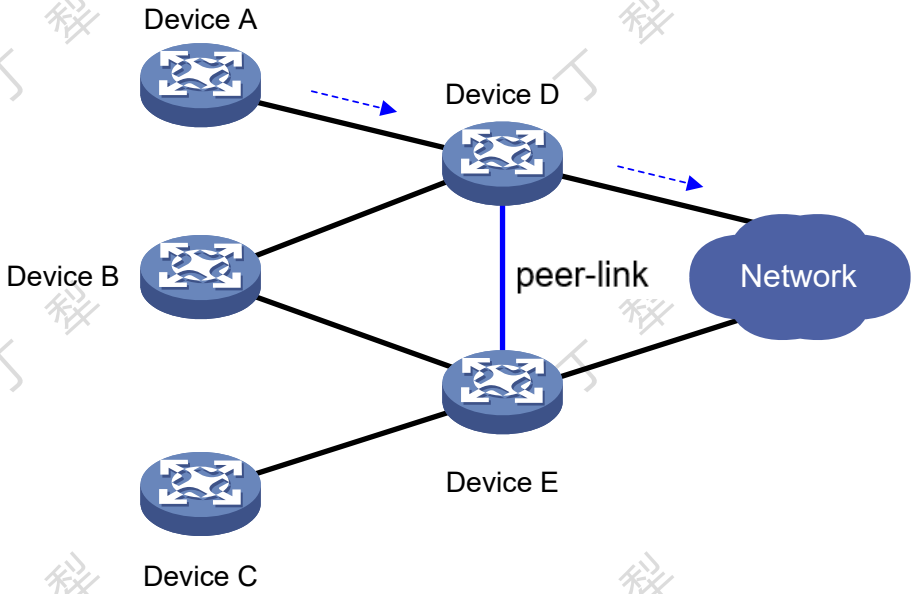
- 通过广播流程，学习MAC/ARP等表项；M-LAG设备间会实时平滑表项，最终M-LAG设备间转发表项将一致
- 根据peer-link接口和M-LAG隔离，广播到peer-link链路的流量不会广播到双活的M-LAG接口，可以去单挂口

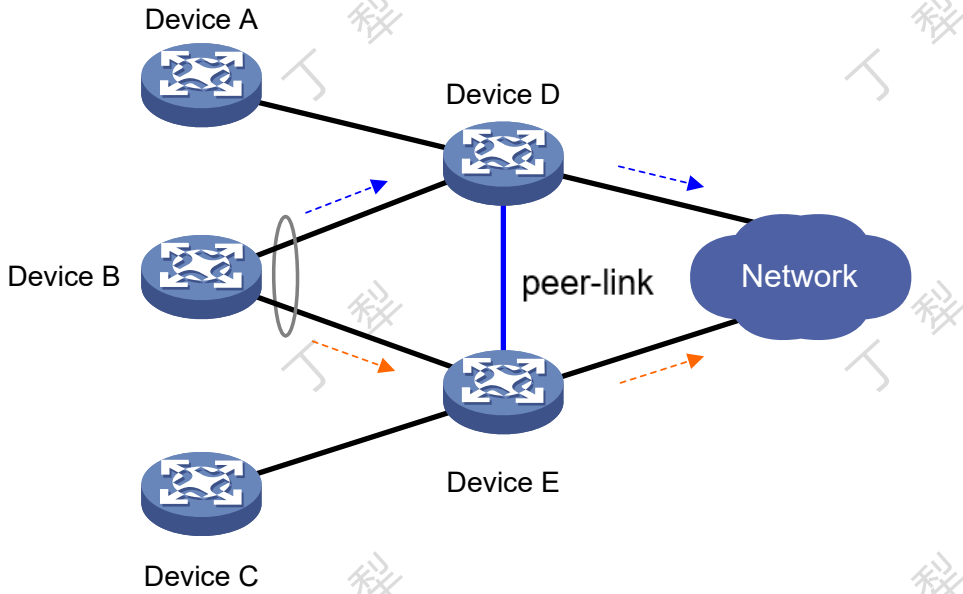




M-LAG的转发-用户侧单播流量

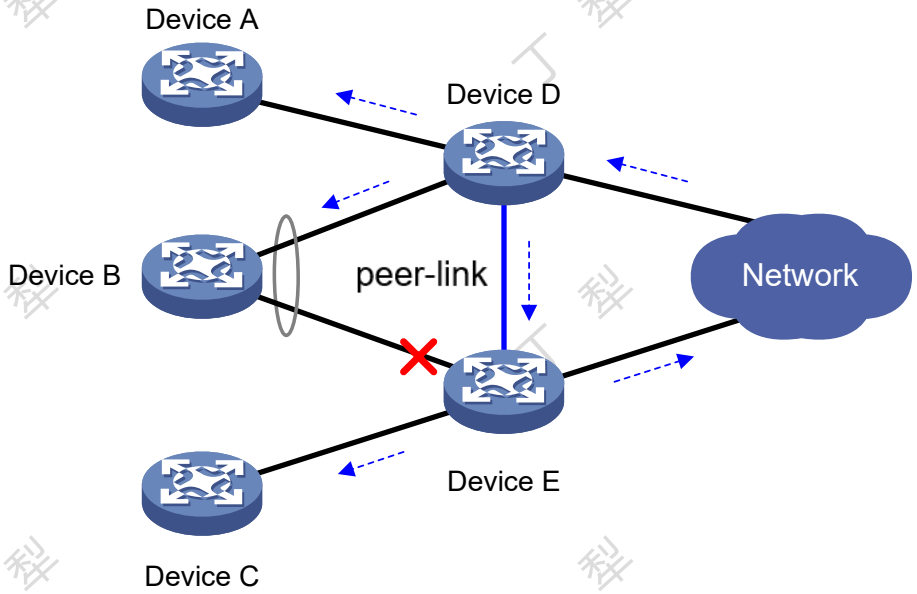
- M-LAG设备对已知单播做本地转发，正常情况下不会跨peer-link链路链路转发，两边会同步M-LAG侧表项
- M-LAG设备按照本地转发优先原则将其转发。本地转发优先是指当接收流量的M-LAG设备上存在对应的转发表项，则只在该设备上转发流量，不向peer-link链路链路上转发
- 为了保证L3流量的本地转发，VLAN虚接口MAC和VRRP实MAC都会同步给Peer M-LAG设备

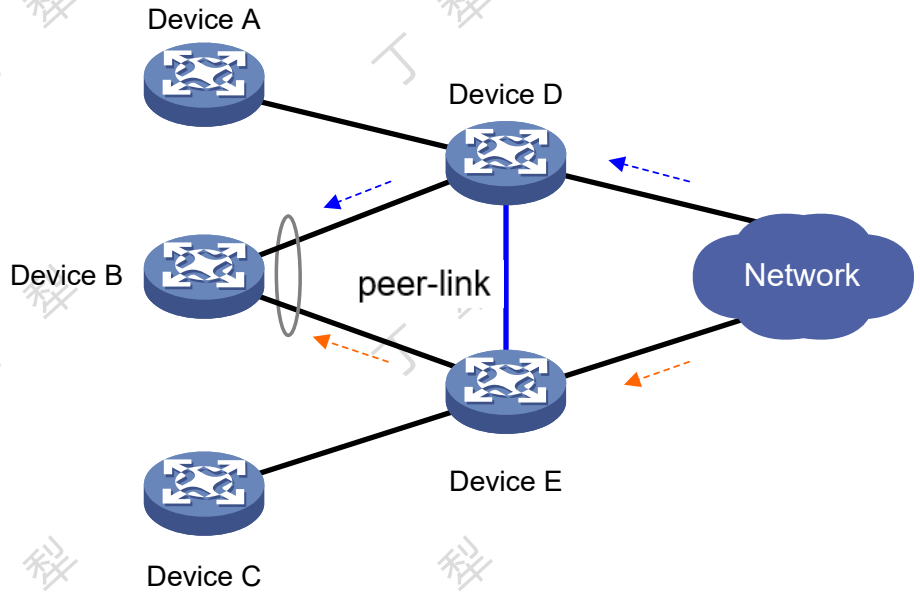




M-LAG的转发-网络侧流量

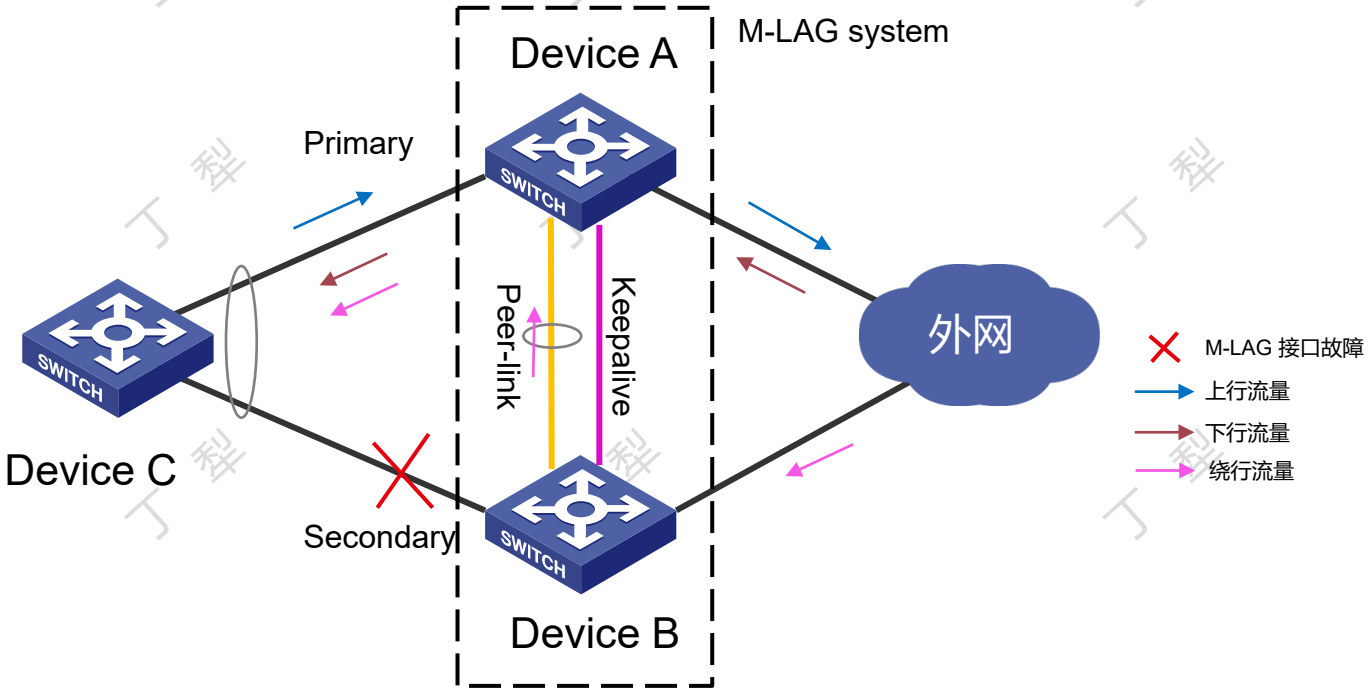
- Device D会将外网的广播流量发送到每一个用户侧端口，且由于peer-link接口与M-LAG组成员端口存在单向隔离机制，到达Device E的流量不会向Device B转发
- 对于外网侧发往M-LAG组成员端口的单播流量，流量会负载分担到Device D和Device E，由于D和E之间表项已经同步，都会发送至Device B





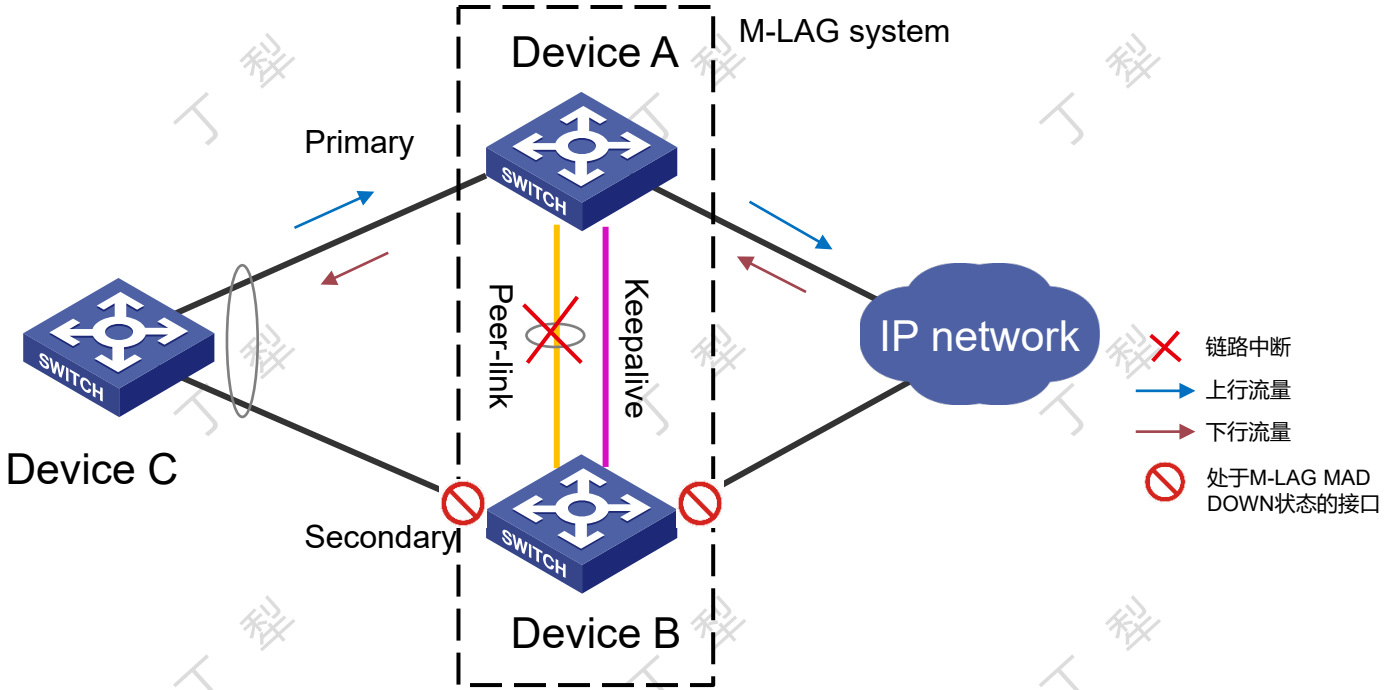
M-LAG的故障处理-M-LAG成员接口故障

- 某M-LAG接口故障，来自外网侧的流量会通过peer-link链路发送给另外一台设备，所有流量均由另外一台M-LAG设备转发，具体过程如下：
  - Device B的某M-LAG接口故障，外网侧不感知，流量依然会发送给所有M-LAG设备。
  - Device A的相同M-LAG接口正常，则Device B收到外网侧访问Device C的流量后，通过peer-link链路将流量交给Device A后转发给Device C。
  - 故障恢复后，Device B的该M-LAG接口up，流量正常转发。



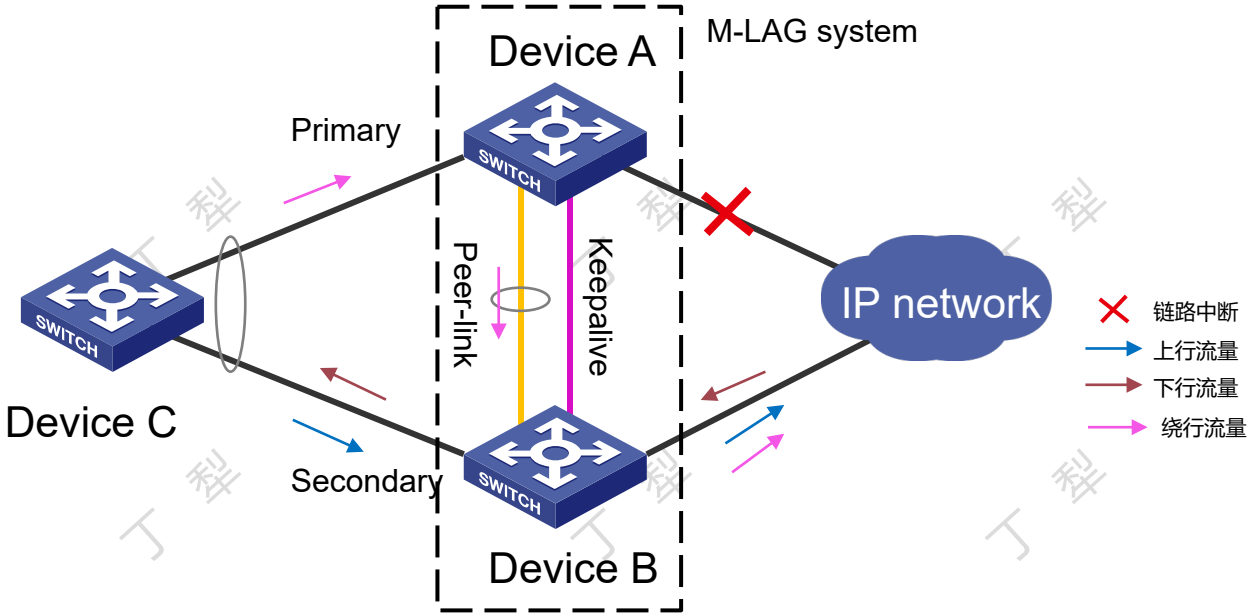
M-LAG的故障处理-peer-link链路故障

- **peer-link链路故障后，Keepalive保活的情况下，缺省自动触发M-LAG MAD机制**：将从设备上除M-LAG保留接口以外的接口置于M-LAG MAD DOWN状态，防止从设备继续转发流量；
- peer-link链路故障恢复，处于M-LAG MAD DOWN状态的接口经过延迟restore-delay恢复时间后（缺省30秒，从设备尽可能在延迟恢复时间内完成表项同步），自动恢复为up状态。



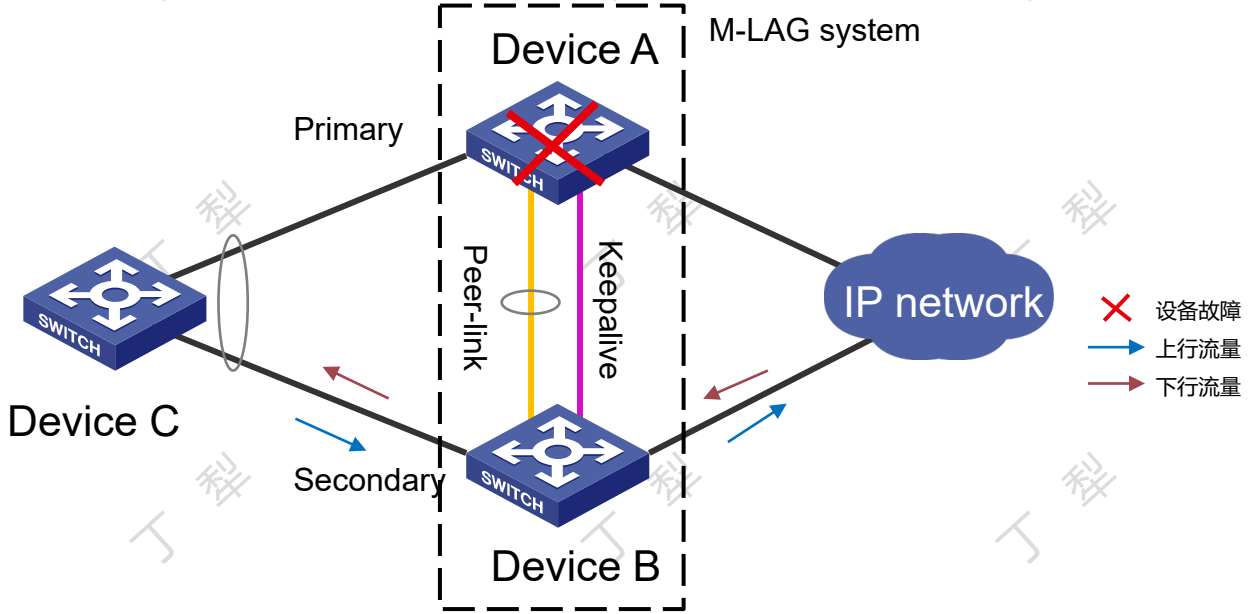
M-LAG的故障处理-上行故障

- 上行链路故障并不会影响M-LAG系统的转发。Device A上行链路虽然故障，但是外网侧的转发相关表项由Device B通过peer-link链路同步给Device A，Device A会将访问外网侧的流量发送给Device B进行转发。而外网侧发送给Device C的流量由于接口故障，自然不会发送给Device A处理。
- 上行链路故障时，如果通过Device A将访问外网侧的流量发送给Device B进行转发，会降低转发效率。此时用户可以配置Monitor Link功能，将M-LAG组成员端口和上行端口关联起来，一旦上行链路故障了，会联动M-LAG组成员端口状态，将其状态变为down，提高转发效率。



M-LAG的故障处理-设备故障处理

- Device A为主设备，Device B为备设备。当主设备故障后，主设备上的聚合链路状态变为down，不再转发流量。备设备将升级为主设备，该设备上的聚合链路状态依旧为up，流量转发状态不变，继续转发流量。主设备故障恢复后，M-LAG系统中由从状态升级为主状态的设备仍保持主状态，故障恢复后的设备成为M-LAG系统的备设备。
- 如果是备设备发生故障，M-LAG系统的主备状态不会发生变化，备设备上的聚合链路状态变为down。主设备上的聚合链路状态为up，流量转发状态不变，继续转发流量。



4

PART 04

M-LAG常见组网介绍

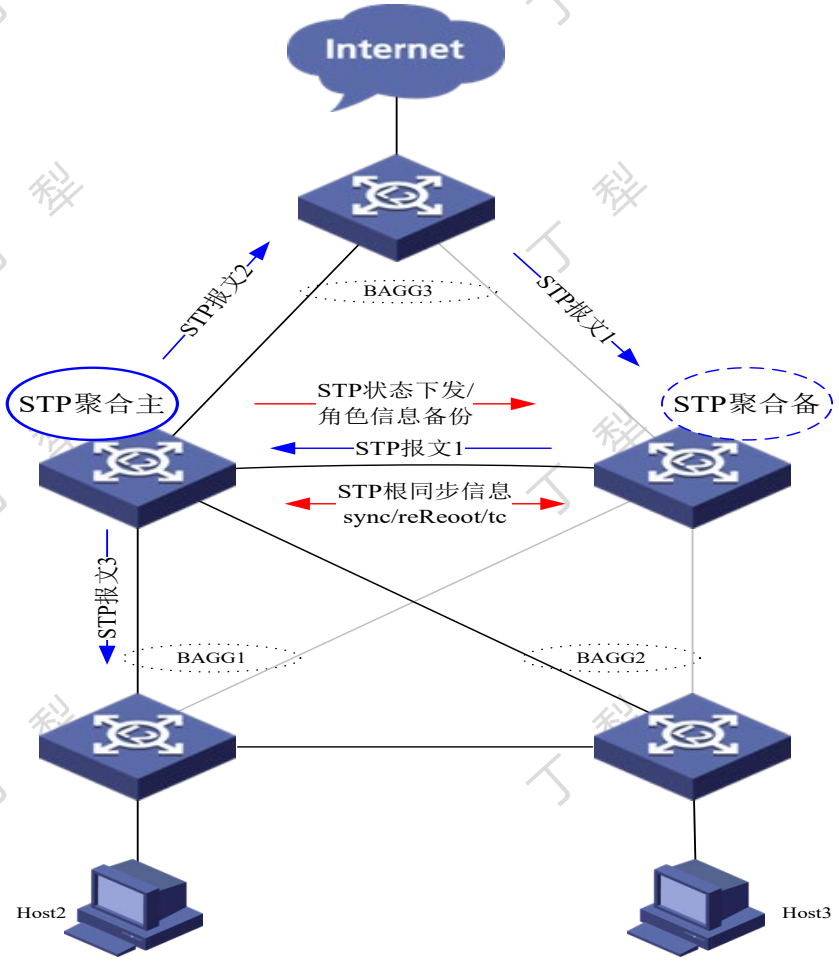
1

PART 4

M-LAG组网STP实现

M-LAG+STP设备处理机制

- 整个M-LAG系统是一个STP桥，STP桥ID使用系统MAC；
- STP对M-LAG接口集中式计算，单挂口独立计算；
- 当M-LAG设备非根设备时，整个M-LAG系统具有唯一 一个根端口；
- peer-link接口端口关闭拓扑协议；



M-LAG系统有唯一根端口

- 当M-LAG设备非根设备时，整个M-LAG系统具有唯一 一个根端口，根端口信息会在两个M-LAG设备之间同步；
- 另一台设备上查看根端口信息会显示在另一台M-LAG设备上；

[S6800-1]dis stp brief

MST ID	Port	Role	STP State	Protection
0	Bridge-Aggregation11 (M-LAG)	DESI	FORWARDING	NONE
0	Ten-GigabitEthernet1/0/1	ROOT	FORWARDING	NONE

[S6800-1]display stp root

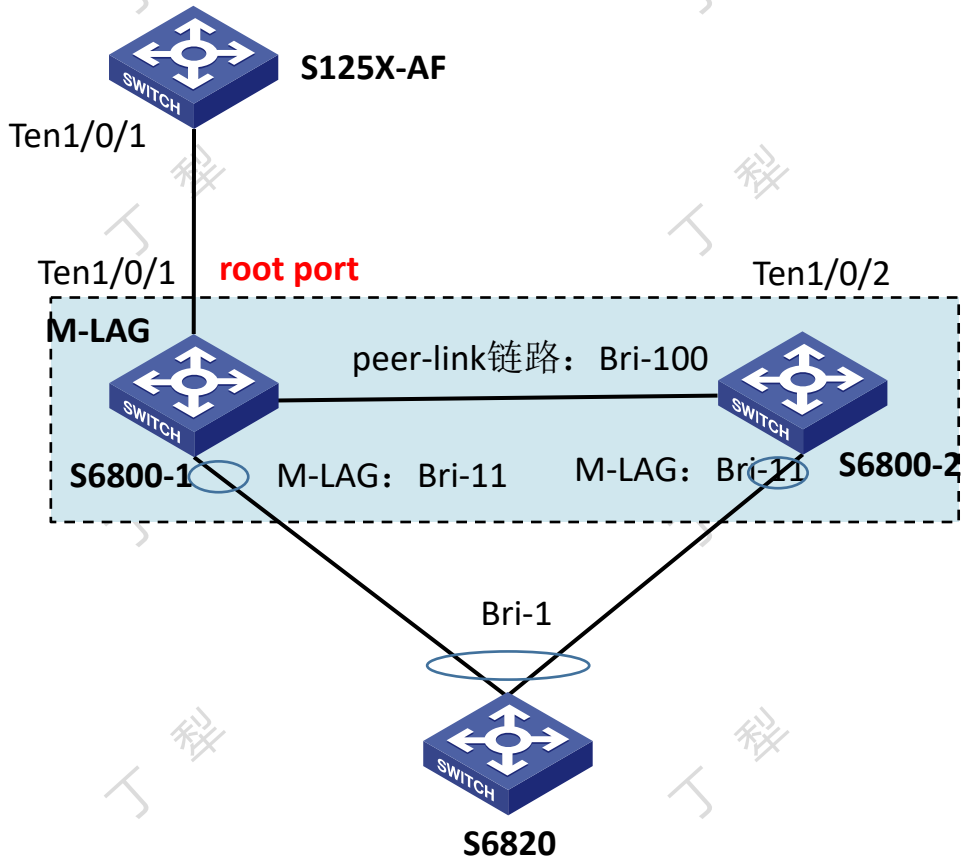
MST ID	Root Bridge ID	ExtPathCost	IntPathCost	Root Port
0	0.586a-b1f4-974f	2	0	XGE1/0/1

[S6800-2]display stp brief

MST ID	Port	Role	STP State	Protection
0	Bridge-Aggregation11 (M-LAG)	DESI	FORWARDING	NONE

[S6800-2]display stp root

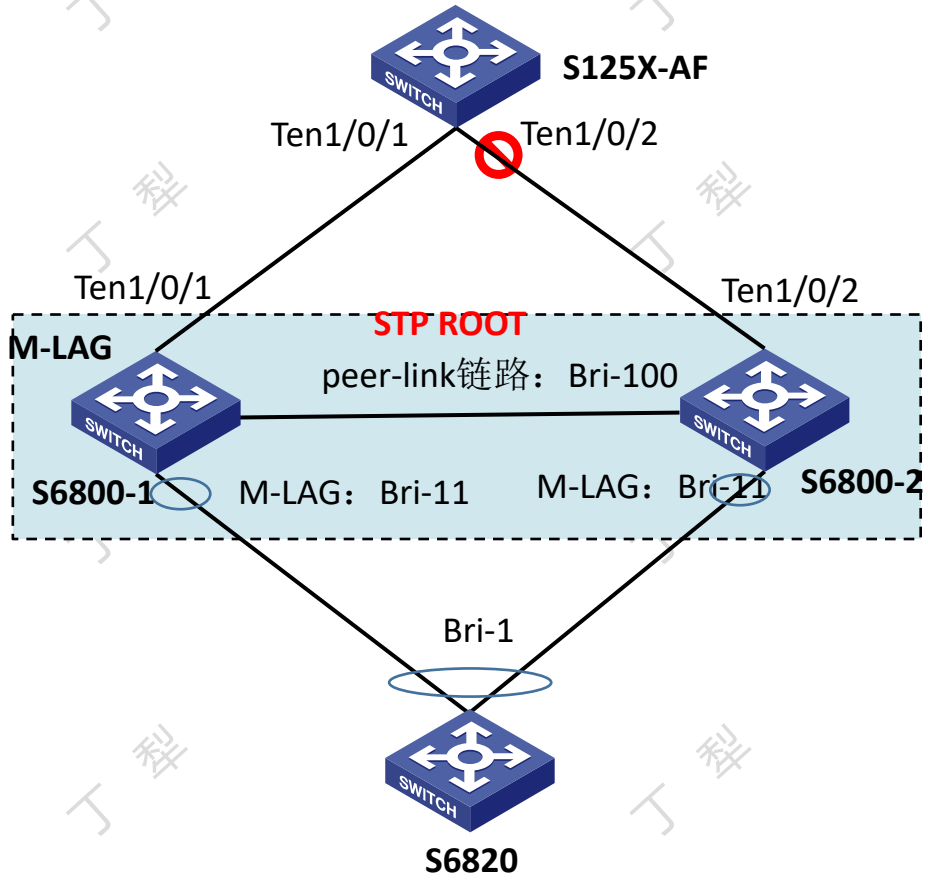
MST ID	Root Bridge ID	ExtPathCost	IntPathCost	Root Port
0	0.586a-b1f4-974f	2	0	On M-LAG peer device



M-LAG设备M-LAG接口与单挂口STP报文处理

- STP对M-LAG接口集中式计算，从M-LAG接口收到的BPDU报文会通过RLINK送到主设备计算，之后将状态同步到备设备；
- 对于单挂的接口，单挂接口的BPDU报文，各个设备单独计算，不会集中式计算；
- 注意：两台M-LAG设备的STP全局配置必须配置一致，否则可能会出现计算错误的情况

```
[S125XAF]dis stp brief
 MST ID   Port                          Role    STP State       Protection
 0        Ten-GigabitEthernet1/0/1      ROOT    FORWARDING      NONE
 0        Ten-GigabitEthernet1/0/2      ALTE    DISCARDING      NONE

 [S6800-1]dis stp brief
 MST ID   Port                          Role    STP State       Protection
 0        Bridge-Aggregation11          DESI    FORWARDING      NONE
 0        Ten-GigabitEthernet1/0/1      DESI    FORWARDING      NONE

[S6800-2]dis stp brief
 MST ID          Port                   Role  STP State    Protection
 0        Bridge-Aggregation11          DESI  FORWARDING   NONE
 0        Ten-GigabitEthernet1/0/2      DESI  FORWARDING   NONE
```



2

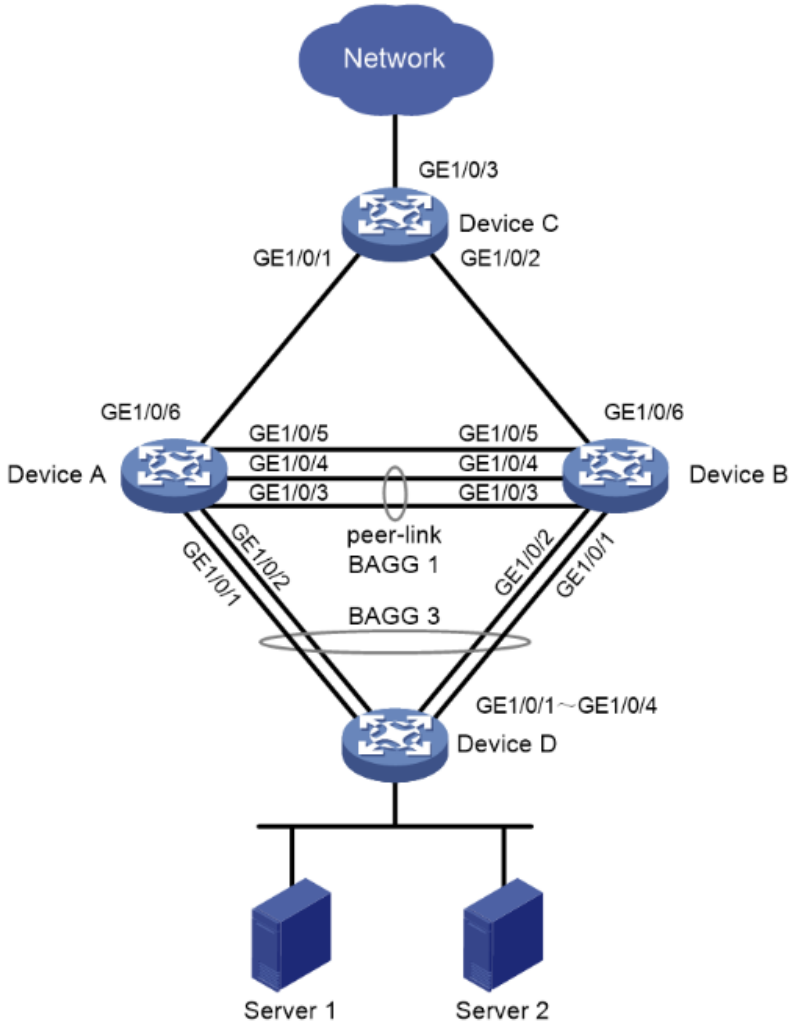
PART 4

M-LAG组网VLAN双活实现

M-LAG+动态OSPF路由实现

用户在服务器上通过三层路由方式接入到M-LAG时，需要满足以下要求：

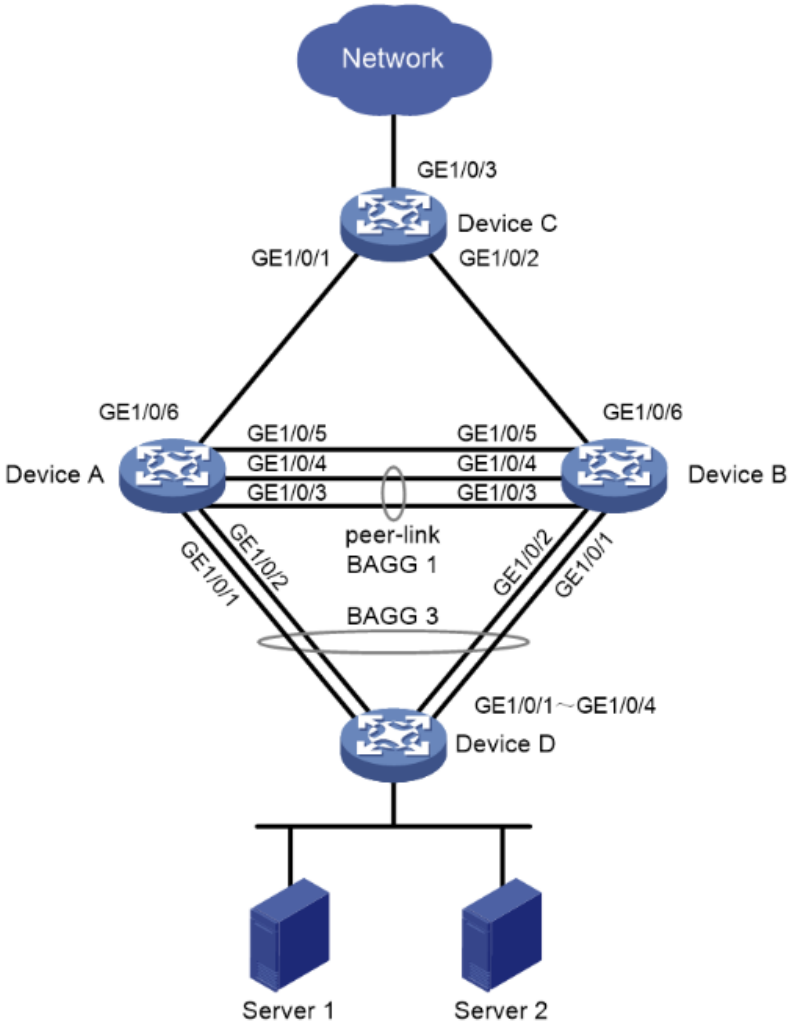
- 当一条接入链路发生故障时，流量可以快速切换到另一条链路，保证可靠性。
- 用户流量在两条接入链路上负载分担。



M-LAG+动态路由实现——M-LAG基础配置

DeviceA

```
#
 m-lag mad exclude interface
GigabitEthernet1/0/5
 m-lag auto-recovery reload-delay 244
 m-lag restore-delay 300
 m-lag system-mac 0002-0002-0002
 m-lag system-number 1
 m-lag system-priority 123
 m-lag mad exclude logical-interfaces
 m-lag keepalive ip destination 21.1.1.2
source 21.1.1.1
#
```



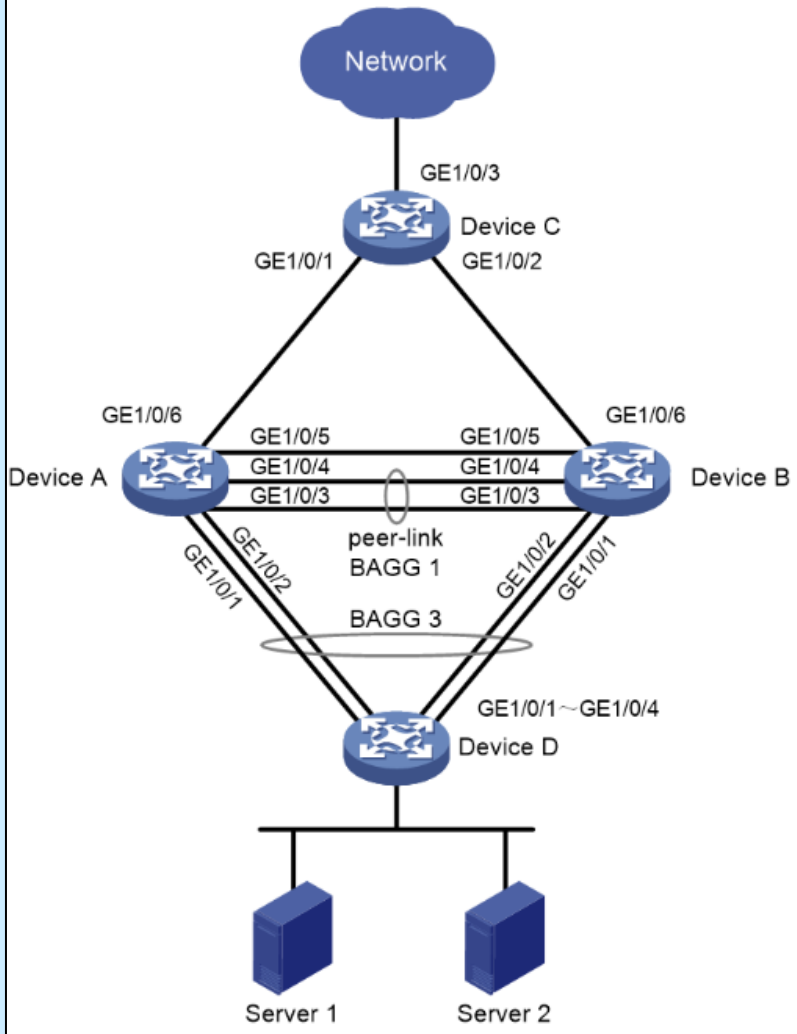
DeviceB

```
#
 m-lag mad exclude interface
GigabitEthernet1/0/5
 m-lag auto-recovery reload-delay 244
 m-lag restore-delay 300
 m-lag system-mac 0002-0002-0002
 m-lag system-number 2
 m-lag system-priority 123
 m-lag mad exclude logical-interfaces
 m-lag keepalive ip destination
21.1.1.1 source 21.1.1.2
#
```

M-LAG+动态路由实现——peer-link & keepalive配置

DeviceA

```
#
interface Bridge-Aggregation1
 port link-type trunk
 port trunk permit vlan all
 link-delay down 1
 link-delay up 1
 link-aggregation mode dynamic
 m-lag drcp period short
 port m-lag peer-link 1
#
interface GigabitEthernet1/0/5
 port link-mode route
 combo enable fiber
 ip address 21.1.1.1 255.255.255.0
#
```



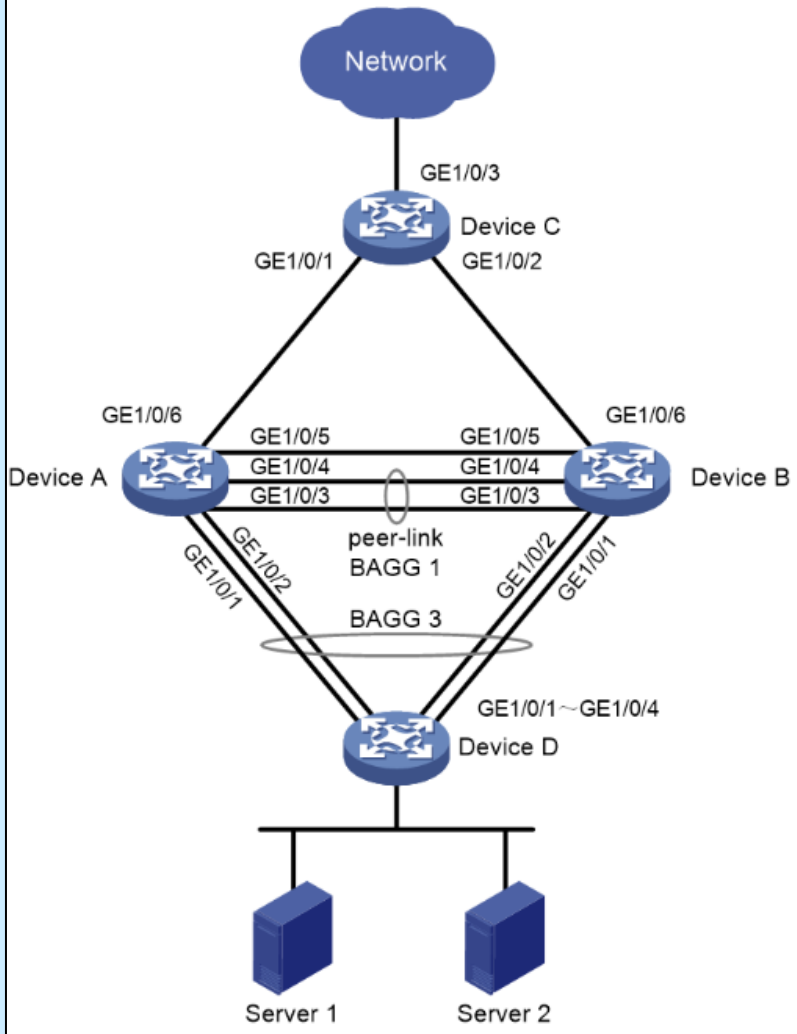
DeviceB

```
#
interface Bridge-Aggregation1
 port link-type trunk
 port trunk permit vlan all
 link-delay down 1
 link-delay up 1
 link-aggregation mode dynamic
 m-lag drcp period short
 port m-lag peer-link 1
#
interface GigabitEthernet1/0/5
 port link-mode route
 combo enable fiber
 ip address 21.1.1.2 255.255.255.0
#
```

M-LAG+动态路由实现——M-LAG接口配置

DeviceA

```
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 102
 link-aggregation mode dynamic
 m-lag drcp period short
 port m-lag group 1
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 102
 lacp period short
 port link-aggregation group 3
#
```



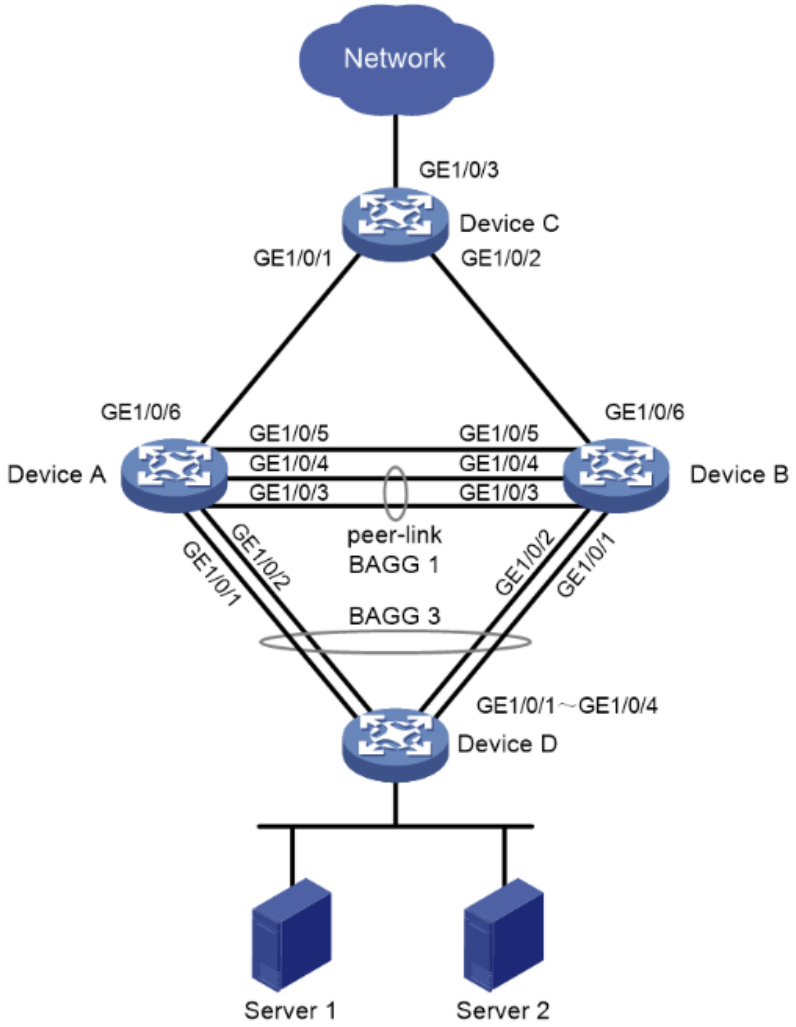
DeviceB

```
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 102
 link-aggregation mode dynamic
 m-lag drcp period short
 port m-lag group 1
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 102
 lacp period short
 port link-aggregation group 3
#
```

M-LAG+动态路由实现——OSPF配置

DeviceA

```
#
interface Vlan-interface102
 ip address 10.102.0.1 255.255.255.0
 ospf bfd enable
 ospf peer sub-address enable
10.102.0.3
 port m-lag virtual-ip 10.102.0.3
255.255.255.0 active
 mac-address 0003-0003-0003
#
interface GigabitEthernet1/0/6
 port link-mode route
 ip address 32.1.1.1 255.255.255.0
 ospf bfd enable
#
ospf 1 router-id 11.1.1.1
 area 0.0.0.0
  network 10.102.0.0 0.0.0.255
  network 32.1.1.1 0.0.0.0
#
```



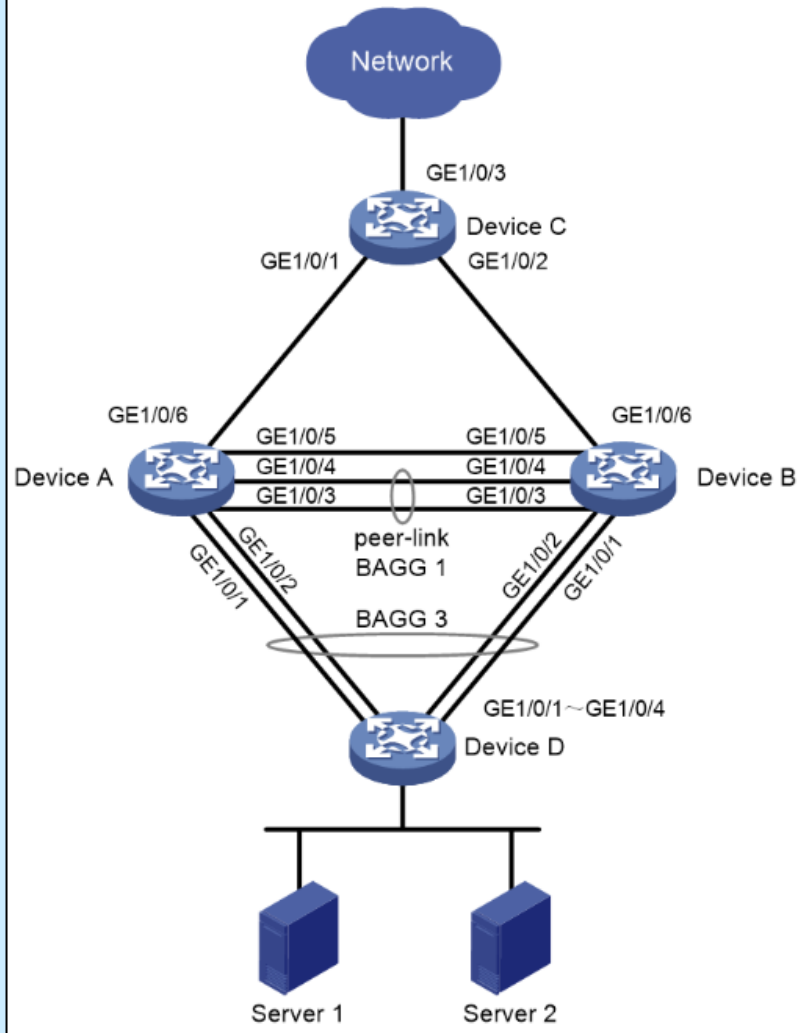
DeviceB

```
#
interface Vlan-interface102
 ip address 10.102.0.1 255.255.255.0
 ospf bfd enable
 ospf peer sub-address enable
10.102.0.4
 port m-lag virtual-ip 10.102.0.4
255.255.255.0 active
 mac-address 0003-0003-0003
#
interface GigabitEthernet1/0/6
 port link-mode route
 ip address 33.1.1.1 255.255.255.0
 ospf bfd enable
#
ospf 1 router-id 11.1.1.2
 area 0.0.0.0
  network 10.102.0.0 0.0.0.255
  network 33.1.1.1 0.0.0.0
#
```

M-LAG+动态路由实现——OSPF配置

DeviceC

```
#
ospf 1 router-id 11.1.1.3
 area 0.0.0.0
  network 13.13.13.13 0.0.0.0
  network 32.1.1.2 0.0.0.0
  network 33.1.1.2 0.0.0.0
#
interface LoopBack13
 ip address 13.13.13.13
255.255.255.255
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 32.1.1.2 255.255.255.0
 ospf bfd enable
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 33.1.1.2 255.255.255.0
 ospf bfd enable
#
```

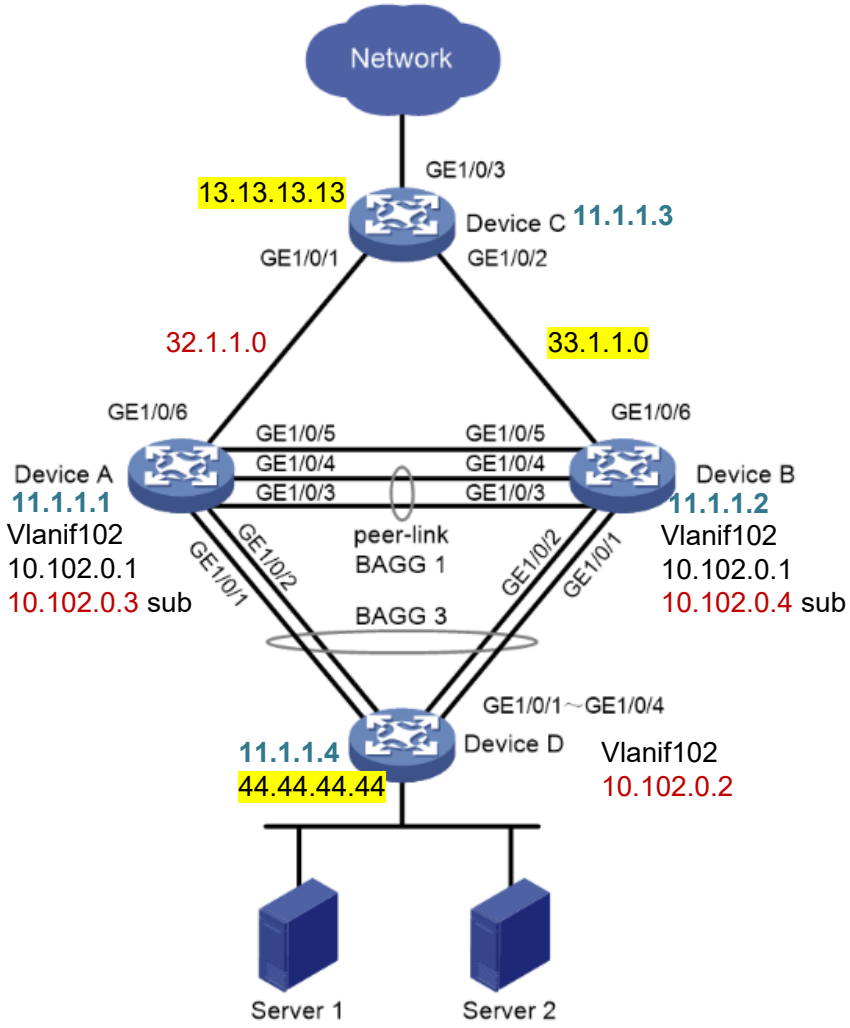


DeviceD

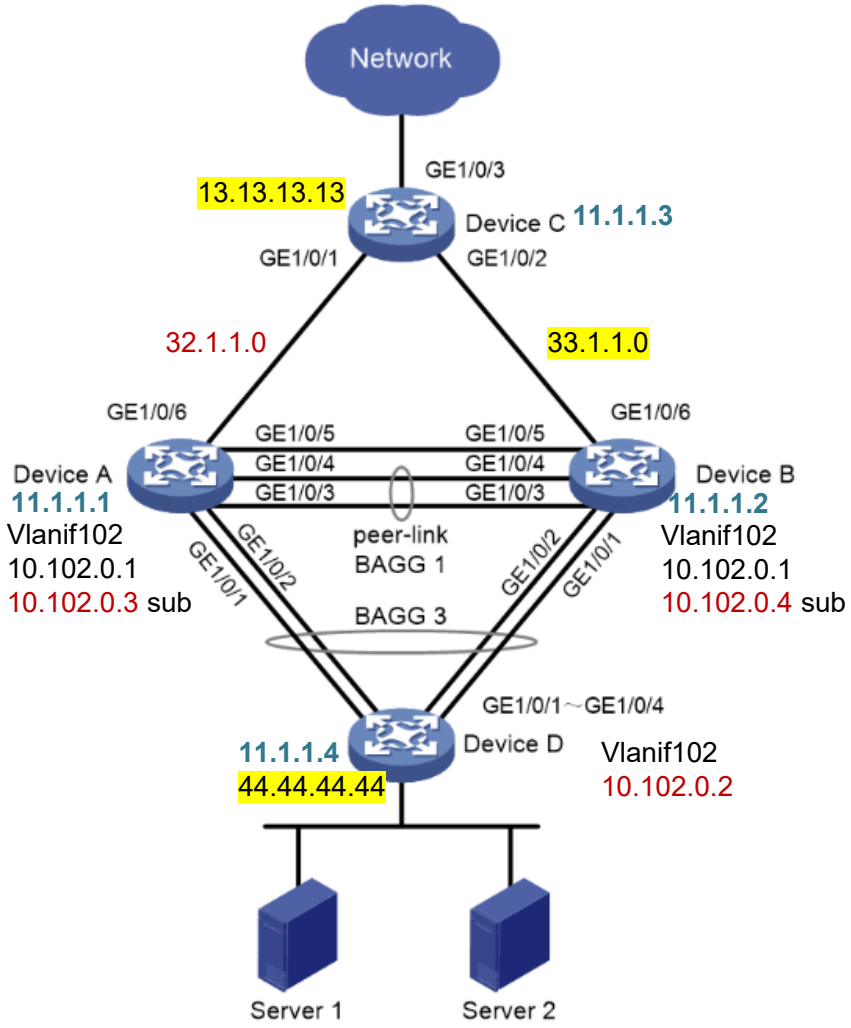
```
#
ospf 1 router-id 11.1.1.4
 area 0.0.0.0
  network 10.102.0.2 0.0.0.0
  network 44.44.44.44 0.0.0.0
#
interface LoopBack0
 ip address 44.44.44.44
255.255.255.255
#
interface Vlan-interface102
 ip address 10.102.0.2 255.255.255.0
 ospf bfd enable
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 102
 link-aggregation mode dynamic
#
```

M-LAG+动态路由实现——OSPF验证


```
<DeviceA>dis ospf peer

        OSPF Process 1 with Router ID 11.1.1.1
              Neighbor Brief Information

 Area: 0.0.0.0
 Router ID       Address         Pri Dead-Time  State           Interface
 11.1.1.3        32.1.1.2        1   35         Full/DR         GE1/0/6
 11.1.1.4        10.102.0.2      1   34         Full/DR         Vlan102
 11.1.1.2        10.102.0.4      1   38         Full/BDR        Vlan102

<DeviceA>dis ip routing-table
Destinations : 30       Routes : 31
Destination/Mask   Proto   Pre Cost       NextHop         Interface
……
13.13.13.13/32     O_INTRA 10  1          32.1.1.2        GE1/0/6
33.1.1.0/24        O_INTRA 10  2          10.102.0.4      Vlan102
                                          32.1.1.2        GE1/0/6
44.44.44.44/32     O_INTRA 10  1          10.102.0.2      Vlan102
```



M-LAG+动态路由实现——OSPF验证




```
<DeviceB>dis ospf peer

        OSPF Process 1 with Router ID 11.1.1.2
            Neighbor Brief Information

 Area: 0.0.0.0
 Router ID      Address        Pri Dead-Time  State           Interface
 11.1.1.3       33.1.1.2       1   31         Full/DR         GE1/0/6
 11.1.1.4       10.102.0.2     1   39         Full/DR         Vlan102
 11.1.1.1       10.102.0.3     1   35         Full/DROther    Vlan102

<DeviceB>dis ip routing-table
Destinations : 30       Routes : 31
Destination/Mask   Proto   Pre Cost      NextHop         Interface
......
13.13.13.13/32     O_INTRA 10  1         33.1.1.2        GE1/0/6
32.1.1.0/24        O_INTRA 10  2         10.102.0.3      Vlan102
                                         33.1.1.2        GE1/0/6
44.44.44.44/32     O_INTRA 10  1         10.102.0.2      Vlan102
```

M-LAG+动态路由实现——OSPF验证

<**DeviceC**>dis ospf peer

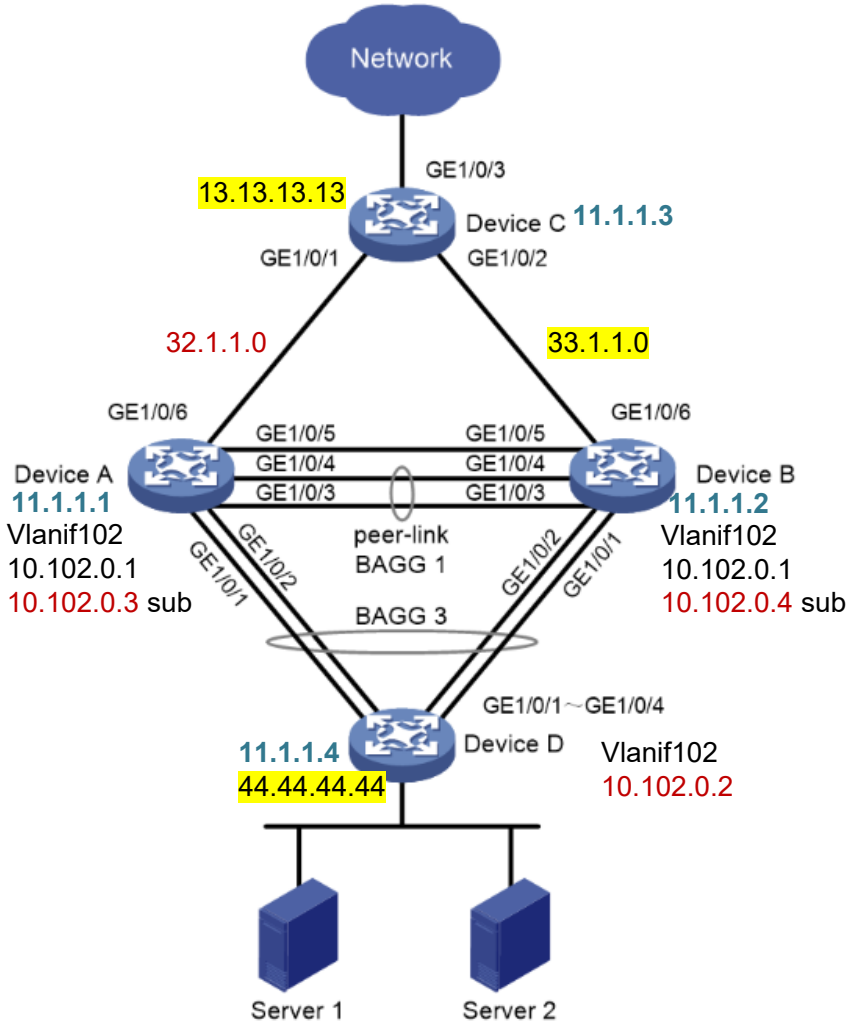
OSPF Process 1 with Router ID 11.1.1.3
Neighbor Brief Information

Area: 0.0.0.0

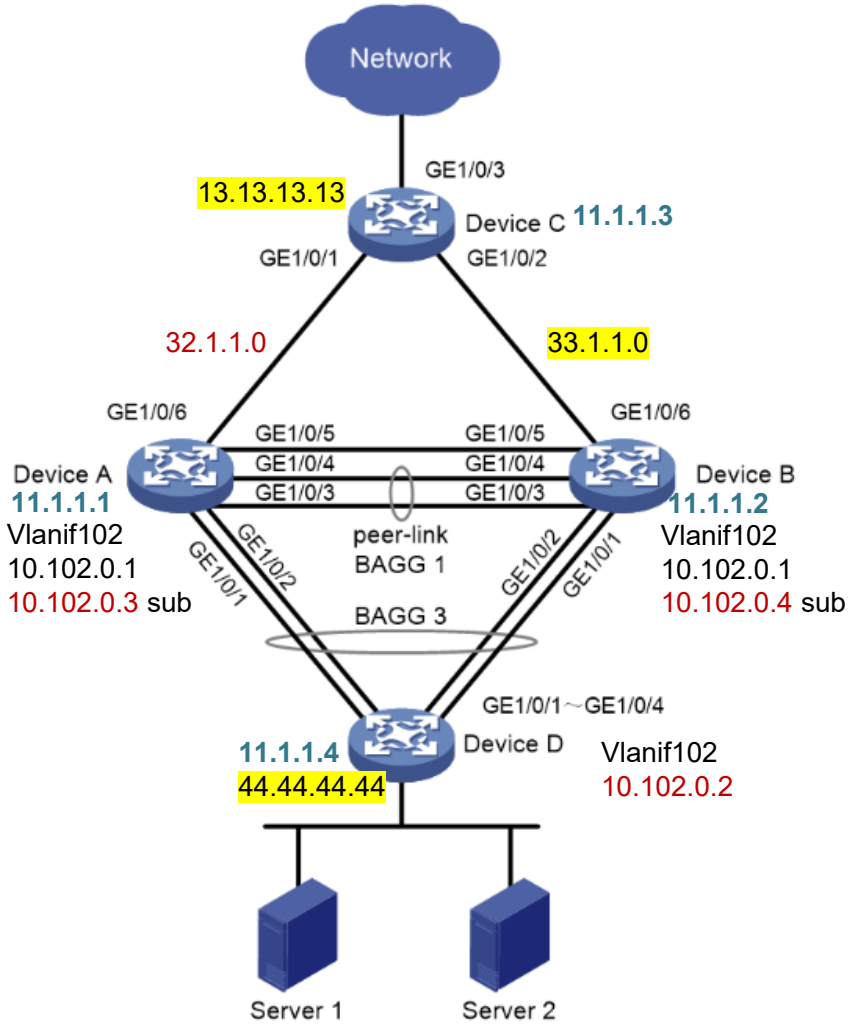
Router ID	Address	Pri	Dead-Time	State	Interface
11.1.1.1	32.1.1.1	1	39	Full/BDR	GE1/0/1
11.1.1.2	33.1.1.1	1	36	Full/BDR	GE1/0/2

<**DeviceC**>dis ip routing-table
Destinations : 21 Routes : 24

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
……					
10.102.0.0/24	O_INTRA	10	2	32.1.1.1	GE1/0/1
				33.1.1.1	GE1/0/2
44.44.44.44/32	O_INTRA	10	2	32.1.1.1	GE1/0/1
				33.1.1.1	GE1/0/2



M-LAG+动态路由实现——OSPF验证




```
<DeviceD>dis ospf peer

        OSPF Process 1 with Router ID 11.1.1.4
            Neighbor Brief Information

 Area: 0.0.0.0
 Router ID     Address         Pri Dead-Time  State           Interface
 11.1.1.1      10.102.0.3      1   38         Full/DROther    Vlan102
 11.1.1.2      10.102.0.4      1   35         Full/BDR        Vlan102

<DeviceD>dis ip routing-table
Destinations : 21       Routes : 23
Destination/Mask   Proto   Pre Cost      NextHop         Interface
……
13.13.13.13/32     O_INTRA 10  2         10.102.0.3      Vlan102
                                         10.102.0.4      Vlan102
32.1.1.0/24        O_INTRA 10  2         10.102.0.3      Vlan102
33.1.1.0/24        O_INTRA 10  2         10.102.0.4      Vlan102
```

附件

注意：使用HCL模拟器验证M-LAG配置时，建议相关M-LAG设备分配的内存大于等于2048M（缺省为512M）

m-lag+vlan双活(ospf+bgp)

THANKS

www.h3c.com