

文/叶玉其

## 1 IRF概述

IRF（Intelligent Resilient Framework，智能弹性架构）是H3C自主研发的软件虚拟化技术。它的核心思想是将多台设备连接在一起，进行必要的配置后，虚拟化成一台设备。使用这种虚拟化技术可以集合多台设备的硬件资源和软件处理能力，实现多台设备的协同工作、统一管理和不间断维护。

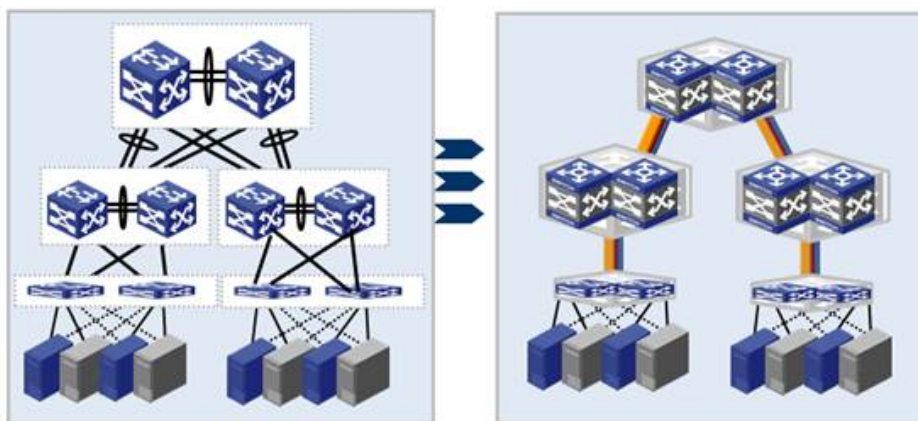


图1 IRF组网应用示意图

## 2 IRF的优点

IRF主要具有以下优点：

- 简化管理。IRF形成之后，用户通过任意成员设备的任意端口都可以登录IRF系统，对IRF内所有成员设备进行统一管理。
- 1:N备份。IRF由多台成员设备组成，其中，Master设备负责IRF的运行、管理和维护，Slave设备在作为备份的同时也可以处理业务。一旦Master设备故障，系统会迅速自动选举新的Master，以保证业务不中断，从而实现了设备的1:N备份。
- 强大的网络扩展能力。通过增加成员设备，可以轻松自如的扩展IRF的端口数、带宽。因为各成员设备都有CPU，能够独立处理协议报文、进行报文转发，所以IRF还能够轻松自如的扩展处理能力。

联系我们

### 3 多IRF冲突检测（MAD功能）

#### 3.1 机制介绍

IRF链路故障会导致一个IRF变成多个新的IRF。这些IRF拥有相同的IP地址等三层配置，会引起地址冲突，导致故障在网络中扩大。为了提高系统的可用性，当IRF分裂时我们就需要一种机制，能够检测出网络中同时存在多个IRF，并进行相应的处理尽量降低IRF分裂对业务的影响。MAD（Multi-Active Detection，多Active检测）就是这样一种检测和处理机制。它主要提供以下功能：

##### (1)分裂检测

通过ARP（Address Resolution Protocol）、ND（Neighbor Discovery Protocol）、LACP（Link Aggregation Control Protocol，链路聚合控制协议）或者BFD（Bidirectional Forwarding Detection，双向转发检测）来检测网络中是否存在多个IRF。

##### (2)冲突处理

IRF分裂后，通过分裂检测机制IRF会检测到网络中存在其它处于Active状态（表示IRF处于正常工作状态）的IRF。

- 对于BFD MAD/ ARP MAD/ND MAD检测，冲突处理会直接让Master成员编号小的IRF处于Active状态，继续正常工作；其它IRF迁移到Recovery状态。
- 对于LACP MAD检测，冲突处理会先比较两个IRF中成员设备的数量，数量多的IRF处于Active状态，继续工作；数量少的迁移到Recovery状态；如果成员数量相等，则Master成员编号小的IRF处于Active状态，继续正常工作；其它IRF迁移到Recovery状态。

IRF迁移到Recovery状态后会关闭该IRF中所有成员设备上除保留端口以外的其它所有物理端口（通常为业务接口），以保证该IRF不能再转发业务报文。缺省情况下，只有IRF链路物理端口是保留端口，用户也可以通过mad exclude interface命令行将其它端口设置为保留端口。

##### (3)MAD故障恢复

IRF链路故障导致IRF分裂，从而引起多Active冲突。因此修复故障的IRF链路，让冲突的IRF重新合并为一个IRF，就能恢复MAD故障。如果在MAD故障恢复前，处于Active状态的IRF出现其他故障，则可以通过命令行先启用Recovery

联系我们



状态的IRF，让它接替原IRF工作，以便保证业务尽量少受影响，再恢复MAD故障。

## 3.2 原理介绍

IRF支持的MAD检测方式有：LACP MAD检测、BFD MAD检测、ARP MAD检测和ND MAD检测。下面针对这四种MAD检测方式进行逐一阐述：

### 3.2.1 ARP MAD检测

#### (一) ARP MAD检测原理

ARP MAD检测是通过扩展ARP协议报文内容实现的，即将ARP协议报文中“Target MAC Address”字段扩展为IRF的DomainID（域编号）、对端ActiveID（即对端Master的成员编号）、自身ActiveID（自身Master的成员编号）和PktType（检测包类型）（如图2所示）。其中检测包类型（PktType）包括如下三种类型：

- a) 0x00，Hello报文，载荷为自身ActiveID
- b) 0x01，Alive检测报文，载荷为自身ActiveID和对端ActiveID
- c) 0x02，Alive确认报文，载荷为自身ActiveID和对端ActiveID

联系我们

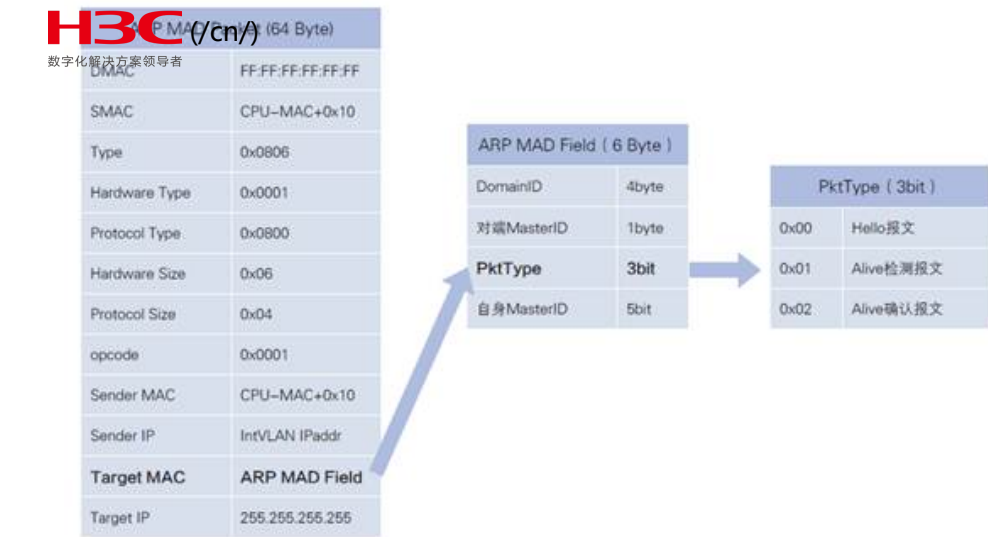


图2 ARP MAD PACKET

使能ARP MAD检测后，成员设备可以通过ARP协议报文和其它成员设备交互DomainID和ActiveID信息。

- 当成员设备收到ARP协议报文后，先比较DomainID。如果DomainID相同，再比较ActiveID；如果DomainID不同，则认为报文来自不同IRF，不再进行MAD处理。
- 如果ActiveID相同，则表示IRF正常运行，没有发生多Active冲突；如果ActiveID值不同，则表示IRF分裂，进行ARP协议报文快速交互确认，检测到多Active冲突。
- 检测到多Active冲突后，会直接让Master成员编号小的IRF处于Active状态，继续正常工作；其它IRF上报MAD冲突事件给IRF模块，IRF模块将该IRF迁移到MAD Recovery状态。

其中IRF分裂后，进行ARP协议报文快速交互确认状态机和流程如下（如图3所示）：

- a) 初始情况下，IRF使能ARP MAD功能后，各成员设备ARP MAD状态机迁移到ARP\_MAD\_STS\_WAIT\_Hello，IRF内所有设备从ARP MAD VLAN对应的接口发送ARP MAD Hello报文（由于ARP MAD需要使能STP功能，相应端口会被STP Blocking，各成员端口发送的ARP MAD Hello报文无法相互接收到）。
- b) 当IRF分裂后，各IRF都重新选择了Master，使用新Master的MAC地址封装STP报文，各IRF之间专门用于ARP MAD检测的链路STP状态迁移到Forwarding状态，ARP MAD模块可以

联系我们



数字化解决方案领导者

收到对端IRF发送的MAD Hello报文，ARP MAD状态机迁移到ARP\_MAD\_STS\_WAIT\_ALIVEACK，并发送ARP MAD Alive检测报文。

- c) 当收到ARP MAD Alive检测报文时，如果当前状态为ARP\_MAD\_STS\_WAIT\_Hello，则迁移到ARP\_MAD\_STS\_WAIT\_ALIVEACK，并发送ARP MAD Alive检测报文；否则发送ARP MAD Alive确认报文。
- d) 如果收到ARP MAD Alive确认报文，且当前为ARP\_MAD\_STS\_WAIT\_ALIVEACK状态，则开始ARP MAD竞选。

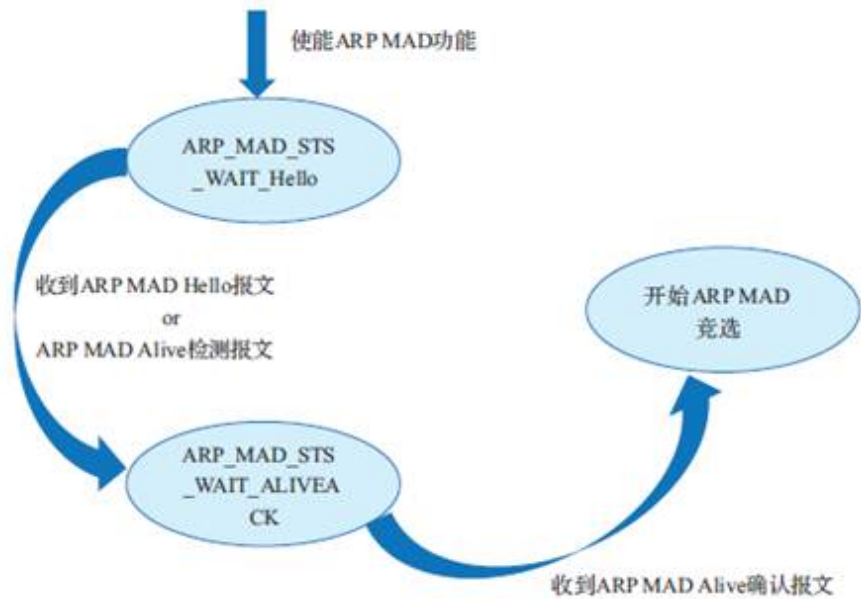


图3 ARP MAD检测状态机

(二) ARP MAD检测组网要求

ARP MAD检测方式可以使用二层交换机设备作为中间设备来进行连接（如图4所示），也可以不使用中间设备（如图5所示）。当ARP MAD检测组网使用中间设备进行连接时，可使用普通的数据链路作为ARP MAD检测链路；当不使用中间设备时，需要在所有的成员设备之间建立两两互联的ARP MAD检测链路。不管是否使用中间设备来连接，都需要在IRF设备和中间设备上配置生成树功能，以防止形成环路。同时还需要在使能ARP MAD检测的IRF设备上配置IRF桥MAC不保留[1]，这样才能在IRF分裂后，触发STP状态快速切换，ARP MAD检测快速生效。

联系我们

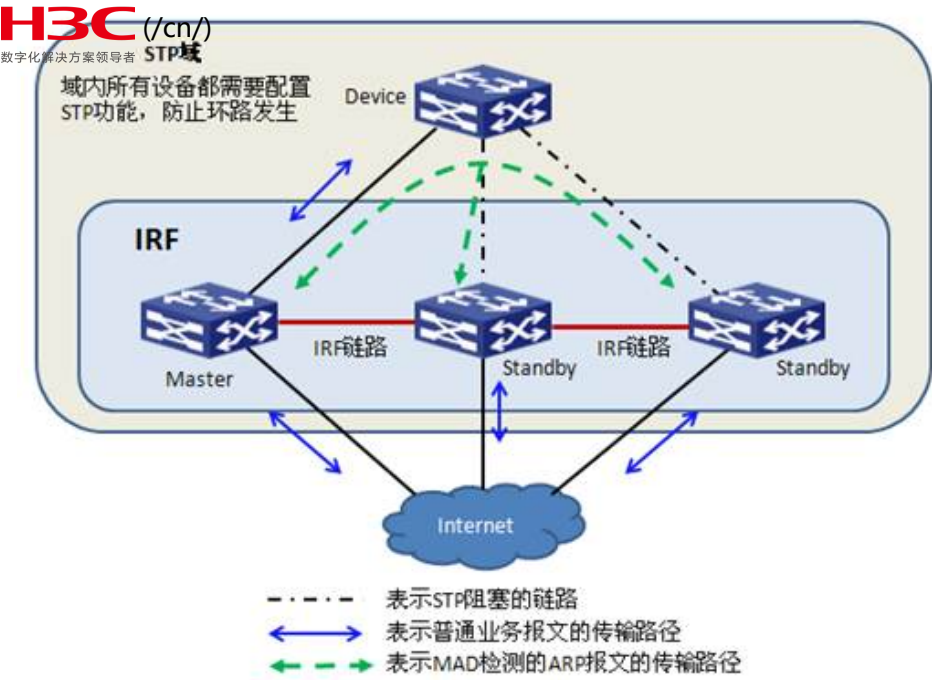


图4 ARP MAD检测组网示意图一

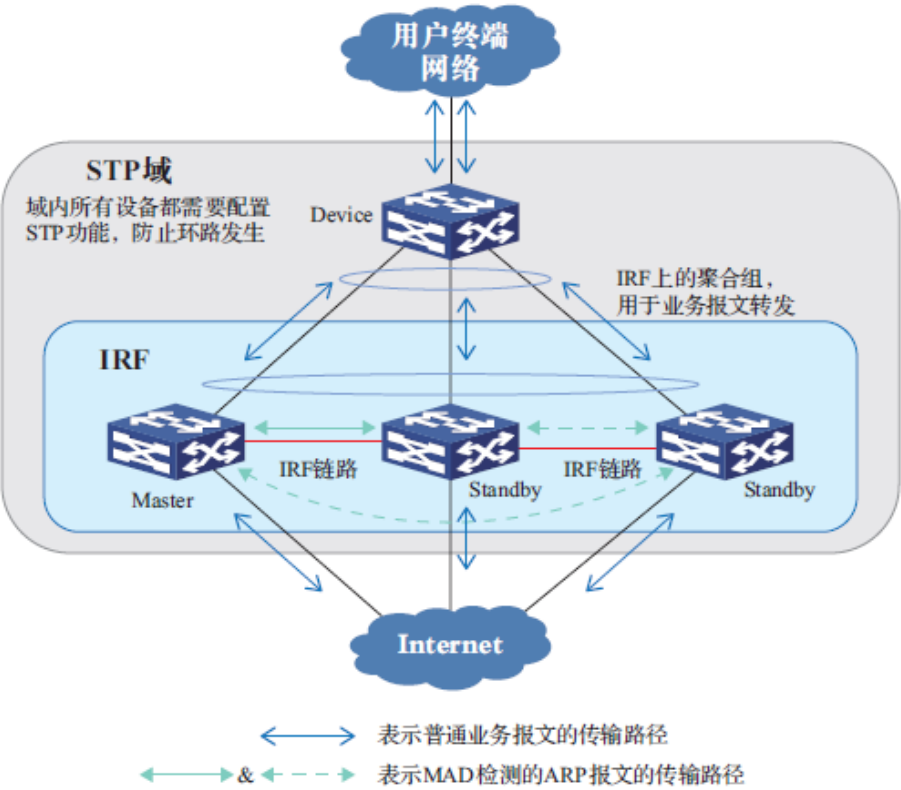


图5 ARP MAD检测组网示意图二

联系我们

### 3.2.2 ND MAD检测

ND MAD检测是通过扩展ND（Neighbor Discovery Protocol）协议报文内容实现的，即使用ND的NS（Neighbor Solicitation）协议报文携带扩展TLV选项数据来交互IRF的DomainID和ActiveID。ND MAD检测原理及组网要求皆和ARP MAD检测类似，请参考ARP MAD检测介绍。

### 3.2.3 BFD MAD检测

#### （一）BFD MAD检测原理

BFD MAD检测是通过BFD协议来实现的。要使BFD MAD检测功能正常运行，除在三层接口下使能BFD MAD检测功能外，还需要在该接口上配置MAD IP地址。MAD IP地址与普通IP地址不同的地方在于：MAD IP地址与成员设备是绑定的，IRF中的每个成员设备上都需要配置，且所有成员设备的MAD IP必须属于同一网段。

- 当IRF正常运行时，只有Master上配置的MAD IP地址生效，Slave设备上配置的MAD IP地址不生效，BFD会话处于down状态；（使用display bfd session命令查看BFD会话的状态。如果Session State显示为Up，则表示激活状态；如果显示为Down，则表示处于down状态）
- 当IRF分裂形成多个IRF系统时，不同IRF中Master上配置的MAD IP地址均会生效，BFD会话被激活，此时会检测到多Active冲突。
- 检测到多Active冲突后，会直接让Master成员编号小的IRF处于Active状态，继续正常工作；其它IRF上报MAD冲突事件给IRF模块，IRF模块将该IRF迁移到MAD Recovery状态。

联系我们



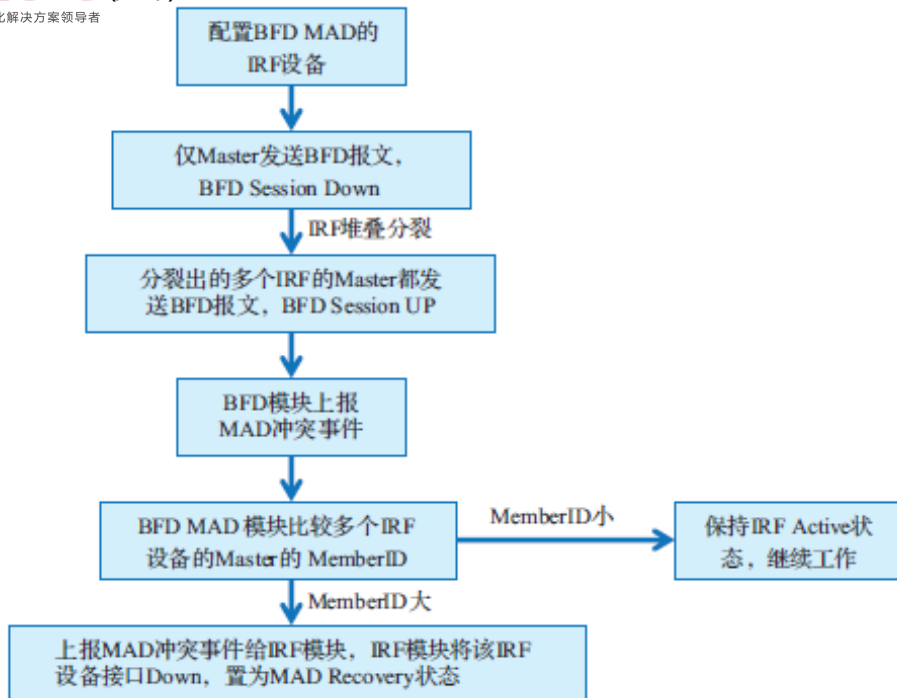
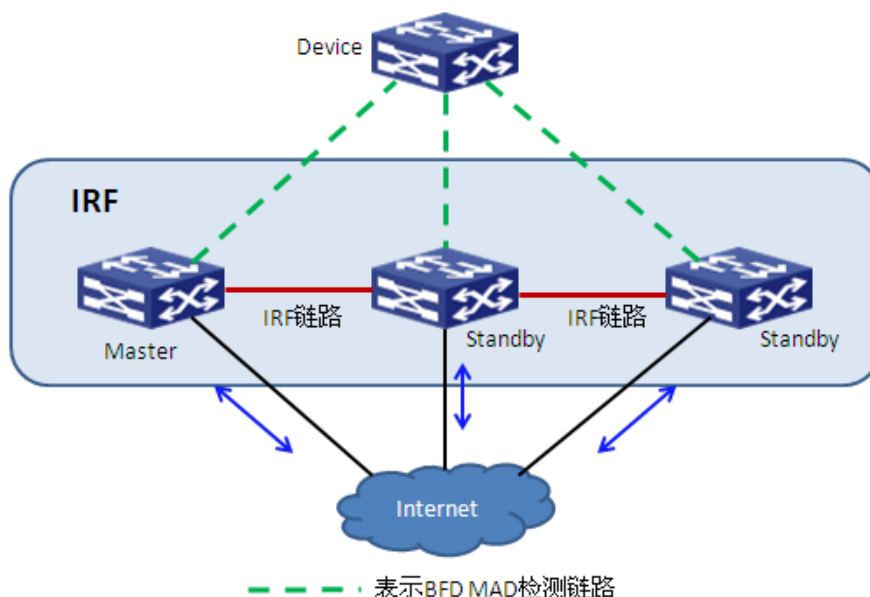


图6 BFD MAD交互流程

## (二) BFD MAD检测组网要求

BFD MAD检测方式需要使用中间设备（如图7所示），每个成员设备都需要连接到中间设备，这些BFD链路专用于MAD检测。这些链路连接的接口必须属于同一VLAN，在该VLAN接口视图下给不同成员设备配置同一网段下的不同IP地址。

在用于BFD MAD检测的接口下必须使用mad ip address命令配置MAD IP地址，而不要配置其它IP地址（包括使用ip address命令配置的普通IP地址、VRRP虚拟IP地址等），以免影响MAD检测功能。



联系我们



### 3.2.4 LACP MAD检测

#### (一) LACP MAD检测原理

LACP MAD检测是通过扩展LACP协议报文内容实现的，即在LACP协议报文的扩展字段内定义一个新的TLV（Type/Length/Value，类型/长度/值）数据域——用于交互IRF的DomainID（域编号）、ActiveMemNum（当前IRF的成员数目）和ActiveID（等于Master的成员编号）。使能LACP MAD检测后，成员设备通过LACP协议报文和其它成员设备交互DomainID、ActiveMemNum和ActiveID信息。

使能LACP MAD检测后，聚合成员端口周期性地发送带有扩展TLV字段的LACP报文（缺省发送周期为30s，该LACP报文和动态LACP报文分开发送），对端设备（中间设备）收到带扩展TLV的LACP报文后，会从聚合组内除接收端口外的所有其他成员端口各转发一份。成员设备通过LACP协议报文和其它成员设备交互DomainID、ActiveMemNum和ActiveID信息。

- 当成员设备收到LACP MAD报文后，先比较DomainID。如果DomainID相同，再比较ActiveID；如果DomainID不同，则认为报文来自不同IRF，不再进行MAD处理，作为中间设备，仍然需要从聚合组内除接收端口外的所有其他成员端口各转发一份。
- 如果ActiveID相同，则表示IRF正常运行，没有发生多Active冲突；如果ActiveID值不同，快速进行LACP报文交互、确认冲突，确认后表示IRF分裂，检测到多Active冲突。
- 如果ActiveMemNum不同，ActiveMemNum大的为优，处于IRF Active状态继续工作，ActiveMemNum小的迁移到Recovery状态（即禁用状态）。
- 如果ActiveMemNum相同，继续比较ActiveID，ActiveID小的为优，处于IRF Active状态继续工作，ActiveID大的上报MAD冲突事件给IRF模块，IRF模块将该IRF迁移到MAD Recovery状态（即禁用状态）。

联系我们

详细流程见下图8，其中当IRF分裂时，快速LACP MAD报文交换、确认冲突过程和ARP MAD类似。



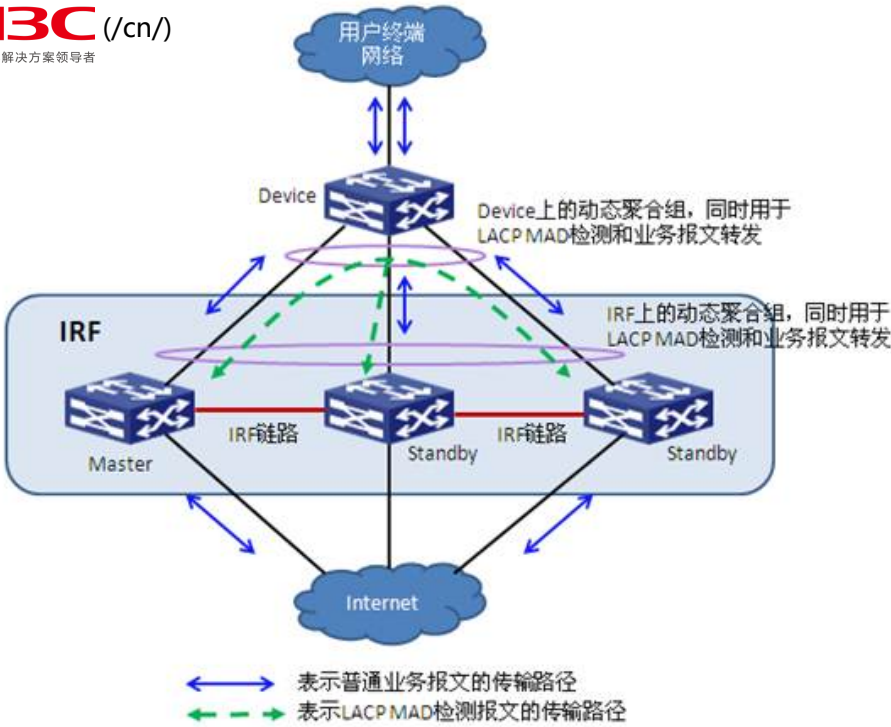


图9 LACP MAD检测组网示意图

4 总结

IRF支持的MAD检测方式有：LACP MAD检测、BFD MAD检测、ARP MAD检测和ND MAD检测。四种MAD检测机制各有特点，用户可以根据现有组网情况进行选择。由于LACP MAD和BFD MAD、ARP MAD、ND MAD冲突处理的原则不同，请不要同时配置。BFD MAD、ARP MAD、ND MAD这三种方式独立工作，彼此之间互不干扰，可以同时配置。

MAD检测方式	优势	限制
LACP MAD	检测速度快，利用现有聚合组网即可实现，无需占用额外接口，利用聚合链路同时传输普通业务报文和MAD检测报文（扩展LACP报文）	组网中需要使用H3C设备作为中间设备，每个成员设备都需要连接到中间设备

联系我们



 <div>H3C MAD (v.cn) 数字化解决方案领导者</div>	检测速度较快，组网形式灵活，对其它设备没有要求	当堆叠设备大于两台时，组网中需要使用中间设备，每个成员设备都需要连接到中间设备，这些BFD链路专用于MAD检测
ARP MAD	非聚合的IPv4组网环境，和MSTP配合使用，无需占用额外端口。在使用中间设备的组网中对中间设备没有要求	检测速度慢于前两种。
ND MAD	非聚合的IPv6组网环境，和MSTP配合使用，无需占用额外端口。在使用中间设备的组网中对中间设备没有要求	检测速度慢于前两种

表1 MAD检测机制的比较

[1] IRF作为一台虚拟设备与外界通信，具有唯一的桥MAC，称为IRF桥MAC。通常情况下使用主设备的桥MAC作为IRF桥MAC。IRF桥MAC不保留是我司IRF桥MAC三种不同保留时间中的一种。IRF桥MAC不保留，即当主设备离开IRF时，系统会立即使用新选举的主设备的桥MAC做IRF桥MAC。

感谢您对本刊物的关注，如果您在阅读时有何感想，请点击  
(/cn/aspx/voteforms/frm50.aspx?doctitle=irf%20mad%u5E94%u7528%u6A21%u578B%u53CA%u6280%u672F%u5206%u6790&magazine=反馈。

- 如何购买
- 关于新华三
- 联系新华三
- 常用链接

联系我们



