



Huawei AR1200 系列企业路由器 V200R002C00

特性描述-组播

文档版本 01

发布日期 2011-12-30

版权所有 © 华为技术有限公司 2011。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本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，本档内容会不定期进行更新。除非另有约定，本档仅作为使用指导，本档中的所有陈述、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： <http://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

前言

读者对象

本文档针对 AR1200 支持的组播相关特性，从基本功能、工作机制和主要应用介绍了 PIM、IGMP、MSDP、组播管理和组播路由管理。

本文档与其它类型手册相结合，便于读者深入掌握特性的实现原理。

本文档主要适用于以下工程师：

- 网络规划工程师
- 调测工程师
- 数据配置工程师
- 系统维护工程师

符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

符号	说明
 危险	以本标志开始的文本表示有高度潜在危险，如果不能避免，会导致人员死亡或严重伤害。
 警告	以本标志开始的文本表示有中度或低度潜在危险，如果不能避免，可能导致人员轻微或中等伤害。
 注意	以本标志开始的文本表示有潜在风险，如果忽视这些文本，可能导致设备损坏、数据丢失、设备性能降低或不可预知的结果。
 窍门	以本标志开始的文本能帮助您解决某个问题或节省您的时间。
 说明	以本标志开始的文本是正文的附加信息，是对正文的强调和补充。

命令行格式约定

格式	意义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用 加粗 字体表示。
<i>斜体</i>	命令行参数（命令中必须由实际值进行替代的部分）采用 <i>斜体</i> 表示。
[]	表示用“[]”括起来的部分在命令配置时是可选的。
{ x y ... }	表示从两个或多个选项选取一个。
[x y ...]	表示从两个或多个选项选取一个或者不选。
{ x y ... } *	表示从两个或多个选项选取多个，最少选取一个，最多选取所有选项。
[x y ...] *	表示从两个或多个选项选取多个或者不选。
&<1-n>	表示符号&的参数可以重复 1 ~ n 次。
#	由“#”开始的行表示为注释行。

修订记录

修改记录累积了每次文档更新的说明。最新版本的文档包含以前所有文档版本的更新内容。

文档版本 01 (2011-12-30)

第一次正式发布。

目录

前言.....	ii
1 IP 组播基础.....	1
1.1 介绍.....	2
1.2 参考标准和协议.....	4
1.3 可获得性.....	4
1.4 原理描述.....	5
1.4.1 基本概念.....	5
1.4.2 基本构架.....	6
1.4.3 组播地址.....	6
1.4.4 组播协议.....	9
1.4.5 组播模型分类.....	11
1.4.6 组播报文转发.....	11
1.5 术语与缩略语.....	12
2 IGMP.....	13
2.1 介绍.....	14
2.2 参考标准和协议.....	14
2.3 可获得性.....	14
2.4 原理描述.....	15
2.4.1 IGMPv1&v2&v3.....	15
2.4.2 IGMP 组兼容.....	17
2.4.3 IGMP 查询器选举.....	17
2.4.4 IGMP 支持 Router-Alert.....	18
2.4.5 IGMP Only-Link.....	18
2.4.6 IGMP On-Demand.....	19
2.4.7 IGMP Prompt-Leave.....	19
2.4.8 IGMP 策略控制.....	20
2.4.9 SSM Mapping.....	22
2.4.10 IGMP 主机地址过滤.....	23
2.4.11 协议的比较.....	24
2.5 应用.....	24
2.5.1 IGMP 典型应用.....	24
2.6 术语与缩略语.....	25

3 PIM	27
3.1 介绍	28
3.2 参考标准和协议	28
3.3 可获得性	28
3.4 原理描述	29
3.4.1 基本概念	29
3.4.2 PIM-SM	30
3.4.3 PIM-SSM	38
3.4.4 PIM-DM	39
3.4.5 协议比较	42
3.4.6 PIM 安全性	42
3.5 应用	43
3.5.1 PIM 典型应用	43
3.5.2 Internet 组播	46
3.6 术语与缩略语	46
4 MSDP	49
4.1 介绍	50
4.2 参考标准和协议	50
4.3 可获得性	50
4.4 原理描述	51
4.4.1 MSDP 实现域间组播	51
4.4.2 MSDP 实现 Anycast RP	52
4.4.3 MSDP 支持 MD5/Key-chain 认证	53
4.4.4 SA 消息的 RPF 检查规则	54
4.5 应用	54
4.6 术语与缩略语	55
5 组播管理	57
5.1 介绍	58
5.2 参考标准和协议	58
5.3 可获得性	59
5.4 原理描述	59
5.4.1 MPing	59
5.4.2 MTrace	60
5.5 术语与缩略语	61
6 组播路由管理	62
6.1 介绍	63
6.2 参考标准和协议	63
6.3 可获得性	63
6.4 原理描述	64
6.4.1 RPF 单播逆向路由检查	64
6.4.2 组播负载分担	65

6.4.3 按照最长匹配选择路由.....	68
6.4.4 指定组播转发边界.....	68
6.5 术语与缩略语.....	69

1 IP 组播基础

关于本章

- 1.1 介绍
- 1.2 参考标准和协议
- 1.3 可获得性
- 1.4 原理描述
- 1.5 术语与缩略语

1.1 介绍

随着 Internet 网络的不断发展，网络中交互的各种数据、语音和视频信息越来越多，同时新兴的电子商务、网上会议、网上拍卖、视频点播、远程教学等服务也在逐渐兴起。这些服务大多符合点对多点的模式，对信息安全性、有偿性、网络带宽提出了较高的要求。

用户主机能够识别的 IP 地址

用户主机能够识别的 IP 地址分为三类：

- 单播 IP 地址
一个单播 IP 地址只能标识一台用户主机，一台用户主机只能识别一个单播 IP 地址。一份使用单播 IP 地址为目的地址的 IP 报文，只能被一台用户主机接收。
- 广播 IP 地址
一个广播 IP 地址能够标识某确定网段内的所有用户主机，一台用户主机只能识别一个广播 IP 地址。一份使用广播 IP 地址为目的地址的 IP 报文，能够被该网段内的所有用户主机接收。IP 广播报文不能跨网段传播。
- 组播 IP 地址
一个组播 IP 地址能够标识网络不同位置的多个用户主机，一台用户主机可以同时识别多个组播 IP 地址。一份使用组播 IP 地址为目的地址的 IP 报文，能够被网络不同位置的多个用户主机接收。

IP 传输三种方式

IP 传输分为三种方式，分别使用以上三类 IP 地址。

- IP 单播（Unicast），简称为单播。
- IP 广播（Broadcast），简称为广播。
- IP 组播（Multicast），简称为组播。

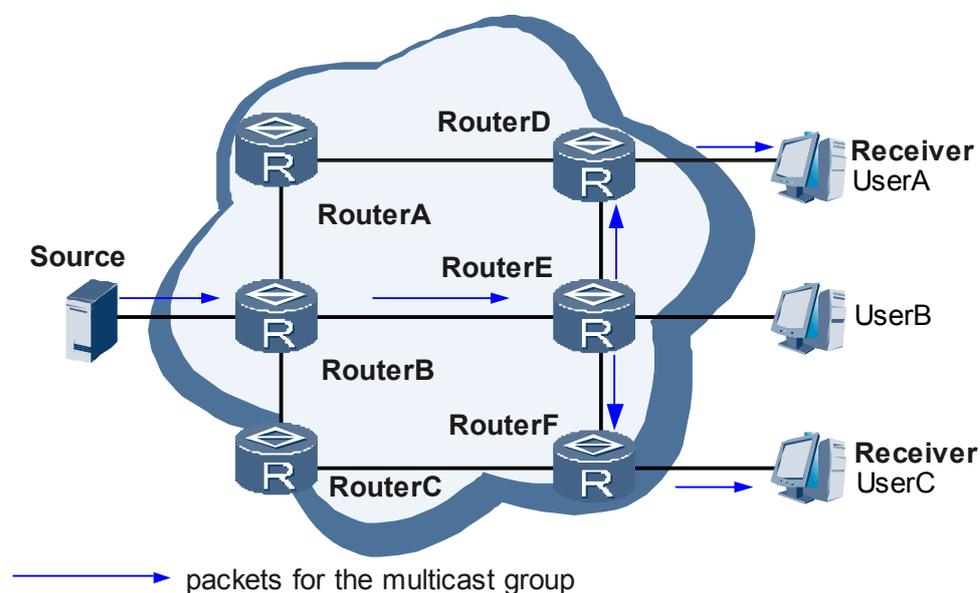
应用以上三种传输方式，分别进行点对多点的数据传输。经过比较可以发现，组播在这方面更具优势。

- 单播方式
 - 单播的特点
 - 一份单播报文，使用一个单播地址作为目的地址。Source 向每个 Receiver 发送一份独立的单播报文。如果网络中存在 N 个接收者，则 Source 需要发送 N 份单播报文。
 - 网络为每份单播报文执行独立的数据转发，形成一条独立的数据传送通路。N 份单播报文形成 N 条相互独立的传输路径。
 - 单播的缺陷
 - 单播方式下，网络中传输的信息量和需求该信息的用户量成正比，当需求该信息的用户量较大时，网络中将出现多份相同信息流，不仅占用处理器资源而且浪费带宽。
 - 单播方式较适合用户稀少的网络，当用户量较大时很难保证网络传输质量。
- 广播方式

- 广播的特点
 - 一份广播报文，使用一个广播地址作为目的地址。Source 向本网段对应的广播地址发送且仅发送一份报文。
 - 不管是否有需求，保证报文被网段中的所有用户主机接收。
- 广播的缺陷
 - 广播方式下，信息发送者与用户主机被限制在一个共享网段中，且该网段所有用户主机都能接收到该信息。
 - 广播方式只适合共享网段，且信息安全性和有偿服务得不到保障。
- 组播方式

如图 1-1 所示，网络中存在信息发送者 Source，UserA 和 UserC 提出信息需求，网络采用组播方式传输信息。

图 1-1 组播方式传输信息



- 组播的特点
 - 一份组播报文，使用一个组播地址作为目的地址。Source（组播源）向一个组播地址发送且仅发送一份报文。如图 1-1 所示：packets for the multicast group。
 - 网络中部署的组播协议为此组播报文建立一棵树型路由，根连接 Source，分支连接所有组播组成员。如图 1-1 所示：Source → RouterB → RouterE [→ RouterD → UserA | → RouterF → UserC]。
- 组播的优势
 - 组播方式下，单一的信息流沿组播分发树被同时发送给一组用户，相同的组播数据流在每一条链路上最多仅有一份。相比单播来说，使用组播方式传递信息，用户的增加不会显著增加网络的负载，减轻了服务器和 CPU 的负荷。

- 组播报文可以跨网段传输，不需要此报文的用户不能收到此报文。相比广播来说，使用组播方式可以远距离传输信息，且只将信息传输到有接收者的地方，保障了信息的安全性。
- 组播技术有效地解决了单点发送多点接收的问题，实现了 IP 网络中点到多点的高效数据传送。

- 组播的应用

组播适用于任何“点到多点”的数据发布，主要包含以下几方面：

- 多媒体、流媒体的应用
- 培训、联合作业场合的通信
- 数据仓库、金融应用（股票）

如今 ISP 提供的互联网信息服务中，已经应用了 IP 组播技术。例如：在线直播、网络电视、远程教育、远程医疗、网络电台和实时视/音频会议等。

1.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RFC1112	Host Extensions for IP Multicasting	-
RFC2236	Internet Group Management Protocol, Version 2	-
RFC3376	Internet Group Management Protocol, Version 3	-
RFC4601	Protocol Independent Multicast - Sparse Mode (PIM-SM)	-
RFC3618	Multicast Source Discovery Protocol (MSDP)	-
RFC3973	Protocol Independent Multicast - Dense Mode (PIM-DM)	-

1.3 可获得性

涉及网元

无需其他网元的配合。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

AR1200 支持该特性的版本如下：

产品	最低支持版本
AR1200	V200R001C00

1.4 原理描述

1.4.1 基本概念

组播组

组播组使用一个 IP 组播地址标识。任何用户主机（或其他接收设备），加入一个组播组，就成为了该组成员，可以识别并接收以该 IP 组播地址为目的地址的 IP 报文。

组播源

以组播组地址为目的地址，发送 IP 报文的信源称为组播源。

- 一个组播源可以同时向多个组播组发送数据。
- 多个组播源可以同时向一个组播组发送报文。

组播组成员

组播组中的成员是动态的，网络中的用户主机可以在任何时刻加入和离开组播组。组成员可能广泛分布在网络中的任何地方。

组播源通常不会同时是数据的接收者，不属于组播组成员。

说明

下文以收看某电视频道的节目为例，可以帮助理解 IP 组播中的概念。

- 组播组是发送者和接收者之间的一个约定，如同电视频道。
- 电视台是组播源，它向某频道内发送数据。
- 电视机是接收者主机，观众打开电视机选择收看某频道的节目，表示主机加入某组播组；然后电视机播放该频道电视节目，表示主机接收到发送给这个组的数据。
- 观众可以随时控制电视机的开关和频道间的切换，表示主机动态的加入或退出某组播组。

组播路由器

网络中支持组播功能的路由器称为组播路由器。

组播路由器的功能：

- 在与用户主机连接的末梢网段，提供组播组成员管理功能。
- 实现组播路由，指导组播报文的转发。

组播分发树

根据组播组成员的分布情况，组播路由协议为多目的端的数据包转发建立树型路由。报文在距离组播源尽可能远的分叉路口才开始复制和分发，最终传送到组播组成员。

1.4.2 基本构架

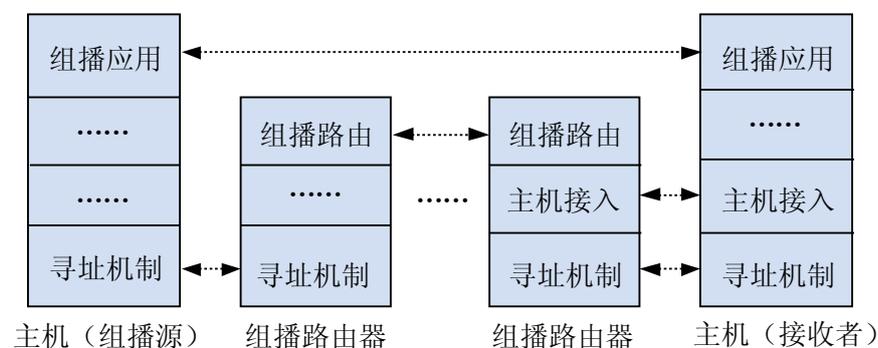
组播模型复杂多样，其目的都是以组播方式将信息从组播源传输到接收者手中，同时满足接收者对信息的各种需求。对于组播，需要关注的事项如表 1-1 所示。

表 1-1 组播技术关注事项

关注事项	组播技术
哪里有组播接收者	主机接入
从哪里可以获得组播数据	组播源发现技术
将组播数据传输到哪里	组播寻址机制
如何传输组播信息	组播路由

组播属于一种端到端服务，按照协议层从下往上划分，IP 组播基本构架包括寻址机制、主机接入、组播路由、组播应用四个部分，如图 1-2 所示。

图 1-2 IP 组播基本构架



- 寻址机制：使用组播地址，将一份数据报文发送给一组接收者。
- 主机接入：基于组播协议实现，允许用户主机动态加入和离开某组播组，实现组播成员管理。
- 组播路由：基于组播协议实现，构建报文分发树进行组播路由，从组播源传输报文到接收者。
- 组播应用：组播源与接收者必须支持视频会议等组播应用软件，TCP/IP 协议栈必须支持组播信息的发送和接收。

1.4.3 组播地址

如果采用组播方式传输信息，信息源该将信息发往何处，组播报文目的地址如何选取？这些问题简而言之就是组播寻址。

- 为了使信息源和组播组成员跨越互联网进行通讯，需要提供网络层组播，使用 IP 组播地址。
- 为了在本地物理网络上实现组播信息的正确传输，需要提供链路层组播，链路层组播使用组播 MAC 地址。
- 同时必须存在一种技术将 IP 组播地址映射为组播 MAC 地址。

IPv4 组播地址

IPv4 地址空间分为五类，即 A 类、B 类、C 类、D 类和 E 类。D 类地址为 IPv4 组播地址，用于标识组播组，使用在 IPv4 组播报文的目的地址字段。

IPv4 组播报文的源地址字段为 IPv4 单播地址，可使用 A、B 或 C 类地址，不能出现 D 类地址。E 类地址保留。

在网络层上，加入同一组播组的所有用户主机能够识别同一个 IPv4 组播组地址。一旦网络中某用户加入该组播组，则此用户就能接收以该组地址为目的地址的 IP 报文。

D 类组播地址范围是从 224.0.0.0 到 239.255.255.255，范围及含义见[表 1-2](#)。

表 1-2 D 类地址的范围及含义

D 类地址范围	含义
224.0.0.0 ~ 224.0.0.255	为路由协议预留的永久组地址。
224.0.1.0 ~ 231.255.255.255 233.0.0.0 ~ 238.255.255.255	用户可用的 ASM 临时组地址，全网范围内有效。
232.0.0.0 ~ 232.255.255.255	用户可用的 SSM 临时组地址，全网范围内有效。
239.0.0.0 ~ 239.255.255.255	用户可用的 ASM 临时组地址，仅在特定的本地管理域内有效，称为本地管理组播地址。本地管理组播地址属于私有地址，在不同的管理域内使用相同的本地管理组播地址不会导致冲突。

- 永久组地址：IANA 为路由协议预留的组播地址（也称为保留组地址），用于标识一组特定的网络设备。具体请参见[表 1-3](#)。永久组地址保持不变，组成员的数量可以是任意的，甚至可以为零。
- 临时组地址：为用户组播组临时分配的 IPv4 地址（也称为普通组地址），组成员的数量一旦为零，即取消。

表 1-3 常见的永久组地址列表

永久组地址	含义
224.0.0.0	不分配
224.0.0.1	网段内所有主机和路由器（等效于广播地址）
224.0.0.2	所有组播路由器的地址
224.0.0.3	不分配

永久组地址	含义
224.0.0.4	DVMRP 路由器
224.0.0.5	OSPF 路由器
224.0.0.6	OSPF DR
224.0.0.7	ST 路由器
224.0.0.8	ST 主机
224.0.0.9	RIP-2 路由器
224.0.0.11	移动代理
224.0.0.12	DHCP 服务器/中继代理
224.0.0.13	所有 PIM 路由器
224.0.0.14	RSVP 封装
224.0.0.15	所有 CBT 路由器
224.0.0.16	指定 SBM
224.0.0.17	所有 SBMS
224.0.0.18	VRRP
224.0.0.19 ~ 224.0.0.21	未指定
224.0.0.22	所有使能 IGMPv3 的路由器
224.0.0.23 ~ 224.0.0.255	未指定

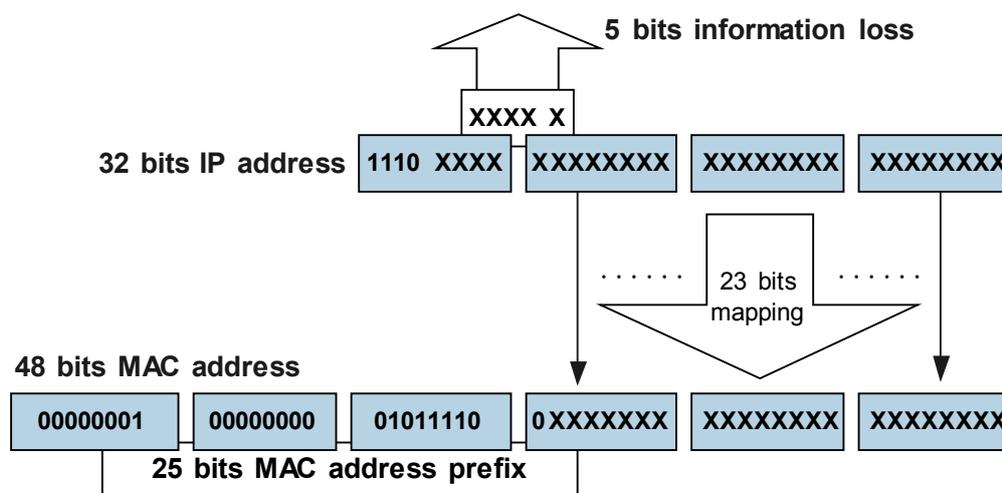
以太网组播 MAC 地址

组播 MAC 地址用于在链路层上标识属于同一组播组的接收者。

网络设备上的以太网接口板可以识别组播 MAC 地址。通过在驱动程序中配置某组播 MAC 地址，设备就可以在以太网上接收和转发该组播组的数据。

IANA 规定，组播 MAC 地址的高 25bit 为 0x01005e，第 25bit 为 0，低 23bit 为组播 IP 地址的低 23bit，映射关系如图 1-3 所示。

图 1-3 组播 IP 地址与组播 MAC 地址的映射关系

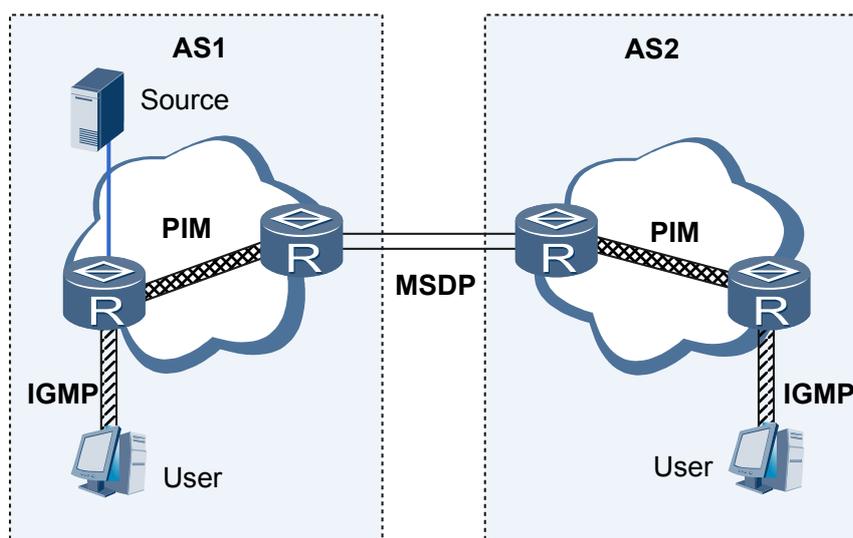


IP 组播地址的前 4bit 是固定的 1110，对应组播 MAC 地址的高 25bit。IP 组播地址的后 28bit 中只有 23bit 被映射到 MAC 地址，因此丢失了 5bit 的地址信息，直接结果是有 32 个 IP 组播地址映射到同一 MAC 地址上。

1.4.4 组播协议

实现一套完整的组播服务，需要在网络各个位置部署多种组播协议相互配合，共同运作，如图 1-4 所示。

图 1-4 IPv4 组播网络



针对不同的应用位置与目的，AR1200 开发出功能各异的多种组播路由协议，如表 1-4 所示。

表 1-4 组播协议

应用位置	目的	组播协议
用户主机与组播路由器之间	将用户主机接入组播网络： <ul style="list-style-type: none">● 在主机侧实现组播组成员动态加入与离开。● 在路由器侧实现组成员关系的维护与管理，同时支持与上层组播路由协议的信息交互。	IGMP（Internet Group Management Protocol）因特网组管理协议，用于 IPv4 网络。
域内组播路由器之间	组播路由与转发： <ul style="list-style-type: none">● 按需创建组播路由。● 动态响应网络拓扑变化，维护组播路由表。● 按照路由表项执行转发。	PIM（Protocol Independent Multicast）协议无关组播。
域间组播路由器之间	域间组播源信息共享： <ul style="list-style-type: none">● 源所在域内的路由器将本地源信息传播给其他域内的路由器。● 不同域的路由器之间传递源信息。	MSDP（Multicast Source Discovery Protocol）组播源发现协议，目前仅用于 IPv4 网络。

从表 1-4 中观察可知，组播协议的功能主要分为以下两类：

组播组成员关系管理

组播组成员关系管理是指在主机与路由器之间建立和维护组成员关系。

IGMP 是用于 IPv4 网络的组播组成员关系管理协议，有以下特点：

- 包含三个版本，分别是 IGMPv1、IGMPv2 和 IGMPv3。新版本完全兼容旧版本。目前应用最广泛的是 IGMPv2。
- 三个版本都支持 ASM 模型；IGMPv3 可以直接支持 SSM 模型，而 IGMPv1 和 IGMPv2 需要结合 SSM-Mapping 技术才能支持 SSM 模型。

建立并维护组播路由

组播路由也称为组播分发树，指从一个组播源到所有组成员的数据传输路径。组播路由单向、无环且路径最短。通过在路由器之间建立和维护组播路由，网络才能够正确、高效地转发组播数据包。

- 域内组播路由协议：用来在自治系统 AS（Autonomous System）内发现组播源并构建组播分发树，将信息传递到接收者。PIM 是典型的域内组播路由协议，有两套独立的模式：
 - DM（Dense Mode）：适用于小规模、接收者分布较为密集的情况，支持 ASM 模型。
 - SM（Sparse Mode）：适用于大规模、接收者分布较为稀疏的情况，同时支持 ASM 模型和 SSM 模型。

- 域间组播路由协议：用来在 AS 之间传递组播源信息，从而跨域建立组播路由，实现域间组播资源共享。MSDP 是典型的域间组播路由协议，通常与 MBGP 协同工作。MSDP 适用于各域内运行 PIM-SM 的情况。

对于 SSM 模型来说，没有域内和域间的划分。由于接收者预先知道组播源的具体位置，因此可以借助 PIM-SM 的部分功能直接创建组播传输路径。

1.4.5 组播模型分类

根据对组播源的控制程度的不同，IP 组播分为三种模型，分别为：

- ASM 模型
- SFM 模型
- SSM 模型

ASM 模型

ASM 全称为 Any-Source Multicast，译为任意源组播。在 ASM 模型中，任意发送者都可以成为组播源，向某组播组地址发送信息。接收者加入该组播组后，能够接收到发往该组播组的所有信息。

在 ASM 模型中，接收者无法预先知道组播源的位置，接收者可以在任意时间加入或离开该组播组。

SFM 模型

SFM 全称为 Source-Filtered Multicast，译为过滤源组播。SFM 模型继承了 ASM 模型，从发送者角度来看，组播组成员关系相同。

同时，SFM 在功能上对 ASM 进行了扩展：上层软件对接收到的组播报文的源地址进行检查，允许或禁止来自某些组播源的报文通过。最终，接收者只能接收到来自部分组播源的数据。从接收者角度来看，只有部分组播源是有效的，组播源经过了筛选。

说明

SFM 在 ASM 的基础上添加了组播源过滤策略，此外基本原理和配置方法相同。本手册中将 SFM 与 ASM 统称为 ASM。

SSM 模型

SSM 全称为 Source-Specific Multicast，译为指定源组播。在现实生活中，用户可能仅对某些源发送的组播信息感兴趣，而不愿接收其它源发送的信息。SSM 模型为用户提供了一种能够在客户端指定信源的传输服务。

SSM 模型和 ASM 模型的根本区别是接收者已经通过其他手段预先知道了组播源的具体位置。SSM 和 ASM 使用不同的组播地址范围，直接在接收者和组播源之间建立组播转发树。

1.4.6 组播报文转发

组播报文转发和单播报文转发相互隔离，互不影响。

在组播模型中，IP 报文的目地址字段为组播组地址，组播源向以此目的地址所标识的主机群组传送信息。因此，转发路径上的路由器为了将组播报文传送到各个方位的接收站点，往往需要将从一个入接口接收到的组播报文，从多个出接口转发出去。与单播模型相比，组播模型的复杂性就在于此。

- 由组播路由表来指导组播报文转发。
- 由 RPF (Reverse Path Forwarding) 机制保证组播路由是一棵最短路径树。RPF 机制是大部分组播路由协议创建组播路由表项、进行组播转发的基础。

组播报文转发的过程为：

- 对于 ASM，设备接收到组播数据时，先查找 MFIB，如果 MFIB 中存在此表项，根据此转发表项进行转发。如果 MFIB 中没有此表项，则转发平面通知 PIM 协议创建组播路由表，由 PIM-SM 协议根据用户的加入信息创建组播路由表下发到 MFIB，然后指导组播数据的转发。
- 对于 SSM，由 PIM-SSM 协议根据用户的加入信息创建组播路由表直接下发到 MFIB。当接收到组播数据时，查找 MFIB，如果 MFIB 中存在此表项，根据此转发表项进行转发；否则直接丢弃。

1.5 术语与缩略语

术语

术语	解释
IGMP	Internet Group Management Protocol，称为因特网组管理协议，是 IP 组播在末端网络上使用的主机对路由器的信令机制。 IGMP 在主机侧实现组播组成员动态加入与离开，在路由器侧实现组成员关系的维护与管理，同时支持与上层组播路由协议的信息交互。
PIM	Protocol Independent Multicast，称为协议无关组播，属于组播路由协议。 网络中单播路由畅通是 PIM 转发的基础。PIM 利用现有的单播路由信息，对组播报文执行 RPF 检查，从而创建组播路由表项，构建组播分发树。
MSDP	Multicast Source Discovery Protocol，称为组播源发现协议。只适用于 PIM-SM 域，仅对 ASM (Any-Source Multicast) 模型有意义。 通过在不同 PIM-SM 域的 RP 之间建立 MSDP 对等体关系，在域间共享组播源信息，实现跨域组播。 通过在同一 PIM-SM 域的多个 RP 之间建立 MSDP 对等体关系，在域内共享组播源信息，实现 Anycast RP。

缩略语

缩略语	英文全称	中文全称
ASM	Any-Source Multicast	任意源组播
SFM	Source-Filtered Multicast	过滤源组播
SSM	Source-Specific Multicast	指定源组播

2 IGMP

关于本章

- 2.1 介绍
- 2.2 参考标准和协议
- 2.3 可获得性
- 2.4 原理描述
- 2.5 应用
- 2.6 术语与缩略语

2.1 介绍

定义

IGMP (Internet Group Management Protocol) 因特网组管理协议, 是 TCP/IP 协议族中负责 IPv4 组播成员管理的协议, 用来在 IP 主机和与其直接相邻的组播路由器之间建立、维护组播组成员关系。

通过在接收者主机和与其直连的组播路由器上配置 IGMP, 可以实现主机动态加入组播组和组播路由器对本地网络组成员信息的管理。

到目前为止, IGMP 有三个版本: IGMPv1 版本 (RFC1112)、IGMPv2 版本 (RFC2236) 和 IGMPv3 版本 (RFC3376)。所有 IGMP 版本都支持 ASM (Any-Source Multicast) 模型。IGMPv3 可以直接应用于 SSM (Source-Specific Multicast) 模型, 而 IGMPv1 和 IGMPv2 则需要通过 SSM-Mapping 技术来支持 SSM 模型。

目的

要使组播报文最终能够到达接收者, 需要将接收者主机接入 IP 组播网络, 并加入到相应的组播组中。IGMP 通过在主机侧和路由器侧交互 IGMP 报文实现组成员管理功能。IGMP 协议可以记录接口下主机的加入和离开等信息, 以确保组播数据能够正确地转发到该接口。

2.2 参考标准和协议

本特性的参考资料清单如下:

文档	描述	备注
RFC1112	Host Extensions for IP Multicasting	-
RFC2236	Internet Group Management Protocol, Version 2	-
RFC3376	Internet Group Management Protocol, Version 3	-
RFC3569	An Overview of Source-Specific Multicast (SSM)	-
RFC4601	Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)	-

2.3 可获得性

涉及网元

无需其他网元的配合。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

AR1200 支持该特性的版本如下：

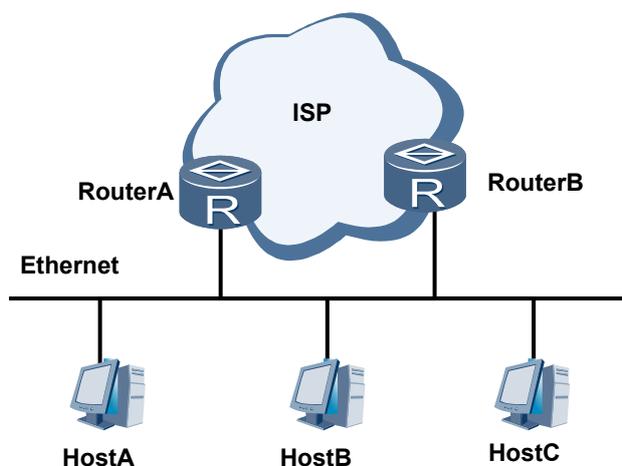
产品	支持版本
AR1200	V200R001C00

2.4 原理描述

2.4.1 IGMPv1&v2&v3

IGMP 协议

图 2-1 IGMP 基本组网图



通过发送查询报文并接收主机反馈的加入报文和离开报文，路由器可以了解与该接口连接的网段上存在哪些组播组的接收者（即组成员）。如果出现组成员，组播路由器应将对应组的组播数据报文转发到这个网段；如果没有组成员则不转发。主机可以自主决定加入或退出某个组播组。

如图 2-1 所示，使能了 IGMP 协议的路由器 RouterA 会自动变为查询器并定时发出 IGMP 查询报文，与 RouterA 在同一网段的所有主机（HostA、HostB、HostC）都能收到它发出的查询报文。

- 主机收到查询报文的处理：
 - 如果主机之前已经加入了组 G，就会在路由器允许的响应时间内随机发送一个组 G 的 IGMP 加入报文。

RouterA 收到组 G 的 IGMP 加入报文会记录组 G 的相关信息，同时对该组启动一个定时器（如果已经启动则刷新定时器），以便长时间没有主机响应时切断该组的组播流量。RouterA 把组 G 的组播数据转发到 HostA 和 RouterA 相连的接口所在的网段上。

- 如果主机不是任何组播组成员，主机在收到 IGMP 查询报文时不作任何响应。

- 主机加入组播组：

一个新加入组播组 G 的主机会主动发送一个该组的 IGMP 加入报文，通知路由器更新组播组信息，后续的加入报文则由路由器的查询报文来驱动。

- 主机离开组播组：

如果一个主机决定离开某个组播组 G，会主动发送组 G 的 IGMP 离开报文。路由器收到后会触发一个指定组 G 的查询，确定该组在当前网段接收者的存在情况。如果在查询结束后仍然没有收到主机针对该组的 IGMP 加入报文，则删除已记录的组信息，停止转发该组数据到对应接口所在的网段上。

IGMPv1 处理 IGMP 报文

IGMPv1 协议主要基于查询和响应机制完成组播组管理，支持查询和加入报文，处理过程与 IGMPv2 相同。IGMPv1 与 IGMPv2 的不同之处是：主机离开组播组时不主动发送离开报文，收到查询消息后不反馈 Report 消息，待维护组成员关系的定时器超时后，路由器删除组记录。

IGMPv2 处理 IGMP 报文

运行 IGMPv2 的主机发送的 IGMP 报告中仅携带组信息。当主机发送一个组的 IGMP 加入报文给路由器后，路由器会通知组播转发模块，以便这个组的组播数据到来时能够正确转发给该主机。

IGMPv2 协议具有报告抑制机制，可以减少网络中的 IGMP 的重复报告。

当一个主机 HostA 加入了某个组播组 G，在收到路由器的查询报文后，HostA 会在 0 ~ 最大响应时间（查询报文中已经指定）之间选择一个随机值作为定时器的超时时间，并在该定时器超时后，向路由器发送组 G 的加入报文。如果在超时时间内，HostA 收到了加入同一个组的主机 HostB 发送的加入报文，则 HostA 不再向路由器发送组 G 的加入报文。

当主机退出某个组 G 时，会向路由器发送一个指定组 G 的 IGMP 离开报文。由于 IGMPv2 报告抑制机制，路由器无法确定是否还有其他主机加入了组 G。这时路由器会触发一个指定组 G 的查询，如果其他主机加入了组 G，就会发送针对组 G 的 IGMP 加入报文。

如果路由器发送了若干次数指定组 G 的查询之后，仍然没有收到主机针对组 G 的 IGMP 加入报文，那么路由器就不再记录组 G 的信息，停止转发该组数据到对应接口所在的网段。

 说明

IGMP 的查询器和非查询器都会处理 IGMP 组加入信息，但是只有查询器负责发送查询报文。IGMP 非查询器不处理 IGMPv2 离开报文。

IGMPv3 处理 IGMP 报文

IGMPv2 报文中只能携带组播组的信息，不能携带组播源的信息。这样运行 IGMPv2 的主机就只能选择加入某个组，而不能选择加入某个组播源/组。IGMPv3 解决了该问题。

运行 IGMPv3 的主机不仅能够选择组，还能根据需要进行组播源/组。主机发送的 IGMPv3 报文中可以包含多个组记录，每个组记录中可以包含多个组播源。

在路由器侧，查询器发送查询报文并接收主机反馈的加入报文和离开报文，以此来了解与该接口连接的网段上有哪些组播组存在接收者，并将组播数据转发到相应的网段。IGMPv3 的组记录有 include 和 exclude 二种组过滤模式。

- 在 include 模式下
 - 处于激活状态的组播源表示需要路由器转发这个源的数据。
 - 不活动的源会被路由器删除并停止转发这个源的数据。
- 在 exclude 模式下
 - 处于激活状态的组播源表示处于冲突域中。也就是说，与该路由器接口同一网段的主机中，有的主机需要该源的数据，有的主机不需要该源的数据，在这种情况下该源的数据仍然需要转发。
 - 不活动的组播源表示不需要转发该源的数据。
 - 组中没有记录的组播源的数据全部都要转发。

在 IGMPv3 中，实现了对采用 Include 模式加入特定源组的 IGMPv3 主机成员信息的跟踪功能。

相对于 IGMPv2，IGMPv3 没有报文抑制机制，所有加入组播组的主机在收到查询时都会响应 IGMP 加入报文。由于有了对组播源的选择，IGMPv3 路由器在通用查询和组查询的基础上增加了指定源组查询，用以在收到特定组播源的数据时确定是否存在该数据的接收者。

2.4.2 IGMP 组兼容

IGMP 组兼容模式是指支持高 IGMP 版本的组播设备可以兼容低版本的主机。例如 v2 版本的组播设备可以正确处理 v1 主机的加入，v3 版本的组播设备可以正确处理 v1 和 v2 版本的主机加入。当组播设备工作在兼容模式时，收到低版本的主机的 IGMP 加入报文后会自动降低组的兼容版本到该主机对应的版本，并工作在该版本上。

工作在 v2 或 v3 版本的组播设备收到 IGMPv1 主机发送的 Report 报文时，会自动把该组的兼容模式设定为 v1 模式。在这种情况下，设备会忽略针对该组的 IGMPv2 Leave 报文。

工作在 v3 版本的组播设备收到 v2 版本的 Report 报文时，会自动把该组的兼容模式设定为 v2 模式。在这种情况下，设备会忽略 IGMPv3 的 BLOCK 报文、IGMPv3 的 TO_IN 报文以及 IGMPv3 的 TO_EX 报文的源列表，即抑制了 IGMPv3 对组播源的选择功能。

通过手工配置把组播设备从低版本升到高版本时，如果有组存在，则这些组继续工作在低版本的兼容模式，直到低版本的主机退出该组播组。

 说明

缺省情况下，IGMP 的版本是 IGMPv2。

2.4.3 IGMP 查询器选举

使能了 IGMP 协议的组播设备在网段中的角色有两个：

- 查询器

负责发送查询报文，并接收主机反馈的加入报文和离开报文，以此来了解与该接口连接的网段上有哪些组播组存在接收者（即组成员）。

- 非查询器

只接收主机反馈的加入报文，了解与该接口连接的网段上有哪些组播组存在接收者，并根据网段中查询器的动作确定当前网段中有哪些组播组成员离开。

通常情况下一个网段只有一个查询器，因此组播设备之间需要用某些方式来选出查询器。查询器选举时采用以下的原则：

- 组播设备 A 使能 IGMP 协议后，在 IGMP 协议启动阶段会默认自己为当前网段中的查询器，向网段中发送查询报文。如果收到 IP 地址比自己小的组播设备 B 发来的查询报文，则 A 由查询器转为非查询器，并启动其他查询器存在定时器，记录 B 为当前网段的查询器。
- 如果组播设备 A 在非查询器状态时，收到查询器组播设备 B 发送的查询报文，则更新其他查询器存在定时器。如果此时收到的查询报文不是先前记录的查询器 B 发来的，而是新的组播设备 C 发来的，且 C 的 IP 地址比 B 的小，则更新查询器为 C，同时更新其他查询器存在定时器。
- 如果组播设备 A 在非查询器状态时，其他查询器存在定时器超时，则由非查询器转为查询器状态，承担起查询器的职责。

 说明

协议规定 IGMPv1 不支持查询器选举，IGMPv1 查询器由上层协议（如 PIM）指定。当前仅支持同网段上同版本的组播设备之间进行查询器选举。为了保证正常工作，需要在同网段所有组播设备上配置相同版本的 IGMP。

2.4.4 IGMP 支持 Router-Alert

通常情况下，只有目的地址属于本设备的接口地址时，报文才会上送给路由协议层处理。Router-Alert 是一种标示协议报文的特殊机制。如果一个报文中带有 Router-alert 选项，则表示该报文需要被上送到路由协议层去处理。在实际使用中，有些协议报文的地址是组播地址或其他特殊地址，如果报文中没有带有 Router-Alert 选项，可能不会被上送。

IGMP 报文的目的地址一般为组地址而非组播设备的接口地址，这样就导致报文可能不会被上送到路由协议层处理，Route-Alert 选项可以解决此类问题。

当组播设备收到 IGMP 报文时：

- 缺省情况下，不检查 Router-alert 选项。无论 IGMP 报文有没有 Route-Alert 选项，都会上送到路由协议层去处理。
- 在配置了检查 Router-Alert 选项的情况下，只有带有 Route-Alert 选项的 IGMP 报文才会被上送到路由协议层处理。

2.4.5 IGMP Only-Link

IGMP Only-Link 是指在组播设备与主机相连的接口上只使能 IGMP 协议，而不使能 PIM 等上层协议，只由 IGMP 协议来指导该网段的数据转发的一种机制。

相比由 PIM 协议指导某一网段数据转发的机制，应用该特性可以减少组播设备对 PIM 邻居、PIM 接口状态机等信息的维护。

应用 IGMP Only-Link 机制时，查询器有以下功能：

- 发送查询报文并接收主机反馈的加入报文和离开报文，以此来了解与该接口连接的网段上有哪些组播组存在接收者。
- 维护 IGMP 组播组加入/离开状态，并根据 IGMP 组播组的加入/离开状态来指导该网段的数据转发。

非查询器上则只维护 IGMP 组播组加入/离开状态。

说明

接口上使能 PIM 协议时由 DR 指导数据转发，详细介绍请参见 [3.4.2 PIM-SM](#) 中的“PIM DR 竞选的基本原理”。

2.4.6 IGMP On-Demand

组播设备通过发送查询报文并接收主机反馈的加入报文和离开报文来了解与该接口连接的网段上有哪些组播组存在接收者。和组播设备相连的可能不是主机，而是一个使能了 IGMP 代理的接入设备。

为了减少组播设备和接入设备间的报文交互，可以做这样的优化：接入设备汇聚其接收的 IGMP 组播组加入/离开状态，只在 IGMP 组播组的状态发生改变时才将该加入/离开状态上报给组播设备。即，接入设备在第一个用户加入组播组时发送加入报文，最后一个用户离开组播组时发送离开报文。在这样的应用场景下，组播设备侧的 IGMP On-Demand 特性应运而生。

使能 IGMP On-Demand 特性的组播设备不主动发送查询来确定某一个组播组在当前网段上是否有接收者，而是通过与该接口相连的 IGMP 代理设备将其汇聚的组播组加入/离开状态主动上报给组播设备的方式来实现 IGMP 组播组维护。

IGMP On-Demand 只适用于 IGMPv2 和 IGMPv3。组播设备使能了 IGMP On-Demand 特性后，与 IGMP 标准协议行为有 3 点不同：

- 不发送查询报文。
- 收到 IGMP 加入报文后创建组播组和源信息，且表项永不超时。
- 只有在收到 IGMP 离开报文后，组播设备才会删除对应表项。

2.4.7 IGMP Prompt-Leave

当主机退出某个组 G 或某个源组 (S, G) 时，会向组播设备发送一个指定组 G 或指定源组 (S, G) 的 IGMP 离开报文。由于 IGMPv2 报告抑制机制，组播设备无法确定是否还有其他主机加入了组 G 或源组 (S, G)。这时设备会触发一个指定组 G 或指定源组 (S, G) 的查询，如果其他主机加入了组 G 或源组 (S, G)，就会发送针对组 G 或源组 (S, G) 的 IGMP 加入报文。如果设备发送了若干次数指定组 G 或指定源组 (S, G) 的查询之后，仍然没有收到主机针对组 G 或源组 (S, G) 的 IGMP 加入报文，那么就不再记录组 G 或源组 (S, G) 的信息，停止转发该组数据到对应接口所在的网段。

如果组播设备只和一个使能了 IGMP 代理的接入设备相连，那么当该接入设备离开某个组播组 G 并向组播设备发送针对该组 G 的 IGMP 离开报文时，组播设备无需触发指定组 G 的查询报文来确定当前网段上该组 G 是否还有其他接收者，可以直接将该组 G 的组记录删除，停止转发该组数据到对应接口所在的网段。IGMP Prompt-Leave（快速离开）特性解决了该问题。

组播设备使能了 IGMP Prompt-Leave 特性后，当其接收到 IGMP 离开报文后，不会触发针对该组的查询报文，而是直接将该组的组记录删除，停止转发该组的数据到对应接口所在的网段。应用该特性可以提升组播设备响应组播组离开的速度。



说明

IGMP Prompt-Leave 特性只在 IGMPv2 版本上支持，其他版本不支持该特性。
IGMP On-Demand 特性已经包含 IGMP Prompt-Leave 特性。

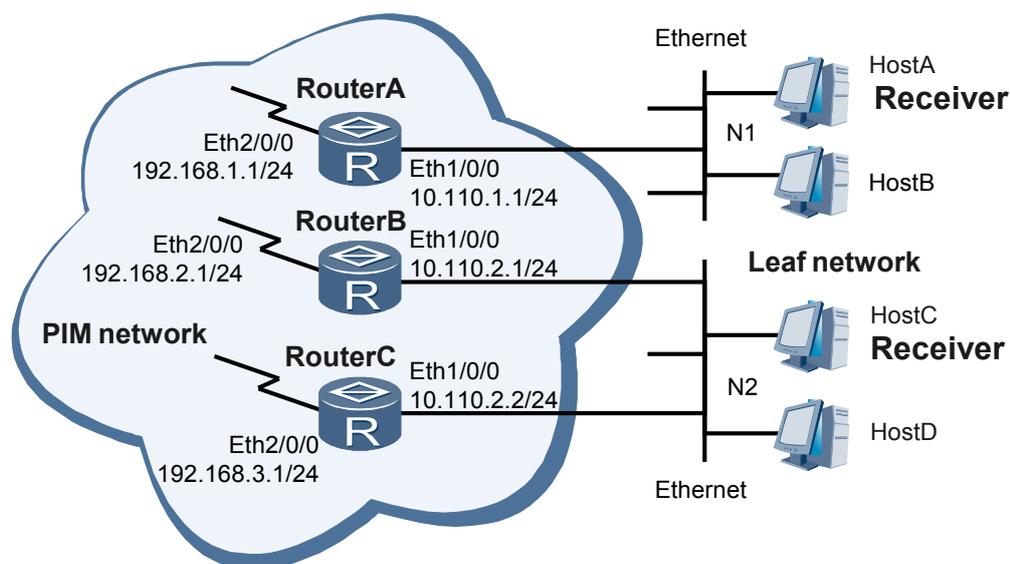
2.4.8 IGMP 策略控制

可控组播是对 IGMP 的行为进行限制或者扩充的附加特殊功能，对 IGMP 协议本身实现没有影响。具体包括 3 个功能：IGMP-Limit、静态组加入和 Group-Policy。

- **IGMP-Limit**
通过配置接口表项限制、单实例和所有实例表项限制，限制组播组或源组数量。
- **静态组加入**
通过在接口上配置静态组播组加入，可以实现快速响应用户请求，将该组播数据转发给接收者，减少用户的频道切换时间。
- **Group-Policy**
是管理员配置在路由器上的一种过滤组策略。配置 Group-Policy 之后，路由器可以对某些特定的组进行限制，不建立对应的表项。

IGMP-Limit

图 2-2 配置 IGMP-Limit 组网图



在大量用户同时收看多套节目时，需要占用路由器的大量带宽，会造成路由器性能下降。为了避免这种情况的发生，需要限制 IGMP 接口和全局下允许加入的组播组个数，使加入的组播组个数在给定的限制之内。这样能够使加入组播组的用户收看更加清晰稳定的节目。

IGMP-Limit 是指通过在路由器上的接口、单实例和所有实例下配置 IGMP 组播组限制个数。当收到 IGMP 加入报文时，首先判断是否超过配置的个数限制，如果没有超过，则建立组成员关系，给用户转发该组的数据流。

- **IGMP 接口表项限制**

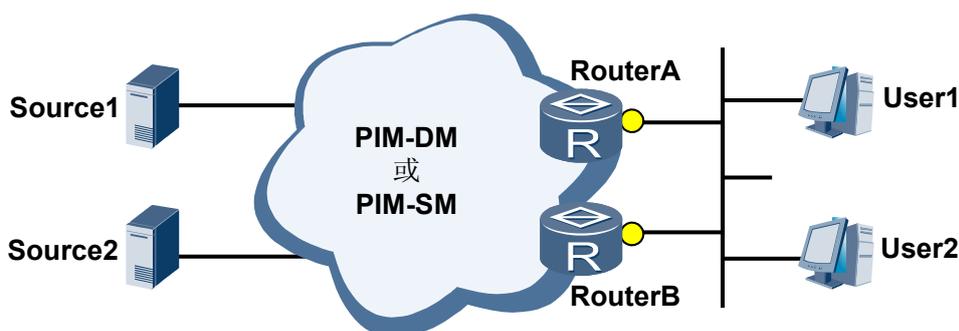
- 支持接口配置 IGMP 加入的接口表项限制。当接口收到 IGMP 协议加入报文时，可以对本接口的表项数量进行限制。
 - 支持通过配置 IGMP 表项限制，使某些组范围或源组范围不受接口加入限制，不参与接口 IGMP 加入的计数。
- IGMP 整路由器表项限制
支持对整路由器的 IGMP 加入进行限制。整路由器的 IGMP 加入限制即对所有实例接口的 IGMP 加入的接口表项总和进行限制。
 - 收到 IGMP 协议加入报文时，判断是否超过整路由器的 IGMP 加入表项限制。
 - 删除接口(*,G)/(S,G)成员关系时，要减少整路由器的 IGMP 加入表项的计数。

以上三种表项限制策略均遵循以下计数规则：

- 每个(*,G)组成员关系和每个(S,G)源组成员关系各计为一个接口表项。
- 用于 SSM-Mapping 的每个(*,G)组成员关系计为一个接口表项，按照映射生成的(S,G)表项不进行计数。

静态组加入

图 2-3 配置静态组加入组网图



- 在路由器接口上配置静态加入组，相当于该网段上存在稳定的组成员

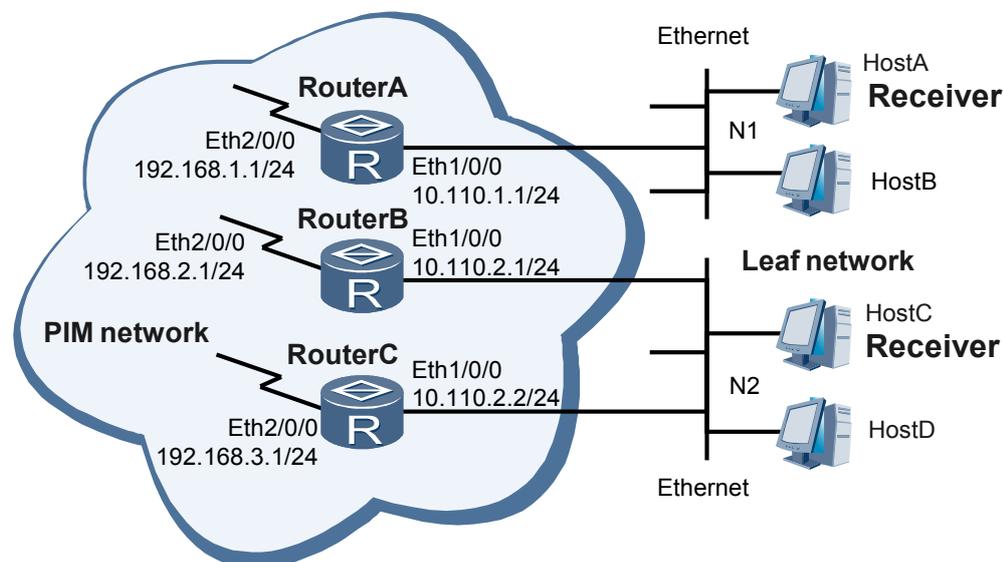
静态组播是通过在接口上配置静态组加入来实现的。配置静态组加入后，路由器创建的表项没有定时器，永远不会超时，因此路由器会持续地给接收者转发数据。接收者不再需要路由器转发该组播组数据时，不能通过表项超时自动删除，只能通过手动删除静态组配置来实现。

在实际应用中，静态组加入配置在接收端路由器与主机相连的接口上，以便将组播数据引到该路由器上。当与该路由器直连的主机或路由器中有接收者想接收对应的组播组数据时，该路由器就能做到快速响应，将相应的组播数据转发给接收者，这样可以减少用户的频道切换时间。

组播静态加入同时支持单条配置和批量配置，即一条静态加入配置支持一个组播组(源组)加入和多个组播组(源组)加入。

IGMP Group-Policy

图 2-4 配置 Group-Policy 组网图



Group-Policy 是管理员配置在路由器接口上的一种过滤策略，配置 Group-Policy 之后，路由器可以对某些特定的组进行限制，不建立对应的表项。

在大量的用户同时收看多套节目组时，会占用路由器大量的带宽，同时也会降低路由器的性能。这时可以使用 Group-Policy 对某些组进行限制，使组播组数量控制在一定的范围内。另外，出于安全考虑或者管理的需要，路由器可能不希望接收某些组的加入报文，不希望转发该组播组的数据，也可以通过 Group-Policy 加以限制。

Group-Policy 的过滤规则统一通过 ACL 配置。

2.4.9 SSM Mapping

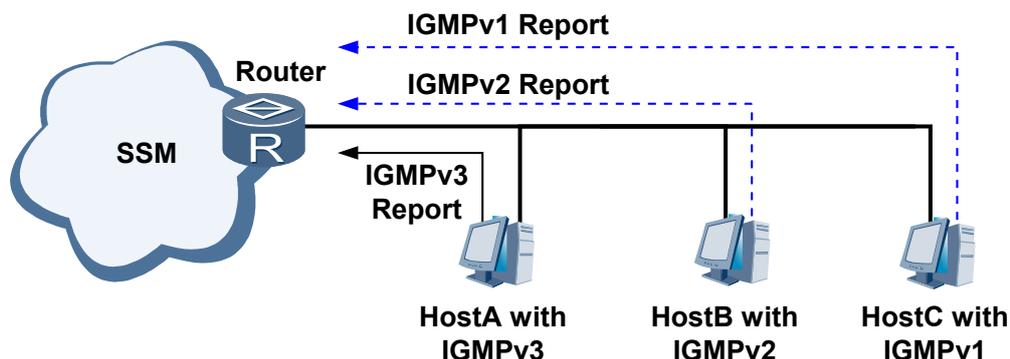
SSM Mapping (Source Specific Multicast Mapping) 机制可以更好的兼容运行 IGMPv3 之前的版本的主机，使其也能够使用组播 SSM 范围的服务。SSM Mapping 机制是：将处于 SSM 范围的 IGMPv1/v2 的(*,G)加入按照配置的转化规则转化成对应的一组(S,G)加入。这样应用低版本 IGMP 的用户也可以获得 SSM 范围的组播服务。

同时，应用 SSM Mapping 机制能够很好的保护组播源服务器，减少其受到攻击的可能。

说明

对于 SSM 范围的组，组播设备不接收整个组的需求，只处理源组的需求。SSM 的详细介绍请参见 [PIM-SSM](#)。

图 2-5 SSM Mapping 应用组网图



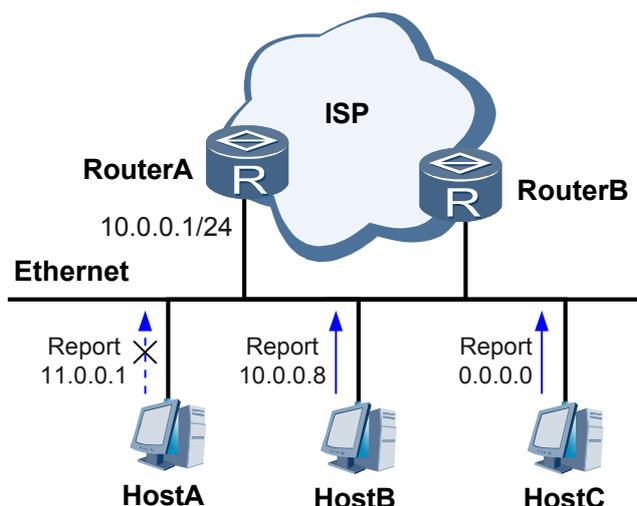
如图 2-5 所示，SSM 网络的用户网段中 HostA 运行 IGMPv3、HostB 运行 IGMPv2、HostC 运行 IGMPv1。在不允许将 HostB 和 HostC 升级为 IGMPv3 的情况下，为该网段中的所有主机提供 SSM 组播服务，需要在组播设备上支持 SSM Mapping。

如果组播设备支持 SSM Mapping，并且配置了相应的转化规则，设备在接收到 HostC 或者 HostB 发出的 IGMP(*,G)加入报文的时候，会分情况进行处理：

- 报文中的组播组为 ASM 范围，则按照 2.4.1 IGMPv1&v2&v3 章节描述的方式进行处理。
- 报文中的组播组为 SSM 范围，则通过查找配置的转化规则将报文中指定的(*,G)加入，转化成对应的一组(S,G)加入。

2.4.10 IGMP 主机地址过滤

图 2-6 IGMP 主机地址过滤应用组网图



为了保证组播流量发送的准确性，允许用户在组播设备的接口配置 IGMP 主机地址过滤策略：

- 如果 IGMP 报文主机地址和接收接口地址在同一网段或者主机地址是网络地址 (0.0.0.0)，那么此报文的主机地址过滤检查通过；
- 如果报文主机地址和接收接口地址不在同一网段，那么此报文的主机地址过滤检查不通过，丢弃此报文。

如图 2-6 所示，RouterA 与 Host 相连的接口地址为(10.0.0.1/24)，HostA 发送的 IGMP Report 报文的主机地址为 11.0.0.1，HostB 发送的 IGMP Report 报文的主机地址为 10.0.0.8，HostC 发送的 IGMP Report 报文的主机地址为 0.0.0.0，那么 RouterA 处理 HostB 和 HostC 的报文，丢弃 HostA 的报文。

2.4.11 协议的比较

IGMPv1 和 IGMPv2 协议的比较：

IGMPv1	IGMPv2	IGMPv2 较 IGMPv1 的优势
报文类型不包含成员离开报文	报文类型包括成员离开报文	可以更及时有效管理组播组成员
只支持普遍组查询	除了支持普遍组查询，还支持特定组查询	可以直接对特定的组播组进行选择，增加了选择的粒度

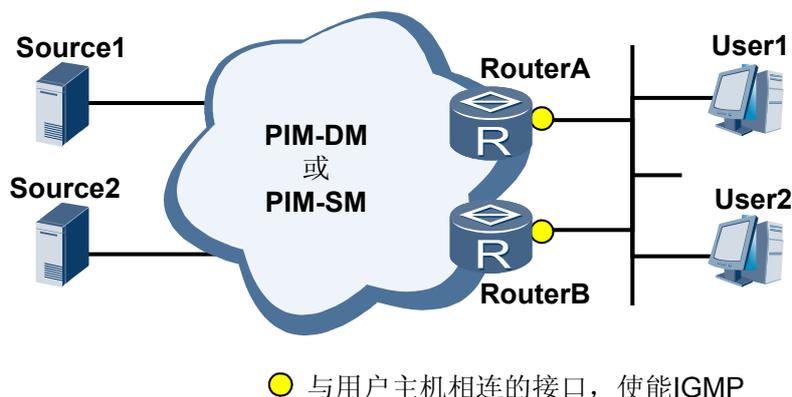
IGMPv2 和 IGMPv3 协议的比较：

IGMPv2	IGMPv3	IGMPv3 较 IGMPv2 的优势
报文中不能携带组播源信息，只能指定组播组信息	报文中除了携带组播组信息，还能携带组播源信息	可以直接对特定的组播源进行选择，增加了选择的粒度
一个报文中只能携带一个组记录	一个报文中可以携带多个组记录	减少了网段中的 IGMP 报文数量
指定组查询报文无重传机制	指定组、指定源组查询报文有重传机制	非查询器和查询器维护的组播组信息能够更好地保持一致

2.5 应用

2.5.1 IGMP 典型应用

图 2-7 IGMP 应用典型组网图



IGMP 是处理主机加入路由网络的协议。因此，该协议应用于路由边界与主机相连的区域。该协议可以处理多主机和多个组播设备分别使用不同版本时的情况。

IGMP On-Demand 特性和 **IGMP Prompt-Leave** 特性只适用于共享网段上只有一个组播设备和单一接入设备的情况。

2.6 术语与缩略语

术语

术语	解释
IGMP	Internet Group Management Protocol，称为因特网组管理协议，是 IP 组播在末端网络上使用的主机对组播设备的信令机制。 主机通过 IGMP 加入或者离开组播组；组播设备通过 IGMP 了解下游网段是否存在组播组成员。
(S,G)	属于组播路由表项，S 表示组播源，G 表示组播组。 源地址为 S、组地址为 G 的组播报文，到达组播设备后，从(S,G)表项中的下游接口转发出去。 通常，将源地址为 S，组地址为 G 的组播报文表示为(S,G)报文。
(*G)	属于 PIM 路由表项，*表示任意源，G 表示组播组。 (*G)表项适用于所有组地址为 G 的组播报文。不论是哪个组播源发出的，只要是发往组播组 G 的组播报文，都应该从(*G)表项中的下游接口转发出去。

缩略语

缩略语	英文全称	中文全称
ASM	Any-Source Multicast	任意源组播

缩略语	英文全称	中文全称
IGMP	Internet Group Management Protocol	因特网组管理协议
SSM	Source-Specific Multicast	指定源组播

3 PIM

关于本章

- 3.1 介绍
- 3.2 参考标准和协议
- 3.3 可获得性
- 3.4 原理描述
- 3.5 应用
- 3.6 术语与缩略语

3.1 介绍

定义

PIM（Protocol Independent Multicast）称为协议无关组播，作为一种组播路由解决方案，主要用于将网络中的组播数据流引入到有组播数据请求的组成员，从而实现组播数据流的转发。

目前在实际网络中应用较为广泛的实现方式主要有以下三种。

- PIM-DM（Protocol Independent Multicast Dense Mode）：协议无关组播—密集模式。
- PIM-SM（Protocol Independent Multicast Sparse Mode）：协议无关组播—稀疏模式。
- PIM-SSM（Protocol Independent Multicast Source-Specific Multicast）：协议无关组播—指定源组播。

目的

组播源向组播组地址发出组播报文，经过中间网络到达组播组所有成员。为使中间网络能够实现组播报文的复制和转发，必须为网络中的路由器配置组播路由协议。PIM 协议是网络中实现组播报文的复制和转发的一个重要协议。

3.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RFC4601	Protocol Independent Multicast - Sparse Mode（PIM-SM）	-
RFC5059	Bootstrap Router（BSR） Mechanism for PIM	-
RFC3973	Protocol Independent Multicast - Dense Mode protocol	-
RFC4607	Source-Specific Multicast for IP	-

3.3 可获得性

涉及网元

需要其他网元也支持 PIM。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

AR1200 支持该特性的版本如下：

产品	支持版本
AR1200	V200R001C00

3.4 原理描述

3.4.1 基本概念

PIM 路由器

支持 PIM 协议的组播路由器称为 PIM 路由器。使能了 PIM 协议的接口称为 PIM 接口。

PIM 域

由 PIM 路由器所组成的网络称为 PIM 网络。

通过在组播设备接口上设置“边界”，可以将一个大的 PIM 网络划分多个 PIM 域。“边界”可以拒绝特定组播报文通过，或者限制 PIM 控制消息的传输。

组播分发树

在 PIM 组播域中，以组播组为单位建立一点到多点的组播转发路径。由于组播转发路径呈现树型结构，也称为组播分发树（MDT，Multicast Distribution Tree）。

- 以组播源为根，组播组成员为叶子的组播分发树称为 SPT（Shortest Path Tree）。SPT 同时适用于 PIM-DM 和 PIM-SM。
- 以 RP（Rendezvous Point）为根，组播组成员为叶子的组播分发树称为 RPT（RP Tree）。RPT 仅适用于 PIM-SM。

组播分发树的特点：

- 无论网络中的组成员有多少，每条链路上相同的组播数据最多只有一份。
- 被传递的组播数据在距离组播源尽可能远的分叉路口才开始复制和分发。

叶子路由器

与用户主机相连的 PIM 路由器称为叶子路由器。

组播源 DR

与组播源直接相连且负责向 RP 发送注册报文的 PIM 路由器。

接收者 DR

与组播组成员（通常为接收者主机）直接相连且负责向该组成员转发组播数据的 PIM 路由器。

中间路由器

组播转发路径上，第一跳路由器与最后一跳路由器之间的 PIM 路由器。

3.4.2 PIM-SM

PIM-SM（Protocol Independent Multicast-Sparse Mode）称为协议无关组播-稀疏模式，主要采用接收者主动加入的方式建立组播转发树，适用于网络中的组成员相对比较稀疏，分布广泛的大型网络。

基本原理

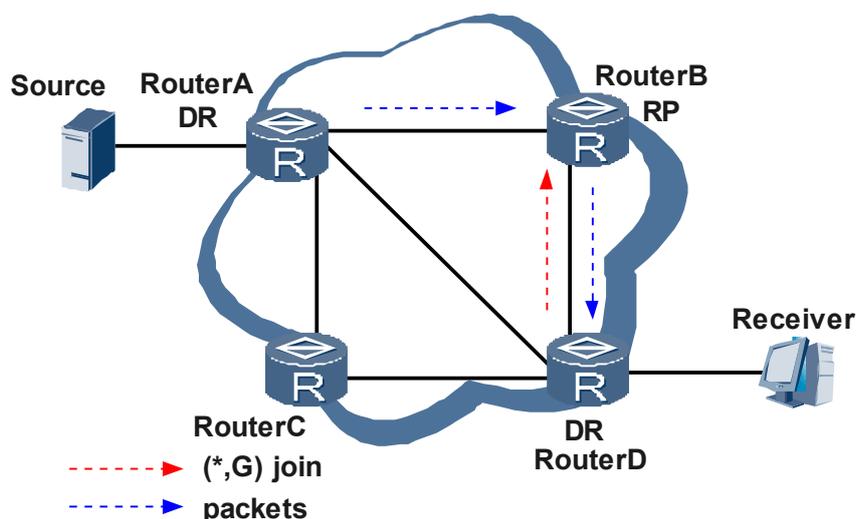
PIM-SM 转发组播数据的关键是建立 RPT（Rendezvous Point Tree，汇聚点树也称共享树）和 SPT（Shortest Path Tree，最短路径树）。

RPT 建立原理

PIM-SM RPT 是一棵以 RP 为根，以存在组成员关系的路由器为叶子的组播分发树。

汇聚点 RP 为网络中一台重要的 PIM 路由器，用于处理组播源 DR 注册信息及组成员加入请求，网络中的所有 PIM 路由器都知道 RP 的位置，RP 类似于一个供求信息汇聚中心。

图 3-1 RPT 建立原理图



建立 RPT 的过程即建立组播数据转发路径的过程。如图 3-1 所示，RPT 的建立及数据转发过程如下：

- 当网络中出现活跃的组播源（组播源向某组播组 G 发送第一个组播数据）时，组播源端 DR 将组播数据封装在 Register 消息中单播发往 RP，在 RP 上创建 (S, G) 表项，注册源信息。
- 当网络中出现组成员（用户主机通过 IGMP 加入某组播组 G）时，接收者 DR 向 RP 发送 Join 消息，在通向 RP 的路径上逐跳创建 (*, G) 表项，生成以一棵以 RP 为根的 RPT。
- 当网络中同时出现组成员和向该组发送数据的组播源时，以 RP 为中转站，组播数据先被封装在 Register 消息中单播发往 RP，再沿 RPT 到达组成员。

RPT 实现了组播数据按需转发的目的，减少无需求数据对网络带宽的占用。

DR (Designated Router) 的分类：

- 在连接组播源的共享网段，由 DR 负责向 RP 发送 Register 注册消息。与组播源相连的 DR 称为组播源端 DR。
- 在连接组成员的共享网段，由 DR 负责向 RP 发送 Join 加入消息。与组成员相连的 DR 称为接收者 DR。

说明

为了减轻 RPT 的转发负担、提高组播数据转发效率，PIM-SM 允许进行 SPT 切换。即建立一条从数据源直接到接收者的转发链路，组播源可以沿 SPT 将组播数据转发到接收者。

邻居发现

PIM 路由器在每个使能了 PIM 的接口上，对外发送 Hello 消息。封装 Hello 消息的组播报文的目的地址是 224.0.0.13（表示同一网段中所有 PIM 路由器）、源地址为接口的 IP 地址、TTL 数值为 1。

Hello 消息的作用：发现邻居、协调各项协议参数、维持邻居关系。

- 发现 PIM 邻居

同一网段中的 PIM 路由器都必须接收目的地址为 224.0.0.13 的组播报文。这样在收到 Hello 报文以后，直接相连的组播路由器之间，就可以彼此知道自己的邻居信息。只有在路由器接收到来自邻居的 Hello 消息后，才会接收其他的 PIM 控制消息或组播报文，从而创建组播路由表项，维护组播分发树。

- 协调各项协议参数

Hello 消息中携带多项协议参数，介绍如下：

- DR_Priority: 表示各路由器接口竞选 DR 的优先级，优先级越高越容易获胜。适用于 PIM-SM。
- Holdtime: 表示保持邻居为可达状态的超时时间。
- LAN_Delay: 表示共享网段内传输 Prune 消息的延迟时间。
- Neighbor-Tracking: 表示邻居跟踪功能。
- Override-Interval: 表示 Hello 消息中携带的否决剪枝的时间间隔。

- 维持邻居关系

PIM 路由器之间周期性地发送 Hello 消息。如果 Holdtime 超时还没有收到该 PIM 邻居发出的新的 Hello 报文，则认为该邻居不可达，将其从邻居列表中清除。

PIM 邻居的变化将导致网络中组播拓扑的变化。如果组播分发树上的某上游邻居或下游邻居不可达，将导致组播路由重新收敛，组播分发树迁移。

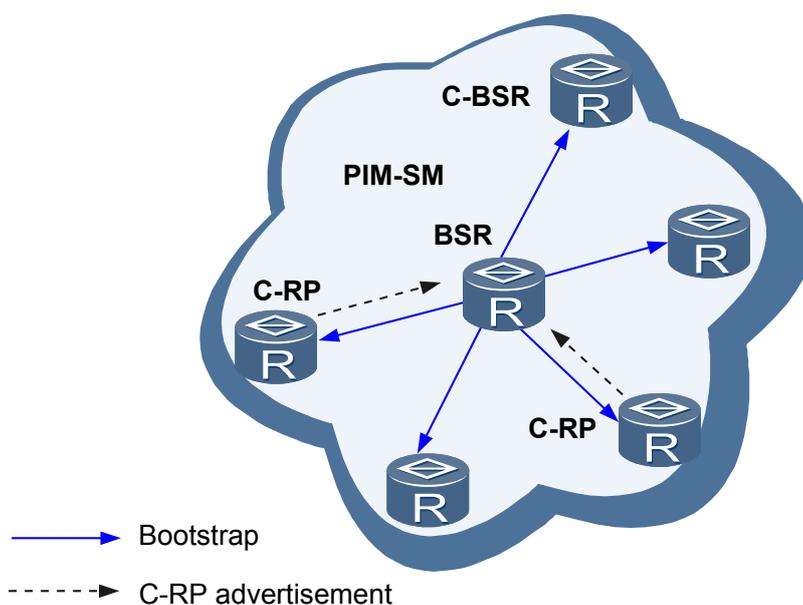
RP 发现机制

- RP 分类

一个 RP 可以同时为多个组播组服务，但一个组播组只能对应一个 RP。RP 是 PIM-SM 网络的核心，网络中的路由器必须知道 RP 的地址，目前可以通过以下方式获取 RP 的地址：

- 静态 RP：用户通过配置命令在网络中的所有路由器上配置相同的 RP 地址；
- 动态 RP：在 PIM 域内选择几台 PIM 路由器，配置 C-RP（Candidate-RP）。从 C-RP 中竞选产生 RP。

图 3-2 动态 RP 竞选机制原理图



如图 3-2 所示：

1. 使用动态 RP，必须同时配置 C-BSR（Candidate-BootStrap Router），由 C-BSR 竞选产生 BSR。网络中的所有路由器都知道 BSR 的地址。
2. C-RP 向 BSR 发送 Advertisement 消息，消息中携带 C-RP 地址、服务的组范围和 C-RP 优先级。
3. BSR 将这些信息汇总为 RP-Set，封装在 Bootstrap 消息中，发布给全网的每一台 PIM-SM 路由器。
4. 各路由器根据 RP-Set，使用相同的规则进行计算和比较，从多个针对特定组的 C-RP 中竞选出该组 RP。规则如下：
 - C-RP 接口地址掩码最长者获胜。
 - C-RP 优先级较高者获胜（优先级数值越大优先级越低）。
 - 如果优先级相同，则执行 Hash 函数，计算结果较大者获胜。
 - 如果以上都相同，则 C-RP 地址较大者获胜。
5. 由于所有路由器使用相同的 RP-Set 和竞选规则，所以得到的“组播组—RP”对应关系也相同。路由器将“组播组—RP”对应关系保存下来，指导后续的组播操作。



说明

为了和使用 Auto-RP 的厂商进行互通，支持作为其 Auto-RP 的 Listening 端。

● Anycast RP

在传统的 PIM-SM 域中，每个组播组只能映射到一个 RP。但当网络负载较大或者流量过于集中时，可能导致 RP 的压力过大、RP 失效后路由收敛较慢、组播转发路径非最优等问题。

针对上述问题实现 Anycast RP，目前有两种配置方式：

- 采用 MSDP 协议：在同一 PIM-SM 域内设置多个具有相同地址的 RP，并在这些 RP 之间通过建立 MSDP（Multicast Source Discovery Protocol）对等体的方式共享组播数据源信息。
- 采用 PIM 协议：在同一个 PIM-SM 域内设置多个具有相同地址的 RP，同时在这些 RP 所在的设备上配置全网唯一标识该 RP 的本地地址，用于这些设备之间相互建立无连接的对等体，对等体之间以注册报文的方式共享组播源信息。



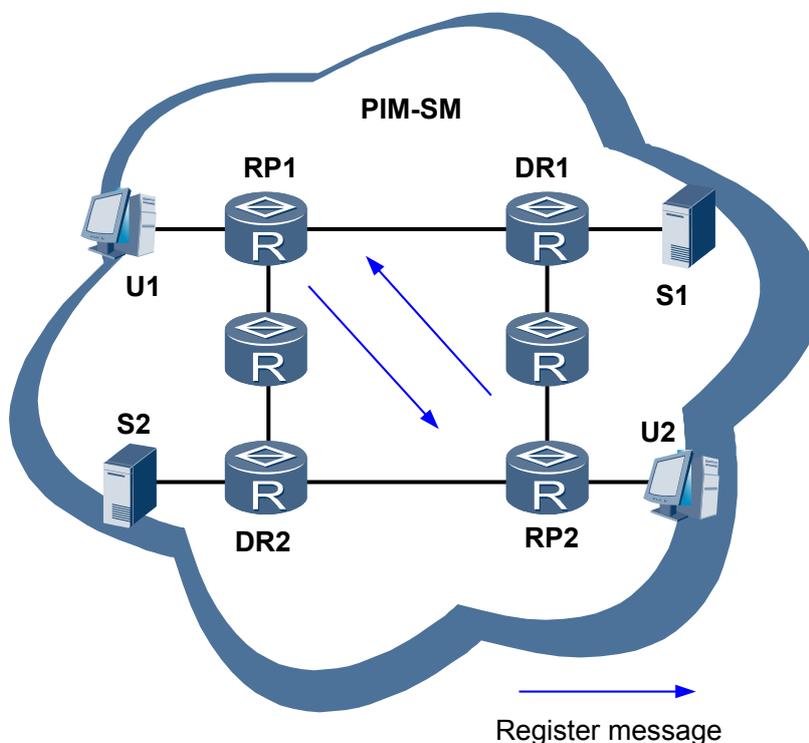
说明

在同一 PIM-SM 域内，不支持同时使用 MSDP 和 PIM 协议两种方式配置同一个 Anycast RP；采用 PIM 协议实现 Anycast RP 支持通过命令行设置有选择地将 MSDP 获得的域外数据源信息，通知给域内的其它对等体（可配置）。

通过以上两种方式之一实现 Anycast RP，接收者和组播源分别选择距离自己最近的 RP 进行 RPT 的创建，当接收者 DR 接收到组播数据后自行决定是否发起 SPT 切换。从而实现 RP 路径最优及负荷分担。

对于采用 MSDP 实现 Anycast RP 的实现原理，请参见 [4.4.2 MSDP 实现 Anycast RP](#)，下面重点介绍采用 PIM 协议实现 Anycast RP 的原理。

图 3-3 采用 PIM 协议实现 Anycast RP 典型组网图



如图 3-3 所示，在 PIM-SM 域内，组播源 S1 和 S2 向组播组 G 发送组播数据，U1 和 U2 是组播组 G 的成员。在 PIM-SM 域内应用 PIM 协议实现 Anycast RP 的配置方法如下：

- 配置 RP1 和 RP2，使用相同的 IP 地址（使用 Loopback 接口，假设为 10.10.10.10）。
- 在 RP1 和 RP2 两个设备间配置无连接的对等体关系（使用网络中唯一标识的 IP 地址，假设 RP1 的 IP 地址为 1.1.1.1，RP2 的 IP 地址为 2.2.2.2）。

采用 PIM 协议实现 Anycast RP 的实现过程如下：

1. 接收者选择距离最近的 RP 发送加入消息构建 RPT 树。
 - U1 加入以 RP1 为根的 RPT，在 RP1 上创建 (*, G)。
 - U2 加入以 RP2 为根的 RPT，在 RP2 上创建 (*, G)。
2. 组播源选择距离最近的 RP 进行注册。
 - DR1 向 RP1 发送注册消息，在 RP1 上创建 (S1, G)。从 S1 发来的组播数据沿 RPT 到达 U1。
 - DR2 向 RP2 发送注册消息，在 RP2 上创建 (S2, G)。从 S2 发来的组播数据沿 RPT 到达 U2。
3. RP 收到源 DR 发过来的注册报文，重新封装成注册报文转给自己的对等体，共享组播源信息。
 - RP1 收到源 DR1 发过来的 (S1, G) 注册报文后，将该注册报文的源地址和目的地址替换为 1.1.1.1 和 2.2.2.2，重新封装后发送到 RP2。RP2 收到该注册报文后检查发现该注册报文来自对等体 1.1.1.1，不再转发给其它对等体，只是处理该注册报文。
 - RP2 收到源 DR2 发过来的 (S2, G) 注册报文后，将该注册报文的源地址和目的地址替换为 2.2.2.2 和 1.1.1.1，重新封装后发送到 RP1。RP1 收到该注册报文后检查发现该注册报文来自对等体 2.2.2.2，不再转发给其它对等体，只是处理该注册报文。
4. RP 加入以源端 DR 为根的 SPT，将组播数据引下来。
 - RP1 向 S2 发送加入消息。从 S2 发来的组播数据先沿 SPT 到达 RP1，再沿 RPT 到达 U1。
 - RP2 向 S1 发送加入消息。从 S1 发来的组播数据先沿 SPT 到达 RP2，再沿 RPT 到达 U2。
5. 接收者 DR 接收到组播数据后，自行决定是否发起 SPT 切换。

BootStrap Router 机制

- BSR 竞选机制

BSR 称为自举路由器 (Bootstrap Router)，负责收集并发布网络中的 C-RP 信息，确保网络中的所有路由器都知道 RP 的位置。

BSR 是从众多 C-BSR 中竞争产生的。最初，每个 C-BSR 都认为自己是 BSR，向全网发送 Bootstrap 消息。Bootstrap 消息中携带 C-BSR 地址、C-BSR 的优先级。每一台路由器都收到所有 C-BSR 发出的 Bootstrap 消息，通过比较这些 C-BSR 信息，竞选产生 BSR。竞选规则如下：

- 优先级较高者获胜（优先级数值越大优先级越高）。
- 如果优先级相同，IP 地址较大者获胜。

由于所有路由器使用相同的竞选规则，所以得到的当选 BSR 也相同。

设备发送 BSR 报文时需要携带网络中所有的 C-RP 信息。当网络中存在大量 C-RP，BSR 报文携带这些 C-RP 信息时，会导致报文长度过大，超过接口 MTU 值，

造成设备无法正确处理 BSR 报文，从而无法选举出 RP 信息，组播业务无法正常传输。此时可以使用 BSR 报文分片功能对 BSR 报文进行分片处理。

推荐使用 BSR 报文分片功能，可以解决 IP 分片时，分片信息丢失而导致所有分片不可用的问题。

- BSR 管理域

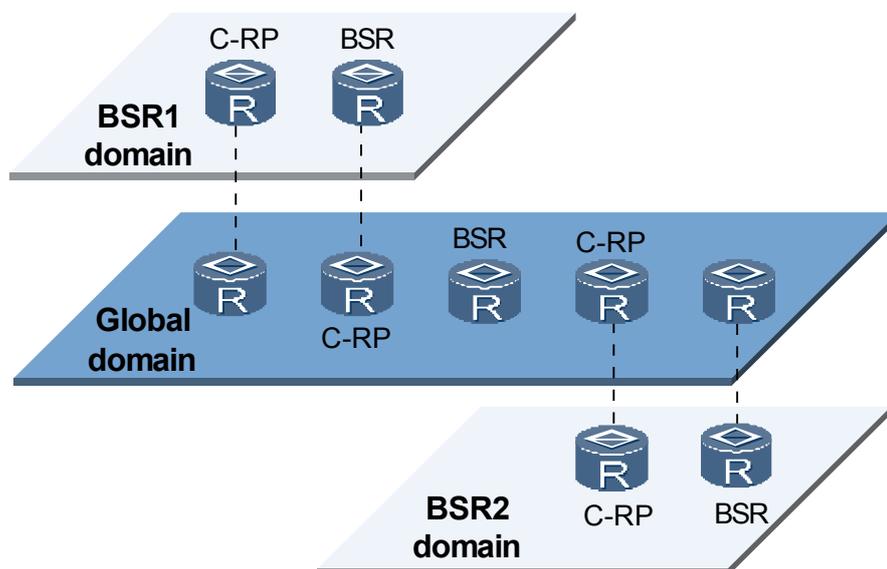
为了实现网络管理精细化，可以选择将一个 PIM-SM 网络划分为多个 BSR 管理域和一个 Global 域。这样一方面可以有效地分担单一 BSR 的管理压力，另一方面可以使用私有组地址为特定区域的用户提供专门服务。

每个 BSR 管理域中维护一个 BSR，为某一特定地址范围的组播组服务。Global 域中维护一个 BSR，为所有剩余的组播组服务。

下文将从地域空间、组地址范围、组播功能三个角度分析 BSR 管理域和 Global 域的关系。

- 地域空间

图 3-4 BSR 管理域_地域空间

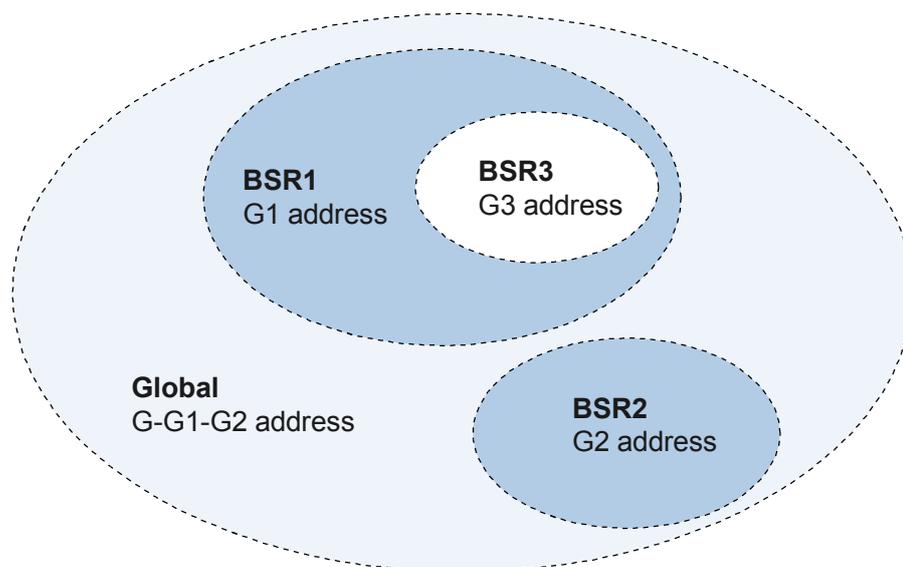


如图 3-4 所示，对于有相同组地址的不同管理域，各 BSR 管理域所包含的路由器互不相同，同一路由器不能从属于多个 BSR 管理域。各 BSR 管理域在地域上相互独立，且相互隔离。BSR 管理域是针对特定地址范围的组播组的管理区域，属于此范围的组播报文只能在本管理域内传播，无法通过 BSR 管理域边界。

Global 域包含 PIM-SM 网络内的全部路由器。不属于任意 BSR 管理域的组播报文，可以在整个 PIM 网络范围内传播。

- 组地址范围

图 3-5 BSR 管理域_地址范围



每个 BSR 管理域为特定地址范围的组播组提供服务，不同的 BSR 管理域服务的组播组范围可以重叠。该组播地址只在本 BSR 管理域内有效，相当于私有组地址。如图 3-5 所示，BSR1 域和 BSR3 域对应的组地址范围出现重叠。

不属于任何 BSR 管理域的组播组，一律属于 Global 域的服务范围。即 Global 域组地址范围是 G-G1-G2。

- 组播功能

如图 3-4 所示，Global 域和每个 BSR 管理域都包含针对自己域的 C-RP 和 BSR 设备，这些设备在行使相应功能时，仅在本域内有效。即 BSR 机制和 RP 竞选在各管理域之间是隔离的。

每个 BSR 管理域都有自己的边界，该管理域的组播信息（C-RP 宣告消息、BSR 自举消息等）不能跨越域传播。同时 Global 域的组播信息可以在整个 Global 域内传递，可以穿越任意 BSR 管理域。

Assert 的基本原理

当满足如下条件时，说明网段上还存在着其他的组播转发者。路由器执行 Assert：

- 该组播报文不能通过 RPF 检查。
- 接收到组播报文的接口是本路由器上（S，G）表项中的一个下游接口。

路由器从该下游接口发送 Assert 消息。同时，该下游接口也接收到了来自该网段上其他组播转发者的 Assert 消息。PIM Assert 消息的目的地址为 224.0.0.13，源地址为下游接口地址，TTL 为 1。Assert 消息中携带：该 PIM 路由器到组播源或 RP 的开销、所采用的单播路由协议的优先级、组播组地址 G。

路由器将自身条件与对方报文中携带的信息进行比较，称为 Assert 竞选。规则如下：

- 单播路由协议优先级较高者获胜。
- 如果优先级相同，则到组播源的开销较小者获胜。
- 如果以上都相同，则下游接口 IP 地址最大者获胜。

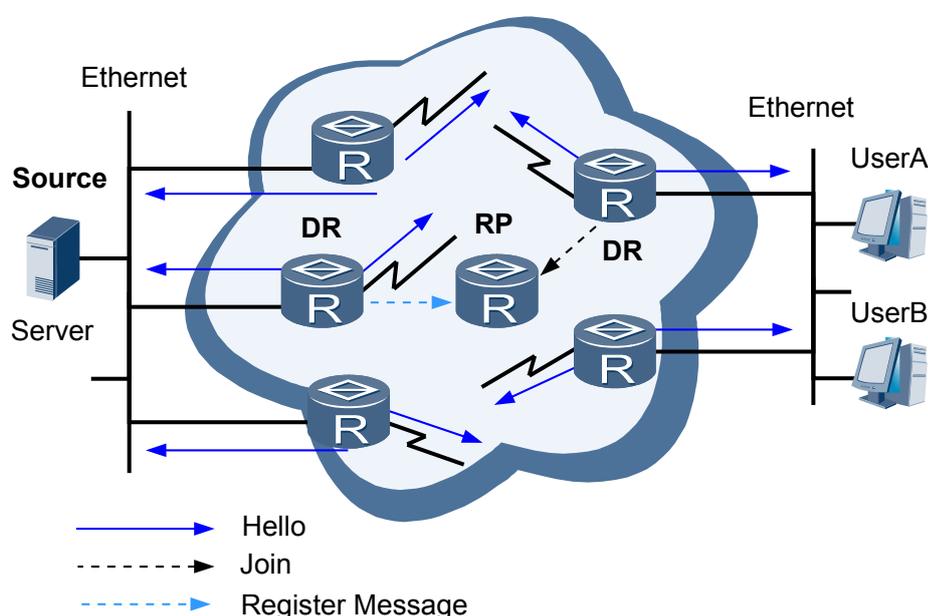
根据 Assert 竞选结果，路由器将执行不同的操作：

- 如果获胜，则该下游接口保持转发状态，路由器负责后续在该网段上的（S，G）转发，该下游接口称为 Assert winner。
- 如果落败，则禁止该下游接口转发组播报文，将其从（S，G）表项下游接口列表中删除。该下游接口称为 Assert loser。

Assert 竞选结束后，该网段上只存在一个有下游接口的上游路由器，只传输一份组播报文。Assert winner 周期性发送 Assert 消息，维持 Assert loser 的状态。若定时器超时后，Assert loser 仍没有收到 Assert winner 的 Assert 消息，则重新添加下游接口转发组播数据。

PIM DR 竞选的基本原理

图 3-6 DR 竞选示意图



如图 3-6 所示，DR（Designated Router）应用在 PIM-SM 网络中的如下两个位置：

- 在连接组播源的共享网段，由 DR 负责向 RP 发送 Register 注册消息。与组播源相连的 DR 称为源端 DR。
- 在连接组成员的共享网段，由 DR 负责向 RP 发送 Join 加入消息。与组成员相连的 DR 称为接收者 DR。

在组播源或组成员所在的网段，通常同时连接着多台 PIM 路由器。这些 PIM 路由器之间通过交互 Hello 消息成为 PIM 邻居，Hello 消息中携带 DR 优先级和该网段接口地址。路由器将自身条件与对方报文中携带的信息进行比较，称为 DR 竞选。规则如下：

- DR 优先级较高者获胜（网段中所有路由器都支持 DR 优先级）。
- 如果 DR 优先级相同或该网段存在至少一台路由器不支持在 Hello 报文中携带 DR 优先级，则 IP 地址较大者获胜。

如果当前 DR 出现故障，将导致 PIM 邻居关系超时，其他 PIM 邻居之间会触发新一轮的 DR 竞选过程。

PIM DR 切换延迟的基本原理

多台路由器处于同一共享网段时，通过一定的 DR 选举策略（通常在不配置 DR 优先级的情况下，共享网段内 IP 地址高的路由器将被选举为 DR）使其中一台路由器被选举为接收者 DR，在接收端负责共享网段内的组播数据转发。

缺省情况下，接口由 DR 变为非 DR 时，路由器会立即停止使用此接口转发数据。如果此时新 DR 的组播数据还未到达，那么将会出现短暂的组播数据断流。

在一个接口配置 PIM DR 切换延迟后，当某个使能 PIM-SM 的接口由于收到一个新邻居的 Hello 报文导致该接口由 DR 变为非 DR 时，该接口在延迟时间超时前仍然具有部分 DR 功能，并继续转发组播数据。

如果在 DR 延迟期间收到新 DR 转发过来的数据，处于 DR 延迟的路由器将会立刻停止转发数据，从而保证不会出现重复数据流的现象。此时在共享网段上若收到新的 IGMP 加入，处于 DR 切换延迟状态的旧 DR 将不会为其向上游发送 PIM 加入报文，而由新 DR 处理。

说明

在 DR 切换延迟期间，如果新 DR 收到原 DR 发送的组播数据，则会触发 Assert 竞选。

PIM Silent 的基本原理

若路由器直连用户主机的接口上使能了 PIM 协议，就可以在该接口上建立 PIM 邻居，处理各类 PIM 协议报文。但此配置同时存在着安全隐患：当恶意主机模拟发送 PIM Hello 报文时，有可能导致路由器瘫痪。

为了避免上述情况，可以在路由器直连用户主机的接口上配置 PIM Silent，用来禁止该接口接收和转发任何 PIM 协议报文。同时，此接口上的 IGMP 功能不受影响。

3.4.3 PIM-SSM

PIM 支持 ASM（Any-Source Multicast）模型和 SSM（Source-Specific Multicast）两种模型，本节介绍 SSM 模型。

SSM 模型是借助 PIM-SM 的部分技术和 IGMPv3/MLDv2 来实现的，其建立组播转发树的过程与 PIM-SM 创建 SPT 树的过程相似，即接收者 DR 在知道组播数据源的具体位置后，直接向组播数据源发送 Join 消息，将组播数据流发送到接收者。

缺省情况下，SSM 组播组地址的范围为 232.0.0.0 ~ 232.255.255.255。当用户加入的组播组属于 SSM 组地址范围内，适用于 SSM 模型；当用户加入的组播组不属于 SSM 组地址范围，则适用 ASM 模型，ASM 模型原理即 PIM-SM 原理。

SSM 的特点是网络用户能够预先知道组播源的具体位置。因此用户在加入组播组时，可以明确指定从哪些源接收信息。组成员端 DR 了解到用户的需求后，直接向组播源的方向发送 Join 消息。Join 消息逐跳向上传输，在源与组成员之间建立 SPT。

SSM 只使用了 PIM-SM 的部分技术：无需维护 RP、无需构建 RPT、无需注册组播源，可以直接在源与组成员之间建立 SPT。

说明

可以通过静态组加入和 SSM-Mapping 建立组播分发树，构建 SPT。

在 SSM 中，DR 仅在与组成员相连的共享网段上有效。由 DR 向组播源的方向发送加入消息，逐跳创建 (S, G) 表项，构建 SPT。

PIM-SSM 支持 PIM DR 切换延迟、PIM Silent 特性。

3.4.4 PIM-DM

应用场景

PIM-DM (Protocol Independent Multicast Dense Mode) 协议无关组播—密集模式，主要采用扩散-剪枝的方式转发组播数据流，对于组播组成员稀疏的网络会产生大量剪枝报文，而对规模较大的网络扩散-剪枝周期会很长。所以适合规模较小、组播组成员相对比较密集的网络。

基本原理

PIM-DM 假设网络中的组成员分布非常稠密，每个网段都可能存在组成员。其设计思路是：首先将数据报文扩散到各个网段，然后再裁剪掉不存在组成员的网段。通过周期性的“扩散—剪枝”，构建并维护一棵连接组播源和组成员的单向无环 SPT (Source Specific Shortest Path Tree)。

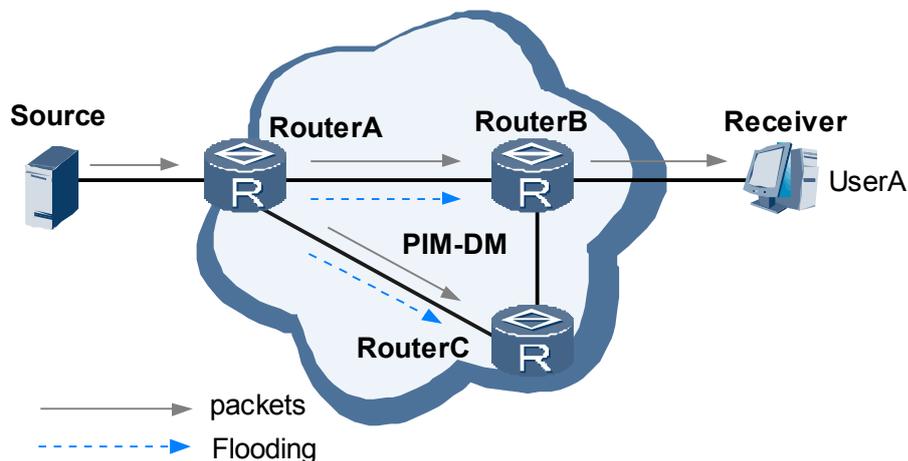
邻居发现

与 PIM SM 协议相同，请参见 3.4.2 PIM-SM。

扩散 Flooding

如图 3-7 所示，组播数据源 (Source) 把数据发送到 RouterA，RouterA 把数据报文发送给它的所有的邻居 (除了给它发送数据的邻居，如 RouterB 和 RouterC，不会把数据发给 RouterA)。这时 RouterB 与 RouterC 也会相互转发数据报文，但 DM 协议采用 RPF 检查机制可以保证数据只从一个方向接收。最后数据被扩散到连接接收者的 RouterB，由 RouterB 把数据发送给它的接收者 UserA。

图 3-7 PIM-DM 扩散示意图

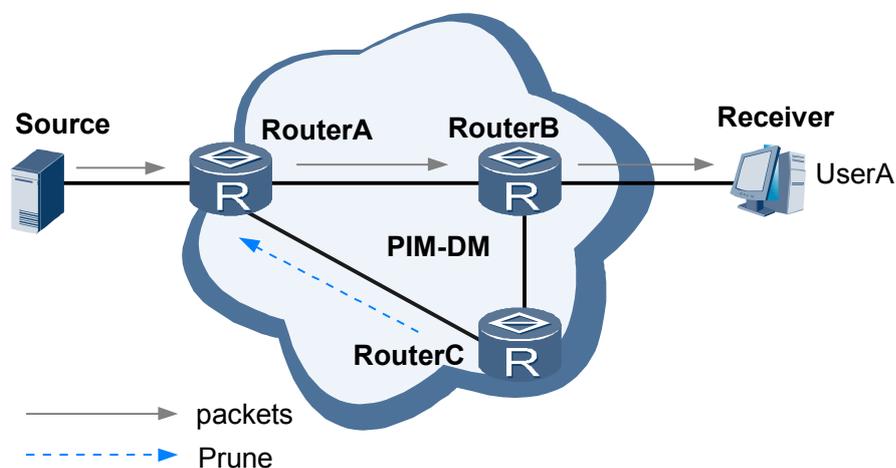


剪枝 Prune

如图 3-8 所示，由于 RouterC 没有接收者，不需要数据，则向上游 RouterA 发送 Prune 消息，通知 RouterA 不必再向该下游网段转发数据。

RouterA 收到 Prune 消息后，停止该下游接口转发，由于 RouterA 上还存在其他处于转发状态的下游接口，剪枝过程停止。后续到达的报文只向 RouterB 转发。从而实现了一棵连接组播源和组成员 UserA 的单向无环最短路径树。

图 3-8 PIM-DM 剪枝示意图

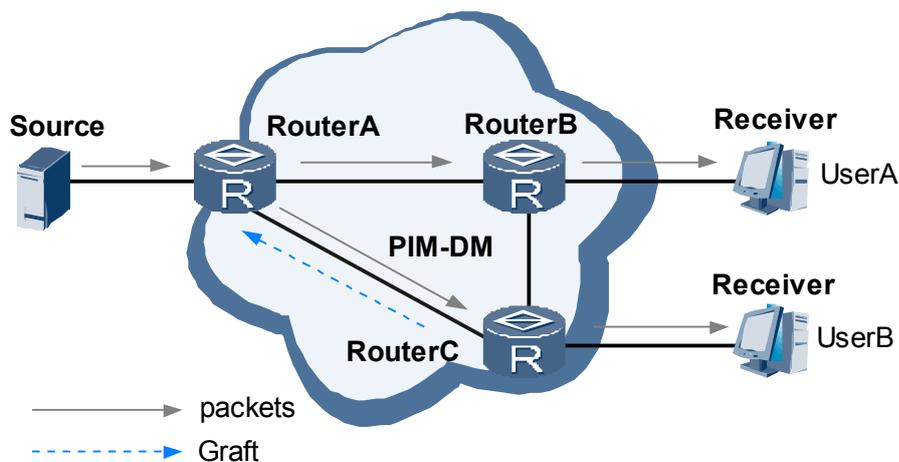


嫁接 Graft

如图 3-9 所示，如果 RouterC 收到接收者 UserB 的 IGMP Report 报文，请求转发组播源数据，即 RouterC 具有转发数据需求。为了避免周期性扩散-剪枝的时间延迟，PIM DM 用嫁接 Graft 方式实现数据的快速转发。

RouterC 发送 Graft 嫁接消息，请求上游 RouterA 恢复对应出接口的转发。RouterA 收到 Graft 消息后，将连接 RouterC 的出接口恢复转发，组播报文由该下游接口到达 RouterC。

图 3-9 PIM-DM 嫁接示意图

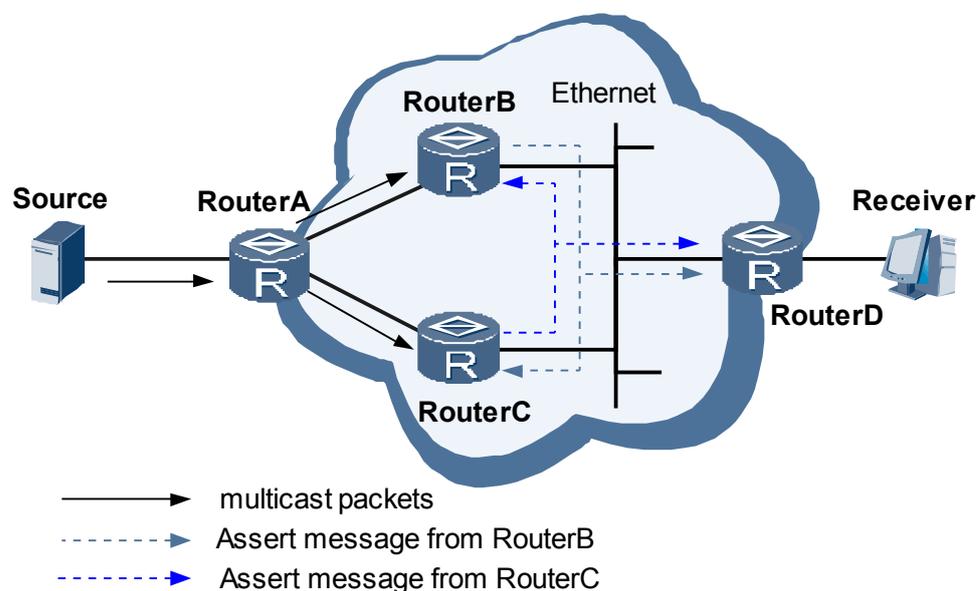


断言 Assert

如图 3-10 所示，如果 RouterB 和 RouterC 都能够接收到组播源 S 发出的组播报文，并且均能通过 RPF 检查，创建 (S, G) 表项。RouterB、RouterC 的下游接口连接在同一网段，RouterB 和 RouterC 就会同时向该网段发送组播数据。Assert 可以保证一个网段只能存在一个组播数据转发者。断言过程如下：

1. RouterB 从下游接口接收到 RouterC 发来的组播报文，RPF 检查失败，报文被丢弃。同时，向该网段发送 Assert 消息。
2. RouterC 将自身的路由信息与对方报文中携带的路由信息进行比较，由于自身到组播源的开销较大而落败。于是禁止该下游接口转发组播报文，将其从 (S, G) 表项的下游接口列表中删除。
3. RouterC 从该网段接收到 RouterB 发来的组播报文，RPF 检查失败，报文被丢弃。Assert 过程结束。

图 3-10 PIM-DM 断言示意图



状态刷新

如图 3-10 所示，若 RouterA 对 RouterC 所在网段处于剪枝状态，RouterA 对 RouterC 的接口会维护一个“剪枝定时器”，当剪枝定时器超时，RouterA 就会恢复对不需要数据的 RouterC 的数据转发，这样会导致不必要的网络资源浪费。

PIM DM 协议采用状态刷新特性解决此问题：离组播源最近的第一跳 RouterA 周期性触发 State Refresh 消息。State Refresh 消息在全网扩散，刷新所有设备上的剪枝定时器状态。

PIM Silent

与 PIM SM 协议相同，请参见 3.4.2 PIM-SM。

3.4.5 协议比较

表 3-1 协议比较

协议	特点
PIM-SM	协议无关组播—稀疏模式。采用接收者主动加入的方式建立组播转发树，适合网络中的组成员相对比较稀疏，分布广泛的大型网络。 需维护 RP、构建 RPT、注册组播源。
PIM-SSM	与 PIM-SM 类似，但无需维护 RP、无需构建 RPT、无需注册组播源。只要存在到数据源的路由，即可以直接在源与组成员之间建立 SPT。在跨域的组播数据流转发方面有很大的优势。
PIM-DM	协议无关组播—密集模式。采用扩散-剪枝的方式转发组播数据流，适合规模较小、组播组成员相对比较密集的局域网。 没有 RP 的概念，通过周期性“扩散-剪枝”维护一棵连接组播源和组成员的单向无环 SPT。

3.4.6 PIM 安全性

源地址过滤

适用于 PIM-DM 和 PIM-SM。

此功能用来实现路由器对接收的组播数据报文根据源或源组进行过滤。通过配置 ACL，路由器可以选择只转发源地址属于过滤规则范围内的组播报文，或转发源地址和组地址都属于过滤规则范围内的组播报文。

配置合法的 BSR 地址范围

适用于 PIM-SM。

此功能用来限定合法 BSR 地址范围，使路由器丢弃来自该地址范围之外的 BSR 报文，从而防止 BSR 欺骗。

配置合法的 C-RP 地址范围

适用于 PIM-SM。

此功能用来限定合法的 C-RP 地址范围及其服务的组播组地址范围，使 BSR 丢弃来自该地址范围之外的 C-RP 消息，从而防止 C-RP 欺骗。

Register 报文过滤

适用于 PIM-SM。

此功能用来实现 RP 过滤由组播源端 DR 发送的注册报文，根据报文过滤规则接受或拒绝和规则匹配的注册报文，从而防止非法注册报文攻击。

PIM 邻居过滤

适用于 PIM-DM 和 PIM-SM。

为了防止与其它未知设备建立 PIM 邻居，阻止未知设备成为 DR，需要对不期望的邻居进行过滤。配置此功能后，接口只与符合过滤规则的地址建立邻居关系，删除不符合过滤规则的邻居。

Join 信息过滤

适用于 PIM-SM。

接口上接收的 Join/Prune 消息中包含 Join 信息和 Prune 信息。此功能可过滤 Join 信息，路由器根据符合过滤规则的 Join 信息建立 PIM 表项，从而防止非法用户加入。

PIM 邻居检查

适用于 PIM-SM。

默认情况下，接收或发送 Join/Prune 消息和 Assert 消息时，不检查该消息是否来自 PIM 邻居或发送给 PIM 邻居。

如果需要配置 PIM 邻居检查功能，建议在与用户相连的设备上配置，在网络内部设备上不推荐使用此功能。接收或发送 Join/Prune 消息和 Assert 消息时，检查该消息是否来自 PIM 邻居或发送给 PIM 邻居，如果是则处理，否则丢弃。

PIM Silent

若路由器直连用户主机的接口上使能了 PIM 协议，就可以在该接口上建立 PIM 邻居，处理各类 PIM 协议报文。但此配置同时存在着安全隐患：当恶意主机模拟发送 PIM Hello 报文时，有可能导致路由器瘫痪。

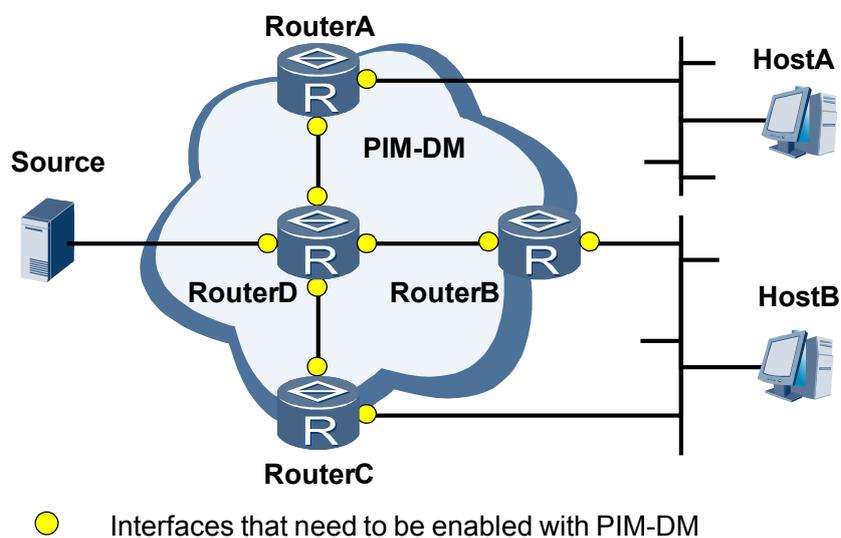
为了避免上述情况，可以在路由器直连用户主机的接口上配置 PIM Silent，用来禁止该接口接收和转发任何 PIM 协议报文。同时，此接口上的 IGMP 功能不受影响。

3.5 应用

3.5.1 PIM 典型应用

小规模试验网组播

图 3-11 PIM-DM 典型组网图



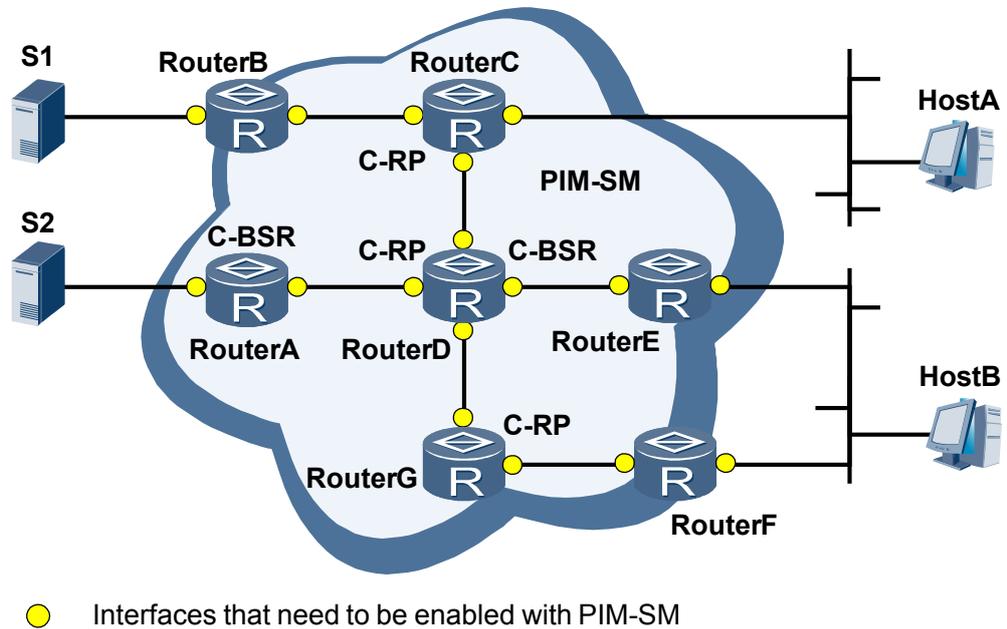
PIM-DM 称为协议独立组播—密集模式。适合规模较小、组播组成员相对比较密集的局域网。

如图 3-11 所示：

- 配置单播路由协议，确保任意网段路由可达。
- 在接收者主机上启用 IGMP 协议。
- 在与接收者主机处于同一网段的路由器接口上启用 IGMP 协议。
- 在所有路由器接口上启用 PIM-DM 协议。

大规模稀疏组播网络

图 3-12 PIM-SM 域典型组网图



PIM-SM 适合网络中的组成员相对比较稀疏，分布广泛的大型网络。

如图 3-12 所示：

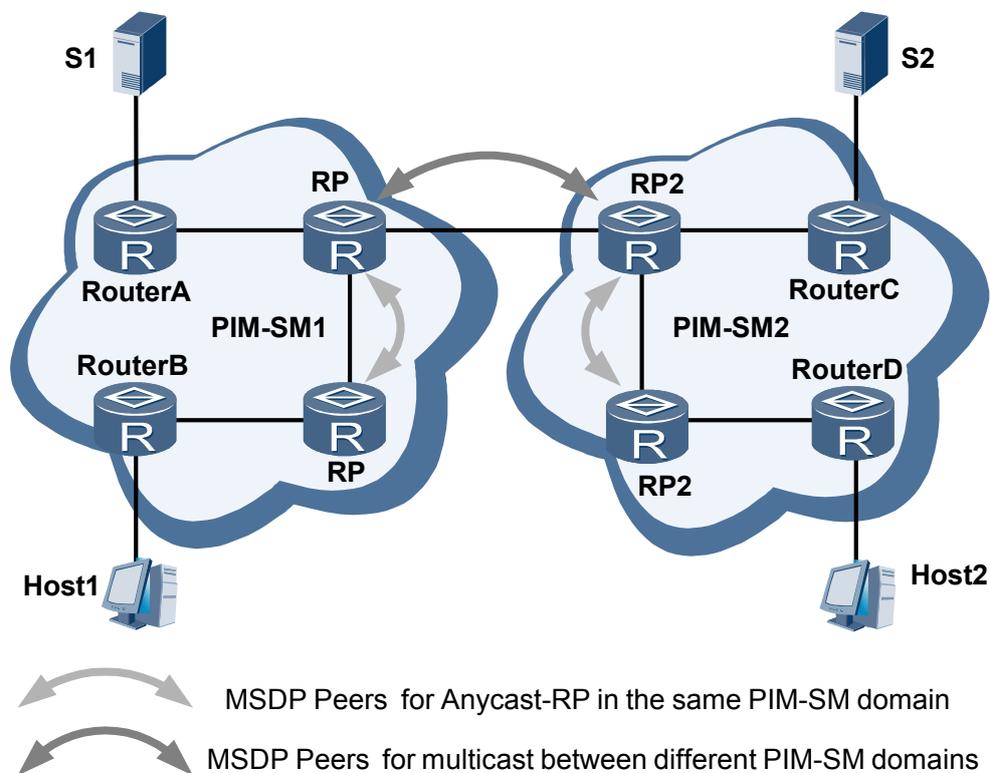
- 配置单播路由协议，确保任意网段路由可达。
- 在用户主机上启用 IGMP 协议。
- 在与接收者主机处于同一网段的路由器接口上启用 IGMP 协议。
- 在所有路由器接口上启用 PIM-SM 协议。
- 在某些 PIM-SM 接口上配置 C-BSR 和 C-RP。

📖 说明

如果需要该 PIM-SM 网络同时支持 ASM 和 SSM，则建议在主机与路由器接口上启用 IGMPv3。
如果主机上运行的版本无法升级，则必须在路由器上配置 SSM Mapping 功能。

3.5.2 Internet 组播

图 3-13 PIM-SM 单 BSR 域典型组网图



PIM-SM 适合网络中的组成员相对比较稀疏，分布广泛的大型网络。ISP 在 Internet 上建立组播路由，适用 PIM-SM。

如图 3-13 所示：

- 配置单播路由协议，确保任意网段路由可达。
- 在 PIM-SM 域内使用 MSDP 建立 Anycast RP，可以实现 RP 负荷分担、组播路由优化和快速收敛。
- 在 ISP 组播服务区域边缘的路由器接口上配置 BSR 域边界。同时使用 MSDP 实现域间组播。

3.6 术语与缩略语

术语

术语	解释
(S, G)	属于组播路由表项，S 表示组播源，G 表示组播组。 源地址为 S、组地址为 G 的组播报文，到达组播设备后，从 (S, G) 表项中的下游接口转发出去。 通常，将源地址为 S，组地址为 G 的组播报文表示为 (S, G) 报文。
Assert	断言。同时在 PIM-DM 和 PIM-SM 中使用。 如果组播设备从下游接口收到组播报文，且 RPF 检查失败，则说明该网段存在其他的组播转发者。组播设备从该下游接口发出 Assert 消息，参与 Assert 竞选，如果落败则将该下游接口从下游接口列表中删除。 Assert 保证了一个网段上最多只存在一个组播转发者，传输一份组播报文。
Flooding	扩散。只在 PIM-DM 中使用。 PIM-DM 假设网络中的组成员分布非常稠密，每个网段都可能存在组成员。基于这一假设，PIM-DM 的设计思路是：首先将数据报文扩散到各个网段，然后再裁剪掉不存在组成员的网段。 PIM-DM 通过周期性的“扩散—剪枝”，构建并维护一棵连接组播源和组成员的单向无环 SPT。
Graft	嫁接。只在 PIM-DM 中使用。 下游接口列表为 NULL 的组播设备，添加第一个下游接口时，从上游接口发出嫁接消息。 如果上游组播设备上收到嫁接消息的接口处于剪枝状态，则立即恢复转发，添加到下游接口列表中。
PIM	Protocol Independent Multicast，称为协议无关组播，属于组播路由协议。 网络中单播路由畅通是 PIM 转发的基础。PIM 利用现有的单播路由信息，对组播报文执行 RPF 检查，从而创建组播路由表项，构建组播分发树。
Prune	剪枝。同时在 PIM-DM 和 PIM-SM 中使用。 当路由表项下游接口列表为 NULL 时，向上游发出剪枝消息，通知其停止向该下游接口转发组播报文。 上游组播设备将收到剪枝消息的接口从下游接口列表中删除。

缩略语

缩略语	英文全称	中文全称
RP	Rendezvous Point	汇聚点
PIM-SM	Protocol Independent Multicast Sparse Mode	协议无关组播—稀疏模式
SSM	Source-Specific Multicast	指定源组播

缩略语	英文全称	中文全称
PIM-DM	Protocol Independent Multicast Dense Mode	协议无关组播—密集模式
PIM	Protocol Independent Multicast	协议无关组播

4 MSDP

关于本章

- 4.1 介绍
- 4.2 参考标准和协议
- 4.3 可获得性
- 4.4 原理描述
- 4.5 应用
- 4.6 术语与缩略语

4.1 介绍

定义

MSDP (Multicast Source Discovery Protocol, 组播源发现协议) 是基于多个 PIM-SM (Protocol Independent Multicast Sparse Mode) 域互连而开发的一种域间组播解决方案, 目前只支持 IPv4。

目的

多个 PIM-SM 相连组成的网络称为 PIM-SM 网络。一个大的 PIM-SM 网络可以由多个 ISP (Internet Service Provider) 联合维护。

PIM-SM 域间 RP 信息隔离, 组播源只向本域内的 RP 注册, 用户主机只向本域内的 RP 发起加入。由于不同 PIM-SM 域的 RP 之间无法通信, 所以 RP 知道且仅知道本域内的组播源, 只能将本域内的组播源发出的数据分发给本地用户。

PIM-SM 网络依靠 RP (Rendezvous Point) 实现组播转发。将一个大的 PIM-SM 网络划分为多个区域, 每个区域维护一个 RP, 可以实现 RP 负荷分担、增强网络的稳定性, 易于管理。每一个这样的区域, 称为一个 PIM-SM 域。

因此, 将一个大的 PIM-SM 网络划分为多个 PIM-SM 域后, 针对如何实现 PIM-SM 域间组播, 使本 PIM-SM 域内的用户主机能够接收到其它域内组播源发出的组播数据, 产生了 MSDP, 使不同 PIM-SM 域的 RP 之间能够互相通信, 共享组播源信息。

说明

本节中描述的 PIM-SM 域指某个 RP 的服务范围, 可以通过 BSR 边界划分出来的域, 也可以是通过在不同路由器上配置不同的静态 RP 形成的域。

4.2 参考标准和协议

本特性的参考资料清单如下:

文档	描述	备注
RFC3618	Multicast Source Discovery Protocol	-
RFC3446	Anycast Rendezvous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)	-

4.3 可获得性

涉及网元

需要其他网元也支持 MSDP。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

AR1200 支持该特性的版本如下：

产品	最低支持版本
AR1200	V200R001C00

4.4 原理描述

4.4.1 MSDP 实现域间组播

MSDP 对等体

MSDP 实现域间组播解决了不同 PIM-SM 域之间 RP 信息隔离、需要共享不同域内组播源信息的问题，实现了 PIM-SM 域间的 RP 之间的通信，共享组播源信息，保证域间组播业务正常运行。

按如下两种关系配置对等体关系：

- 在属于同一 AS，但属于不同 PIM-SM 域的 RP 之间建立 MSDP 对等体。
- 在跨 AS 的 RP 之间建立 MSDP 对等体。并采用 MBGP，将 MBGP 对等体和 MSDP 对等体建立在相同的接口上。

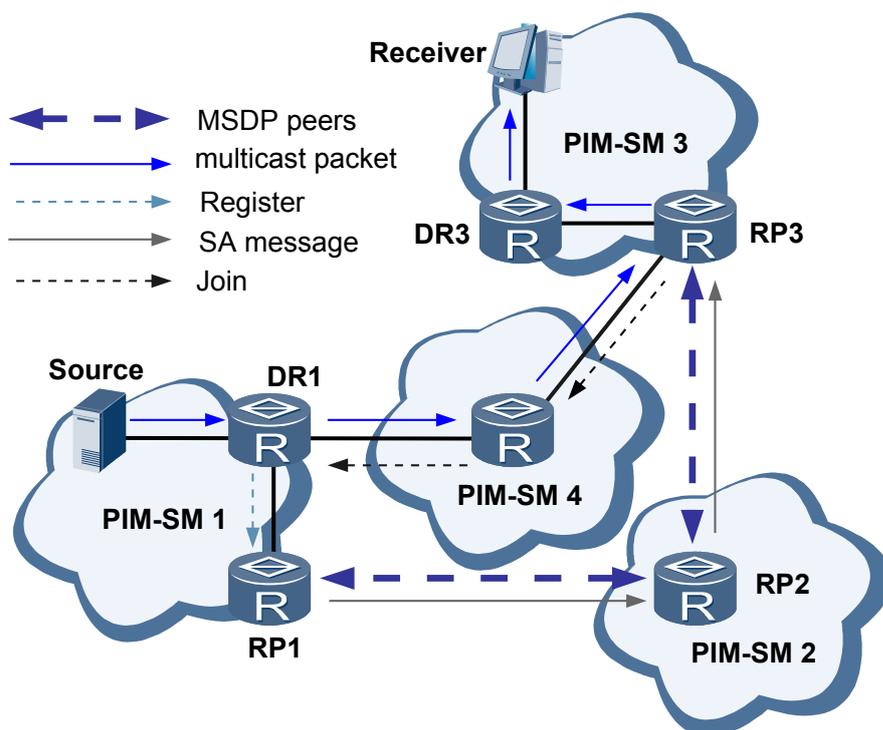
基本原理

通过配置 MSDP Peer 使各个 PIM-SM 域的 RP 之间建立 MSDP 对等体关系，各 MSDP Peer 之间彼此首尾相连，形成一张“MSDP 连通图”，连接各 PIM-SM 域的 RP。

MSDP 对等体之间交互 SA（Source Active）消息，SA 消息中携带组播源 DR 在 RP 上注册时的（S，G）信息。通过这些 MSDP 对等体之间的信息传递，任意一个 RP 发出的 SA 消息能够发送到其他所有的 RP。

如图 4-1 所示，PIM-SM 网络被划分为 4 个 PIM-SM 域。PIM-SM1 域内的组播源 Source 向组 G 发送数据。PIM-SM3 域内的 Receiver 为组 G 成员，RP3 和 Receiver 之间维护了一棵关于组 G 的 RPT（RP-rooted Shared Tree）。

图 4-1 MSDP 实现域间组播



如图 4-1 所示，通过在 RP1、RP2 和 RP3 之间建立 MSDP 对等体关系，可以使 Receiver 接收到 Source 发出的组播数据，具体过程如下：

1. Source 向组 G 发送组播数据。DR1（Designated Router）将组播数据封装在 Register 消息中，发给 RP1。RP1 作为源端 RP，创建 SA 消息，携带 Source 的 IP 地址、组 G 地址和 RP1 地址，发送给对等体 RP2。
2. RP2 接收到该 SA 消息后，执行 RPF（Reverse Path Forwarding）检查。检查通过，向 RP3 转发。
3. RP3 接收到该 SA 消息后，执行 RPF 检查，检查通过。由于 RP3 上存在 (*, G) 表项，表示本域内存在组 G 成员。
4. RP3 创建 (S, G) 表项，向 Source 逐跳发送 (S, G) 加入消息，创建一条从 Source 到 RP3 的组播路径（源树）。组播数据沿源树到达 RP3 后，再沿 RPT 向接收者转发。
5. 接收者接收到组播数据后，自行决定是否发起 SPT 切换。

4.4.2 MSDP 实现 Anycast RP

应用场景

在传统的 PIM-SM 域中，每个组播组只能映射到一个 RP。当网络负载较大或者流量过于集中时，可能导致 RP 路由器的压力过大、RP 失效后收敛较慢、组播转发路径非最优等问题。

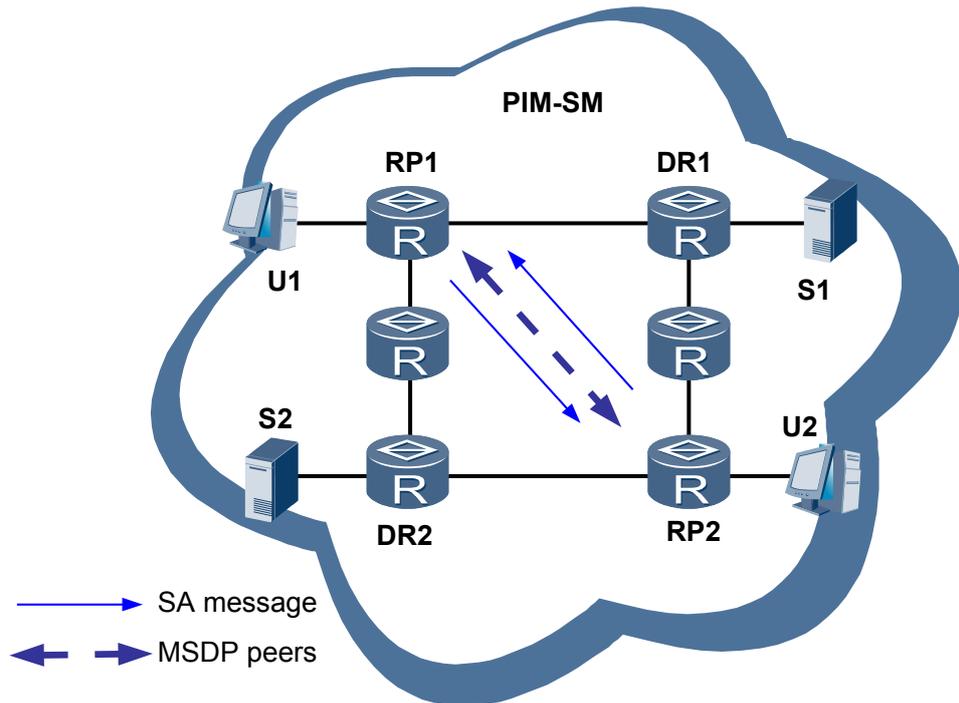
MSDP 实现 Anycast RP 是指在同一 PIM-SM 域内配置多个具有相同 IP 地址的 RP，这些相同的 IP 地址都配置在 loopback 接口上，且在這些 RP 之间建立 MSDP 对等体关系，从而实现 RP 路径最优及负荷分担。

在 PIM-SM 域内应用 Anycast RP 既可以解决组播源信息和组播组加入信息都需要向同一 RP 汇聚，导致单 RP 负荷重的问题。同时，接收者和组播源分别选择最近的 RP 发送加入和进行注册，保证了 RP 路径最优。

实现原理

如图 4-2 所示，在 PIM-SM 域内，组播源 S1 和 S2 向组播组 G 发送组播数据，U1 和 U2 是组播组 G 的成员。

图 4-2 Anycast RP 典型组网图



在 PIM-SM 域内应用 Anycast RP 的实现过程如下：

1. 在 RP1 和 RP2 两个路由器之间建立 MSDP 对等体关系，通过 MSDP 对等体进行域内组播。
2. 接收者选择距离最近的 RP 发送加入消息以构建 RPT 树。组播源选择距离最近的 RP 进行注册，RP 之间通过 MSDP 交互 SA 消息，共享组播源信息。
3. RP 加入以源端 DR 为根的 SPT，接收组播数据并转发，接收者接收到组播数据后，自行决定是否发起 SPT 切换。

4.4.3 MSDP 支持 MD5/Key-chain 认证

MSDP 支持 MD5 或 Key-chain 认证，用于提高 MSDP 报文转发的安全性和可靠性，其应用场景同 MSDP 基本应用场景。当前支持 MD5 和 Key-chain 两种加密方式，MSDP peer 之间同时只能选择 MD5 和 Key-chain 两种加密方式之一，二者在功能上互斥。

Key-chain 能够集中的为所有应用提供加密认证功能，支持多种加密算法，且支持动态更新加密算法的 key 值。关于 Key-chain 的详细介绍，请参考 Key-chain 的特性描述文档。

4.4.4 SA 消息的 RPF 检查规则

为了防止 SA 消息在 MSDP Peer 之间被循环转发，MSDP 对接收到的 SA 消息执行 RPF 检查，在消息传递的入方向上进行严格的控制。不符合 RPF 规则的 SA 消息，将被丢弃。

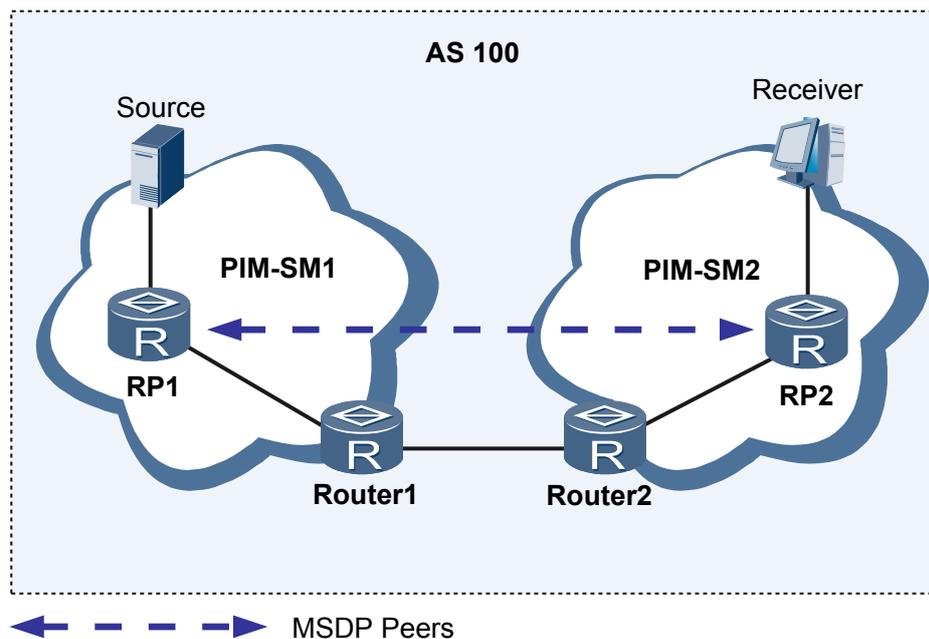
SA 消息的 RPF 规则主要有以下 6 点：

- 规则 1：发出 SA 消息的 Peer 就是源 RP（即创建该 SA 消息的 RP），则接受该 SA 消息并向其他对等体转发。
- 规则 2：接收从静态 RPF 对等体到来的 SA 消息。一台设备可以同时与多个设备建立 MSDP 对等体关系。用户可以从这些远端对等体中选取一个或多个，配置为静态 RPF 对等体。
- 规则 3：如果一台设备只拥有一个远端 MSDP 对等体，则该远端对等体自动成为 RPF 对等体，设备接受从该远端对等体发来的 SA 消息。如果 PIM-SM 域只存在一个域外远端 MSDP 对等体时，该域被称为 STUB 域。
- 规则 4：发出 SA 消息的 Peer 与本地设备属于同一 Mesh Group，则接受该 SA 消息。来自 Mesh group 的 SA 消息不再向属于该 Mesh group 的成员转发，但向该 mesh group 之外的所有对等体转发。
- 规则 5：发出 SA 消息的 Peer 是到源 RP 的“路由”下一跳或“路由”转发者，则接受该 SA 消息并向其他对等体转发。“路由”包括：MBGP、组播静态路由、单播路由（包括 BGP、IGP）。
- 规则 6：到达源 RP 的路由需要跨越多个 AS 时，接收从 AS-path（以 AS 为单位）中的 Peer 发出的 SA 消息。

4.5 应用

域间组播

图 4-3 AS 内 PIM-SM 域间组播

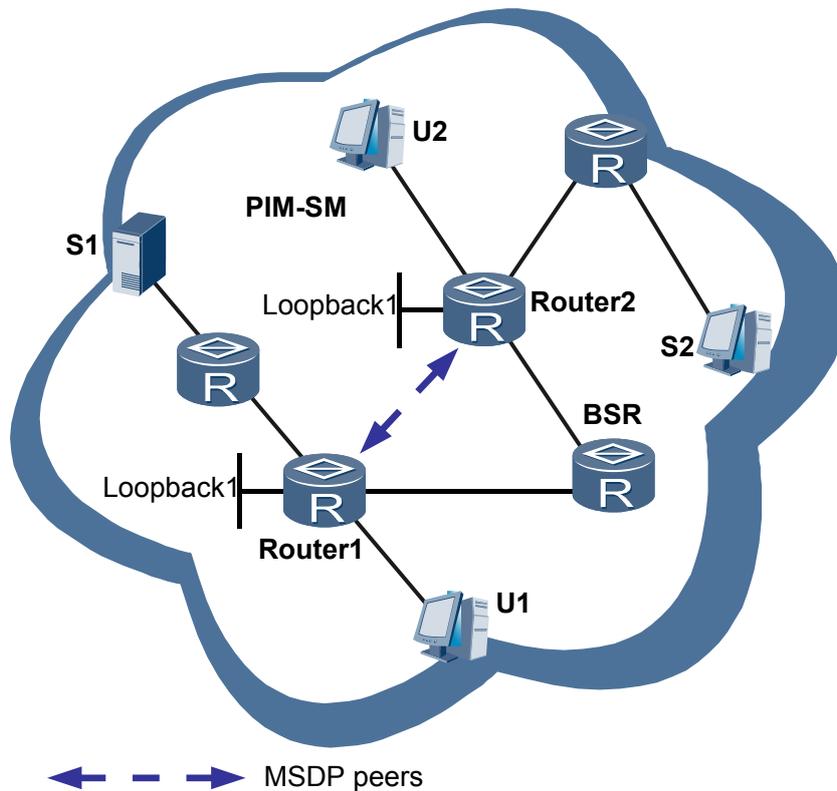


如图 4-3 所示:

- 在两个 PIM-SM 域的 RP 之间配置 MSDP 对等体,共享两个域内的组播源信息。
- 组播数据到达组播源侧 RP1, 通过 SA 消息将对应的组播源信息发送给 RP2。
- RP2 将该组播数据转发给本域内的接收者。
- 接收者接收到组播数据后, 自行决定是否发起 SPT 切换。

Anycast RP

图 4-4 Anycast RP 应用



如图 4-4 所示:

- Router1 和 Router2 作为 RP, 两者之间建立 MSDP 对等体关系。
- 借助 MSDP 对等体进行域内组播, 接收者选择距离最近的 RP 发送加入消息以构建 RPT 树。
- 组播源选择距离最近的 RP 进行注册, RP 之间交互 SA 消息, 共享源信息。
- RP 加入以源端 DR 为根的 SPT, 引入组播数据。
- 接收者接收到组播数据后, 自行决定是否发起 SPT 切换。

4.6 术语与缩略语

术语

术语	解释
MSDP	<p>Multicast Source Discovery Protocol, 称为组播源发现协议。只适用于 PIM-SM 域, 仅对 ASM (Any-Source Multicast) 模型有意义。</p> <p>通过在不同 PIM-SM 域的 RP 之间建立 MSDP 对等体关系, 在域间共享组播源信息, 实现跨域组播。</p> <p>通过在同一 PIM-SM 域的多个 RP 之间建立 MSDP 对等体关系, 在域内共享组播源信息, 实现 Anycast RP。</p>
PIM	<p>Protocol Independent Multicast, 称为协议无关组播, 属于组播路由协议。网络中单播路由畅通是 PIM 转发的基础。PIM 利用现有的单播路由信息, 对组播报文执行 RPF 检查, 从而创建组播路由表项, 构建组播分发树。</p>
SA	<p>Source Active, MSDP 消息类型。SA 消息中包含多组 (S, G) 信息, 或封装一个 Register 消息。MSDP 对等体之间通过交互 SA 消息, 共享组播源信息。</p>
SPT	<p>Shortest Path Tree, 称为最短路径树。以组播源为根, 组播组成员为叶子的组播分发树称为 SPT。SPT 同时适用于 PIM-DM、PIM-SM 和 PIM SSM。</p>
BSR	<p>BootStrap Router, 称为自举路由器。是 PIM-SM 网络的管理核心。BSR 负责收集网络中的 C-RP 信息, 汇集成为 RP-Set, 封装在 Bootstrap 消息中, 发布给全网的每一台 PIM-SM 设备。各设备根据 RP-Set 计算出特定组播组对应的 RP。</p>

缩略语

缩略语	英文全称	中文全称
AS	Autonomous System	自治系统
BGP	Border Gateway Protocol	边界网关协议
BSR	BootStrap Router	自举路由器
MSDP	Multicast Source Discovery Protocol	组播源发现协议
PIM-SM	Protocol Independent Multicast Sparse Mode	协议无关组播—稀疏模式
RP	Rendezvous Point	汇聚点

5 组播管理

关于本章

- 5.1 介绍
- 5.2 参考标准和协议
- 5.3 可获得性
- 5.4 原理描述
- 5.5 术语与缩略语

5.1 介绍

定义

随着 Internet 网络的不断发展，网络中交互的各种数据、语音和视频信息越来越多，组播业务随之快速发展起来，组播管理是对组播业务探测、故障诊断等工具的管理，特性列表如下：

- 组播 Ping（Multicast Ping，以下简称 MPing）是一种组播业务的探测工具，通过发送 ICMP Echo Request 报文促进组播转发树的建立和检测网络中的保留组成员。

说明

保留组：保留的本地组播组地址 224.0.0.0-224.0.0.255 网段。例如：224.0.0.5 是 OSPF 协议组地址，224.0.0.13 是 PIMv2 协议组地址。

- 组播 Tracert（Multicast trace route，以下简称 MTrace），是一种组播路径的追踪工具，可以追踪某一接收者沿着组播转发树到组播源的路径。

目的

在组播应用日益增加的今天，缺乏组播 Ping 和 Tracert，已经不能满足组播业务的维护和故障定位的需要。在选择支持组播的网络设备时，用户不仅仅要求设备支持组播转发和组播路由协议，还要求支持组播故障诊断工具。伴随着组播业务的开展，组播维护和故障定位自然成为必要的需求。

MPing 主要有以下几种用途：

- 发起普通组播组的 MPing
- 通过查看组播设备上的组播路由表信息，检查协议运行状态是否正常，确认组播分发树是否正确建立。
- 通过对目的主机反馈的 ICMP Echo Reply 报文进行统计处理，计算从 MPing 发起者到组播组成员的 TTL、响应时间等。
- 按照一定时间间隔连续执行多次 MPing，计算网络时延和路由抖动。
- 发起保留组播组的 MPing
- 检测网络中的保留组成员。

MTrace 主要有以下几种用途：

- 在组播故障处理和日常维护中使用 MTrace，有助于定位故障结点和检测配置错误。
- 追踪报文实际转发路径，在追踪过程中收集流量信息。循环执行追踪过程，可以统计组播流速率。
- 网管通过分析 MTrace 输出的异常节点信息，产生告警信息。

5.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
draft-sarac-mPing-00.txt	This document describes a mechanism for discovering multicast reachability between end systems within/between multicast enabled networks. It uses request/response messages to verify multicast reachability between the local site and a remote site. With this utility, multicast users can test if they can successfully join a multicast group of a remote source and receiver its data.	-
draft-fenner-traceroute-ipm-01	This draft describes the IGMP multicast traceroute facility. Unlike unicast traceroute, multicast traceroute requires a special packet type and implementation on the part of routers. This specification describes the required functionality in multicast routers, as well as how management applications can use the new router functionality.	-

5.3 可获得性

涉及网元

无需其他网元的配合。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

AR1200 支持该特性的版本如下：

产品	最低支持版本
AR1200	V200R001C00

5.4 原理描述

5.4.1 MPing

MPing 使用标准的 ICMP 消息，查询设备（用户发起检测命令的设备）构造 ICMP Echo Request 报文，其内部封装的 IP 报文的地址为组播地址，包括保留组地址和普通组地址。

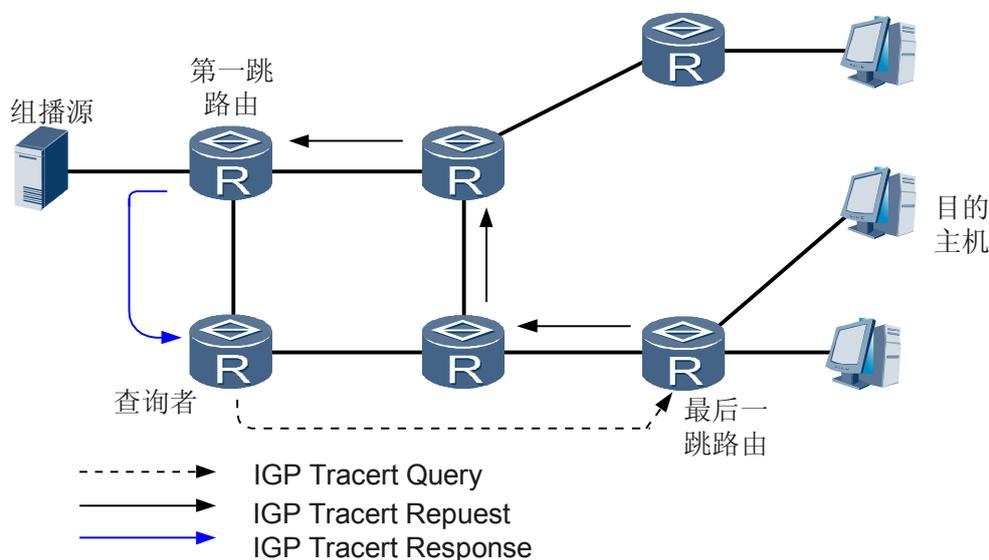
- 目的地址为保留组地址时，查询设备必须指定 ICMP Echo Request 报文的出接口，保留组成员（组播设备）收到目的组地址为保留组地址的 ICMP Echo Request 消息后，回应 ICMP Echo Reply 消息。因此，可以用来检测网络中的保留组成员。
- 目的地址为普通组地址时，查询设备不能指定 ICMP Echo Request 报文的出接口。ICMP Echo Request 报文作为有限的组播流量，在组播网络中正常转发，触发组播路由的建立。通过 NQA 调度多次 MPing 组播组，可以得到时延抖动数据以实现组播业务的正常维护和故障定位。

5.4.2 MTrace

MTrace 遵循 IETF 的协议标准 draft-fenner-traceroute-ipm-01.txt。

该标准描述了一种实现跟踪组播数据从组播源到特定目的接收者所经过的网络路径的机制。

图 5-1 MTrace 组网图



MTrace 基于的网络必须使能了组播协议，例如 PIM 协议（PIM-DM、PIM-SM），并建立了组播转发树。MTrace 通过发送查询报文探测组播路径。报文分为三类：IGMP Traceroute Query、IGMP Traceroute Request 和 IGMP Traceroute Response。

- IGMP Traceroute Request 消息完全继承了 IGMP Traceroute Query 消息，并在消息尾部添加了 Response 数据块。
- IGMP Traceroute Response 消息完全继承了 IGMP Traceroute Request 消息，只修改了消息类型。

实现原理如下：

1. 在查询设备上（如图查询者）输入 MTrace 命令，指定源地址、目的主机地址、组播组。
2. 查询设备向目的主机所连接的最后一跳设备发送 IGMP Traceroute Query 报文。

3. 最后一跳设备收到该报文后，增加包含本设备接口地址信息的响应数据块，然后沿着到组播源的逆向组播路径向上一跳设备发送 IGMP Tracert Request 报文。
4. 每一跳设备都会增加本跳的响应数据块并向组播源方向转发 IGMP Tracert Request 报文。
5. 当连接组播源的第一跳组播设备收到 IGMP Tracert Request 报文时，添加本设备响应数据块，然后向查询设备发送 IGMP Tracert Response 报文。
6. 在查询设备上通过解析 Response 报文获取转发的路径信息，显示从组播源到目的主机的组播路径信息。
7. 如果因为某种错误原因，请求消息在转发时无法到达第一跳组播设备，则直接向查询设备返回 Response 报文，根据解析数据块信息，了解故障节点，从而实现故障点监测的目的。

MTrace 发起方式：

MTrace 支持 4 种封装方式（也称为 MTrace 的发起方式），适用于不同的网络环境。

- **all-router**: 当前组播设备与目的主机直接相连（但不是最后一跳设备），使用 224.0.0.2 为报文目的地址，目的主机网段接口地址为报文源地址。该报文能够被连接在目的主机网段上的所有组播设备（包括最后一跳设备）接收。
- **last-hop**: 使用最后一跳组播设备地址为报文目的地址。这种方法要求用户输入最后一跳设备地址。
- **destination**: 使用目的主机地址为报文的地址。当与用户主机相连的组播设备接收到该报文时，判断自己是不是最后一跳设备，如果不是，则使用 **all-router** 方式重新封装 IGMP Tracert Query 消息。
- **multicast-tree**: 查询设备正好位于从组播源到目的主机的组播路径上（比如第一跳组播设备），使用被追踪的组地址为报文的地址，组播源地址为报文的源地址。该报文沿组播路径下发，到达最后一跳组播设备。

5.5 术语与缩略语

术语

无

缩略语

缩略语	英文全称	中文全称
MPing	Multicast Ping	组播 Ping
MTrace	Multicast trace route	组播 trace route
NQA	Network Quality Analysis	网络质量分析

6 组播路由管理

关于本章

- 6.1 介绍
- 6.2 参考标准和协议
- 6.3 可获得性
- 6.4 原理描述
- 6.5 术语与缩略语

6.1 介绍

定义

组播路由管理（Multicast Route Management）用于管理组播路由表，能够控制组播路由创建或改变组播路由。

组播路由管理包括以下 4 个功能：

- RPF 单播逆向路由检查
- 组播负载分担
- 按照最长匹配选择路由
- 指定组播转发边界(multicast boundary)

目的

- RPF 单播逆向路由检查功能
此功能用于查找到组播源的最优单播路由，创建组播转发树。单播路由的出接口作为转发表项中的数据入口。当转发模块收到组播数据时，在匹配转发表项的同时，还会匹配数据的入接口是否正确。如果组播数据报文实际到达接口和单播路由的出接口相同，则 RPF 检查通过；否则 RPF 检查失败，丢弃该报文。RPF 单播逆向路由检查功能能够防止组播数据在转发过程中出现流量环路。
- 组播负载分担
在组播选路时，使用组播负载分担策略可对不同转发表项从多条等价路由中选取不同的等价路由作为 RPF 路由，指导数据转发。由于转发表项依赖的 RPF 路由能够分布在多条等价路由上，因此此功能可以达到组播数据分流的效果。
- 按照最长匹配选择路由功能
在组播选路时，能够优先选取掩码长度最长的路由，以实现路由的精确匹配。
- 指定组播转发边界功能
在接口上配置组播转发边界可阻塞相应的组播数据，使组播数据流无法从该接口转发。

6.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RPF4601	Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)	-

6.3 可获得性

涉及网元

无需其他网元的配合。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

AR1200 支持该特性的版本如下：

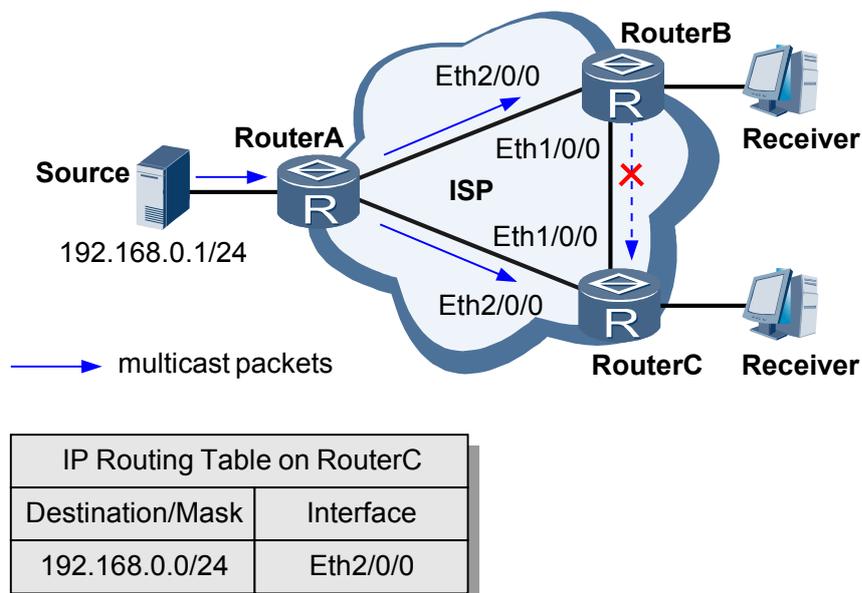
产品	最低支持版本
AR1200	V200R001C00

6.4 原理描述

6.4.1 RPF 单播逆向路由检查

RPF 检查的规则是：依据“报文源”，查找单播路由表、MBGP 路由表、MIGP 路由表和组播静态路由表，从这些路由表中选出一条最优路由，作为 RPF 路由。如果报文实际到达接口与 RPF 接口相同，则 RPF 检查通过；否则 RPF 检查失败。

图 6-1 RPF 检查过程



如图 6-1 所示，组播报文从 GE1/0/0 到达 RouterC，RouterC 对数据报文进行 RPF 检查，发现数据到达接口与转发表项入接口不符，则 RPF 检查失败。如图 6-1 中的路由表，发

现到达 Source 的最短路径出接口是 GE2/0/0，与 (S, G) 表项的入接口相同。于是判定当前 (S, G) 表项正确，该报文从错误的路径而来，丢弃该报文。

6.4.2 组播负载分担

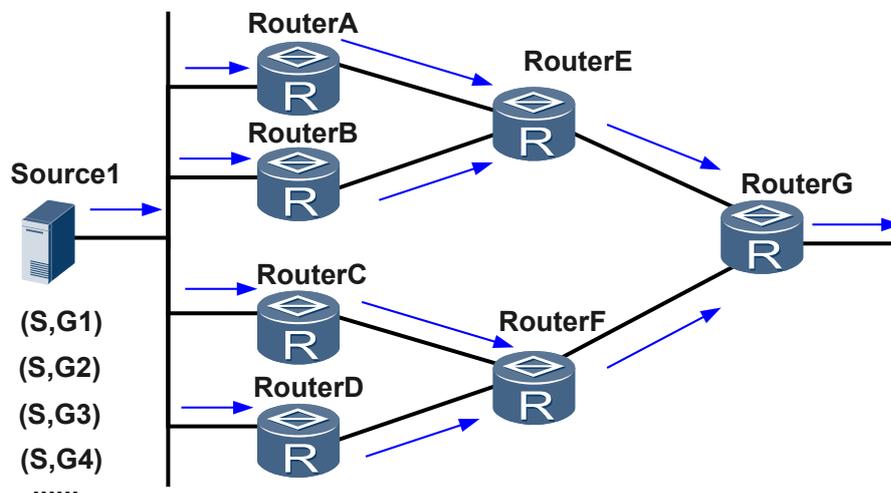
组播负载分担共支持 5 种策略：

- 基于组播组的负载分担
基于组播组的负载分担策略主要应用在网络中存在大量不同的组播组的场景。
- 基于组播源的负载分担
基于组播源的负载分担策略主要应用在网络中存在大量不同的组播源的场景。
- 基于组播源组的负载分担
基于组播源组的负载分担策略应用在网络中既存在大量不同的组播组，又存在大量不同的组播源的场景。
- 稳定优先负载分担
稳定优先负载分担策略可以应用于上述三种负载分担场景，还可以应用于共享网段的应用场景。
- 均衡优先负载分担
均衡优先负载分担的应用场景同稳定优先。

基于组播组的负载分担

如图 6-2 所示：

图 6-2 基于组播组的负载分担

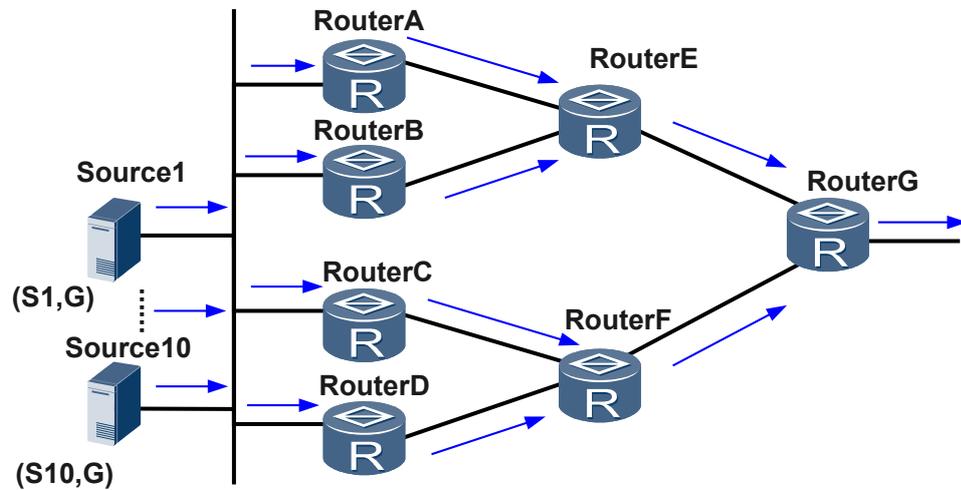


组播路由器根据组播组 G 的不同，经过一系列算法，为每个组播组 G 从多条等价路由中选取一条合适的路由，作为该组播组 G 的转发路由，最终达到不同的转发路径上的流量属于不同的组播组集合。

基于组播源的负载分担

如图 6-3 所示：

图 6-3 基于组播源的负载分担

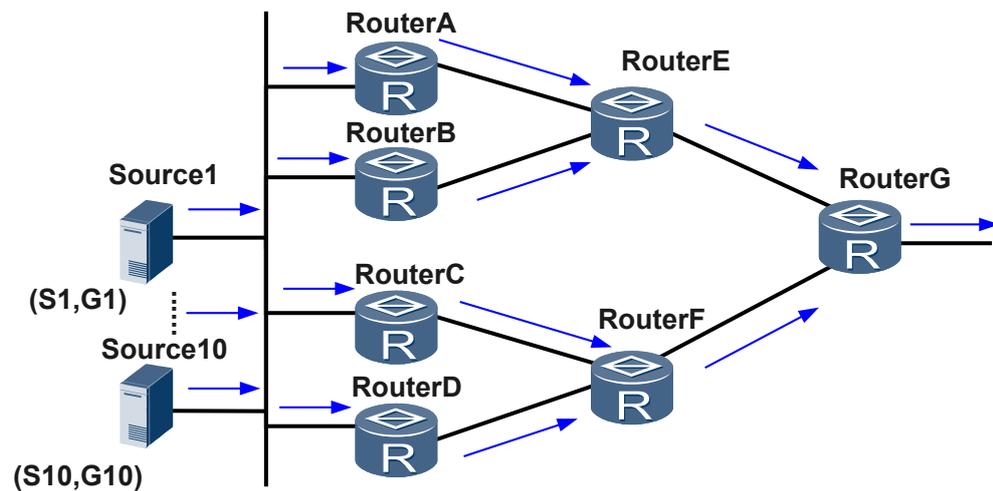


组播路由器根据组播源 S 的不同，经过一系列算法，为每个组播源 S 从多条等价路由中选取一条合适的路由，作为该组播源 S 的转发路由，最终达到不同的转发路径上的流量分属于不同的组播源集合。

基于组播源组的负载分担

如图 6-4 所示：

图 6-4 基于组播源组的负载分担



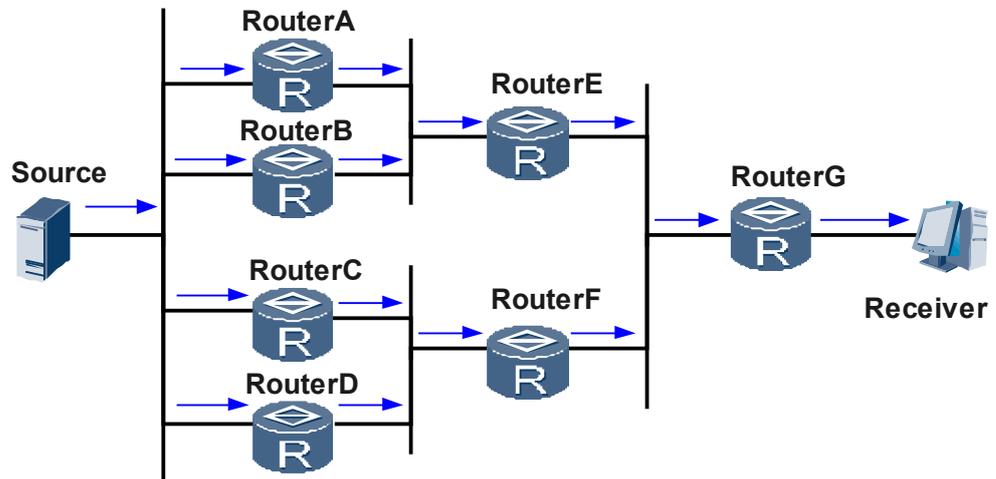
组播路由器根据组播组 G 和组播源 S 的不同，为每一个组播源组(S,G)经过一系列算法，从多条等价路由中选取一条合适的路由，作为该组播源组(S,G)的转发路由，最终达到分配到不同的转发路径上的流量分属于不同的组播源组集合。

稳定优先负载分担

- 应用场景

该负载分担策略可以应用于上述三种负载分担场景，如图 6-2、图 6-3 和图 6-4 所示。负载分担还可以应用于共享网段的应用场景，如图 6-5 所示。

图 6-5 稳定优先负载分担



- 实现原理

在配置稳定优先负载分担的路由器上，对于新的表项加入，会选择最合适的路由，即当前依赖表项最少的路由。在网络拓扑稳定同时表项稳定的情况下，依赖同一个网段源的所有表项会均衡分布在各个等价路由之上。

如果表项退出或者路由权值变化，导致负载不均衡，在配置稳定优先负载分担策略的情况下，会通过后续新的加入选择最合适路由来慢慢“愈合”这种不均衡。

在稳定优先负载分担场景中，若出现表项分布不均衡，经过一定的延迟时间，设备会对所有表项进行均衡调整，均衡调整的延迟时间用来防止频繁变化对设备的冲击。

目前，可通过配置组播负载分担均衡调整定时器控制系统开始调整表项分布不均衡状态的延迟时间。

均衡优先负载分担

在配置均衡优先负载分担的路由器上，对于新的表项加入，会选择最合适的路由，即当前依赖表项最少的路由。均衡优先策略即不论在何种情况下，都要求依赖同一网段源的表项最终均衡的分布在其等价路由上，包括表项退出、权值变化和等价路由发生变化。在发生不均衡以后，均衡调整的动作将会有一定的延迟时间，用来防止频繁变化，同时还可以在延迟时间内，通过新加入表项来“愈合”不均衡。

目前，可通过配置组播负载分担均衡调整定时器控制系统开始调整表项分布不均衡状态的延迟时间。

不均衡负载分担

- 应用场景

不均衡负载分担是对均衡优先负载分担和稳定优先负载分担的功能的一个补充，不会改变这两种策略的基本行为，只是会让分布在等价路由上的表项数呈现一种比例

关系。它的应用场景图同上述两种负载分担（除图 6-5 所示的共享网段场景）。两种典型应用场景为：

- 当几条等价路由之间转发能力存在较大差异，或者流量拥塞程度存在较大差异时，可以通过配置路由负载分担权值来调节等价路由上分配的表项，权值大的路由将会分配更多的表项。稳定优先负载分担策略，只会对后来的加入表项起到作用；均衡优先负载分担策略，则会对已经存在的表项按照权值的比例进行调整。
 - 当路由器的某条等价路由路径上，有路由器需要进行版本升级时，可以通过在接口配置权值为 0，让流量从这条等价路由上转移到其它等价路由上（对均衡优先负载分担策略起作用）。
- 实现原理
单播各条等价路由的转发能力、实际网络负载情况和链路的用途均可能存在差异，因此在特定场景下对组播表项还继续进行均衡负载，很难满足需求。不均衡负载分担是一种针对上述情况的解决策略，允许用户在接口上配置权值，权值越大的接口所在路由可以依赖的表项越多。

6.4.3 按照最长匹配选择路由

选路时，分别从域内单播路由表、域间单播路由表和组播静态路由表中各选出一条最优路由，并从中选择一条作为组播数据转发路径。

按照最长匹配选择路由时，组播路由器会优先选取目的地址与报文源地址匹配“掩码”最长的路由。若存在多条路由由掩码匹配长度相同，则按照组播静态路由、域间单播路由、域内单播路由的顺序选择一条路由作为组播数据的转发路径。

例如：组播源地址为 10.1.1.1，地址为 192.168.1.1 的主机需要接收组播源的组播数据。查找组播静态路由表和域内单播路由表有两条可到达组播源的路由。到达的网段分别为：10.1.1.0/16 和 10.1.1.0/24。按照最长匹配选择路由，选择到达网段为 10.1.1.0/24 的路由作为组播数据的转发路径。若掩码匹配长度相同，则按照路由的优先级顺序选择一条最优路由作为组播数据的转发路径。

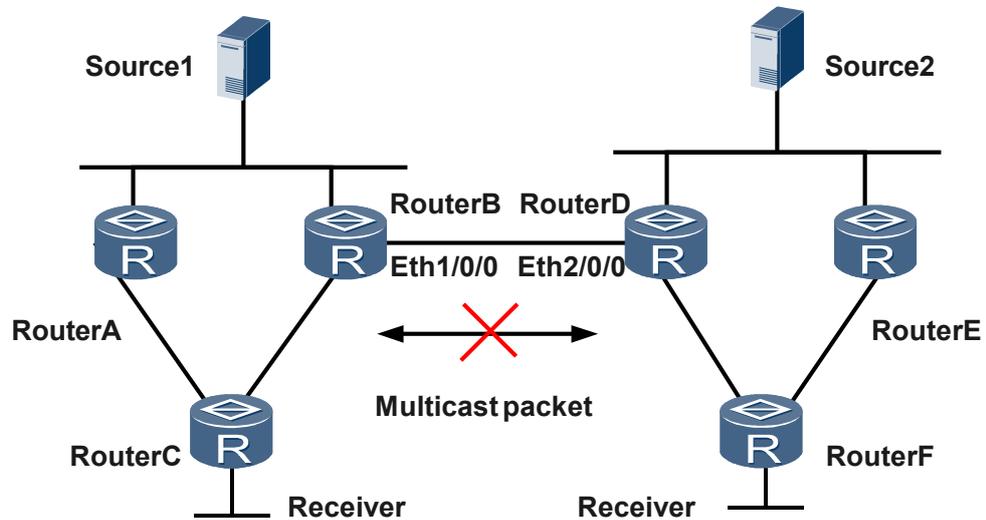
6.4.4 指定组播转发边界

应用场景

应用组播转发边界（multicast boundary）功能可以实现每个组播组对应的组播信息在一个确定的范围内传递。在接口上配置组播边界，以形成一个封闭的组播转发区域。当组播设备接口配置了针对某组播组的转发边界以后，该接口将不能再发出或接收对应的组播组的报文。

实现原理

图 6-6 指定组播转发边界组网图



如图 6-6 所示，RouterA，RouterB，RouterC 构成组播域 1，RouterD，RouterE，RouterF 构成组播域 2，两个组播域通过 RouterB 和 RouterD 进行互通。

如果要两个组播域将组播组 G 的数据隔离开，则只需在接口 GE1/0/0 或接口 GE2/0/0 上配置该组 G 为组播转发边界，对应的接口在配置组播边界后将不再转发和接收对应组播组 G 的报文。

6.5 术语与缩略语

术语

术语	解释
组播负载分担	组播负载分担不等同于负载均衡，负载分担是指组播表项能够分布在各条等价路由上，且每条等价路由上的组播表项数目可以不同。
MBGP 路由表	Multicast BGP，组播边界网关协议路由表。 MBGP 是 MP-BGP 在组播上的应用。MBGP 路由表是在 BGP 协议中定义一个地址族，用于发布组播使用的单播路由。
MIGP 路由表	Multicast IGP，组播内部网关协议路由表。 MIGP 路由表是对单播路由表中出接口为 Shortcut Tunnel 的路由，计算出不使用 shortcut tunnel 作为出接口的路由表。 若组播在单播路由表中选路时选中单向 TE Tunnel 接口，会在 MIGP 路由表中重新选取一条最优路由作为组播数据转发的路径。

缩略语

缩略语	英文全称	中文全称
RPF	Reverse Path Forwarding	反向转发路径