# HCRSE110-PIM 双栈原理

## 组播协议

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

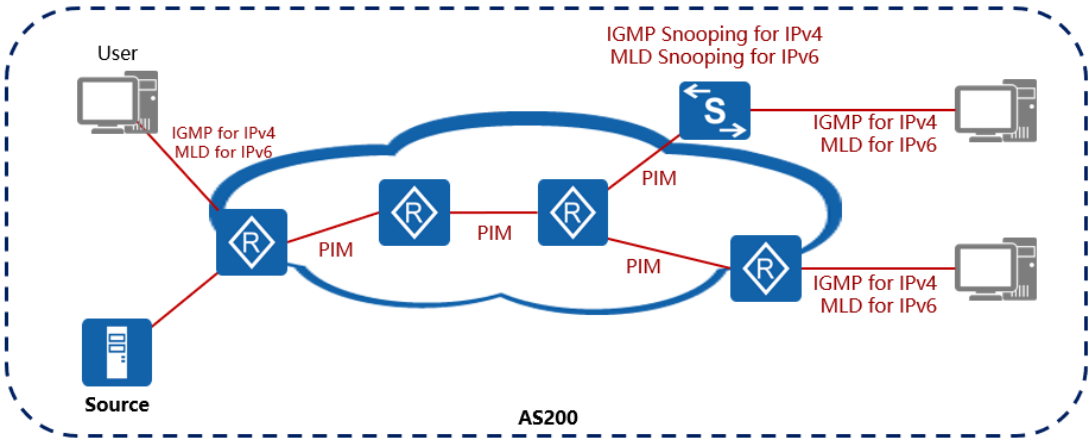
## PIM-SM Silent 设置

设备直连用户主机的接口上需要使能 PIM 协议，当恶意主机模拟 PIM Hello 报文，大量发送时，有可能导致设备瘫痪。为了避免这样的情况发生，可以将该接口设置为 PIM Silent 状态。在接入层上，设备直连用户主机的接口上如果需要使能 PIM 协议，在该接口上可以建立 PIM 邻居，处理各类 PIM 协议报文。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，有可能导致设备瘫痪。
为了避免这样的情况发生，可以将该接口设置为 PIM Silent 状态（即 PIM 消极状态）。当接口进入 PIM 消极状态后，禁止接收和转发任何 PIM 协议报文，删除该接口上的所有 PIM 邻居以及 PIM 状态机，该接口作为静态 DR 立即生效。

```
int g0/0/0
pim silent
```

组播协议包括用于主机注册的组播组管理协议，和用于组播选路转发的组播路由协议。

IGMP（Internet Group Management Protocol）在接收者主机和组播路由器之间运行，该协议定义了主机与路由器之间建立和维护组播成员关系的机制。
组播路由器之间运行组播路由协议，组播路由协议用于建立和维护组播路由，并正确、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。
域内组播路由协议包括：DVRMP、MOSPF、PIM。
DVRMP 是距离矢量组播路由协议（Distance Vector Multicast Routing Protocol）是一种密集模式协议。该协议有跳数限制，最大跳数 32 跳。
MOSPF 是 OSPF 路由协议的扩展协议。它通过定义新的 LSA 来支持组播。
PIM（Protocol Independent Multicast）是典型的域内组播路由协议，分为 DM（Dense Mode）和 SM（Sparse Mode）两种模型。当接收者在网络中的分布较为密集时，适用 DM；较为稀疏时，适用 SM。PIM 必须和单播路由协议协同工作。

## PIM-SM

相对于 PIM-DM 的“推（Push）模式”，PIM-SM 使用“拉（Pull）模式”转发组播报文。PIM-SM 假设网络中的组成员分布非常稀疏，几乎所有网段均不存在组成员，直到某网段出现组成

员时，才构建组播路由，向该网段转发组播数据。一般应用于组播组成员规模相对较大、相对稀疏的网络。

基于这一种稀疏的网络模型，它的实现方法是：
在网络中维护一台重要的 PIM 路由器：汇聚点 RP（Rendezvous Point），可以为随时出现的组成员或组播源服务。网络中所有 PIM 路由器都知道 RP 的位置。
当网络中出现组成员（用户主机通过 IGMP 加入某组播组 G）时，最后一跳路由器向 RP 发送 Join 报文，逐跳创建（*，G）表项，生成一棵以 RP 为根的 RPT。
当网络中出现活跃的组播源（信源向某组播组 G 发送第一个组播数据）时，第一跳路由器将组播数据封装在 Register 报文中单播发往 RP，在 RP 上创建（S，G）表项，注册源信息。

PIM-SM 的关键机制包括邻居建立、DR 竞选、RP 发现、RPT 构建、组播源注册、SPT 切换、Assert；同时也可通过配置 BSR（Bootstrap Router）管理域来实现单个 PIM-SM 域的精细化管理。PIM-SM 中 PIM 邻居建立过程以及 Assert 机制与 PIM-DM 相同。

## IPv6 PIM-SM

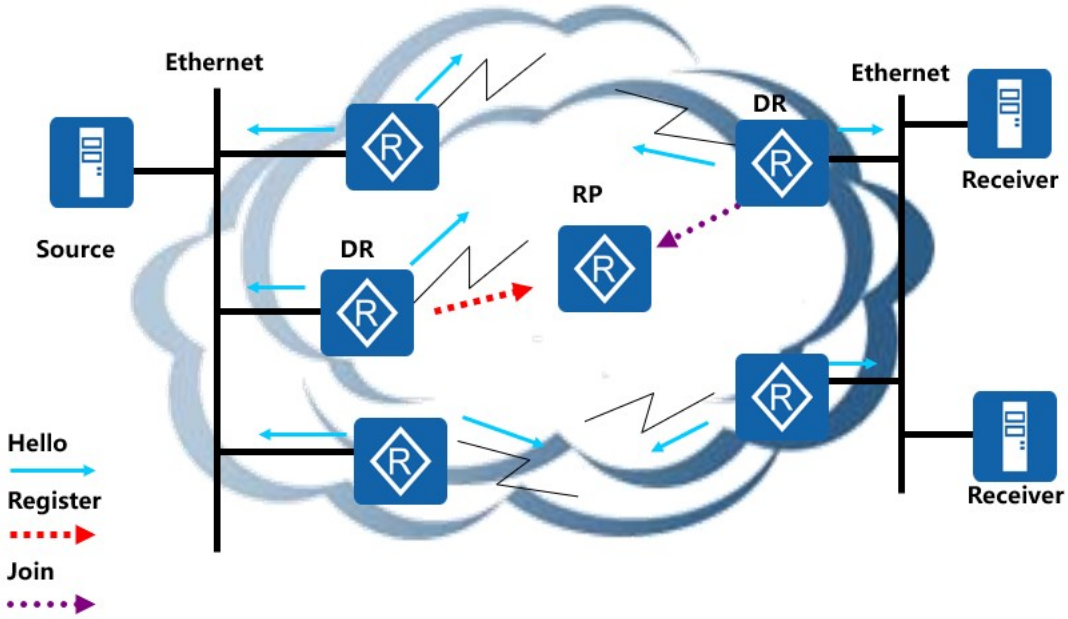
PIM IPv6 是与静态路由、RIPng、OSPFv3、IS-ISv6、BGP4+等 IPv6 单播路由协议类型无关的组播路由协议，其借助上述单播路由协议生成的路由项和 RPF（Reverse Path Forwarding）机制创建组播路由表，实现组播报文转发。PIM IPv6 域是指由支持 PIM IPv6 协议的组播路由器构成的网络。

目前，存在两种组播模型：任意源组播 ASM（Any-Source Multicast）和指定源组播 SSM（Source-Specific Multicast）。IPv6 中，ASM 模型包括 IPv6 PIM-DM（Protocol Independent

Multicast-Dense Mode）和 IPv6 PIM-SM（Protocol Independent Multicast Sparse Mode）；SSM 模型则使用 MLDv2 和 IPv6 PIM-SM 的部分机制来实现。
在组成员稀疏分布的大规模 IPv6 网络中，使用 IPv6 PIM-SM，其主要特性是接收者需要显式加入。缺省情况下，IPv6 PIM-SM 假设网络中的所有节点都不需要接收组播报文，上游节点只有接收到下游节点的加入消息后才进行组播数据的转发。
IPv6 PIM-SM 中，汇集点 RP（Rendezvous Point）只向拥有接收者的下游分枝转发组播信息。这样可以节省数据报文和控制报文占用的网络带宽，减少路由器的处理开销。
当主机希望从指定组播组接收数据时，与之相连的路由器向这个组的 RP 发送加入（Join）消息，沿途建立以 RP 为根节点的共享树 RPT（Rendezvous Point Tree）。共享树的含义是不同组播源向相同组播组转发组播数据时，都使用此共享路径。
当组播源向组播组发送数据时，与源相连的 DR 把组播数据封装在注册消息中以单播方式向 RP 发送。注册消息到达 RP 后，由 RP 对组播数据解封装，再沿 RPT 向接收者发送。当以注册消息方式发送的组播数据达到一定速率后，RP 向组播源发送加入消息，建立组播源与 RP 之间的组播分发树。然后，RP 向组播源的 DR 发送注册停止消息，指示 DR 直接以非封装方式根据组播转发表发送组播数据。

DR 选举 ：最高优先级--最大 IPv6 地址

借助 Hello 消息可以为共享网络（如 Ethernet）选举 DR（Designated Router），DR 将作为本网段中组播信息的唯一转发者。无论是和组播源 S 连接的网络，还是和接收者连接的网络，只要网络为共享媒介则需要选举 DR，接收者侧 DR 向 RP 发送 Join 加入消息；组播源侧 DR 向 RP 发送 Register 注册消息。




DR 的选举过程如下所述：

共享网段上的各路由器相互间发送带有 DR 优先级选项的 Hello 消息；

具有最高优先级的路由器被选举为此网段的 DR。如果路由器具有相同的优先级，则 IPv6 地址最大的路由器被选举为 DR。

当 DR 出现异常，其他路由器将接收不到其发出的 Hello 消息。在此 DR 超时后，会触发共享网段上新一轮的 DR 选举。

如果网络中至少有一台路由器不支持在 Hello 报文中携带 DR 优先级，由 IPv6 链路本地地址最大的路由器充当 DR。

## RP (Rendezvous Point)

如何发现 RP？对于小规模的简单网络，一个 RP 用于全网转发信息就足够了，此 RP 的位置可通过静态指定，在 DR 和叶子路由器以及组播数据流将要经过的所有路由器上手工指定 RP 的 IP 地址。然而，在大多数应用中，IPv6 PIM-SM 网络覆盖了很大的区域，需要通过 RP 转发大量的组播流量。因此，不同的组播组应该具有各自的 RP。为了减少配置多个静态 RP 的工作量以及更好的适应网络的实时变化，采用自举（Boot

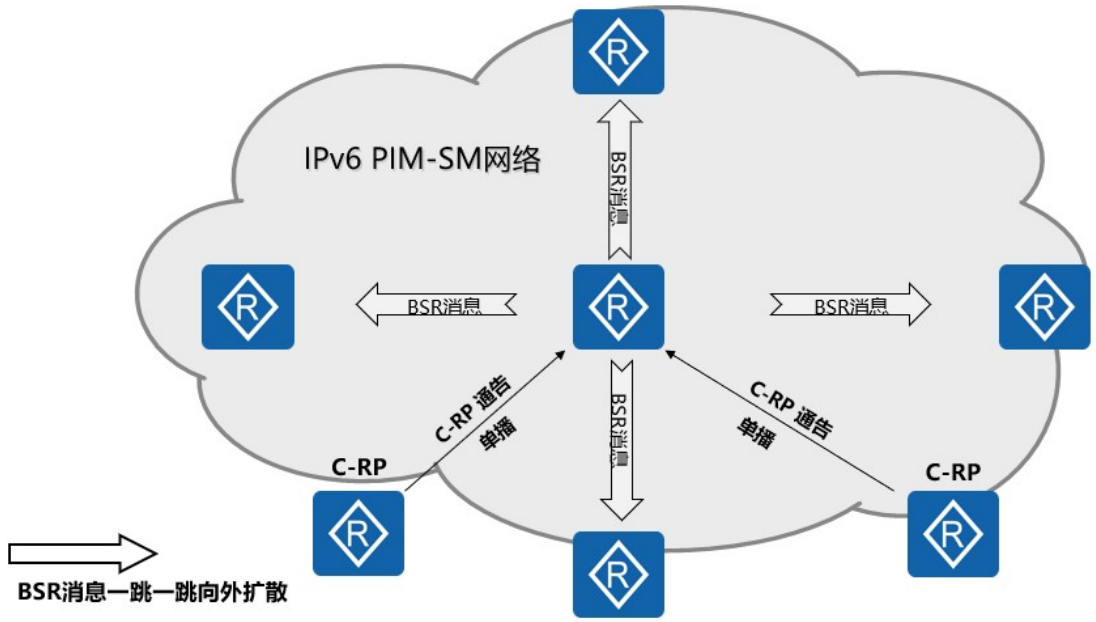
strap）机制来动态选举 RP。

RP 配置方式建议：


中小型网络：建议选择静态 RP 方式，对设备要求低，也比较稳定。

如果网络中只有一个组播源，建议选择直连组播源的设备作为静态 RP，这样可以省略源端 DR 向 RP 注册的过程。采用静态 RP 方式要确保域内所有路由器（包括 RP 本身）的 RP 信息以及服务的组播组范围全网一致。大型网络：可以采用动态 RP 方式，可靠性高，可维护性强。如果网络中存在多个组播源，且分布密集，建议选择与组播源比较近的核心设备作为 C-RP；如果网络中存在多个用户，且分布密集，建议选择与用户比较近的核心设备作为 C-RP。

RP 选举




BootStrap router 工作的原理和过程：

先比优先级大的（默认为 0），再比 IP 地址大的

首先要在网络中选择合适的路由器把它配置成候选 BSR（C-BSR，Candidate BootStrap Router），每个 C-BSR 都有优先级，当它得知自己是 C-BSR 后，首先启动一个定时器（默认为 150 秒），监听网络中的 BootStrap Message。BootStrap Message 初始时通告发送路由器的优先级、BSR 的 IPv6 地址，当 C-BSR 收到一个 BootStrap Message 后，它会把自己的优先级和报文里的优先级做比较，如果对方的优先级高，它就把自己的定时器重置，继续监听 BootStrap Message；如果是自己的高，那么它就发送 BootStrap Message 声明自己是 BSR 如果优先级相等，则比较 IPv6 地址，谁的 IPv6 地址大谁就是 BSR。BSR 消息发送的目的地址是 FF02::13，所有的 PIM IPv6 路由器都能接收到这个报文，该报文 TTL 一般被置为 1，但每个 PIM IPv6 路由器收到此报文后都是把它以泛洪的方式从自己所有的使能 PIM IPv6 的接口上发送出去，这就能保证网络中的每台 PIM IPv6 设备都能收到 BootStrap Message。

c-bsr priority 9

缺省情况下，C-BSR 的全局优先级是 0。优先级数值越大，优先级越高。

RP ：优先级（ 越小越优），Hash 值，IP 地址大

RP 就像 C-BSR 一样需要在设备上手工配置，首先配置 C-RP（Candidate Rendezvous Point），包括 RP IPv6 地址、优先级和它所能服务的组。正如上文所述，一个 RP 可以给所有的 IPv6 组播组提供服务，也可以只给部分组提供服务。当 C-RP 收到 BootStrap Message 后，它可以从该 message 中得知网络中谁是 BSR，然后 C-RP 通过 Candidate-RP-Advertisement Message 把自己所能服务的组单播给 BSR，每个 C-RP 都这么做的话那么 BSR 就收集到了网络中所有 C-RP 的信息并把这些信息整理成一个集 RP-Set。此后 BSR 通过 BootStrap Message 把 RP-Set 的信息通告给全网所有的路由器。

RP 的选举规则：

如果 RP-Set 对应该 IPv6 组地址的 C-RP 只有一个,那么 DR 就选该 C-RP 做 RP；如果对应该 IPv6 组地址的有多个 C-RP，那么优先级最高的是 RP（优先级数越小优先级越高）；如果大家优先级相等，那么 DR 将开始 Hash 运算，把组地址、hash 掩码、和 C-RP 的地址做为输入参数，输出是一些数字，数字高的 C-RP 将是该组的 RP；如果 hash 的结果大家也相等，那么 IPv6 地址最大的 C-RP 将成为该组的 RP。

c-rp 在 IPv4 环境下默认优先级为 0，越小越优

c-rp 在 IPv6 环境下默认优先级为 192，越小越优

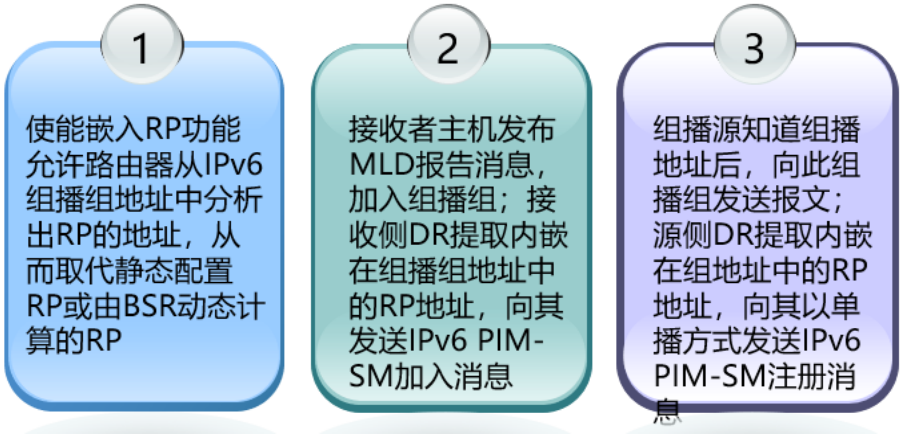
C-RP 竞选 RP 的规则

与用户加入的组地址匹配的 C-RP 服务的组范围掩码最长者获胜；

C-RP 优先级较高者获胜；

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

## 嵌入式 RP




使能嵌入 RP 功能允许路由器从 IPv6 组播组地址中分析出 RP 的地址，从而取代静态配置 RP 或由 BSR 动态计算的 RP。使用嵌入式 RP 的组播组地址范围是 FF7x::/16 和 FFFx::/16，x 表示 0～F 的任意一个十六进制数。

在接收侧：接收者主机发布 MLD 报告消息，加入组播组；接收侧的 DR 提取内嵌在组播组地址中的 RP 地址，向其发送 IPv6 PIM-SM 加入消息。
在组播源侧：组播源知道组播地址后，向此组播组发送报文；组播源侧的 DR 提取内嵌在组播地址中的 RP 地址，向其以单播方式发送 IPv6 PIM-SM 注册消息。

使能了 PIM-SM（IPv6）的设备都会默认使能嵌入式 RP 功能。当设备收到组播报文后，直接从 IPv6 组播地址中解析出 RP 地址，而无需再预先知道 RP 的信息。

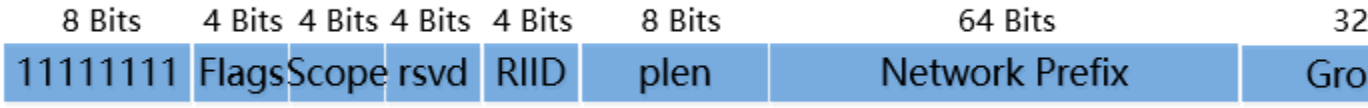
embedded-rp 命令用来使能嵌入式 RP 功能。
undo embedded-rp 命令用来关闭嵌入式 RP 功能。

缺省情况下，设备已使能嵌入式 RP 功能。

使能组播组 ff73::对应的嵌入式 RP 功能。
acl ipv6 number 2000
rule permit source ff73:: 12

pim-ipv6
embedded-rp 2000

一个 128bits 的 RP 地址如何嵌入到一个 128bits 的 IPv6 组播组地址中去？
定义特殊的组播地址：




头 8bits 为 FF 说明是 IPv6 组播地址。
Flags 字段的范围是 7-F，说明是一个嵌入了 RP 地址的 IPv6 组播组地址。
RIID 字段：RP Interface ID，抽取出来填充在 RP 地址的最后 4bits。
Plen 字段：RP 地址的前缀长度，换算成十进制数后不能为 0，也不能大于 64。
Network Prefix 字段：RP 的地址前缀。
Group ID：组 ID。

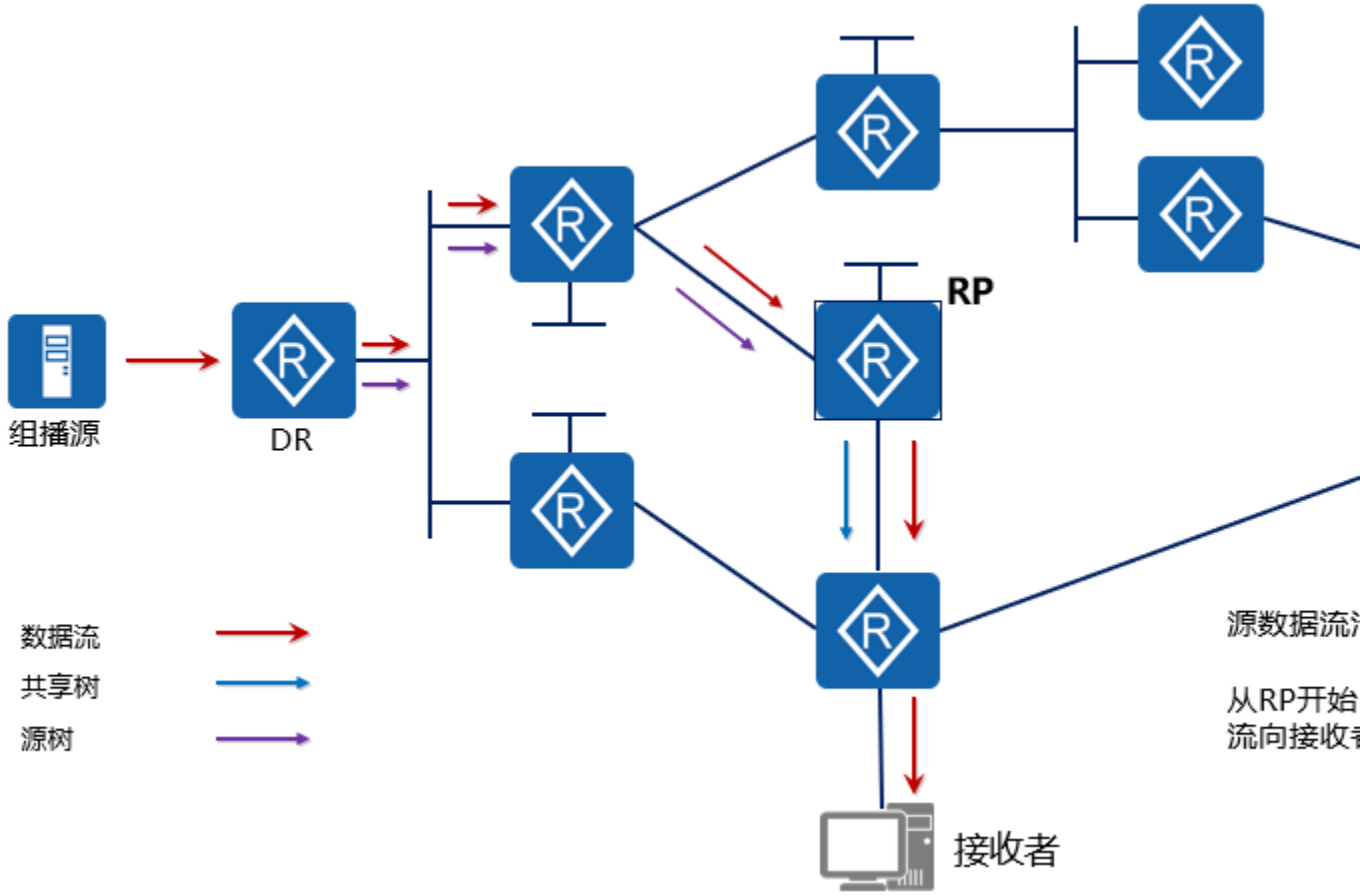
## RP 地址转换

如何由组播组地址得到 RP 地址？
提取“plen”字段，转换为十进制数
将“Network Prefix”字段的前“plen” bits 提取出来作为 RP 地址的地址前缀

将“RIID”字段提取出来作为 RP 地址的 Interface ID 的最后 4bits，Interface ID 其余部分用 0 补齐
嵌入 RP 实例
组播地址 FF70:140:2001:DB8:BEEF:FEED::/96，则从组播地址中获取的 RP 地址为 2001:DB8:BEEF:FEED::1/64

## 组播流转发过程

源数据流延源树（SPT）流向 RP，从 RP 开始，数据流延共享树（RPT）流向接收者。



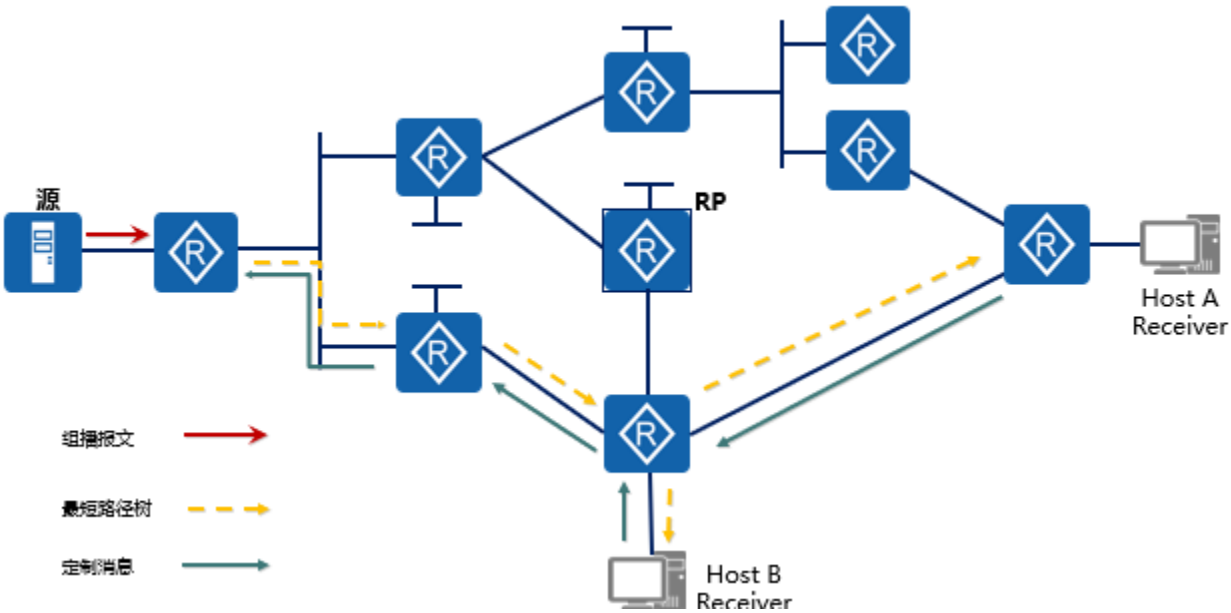

## IPv6 PIM-SSM

SSM 模型提供了指定源组播的解决方案，配合 MLDv2 采用 IPv6 PIM-SM 的部分机制来实现。由于最后一跳路由器通过 MLDv2 协议已经知道了组播源的地址，可以直接发起指定源-组

的加入过程，在 SSM 网络中创建组播源到接收者的 SPT
定义了特殊的组播地址：FF3x::/32，不存在源发现问题
需要和 MLDv2 配合使用
扩展了 PIM SM 协议，PIM-SSM 不涉及 RP、BSR、RPT 生成、组播源注册等复杂机制
基于组播源的单播路由直接生成 SPT 树，可以实现跨域组播

## 工作原理

SSM 模型中，用信道（Channel）概念来表示（S, G）组合，用定制（Subscribed）消息概念来表示加入消息。




假定网络中的 User A 和 User B 需要接收组播源 S 的信息，就通过 MLDv2 向最近的查询器发送一个标为（include S, G）的报告信息。如果 User A 和 User B 不需要接收组播源 S 的信息，发送一个标为（exclude S, G）或包含其他组播源的报告消息。无论使用上述哪个报告消息，接收者是明确指定组播源 S 的。
接收到报告消息的查询器检查此消息的组播地址是否在 SSM 组地址的范围内。如果是，则路由器根据 SSM 模型建立组播分发树，随后向指定源逐跳发送定制消息（也称加入消息）。

沿途上的所有路由器创建（S, G）项。以源 S 为根节点、接收者为叶子的 SPT 树就生成了。SSM 模型使用此 SPT 树作为传输路径。
如果查询器发现组播地址在 SSM 组范围外，就在 IPv6 PIM-SM 基础上建立组播分发树。

## 组播路由管理

为了实现组播路由转发路径的控制与维护，组播路由管理提供了一系列特性。主要分为 RPF（ Reverse Path Forwarding）和组播负载分担。

RPF 检查：用于保证组播数据沿正确的转发路径进行传输。
组播负载分担:用于选取不同的等价路由进行组播数据转发，分流组播数据。

组播路由和转发与单播路由和转发类似，首先每个组播路由协议都各自建立并维护了一张协议路由表。各组播路由协议的组播路由信息经过综合形成一个总的组播路由表（Multicast Routing-Table）。最后，路由器根据组播路由和转发策略，从组播路由表中选出最优的组播路由，并下发到组播转发表（Multicast Forwarding-Table），直接用于控制组播数据的转发。通过组播转发表，整个网络建立了一条以组播源为根，组成员为叶子的一点到多点的转发路径。为了实现转发路径的控制与维护，组播路由管理提供了一系列特性。

组播协议路由表是运行各种组播路由协议时由各个协议自己维护的表项，是组播路由和转发的基础。PIM 路由表项信息如下：

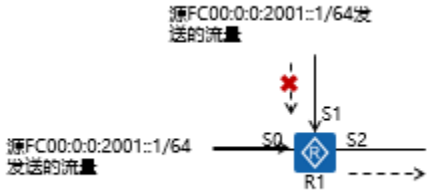
```
<HUAWEI> display pim ipv6 routing-table
VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (FC00::2, FFE3::1)
    Protocol: pim-sm, Flag: SPT LOC ACT
    UpTime: 00:04:24
    Upstream interface: Vlanif20
        Upstream neighbor: FE80::A01:100:1
        RPF prime neighbor: FE80::A01:100:1
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: Vlanif10
            Protocol: pim-sm, UpTime: 00:04:24, Expires: 00:02:47
```

(FC00::2, FFE3::1) (S, G)表项。
Protocol: pim-sm 协议类型。第一个 Protocol 表示生成表项的协议类型，第二个 Protocol 表示生成下游接口的协议类型。
Flag: SPT LOC ACT PIM 路由表项的标志。
UpTime: 00:04:24 存在时间。第一个 UpTime 表示表项已存在的时间，第二个 UpTime 表示下游接口已存在的时间。
Upstream interface: Vlanif20 上游接口。
Upstream neighbor: FE80::A01:100:1 上游邻居。NULL 表示不存在上游邻居。
RPF prime neighbor: FE80::A01:100:1 RPF 邻居。NULL 表示不存在 RPF 邻居。
Downstream interface(s) information: 下游接口信息。
Total number of downstreams: 1 下游接口数量。
Expires: 00:02:47 下游接口老化时间。

## RPF 检查

确保组播流量沿正确路径转发，避免环路
基于单播路由表
检查过程：
路由器确认组播报文是从自身连接到组播源的接口上收到的，才进行转发，否则丢弃

| IPv6路由表 | |
|---|---|
| 网段 | 接口 |
| FC00:0:0:2001::1/64 | S0 |
| FC00:0:0:3001::1/64 | S1 |
| FC00:0:0:4001::1/64 | S2 |




路由器收到一份组播报文后，会根据报文的源地址通过单播路由表查找到达“报文源”的路由，查看到“报文源”的路由表项的出接口是否与收到组播报文的入接口一致。如果一致，则认为该组播报文从正确的接口到达，从而保证了整个转发路径的正确性和唯一性。这个过程就被称为 RPF 检查。
如果这几条等价路由都是来自同一张路由表项，则选取下一跳地址最大的路由作为 RPF 路由。
RPF 检验可以基于单播路由、MBGP 路由和组播静态路由。他们之间的优先顺序为组播静态路由、MBGP 路由、单播路由。

拓扑描述
来自组播源 FC00:0:0:2001::1/64 的组播流从 S1 口到达路由器，路由器检查路由表，发现可以转发该组播流的端口为 S0，RPF 检查失败。因此达到 S1 口的数据流被丢弃。
来自组播源 FC00:0:0:2001::1/64 的组播流从 S0 口达到路由器，检查路由表发现入接口与接收该组播流的接口 S0 一致，RPF 检查成功。因此组播流将被正确的转发。
组播路由协议通过已有的单播路由信息来确定上、下游邻居设备，创建组播路由表项。运用 RPF 检查机制，来确保组播数据流能够沿组播分发树（路径）正确的传输，同时可以避免转发路径上环路的产生。
在实际组播数据转发过程中，如果对每一份接收到的组播数据报文都通过单播路由表进行 RPF 检查，会给路由器带来很大

负担。因此，路由器在收到一份来自源 S 发往组 G 的组播数据报文之后，首先会在组播转发表中查找有无相应的（S，G）组播转发表项：
如果不存在（S，G）转发表项，则对该报文执行 RPF 检查，将检查到的 RPF 接口作为入接口，创建组播路由表项，下发到组播转发表中。其中，对 RPF 检查结果的处理方式为：如果检查通过，表明接收接口为 RPF 接口，向转发表项的所有出接口转发；如果检查失败，表明报文来源路径错误，丢弃该报文。
如果存在（S，G）转发表项，并且接收该报文的接口与转发表项的入接口一致，则向所有的出接口转发该报文。
如果存在（S，G）转发表项，但是接收该报文的接口与转发表项的入接口不一致，则对此报文进行 RPF 检查。对 RPF 检查结果的处理方式为：
若 RPF 检查选取出的 RPF 接口与转发表项的入接口一致，则说明（S，G）表项正确，报文来源路径错误，将其丢弃。
若 RPF 检查选取出的 RPF 接口与转发表项的入接口不符，则说明（S，G）表项已过时，于是把表项中的入接口更新为 RPF 接口。然后再根据 RPF 检查规则进行判断：如果接收该报文的接口正是其 RPF 接口，则向转发表项的所有出接口转发该报文，否则将其丢弃。

## 组播负载分担

“负载分担”与“负载均衡”是不同的概念。
“负载分担”是指如果发往某一目的地的数据流存在多条等价的转发路径，就将数据在这多条路径上转发，达到分流的目的。在进行数据转发时，每一条路径上转发的数据流量并不一定相同，转发流量多少需要根据负载分担方式来决定。
“负载均衡”属于“负载分担”的一种特殊形式，不仅将数据流在这多条路径上转发，并且每条路径转发等量的数据流量。

缺省情况下，组播报文转发过程中如果存在多条等价的最优转发路径，按照 RPF 检查对等价路由的处理规则，只会从 IGP 路由表中选取出下一跳地址最大的路由作为 RPF 路由。

## ipv4 地址

配置稳定优先组播负载分担,静态方式加入组播组 225.1.1.1～225.1.1.3。

```
multicast load-splitting stable-preferred

int g0/0/0
igmp static-group 225.1.1.1 inc-step-mask 32 number 3
```

## ipv6 地址

```
multicast ipv6 load-splitting balance-preferred
```

balance-preferred：表示均衡优先负载分担。组播业务频繁加入和退出，要自动调整负载均衡的场景。
group：表示基于组地址进行负载分担。该策略适用于一源多组的场景。
source：表示基于源地址进行负载分担。该策略适用于一组多源的场景。
source-group：表示同时基于源地址和组地址进行负载分担。该策略适用于多个源和多个组的场景。
stable-preferred：表示稳定优先负载分担。该策略适用于组播业务稳定的场景。

接口配置不同的组播负载分担权值，实现组播不均衡负载分担。

```
int g0/0/0
multicast ipv6 load-splitting weight 3
```

int g0/0/1
multicast ipv6 load-splitting weight 2

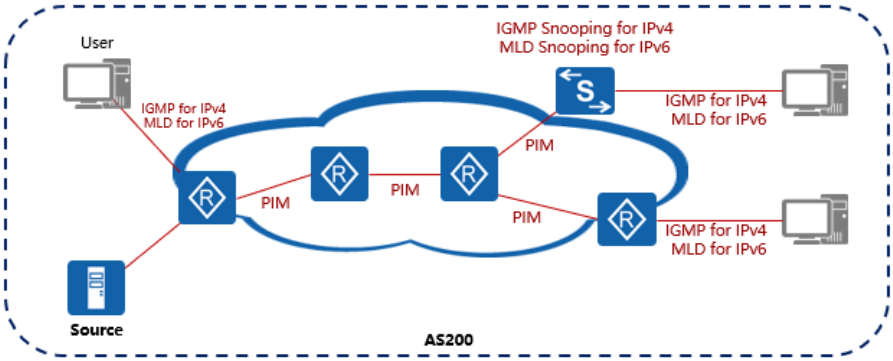
## 前言

- 作为IP传输三种方式之一，IP组播通信指的是IP报文从一个源发出，而被转发到一组特定的接收者。相较于传统的单播和广播，IP组播可以有效地节约网络带宽、降低网络负载，所以组播在IPTV、多媒体会议等诸多方面都有广泛的应用。
- 本章主要介绍与IP组播知识，组播路由和转发原理、各种IP组播协议的主要功能及工作原理，以及各种组播协议的主要应用。

- 现代网络传输技术对以下两项目标给予更高的关注：
- 资源发现
- 点对多点的 IP 传输
- 实现这两项目标有三种解决方案：单播（Unicast）、广播（Broadcast）、组播（Multicast）
- 通过比较三种解决方案的数据传输方式，说明组播方式更适合点对多点的 IP 传输。

## 组播相关协议

- 组播协议包括用于主机注册的组播组管理协议，和用于组播选路转发的组播路由协议。




- 组播协议包括用于主机注册的组播组管理协议，和用于

组播选路转发的组播路由协议。各种组播协议在网络中的如图所示。

- IGMP（Internet Group Management Protocol）在接收者主机和组播路由器之间运行，该协议定义了主机与路由器之间建立和维护组播成员关系的机制。
- 组播路由器之间运行组播路由协议，组播路由协议用于建立和维护组播路由，并正确、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。域内组播路由协议包括：DVRMP、MOSPF、PIM。
- DVRMP 是距离矢量组播路由协议（Distance Vector Multicast Routing Protocol）是一种密集模式协议。该协议有跳数限制，最大跳数 32 跳。
- MOSPF 是 OSPF 路由协议的扩展协议。它通过定义新的 LSA 来支持组播。
- PIM（Protocol Independent Multicast）是典型的域内组播路由协议，分为 DM（Dense Mode）和 SM（Sparse Mode）两种模型。当接收者在网络中的分布较为密集时，适用 DM；较为稀疏时，适用 SM。PIM 必须和单播路由协议协同工作。

## 组播网络中应用的协议

| 协议 | 部署位置 | 作用 |
| --- | --- | --- |
| IGMP，用于 IPv4 网络<br>MLD，用于 IPv6 网络 | 组播路由器与用户主机之间。 | 在主机侧实现组播组成员动态加入与离开。<br>在路由器侧实现组成员 关系的维护与管理，同时与上层组 播路由协议进行信息交互。 |
| PIM | 所有组播路由器上，配置在所有接口上 。 | 实现组播路由与转发，并可以动态响应网络拓扑变化，维护组播路由 表。 |
| IGMP Snooping，用于 IPv4 网络<br>MLD Snooping，用于 IPv6 网络 | 组播路由器和用户主机之间的二层交换机上，配置在VLAN内 。 | 侦听路由器和主机之间发送的 IGMP/MLD 报文建立组播数据的 二层转发表，从而管理和控制组播 数据在二层网络中的转发。 |

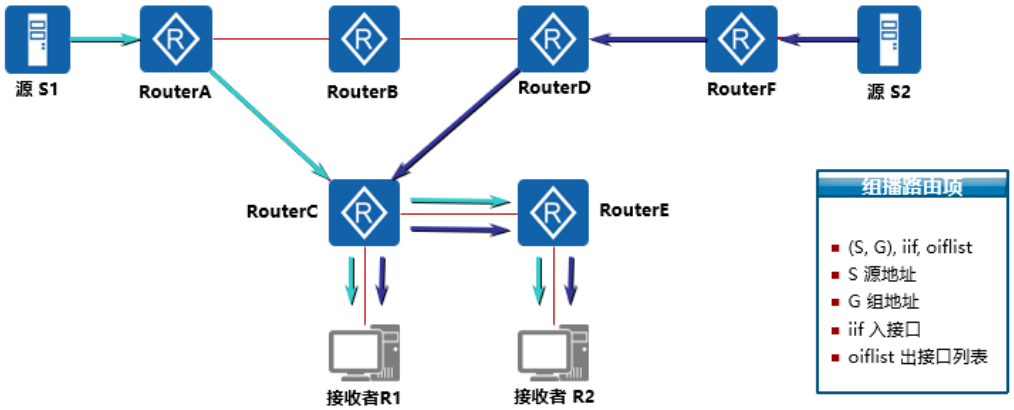
## PIM-SM基本概述

- PIM-SM的关键任务：
  - 建立RPT（Rendezvous Point Tree，汇聚点树也称共享树）。
  - 建立SPT（Shortest Path Tree，最短路径树）。
- 适用于组播成员分布较为稀疏的网络环境。

- 相对于 PIM-DM 的“推（Push）模式”，PIM-SM 使用“拉（Pull）模式”转发组播报文。PIM-SM 假设网络中的组成员分布非常稀疏，几乎所有网段均不存在组成员，直到某网段出现组成员时，才构建组播路由，向该网段转发组播数据。一般应用于组播组成员规模相对较大、相对稀疏的网络。
- 基于这一种稀疏的网络模型，它的实现方法是：
- 在网络中维护一台重要的 PIM 路由器：汇聚点 RP（Rendezvous Point），可以为随时出现的组成员或组播源服务。网络中所有 PIM 路由器都知道 RP 的位置。
- 当网络中出现组成员（用户主机通过 IGMP 加入某组播组 G）时，最后一跳路由器向 RP 发送 Join 报文，逐跳创建（*，G）表项，生成一棵以 RP 为根的 RPT。
- 当网络中出现活跃的组播源（信源向某组播组 G 发送第一个组播数据）时，第一跳路由器将组播数据封装在 Register 报文中单播发往 RP，在 RP 上创建（S，G）表项，注册源信息。
- PIM-SM 的关键机制包括邻居建立、DR 竞选、RP 发现、RPT 构建、组播源注册、SPT 切换、Assert；同时也可通过配置 BSR（Bootstrap Router）管理域来实现单个 PIM-SM 域的精细化管理。PIM-SM 中 PIM 邻居建立过程以及 Assert 机制与 PIM-DM 相同。

## 源路径树

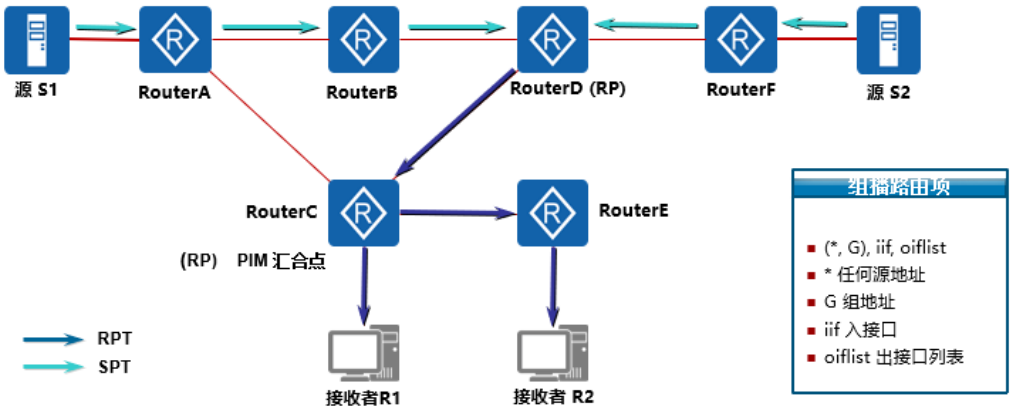
- 每一个组播源与接收者之间建立一棵独立的SPT。




- 源路径树以组播源作为树根，将组播源到每一个接收者的最短路径结合起来构成的转发树。
- 源路径树使用的是从组播源到接收者的最短路径，也称为最短路径树（shortest path tree，SPT）。对于某个组，网络要为任何一个向该组发送报文的组播源建立一棵树。
- 本例中有两个组播源（源 S1 和源 S2），接收者 R1 和 RT2。所以本例中有两棵源路径树，分别是：
- S1—A---C（R1）-----E（R2）
- S2---F----D---C（R1）------E（R2）

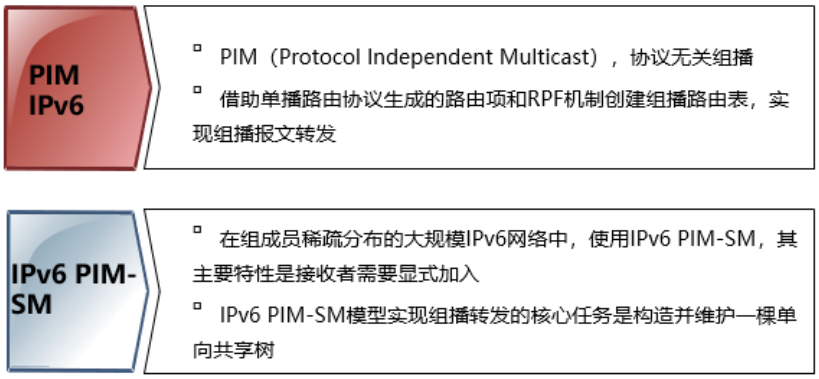
## 共享树

- 对应某个组，网络中只有一棵树。



- 共享树以某个路由器作为路由树的树根，该路由器称为汇集点（Rendezvous Point，RP），将 RP 到所有接收者的最短路径结合起来构成转发树。使用共享树时，对应某个组，网络中只有一棵树。所有的组播源和接收者都使用这棵树来收发报文，组播源先向树根发送数据报文，之后报文又向下转发到达所有的接收者。
- 本例中两个源 S1 和 S2 共享一颗树 D（RP）----C（R1）----E（R2）

## IPv6 PIM-SM概述

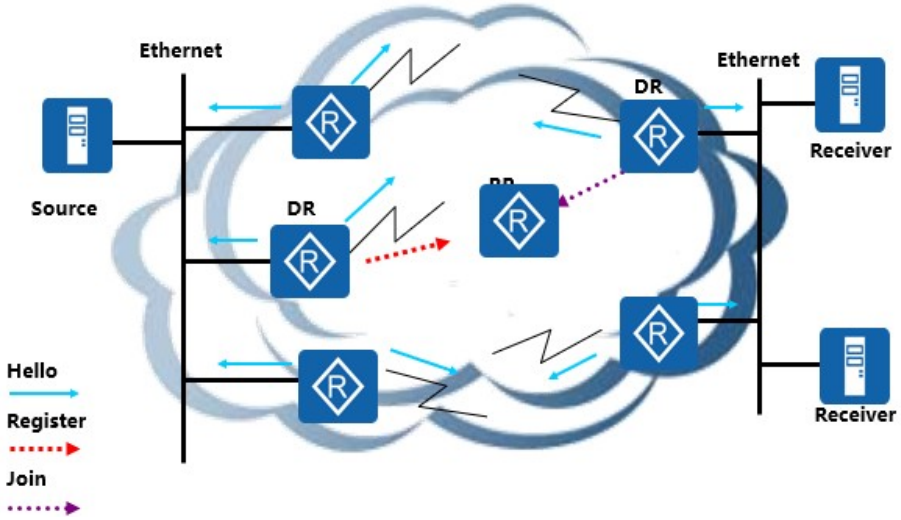



- PIM IPv6 是与静态路由、RIPng、OSPFv3、IS-ISv6、BGP4+等 IPv6 单播路由协议类型无关的组播路由协议，其借助上述单播路由协议生成的路由项和 RPF（Reverse Path Forwarding）机制创建组播路由表，实现组播报文转发。PIM IPv6 域是指由支持 PIM IPv6 协议的组播路由器构成的网络。
- 目前，存在两种组播模型：任意源组播 ASM（Any-Source Multicast）和指定源组播 SSM（Source-Specific Multicast）。IPv6 中，ASM 模型包括 IPv6 PIM-DM（Protocol Independent Multicast-Dense Mode）和 IPv6 PIM-SM（Protocol Independent Multicast Sparse Mode）；SSM 模型则使用 MLDv2 和 IPv6 PIM-SM 的部分机制来实现。
- 在组成员稀疏分布的大规模 IPv6 网络中，使用 IPv6 PI

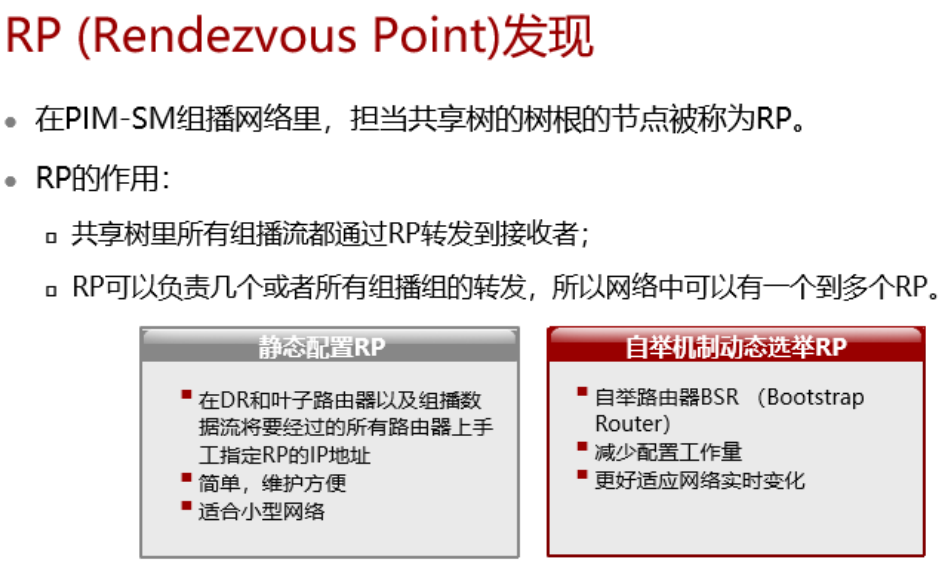
M-SM，其主要特性是接收者需要显式加入。缺省情况下，IPv6 PIM-SM 假设网络中的所有节点都不需要接收组播报文，上游节点只有接收到下游节点的加入消息后才进行组播数据的转发。

- IPv6 PIM-SM 中，汇集点 RP（Rendezvous Point）只向拥有接收者的下游分枝转发组播信息。这样可以节省数据报文和控制报文占用的网络带宽，减少路由器的处理开销。
- 当主机希望从指定组播组接收数据时，与之相连的路由器向这个组的 RP 发送加入（Join）消息，沿途建立以 RP 为根节点的共享树 RPT（Rendezvous Point Tree）。共享树的含义是不同组播源向相同组播组转发组播数据时，都使用此共享路径。
- 当组播源向组播组发送数据时，与源相连的 DR 把组播数据封装在注册消息中以单播方式向 RP 发送。注册消息到达 RP 后，由 RP 对组播数据解封装，再沿 RPT 向接收者发送。当以注册消息方式发送的组播数据达到一定速率后，RP 向组播源发送加入消息，建立组播源与 RP 之间的组播分发树。然后，RP 向组播源的 DR 发送注册停止消息，指示 DR 直接以非封装方式根据组播转发表发送组播数据。

# DR选举




- 借助 Hello 消息可以为共享网络（如 Ethernet）选举 DR（Designated Router），DR 将作为本网段中组播信息的唯一转发者。无论是和组播源 S 连接的网络，还是和接收者连接的网络，只要网络为共享媒介则需要选举 DR，接收者侧 DR 向 RP 发送 Join 加入消息；组播源侧 DR 向 RP 发送 Register 注册消息。
- DR 的选举过程如下所述：
- 共享网段上的各路由器相互间发送带有 DR 优先级选项的 Hello 消息；
- 具有最高优先级的路由器被选举为此网段的 DR。如果路由器具有相同的优先级，则 IPv6 地址最大的路由器被选举为 DR。
- 当 DR 出现异常，其他路由器将接收不到其发出的 Hello 消息。在此 DR 超时后，会触发共享网段上新一轮的 DR 选举。
- 如果网络中至少有一台路由器不支持在 Hello 报文中携带 DR 优先级，由 IPv6 链路本地地址最大的路由器充当 DR。

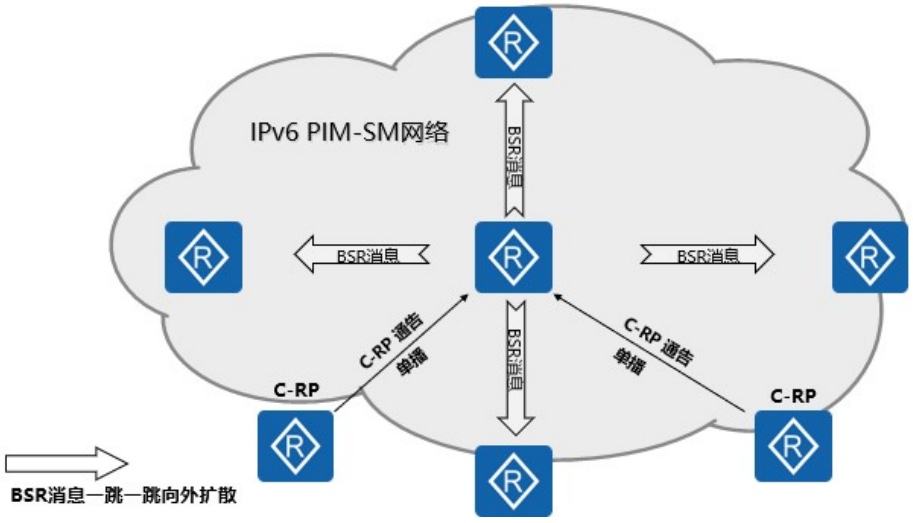



- 如何发现 RP？对于小规模的简单网络，一个 RP 用于全网转发信息就足够了，此 RP 的位置可通过静态指定，在 DR 和叶子路由器以及组播数据流将要经过的所有路由器上手工指定 RP 的 IP 地址。然而，在大多数应用中，IPv6 PIM-SM 网络覆盖了很大的区域，需要通过 RP 转发大量的组播流量。因此，不同的组播组应该具有各自的 RP。为了减少配置多个静态 RP 的工作量以及更好的适应网络的实时变化，采用自举（Bootstrap）机制来动态选举 RP。
- 自举路由器 BSR（BootStrap Router）是 IPv6 PIM-SM 网络的管理核心。BSR 收集来自各候选 RP 即 C-RP（Candidate-RP）的通告（Advertisement）消息，选择合适的 C-RP 来构成组播组的 RP-Set 信息。RP-Set 是各组播组与对应 C-RP 的数据库。BSR 通过自举消息向整个 IPv6 PIM-SM 网络通告 RP-Set 信息。包括 DR 在内的所有路由器学习到各组播组对应的 C-RP 后，根据哈希算法计算出各组播组对应的唯一 RP。
- 一个网络（或一个管理域）只能有一个 BSR，但可以有多个候选 BSR 即 C-BSR（Candidate-BSR）。一旦 BSR 发生故障后，可以通过自举机制从 C-BSR 中选举出新的 BSR，

从而避免业务的中断。IPv6 PIM-SM 域中可以配置多个 C-RP。BSR 负责收集并发送各组播组的 RP-Set 信息。

- RP 配置方式建议：
- 中小型网络：建议选择静态 RP 方式，对设备要求低，也比较稳定。
- 如果网络中只有一个组播源，建议选择直连组播源的设备作为静态 RP，这样可以省略源端 DR 向 RP 注册的过程。
- 采用静态 RP 方式要确保域内所有路由器（包括 RP 本身）的 RP 信息以及服务的组播组范围全网一致。
- 大型网络：可以采用动态 RP 方式，可靠性高，可维护性强。
- 如果网络中存在多个组播源，且分布密集，建议选择与组播源比较近的核心设备作为 C-RP；如果网络中存在多个用户，且分布密集，建议选择与用户比较近的核心设备作为 C-RP。

RP选举



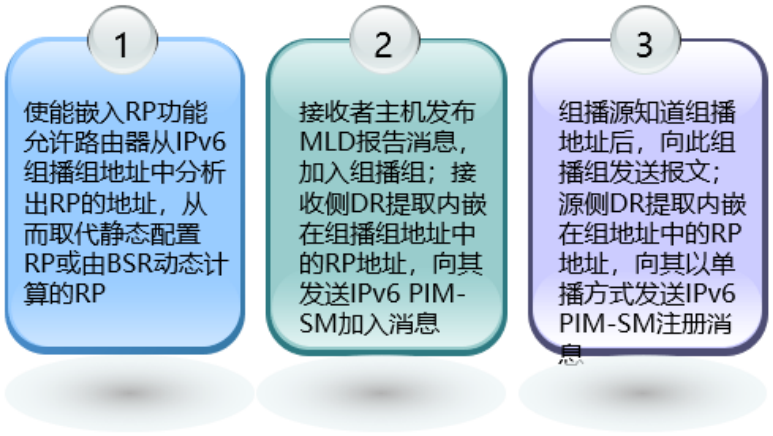

- BootStrap router 工作的原理和过程：首先要在网络中选择合适的路由器把它配置成候选 BSR（C-BSR，Candidate BootStrap Router），每个 C-BSR 都有优先级，当它得知自己是 C-BSR 后，首先启动一个定时器（默认为 150 秒），监听

网络中的 BootStrap Message。BootStrap Message 初始时通告发送路由器的优先级、BSR 的 IPv6 地址，当 C-BSR 收到一个 BootStrap Message 后，它会把自己的优先级和报文里的优先级做比较，如果对方的优先级高，它就把自己的定时器重置，继续监听 BootStrap Message；如果是自己的高，那么它就发送 BootStrap Message 声明自己是 BSR，如果优先级相等，则比较 IPv6 地址，谁的 IPv6 地址大谁就是 BSR。BSR 消息发送的目的地址是 FF02::13，所有的 PIM IPv6 路由器都能接收到这个报文，该报文 TTL 一般被置为 1，但每个 PIM IPv6 路由器收到此报文后都是把它以泛洪的方式从自己所有的使能 PIM IPv6 的接口上发送出去，这就能保证网络中的每台 PIM IPv6 设备都能收到 BootStrap Message。

- RP 就像 C-BSR 一样需要在设备上手工配置，首先配置 C-RP（Candidate Rendezvous Point），包括 RP IPv6 地址、优先级和它所能服务的组。正如上文所述，一个 RP 可以给所有的 IPv6 组播组提供服务，也可以只给部分组提供服务。当 C-RP 收到 BootStrap Message 后，它可以从该 message 中得知网络中谁是 BSR，然后 C-RP 通过 Candidate-RP-Advertisement Message 把自己所能服务的组单播给 BSR，每个 C-RP 都这么做的话那么 BSR 就收集到了网络中所有 C-RP 的信息并把这些信息整理成一个集 RP-Set。此后 BSR 通过 BootStrap Message 把 RP-Set 的信息通告给全网所有的路由器。
- RP 的选举规则：
- 如果 RP-Set 对应该 IPv6 组地址的 C-RP 只有一个,那么 DR 就选该 C-RP 做 RP；
- 如果对应该 IPv6 组地址的有多个 C-RP，那么优先级最高的是 RP（优先级数越小优先级越高）；
- 如果大家优先级相等，那么 DR 将开始 Hash 运算，把组地址、hash 掩码、和 C-RP 的地址做为输入参数，输出是一些数字，数字高的 C-RP 将是该组的 RP；

- 如果 hash 的结果大家也相等，那么 IPv6 地址最大的 C-RP 将成为该组的 RP。

## 嵌入式RP的原理




- 使能嵌入 RP 功能允许路由器从 IPv6 组播组地址中分析出 RP 的地址，从而取代静态配置 RP 或由 BSR 动态计算的 RP。
- 使用嵌入式 RP 的组播组地址范围是 FF7x::/16 和 FFFx::/16，x 表示 0～F 的任意一个十六进制数。
- 在接收侧：
- 接收者主机发布 MLD 报告消息，加入组播组；
- 接收侧的 DR 提取内嵌在组播组地址中的 RP 地址，向其发送 IPv6 PIM-SM 加入消息。
- 在组播源侧：
- 组播源知道组播地址后，向此组播组发送报文；
- 组播源侧的 DR 提取内嵌在组播地址中的 RP 地址，向其以单播方式发送 IPv6 PIM-SM 注册消息。

## 嵌入RP地址的组播组地址

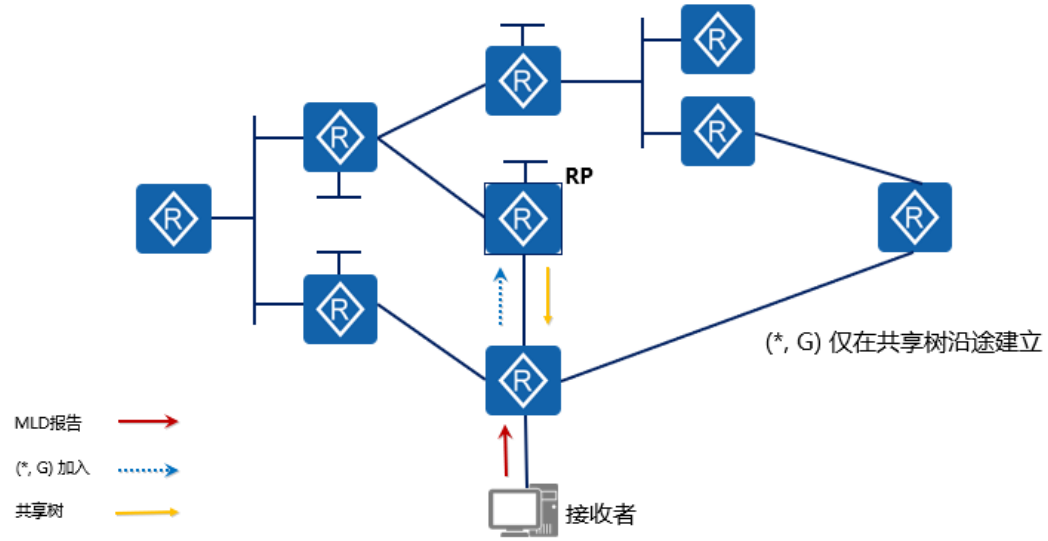
- 一个128bits的RP地址如何嵌入到一个128bits的IPv6组播组地址中去?
- 定义特殊的组播地址:


- 头 8bits 为 FF 说明是 IPv6 组播地址。
- Flags 字段的范围是 7-F，说明是一个嵌入了 RP 地址的 IPv6 组播组地址。
- RIID 字段：RP Interface ID，抽取出来填充在 RP 地址的最后 4bits。
- Plen 字段：RP 地址的前缀长度，换算成十进制数后不能为 0，也不能大于 64。
- 字段：RP 的地址前缀。
-

## RP地址转换

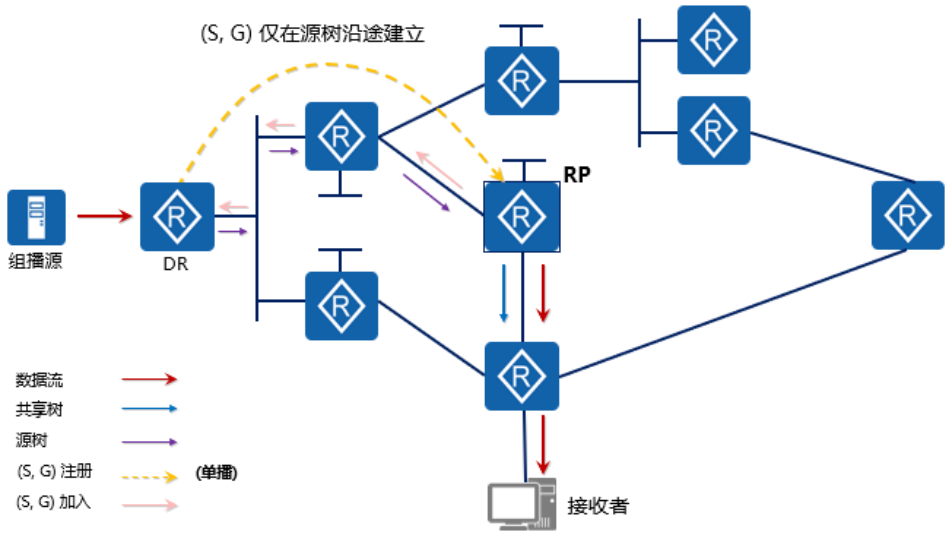
- 如何由组播组地址得到RP地址?
  - 提取“plen”字段，转换为十进制数
  - 将“Network Prefix”字段的前“plen” bits提取出来作为RP地址的地址前缀
  - 将“RIID”字段提取出来作为RP地址的Interface ID的最后4bits，Interface ID其余部分用0补齐
- 嵌入RP实例
  - 组播地址FF70:140:2001:DB8:BEEF:FEED::/96，则从组播地址中获取的RP地址为2001:DB8:BEEF:FEED::1/64

## 加入共享树



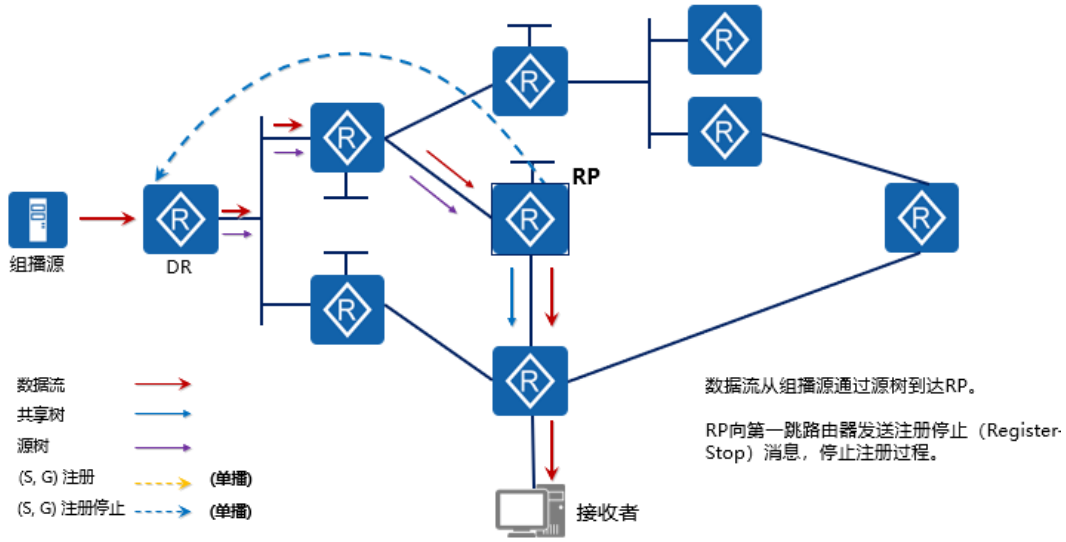

- 当接收者主机加入一个组播组 G 时，通过 MLD 报文知会与该主机直接相连的叶子路由器，叶子路由器掌握组播组 G 的接收者信息，然后朝着 RP 方向往上游节点发送（*，G）的 join 消息。
- 从叶子路由器到 RP 之间途经的每个路由器都会在转发表中生成（*，G）表项，这些沿途经过的路由器就形成了 RP 共享树（RPT）的一个分支。其中（*，G）表示从任意源来的信息去往组播组 G。RPT 共享树以 RP 为根，以接收者为叶子。
- 当从组播源 S 来的发往组播组 G 的报文流经 RP 时，报文就会沿着已经建立好的 RPT 共享树路径到达叶子路由器，进而到达接收者主机。
- 当某接收者对组播信息不再感兴趣时，离该接收者最近的组播路由器会逆着 RPT 树朝 RP 方向逐跳发送 Prune 剪枝消息。第一个上游路由器接收到该剪枝消息，在接口列表中删除连接此下游路由器的接口，并检查自己是否拥有感兴趣的接收者，如果没有则继续向上游转发该剪枝消息。

## 组播源注册



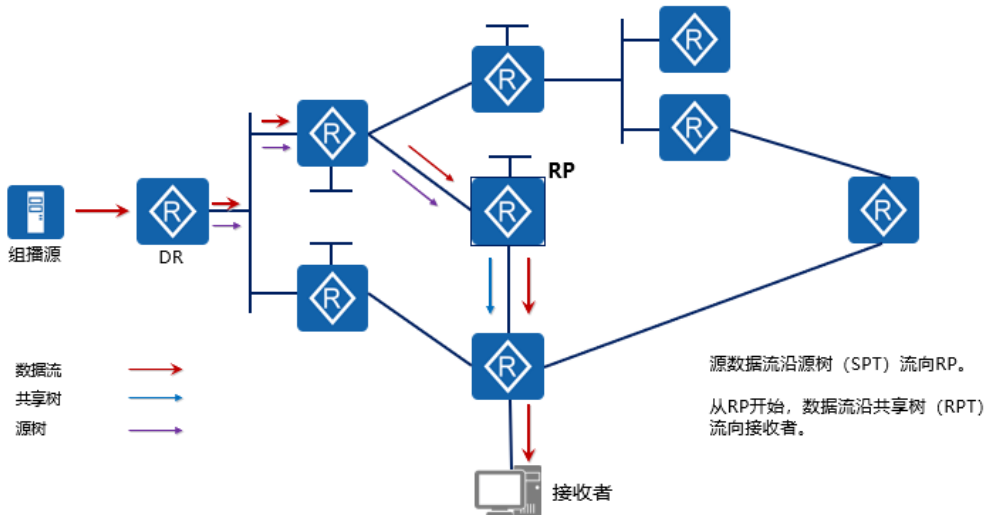

- 为了向 RP 通知组播源 S 的存在，当组播源 S 向组播组 G 发送了一个组播报文时，与组播源 S 直接相连的路由器接收到该组播报文后，就将该报文封装成 IPv6 PIM Register 注册报文，并单播发送给对应的 RP。
- 当 RP 接收到来自组播源 S 的注册消息后，一方面解封装注册消息并将组播信息沿着 RPT 树转发到接收者，另一方面朝组播源 S 逐跳发送（S，G）加入消息，从而让 RP 和组播源 S 之间的所有路由器上都生成了（S，G）表项，这些沿途经过的路由器就形成了 SPT 树的一个分支。SPT 源树以组播源 S 为根，以 RP 为目的地。

## 停止注册过程

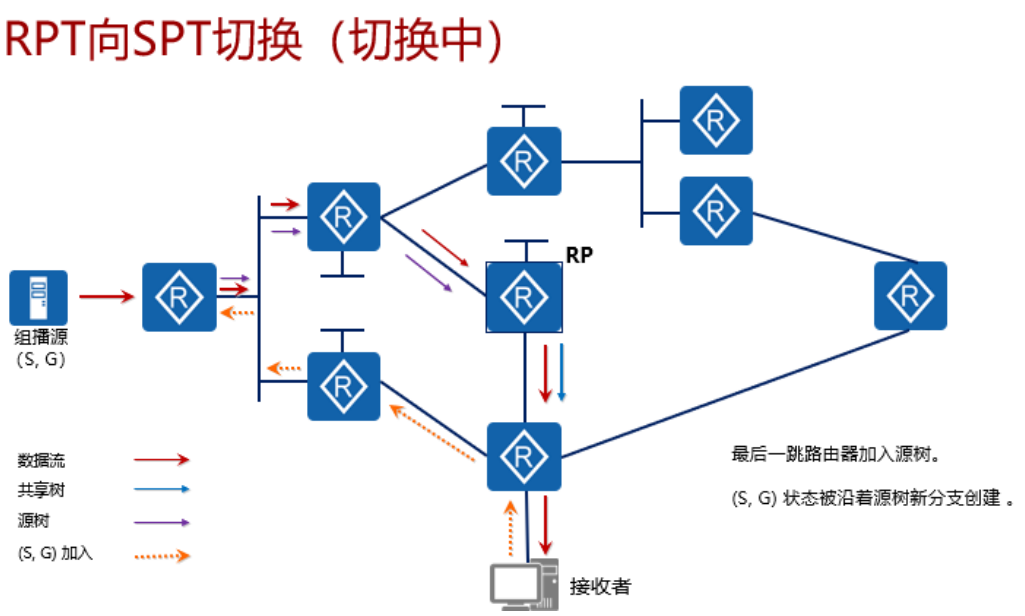



- 组播源 S 发出的组播信息沿着已经建立好的 SPT 树到达 RP，然后由 RP 将信息沿着 RPT 共享树进行转发。当 RP 收到沿着 SPT 树转发的组播流量后，向与组播源 S 直连的路由器单播发送注册停止报文。组播源注册过程结束。

## 组播流转发过程

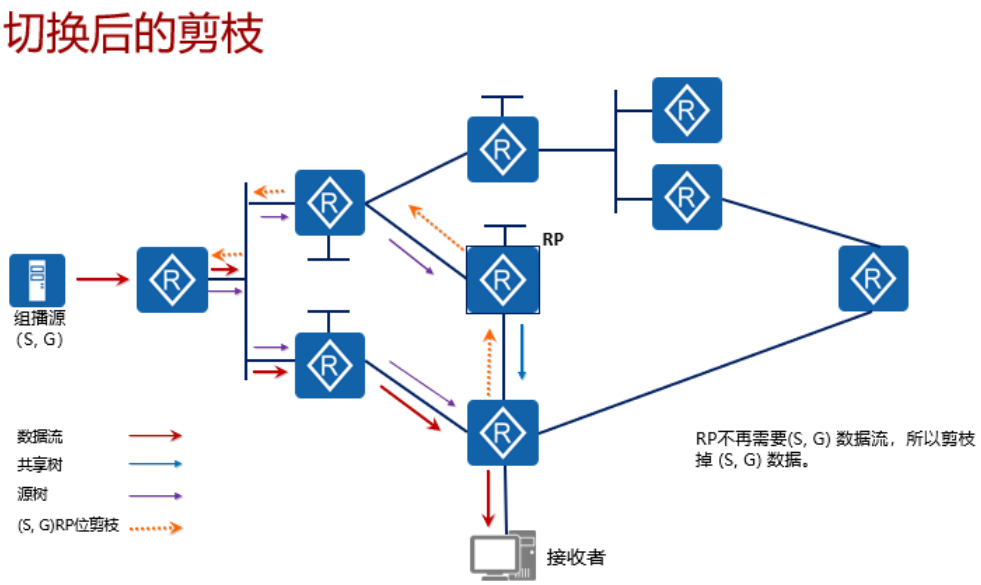



- 源数据流延源树（SPT）流向 RP，从 RP 开始，数据流延共享树（RPT）流向接收者。



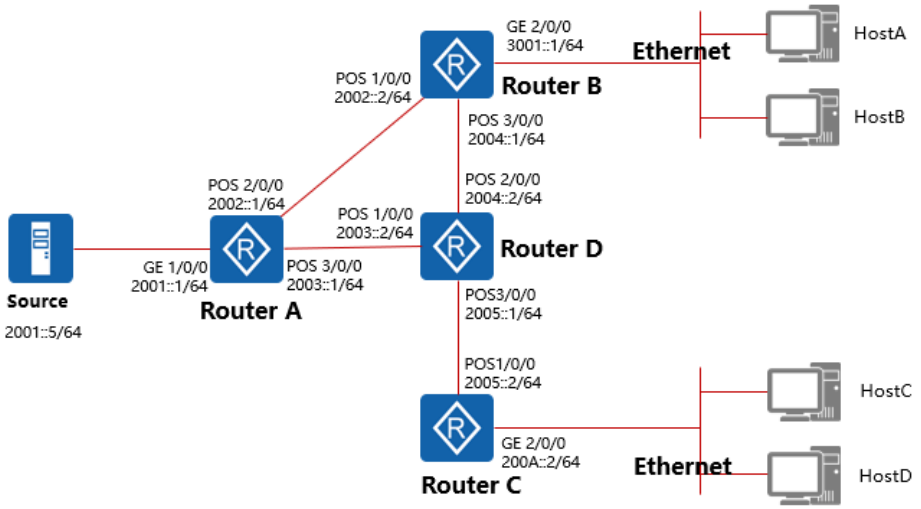

- 针对特定的源，PIM-SM 通过指定一个利用带宽的 SPT 阈值可以实现将最后一跳路由器(即离接收者最近的 DR)从 RPT 切换到 SPT。当最后一跳路由器发现从 RP 发往组播组 G 的组播报文速率超过了该阈值时，就向单播路由表中到组播源 S 的下一跳路由器发送（S，G）加入消息，Join 加入消息经过一个个路由器后到达第一跳路由器（即离组播源最近的 DR），沿途经过的所有路由器都拥有了（S，G）表项，从而建立了 SPT 树分支。
- 用户端 DR 周期性检测组播报文的转发速率，一旦发现从 RP 发往组播组 G 的报文速率超过阈值，则触发 SPT 切换：
- 用户端 DR 逐跳向源端 DR 发送（S，G）Join 报文并创建（S，G）表项，建立源端 DR 到用户端 DR 的 SPT。
- SPT 建立后，用户端 DR 会沿着 RPT 逐跳向 RP 发送剪枝报文，收到剪枝报文的路由器将（*，G）复制成相应的（S，G）,并将相应的下游接口置为剪枝状态。剪枝结束后，RP 不再沿 RPT 转发组播报文到组成员端。
- 如果 SPT 不经过 RP，RP 会继续向源端 DR 逐跳发送剪枝报文，删除（S，G）表项中相应的下游接口。剪枝结束后，源端 DR 不再沿“源端 DR-RP”的 SPT 转发组播报文到 RP。

- 在 VRP 中，缺省情况下连接接收者的路由器在探测到组播源之后（即接收到第一个数据报文），便立即加入最短路径树，即从 RPT 向 SPT 切换。



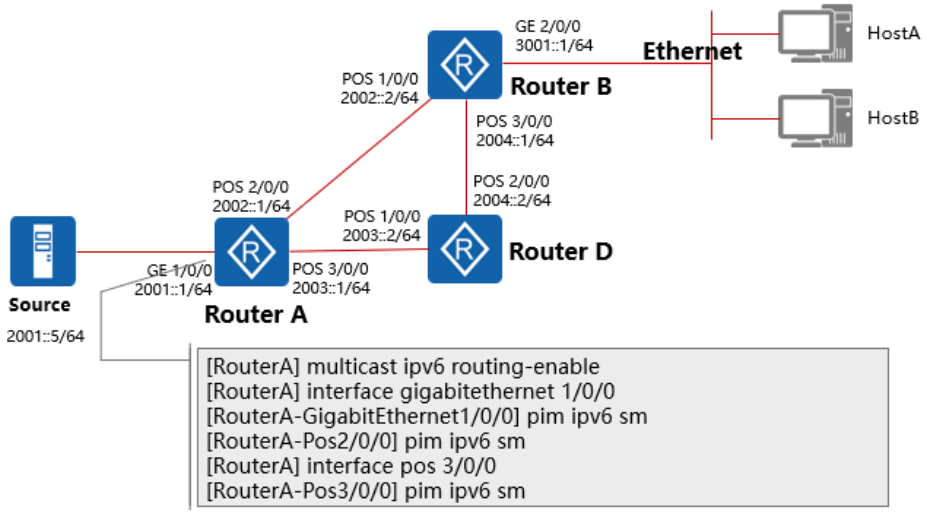

- 当路由器在不同接口接收到 RPT 和 SPT 两条路径上传输的相同组播数据时，丢弃沿 RPT 接收的数据，并向 RP 逐跳发送剪枝消息。RP 接收到剪枝消息后，更新转发状态，并停止沿 RPT 转发（S, G）的组播流量；同时 RP 向组播源发送剪枝消息删除或更新相关的（S, G）转发项。通过这种方法，组播数据从 RPT 切换到 SPT。

# IPv6 PIM-SM配置实例



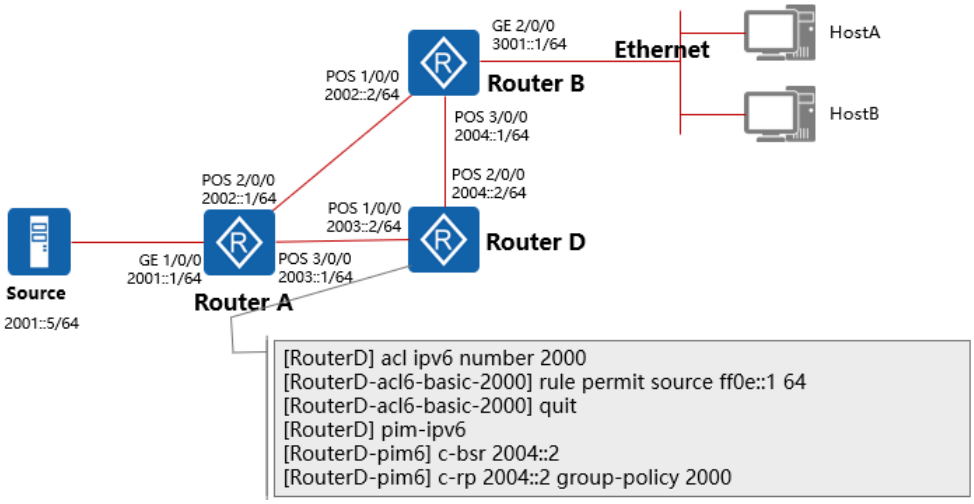

- Host A 和 Host C 分别是两个叶子网络中的组播信息接收者。这些接收者通过 Router A、Router B、Router C 和 Router D 连接到组播源。
- 配置思路：
- 配置各路由器接口的 IPv6 地址和 IPv6 单播路由协议。
- 配置各路由器接口的 IPv6 地址和掩码。
- 配置各路由器之间采用 OSPFv3 进行互连，进程号为 1，区域号为 0，确保网络中各路由器 Router A、Router B、Router C 和 Router D 之间能够在网络层互通。
- 在各路由器上使能 IPv6 组播功能，在路由器各接口上使能 IPv6 PIM-SM，主机侧接口上配置 MLD，采用缺省版本 2。
- 配置 C-BSR 和 C-RP，充当 C-BSR 和 C-RP 的 IPv6 全球单播地址，此例中都为 Router D 上的 2004::2。
- 检查配置结果。

# 配置Router A




- 进入系统视图
- **system-view**
- 使能 IPv6 组播路由
- **multicast ipv6 routing-enable**
- 进入接口视图
- **interface** *interface-type interface-number*
- 使能 IPv6 PIM-SM 功能
- **pim ipv6 sm**
- IPv6 PIM-SM 只有在使能 IPv6 组播后才能配置。在接口上配置 IPv6 PIM-SM 后，路由器周期性发送 Hello 消息来发现 PIM IPv6 邻居，并处理邻居发来的报文。当路由器加入 IPv6 PIM-SM 域时，建议在非边界路由器的所有接口上使能 IPv6 PIM-SM。
- 不能在一个接口上同时使能 IPv6 PIM-SM 和 IPv6 PIM-DM。同一路由器上所有接口的 PIM IPv6 模式必须相同。
- Router B、Router C 和 Router D 上的配置过程与 Router A 上的配置相似。
- 与接受者相连的路由器上需要使能 MLD。

## 配置Router D




- 在 PIM 域中可配置一个或多个 C-BSR。从 C-BSR 选举出的 BSR 负责收集和通告 C-RP 信息。由于 BSR 和域中的其他设备需要交换大量信息，因此 C-BSR 与域中的其它设备之间应预留较大的带宽。所以由骨干网的路由器来充当 C-BSR。
- 在指定本路由器的某接口地址作为 C-BSR 时，必须同时使能接口的 IPv6 PIM-SM。C-BSR 之间自动选举 BSR 的过程简要描述如下：
- 初始时，每个 C-BSR 都认为自己是 PIM-SM 域内的 BSR，使用自己接口的 IPv6 地址作为 BSR 地址来发送自举报文。
- 当 C-BSR 从其他路由器接收到自举报文时，把报文中的 BSR 地址与自己的 BSR 地址相比较。比较的标准包括优先级和 BSR 地址。如果优先级相同，优选 BSR 地址大的，即如果接收的自举报文中的 BSR 地址更高，则用此地址来代替自己的 BSR 地址，并不再认为自己是 BSR；如果自举报文的 BSR 地址并不大于自己的 BSR 地址，则继续认为自己是 BSR。
- 进入 PIM IPv6 视图
- **pim-ipv6**
- 配置自己的接口地址为 C-BSR
- **c-bsr** *ipv6-address* [ *hash-length* ] [ *priority-value* ]

- 配置 C-RP
- **c-rp** *ipv6-address* [ **priority** *priority* ]
- 配置静态 RP
- **static-rp** *rp-address* [ *basic-acl6-number* ] [ **preferred** ]
- 配置嵌入式 RP
- **embedded-rp [** *basic-acl6-number* **]**
- **c-bsr** 命令用来在希望自己成为 BSR 的路由器上配置自身某接口地址为 C-BSR 的地址。
- *ipv6-address*：指定候选自举路由器 C-BSR 的 IPv6 全球单播地址。
- *hash-length*：指定计算 RP 的哈希函数的掩码长度。整数形式，取值范围是 0～128。
- **pim ipv6 bsr-boundary** 命令用来设置作为 BSR 域边界的接口。在接口上配置此命令后，自举消息不能通过此接口，而其他 PIM 报文可以通过。
- **c-rp** 命令用来配置路由器向 BSR 通告自身是 C-RP。配置成为 C-RP 的路由器和其他设备之间需要保留相对大的带宽。通过配置希望成为 RP 的接口地址，并选择合适的优先级。
- *ipv6-address*：指定 C-RP 的 IPv6 全球单播地址。
- 如果网络中仅有一个动态 RP，配置静态 RP 能避免由于单个 RP 发生故障引起的通信中断。使用静态 RP 转发组播数据时，应该在 IPv6 PIM-SM 域中所有路由器上配置完全相同的静态 RP 命令。
- **static-rp** 命令用来配置静态 RP。
- *rp-address*：指定静态 RP 地址。此地址必须是有效的 IPv6 全球单播地址。
- *basic-acl6-number*：指定用于控制静态 RP 服务的组播组范围的基本访问控制列表号。取值范围是 2000～2999。
- **preferred**：表示配置的静态 RP 和由 BSR 机制选择的 R

P 不同时，优先选择此静态 RP。不指定此参数时，优先选择 BSR 机制选择的 RP。

- 嵌入式 RP 用于路由器从组播地址中获取 RP 地址，从而取代静态 RP 或从 BSR 机制选举的动态 RP。使用嵌入式 RP 的组播地址范围是 FF7x::/16～FFFx::/16，x 表示 0～F 的任意一个十六进制数。

## 查看接口的PIM配置和BSR选举信息

```
[RouterB] display pim ipv6 interface
Vpn-instance: public net
Interface    NbrCnt HelloInt   DR-Pri     DR-Address
Pos1/0/0       1     30         1      FE80::A01:10E:1(local)
GEth2/0/0      0     30         1       FE80::200:AFF:FE01:10E(local)
Pos3/0/0       1     30         1      FE80::9D62:0:FDC5:2
```

```
[RouterB] display pim ipv6 bsr-info
  Vpn-instance: public net
    Elected BSR Address: 2004::2
      Priority: 0
      Hash mask length: 126
      State: Accept Preferred
      Uptime: 00:04:22
      Expires: 00:01:46
```

- 查看接口上的 PIM IPv6 信息
- **display pim ipv6 interface** [ *interface-type interface-number* ]
- 查看 IPv6 PIM-SM 域中 BSR 自举路由器的信息
- **display pim ipv6 bsr-info**
- 此网络中的 BSR 是 Router D 的 POS2/0/0 接口。

## 查看RP信息

```
[RouterB] display pim ipv6 rp-info
  Vpn-instance: public net
    PIM-SM BSR RP information:
    Group/MaskLen: FF0E::1/64
      RP: 2004::2
      Priority: 0
      Uptime: 00:05:19
      Expires: 00:02:11
```

- 查看 IPv6 PIM-SM 域中的 RP 信息
- **display pim ipv6 rp-info** [ *ipv6-group-address* ]
- 此网络中的 RP 是 Router D 的 POS2/0/0 接口。

## 查看IPv6组播路由信息 - 源侧DR

```
[RouterA] display pim ipv6 routing-table
Vpn-instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry
(2001::5, FF0E::1)
    RP: 2004::2
    Protocol: pim-sm, Flag: SPT LOC ACT
    UpTime: 00:02:15
    Upstream interface: GigabitEthernet1/0/0
        Upstream neighbor: FE80::200:AFF:FE01:10D
        RPF prime neighbor: FE80::200:AFF:FE01:10D
    Downstream interface(s) information:
    Total number of downstreams: 3
        1: Register
            Protocol: pim-sm, UpTime: 00:02:15, Expires:  -
        2: Pos2/0/0
            Protocol: pim-sm, UpTime: 00:02:15, Expires: 00:03:15
        3: Pos3/0/0
            Protocol: pim-sm, UpTime: 00:02:15, Expires: 00:03:15
```

- 假设 HostA 加入组 G（FF0E::1），RouterD 和 RouterB 之间建立 RPT 树，在 RPT 路径上的路由器（RouterD 和 RouterB）生成（*, G）项。组播源 S（2001::5）向组播组 G 发送组播报文后，在源树路径上的路由器（RouterA 和 RouterD）生成（S, G）项。

## 查看IPv6组播路由信息 - RP

```
[RouterD] display pim ipv6 routing-table
Vpn-instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::1)
     RP: 2004::2 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:16:56
     Upstream interface: Register
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 2
         1: Pos2/0/0
             Protocol: pim-sm, UpTime: 00:16:56, Expires: 00:02:34
         2: Pos3/0/0
             Protocol: pim-sm, UpTime: 00:07:56, Expires: 00:02:35
```

## 查看IPv6组播路由信息 - RP(续)

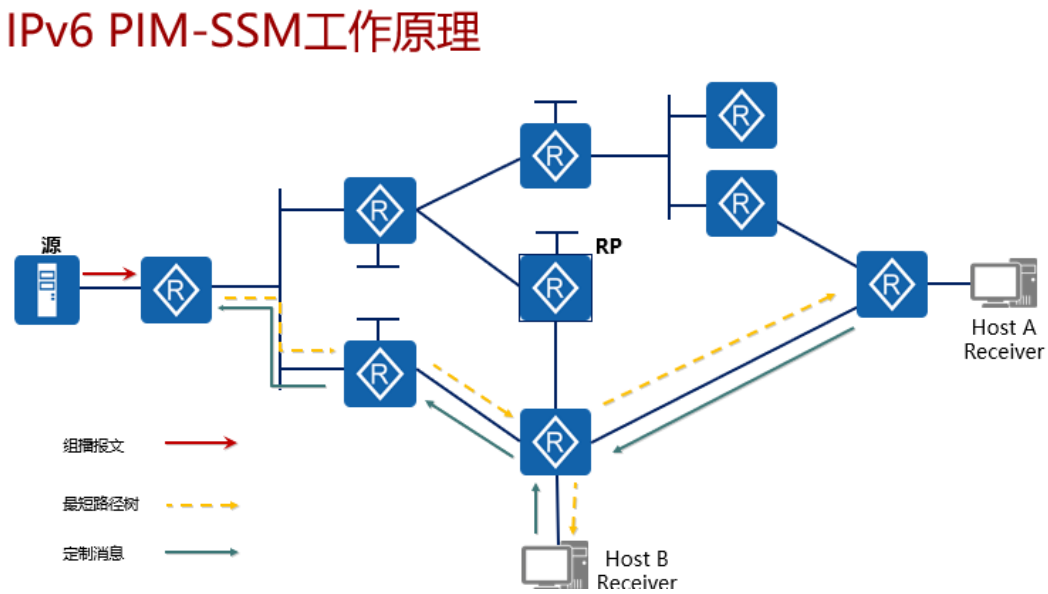
```
(2001::5, FF0E::1)
    RP: 2004::2 (local)
    Protocol: pim-sm, Flag: SWT ACT
    UpTime: 00:02:54
    Upstream interface: Register
        Upstream neighbor: NULL
        RPF prime neighbor: NULL
    Downstream interface(s) information:
    Total number of downstreams: 2
        1: Pos2/0/0
            Protocol: pim-sm, UpTime: 00:02:54, Expires:  -
        2: Pos3/0/0
            Protocol: pim-sm, UpTime: 00:02:54, Expires: 00:02:36
```

## IPv6 PIM-SSM概述

- SSM模型提供了指定源组播的解决方案，配合MLDv2采用IPv6 PIM-SM的部分机制来实现。由于最后一跳路由器通过MLDv2协议已经知道了组播源的地址，可以直接发起指定源-组的加入过程，在SSM网络中创建组播源到接收者的SPT
  - 定义了特殊的组播地址：FF3x::/32，不存在源发现问题
  - 需要和MLDv2配合使用
  - 扩展了PIM SM协议，PIM-SSM不涉及RP、BSR、RPT生成、组播源注册等复杂机制
  - 基于组播源的单播路由直接生成SPT树，可以实现跨域组播

- IPv6 PIM-SSM 的实现可以概括成邻居发现、DR 选举和 SPT 生成：

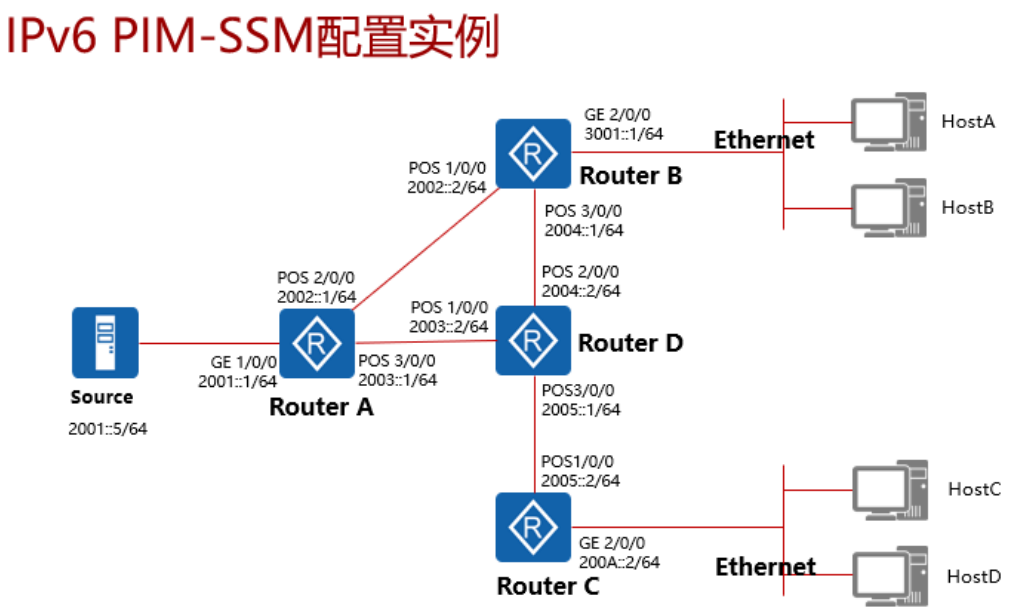
- 邻居发现和 DR 选举过程与 IPv6 PIM-SM 中的描述相同，是通过在路由器间发送 Hello 消息来实现的；
- 由于 PIM-SSM 也使用 PIM-SM 协议，路由器生成 RPT 还是生成 SPT 的判决取决于组播地址是否在 SSM 组地址范围内。




- SSM 模型中，用信道（Channel）概念来表示（S, G）组合，用定制（Subscribed）消息概念来表示加入消息。
- 假定网络中的 User A 和 User B 需要接收组播源 S 的信息，就通过 MLDv2 向最近的查询器发送一个标为（include S, G）的报告信息。如果 User A 和 User B 不需要接收组播源 S 的信息，发送一个标为（exclude S, G）或包含其他组播源的报告消息。无论使用上述哪个报告消息，接收者是明确指定组播源 S 的。
- 接收到报告消息的查询器检查此消息的组播地址是否在 SSM 组地址的范围内。如果是，则路由器根据 SSM 模型建立组播分发树，随后向指定源逐跳发送定制消息（也称加入消息）。沿途上的所有路由器创建（S, G）项。以源 S 为根节点、接收者为叶子的 SPT 树就生成了。SSM 模型使用此 SPT

树作为传输路径。

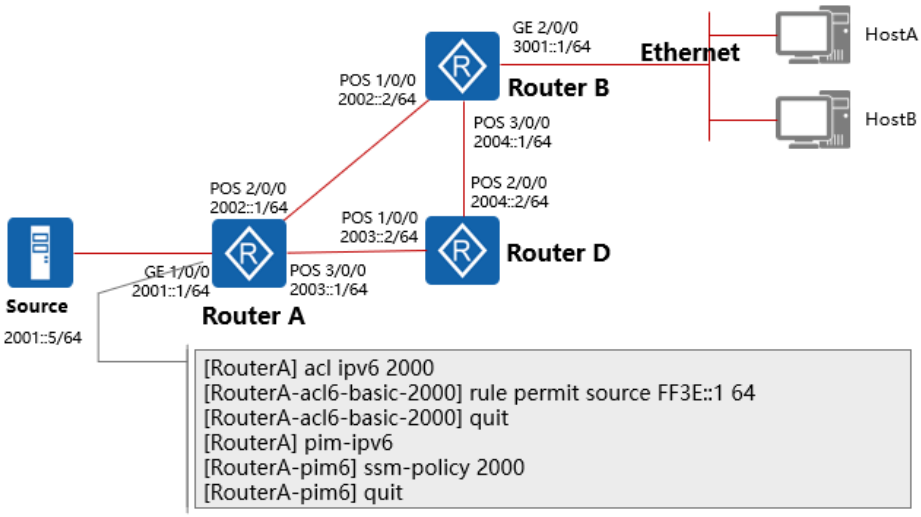
- 如果查询器发现组播地址在 SSM 组范围外，就在 IPv6 PIM-SM 基础上建立组播分发树。




- Host A 和 Host C 分别是两个叶子网络中的组播信息接收者。这些接收者通过 Router A、Router B、Router C 和 Router D 连接到组播源。
- Router B 和 N1、Router C 和 N2 之间的主机侧接口必须运行 MLDv2 协议。
- 配置思路：
- 配置各路由器接口的 IPv6 地址和 IPv6 单播路由协议。
- 配置各路由器接口的 IPv6 地址和掩码。
- 配置各路由器之间采用 OSPFv3 进行互连，进程号为 1，区域号为 0，确保网络中各路由器 Router A、Router B、Router C 和 Router D 之间能够在网络层互通。
- 在各路由器上使能 IPv6 组播功能，在路由器各接口上使能 IPv6 PIM-SM。
- 在各路由器配置 IPv6 PIM-SSM 组播组的地址范围。
- 在路由器的主机侧接口上配置 MLDv2。

- 检查配置结果。

配置Router A




- 在 Router A 上配置 IPv6 PIM-SSM 组播组的地址范围是 FF3E::1。
- Router B、Router C 和 Router D 上的配置过程与 Router A 上的配置相似。
- SSM 模型采用的是 IPv6 PIM-SM 的子集，所以必须先在网络中所有的路由器上使能 IPv6 PIM-SM 功能。同时，配置 SSM 组的地址范围。缺省情况下，采用 IANA 定义的 SSM 组范围。
- 如果用户希望从指定源 S 接收信息，或从指定源外的所有源 S 接收信息，必须发送含有信道（S, G）的 MLDv2 报告消息。接收侧的 DR 接收到此消息后，判断消息中的组播地址 G 是否在 SSM 组的地址范围内。如果是，DR 向组播源 S 发送加入消息，并在沿途各路由器上创建（S, G）项，从而建立 SPT 树，SSM 模型就此建立。如果组播地址 G 在 SSM 组地址范围外，或用户没有显式指定源地址 S，DR 触发建立 IPv6 PIM-SM 基础上的 ASM 模型。

- SSM 模型通过 IPv6 PIM-SM 的子集来实现。使能 IPv6 PIM-SM 的同时也就使路由器具有 SSM 处理能力。路由器周期性发送 Hello 报文来发现 PIM IPv6 邻居，并处理邻居发来的报文。当路由器加入 IPv6 PIM-SSM 时，建议在非边界路由器所有接口上使能 IPv6 PIM-SSM。
- 组播源的信息通过 IPv6 PIM-SSM 模式还是 IPv6 PIM-SM 模式传递到接收者，取决于信道（S, G）的组播地址是否在 SSM 组地址范围内。因此，IPv6 PIM-SSM 模式中，组地址信息十分重要。
- 如果没有指定 SSM 组地址范围，系统采用 IANA 为 SSM 保留的 FF3x::/12 网段作为缺省的地址范围。
- 进入系统视图
- **system-view**
- 使能 IPv6 组播路由
- **multicast ipv6 routing-enable**
- 进入接口视图
- **interface** *interface-type interface-number*
- 使能 IPv6 PIM-SSM 功能
- **pim ipv6 sm**
- 进入 PIM IPv6 视图
- **pim-ipv6**
- 配置 IPv6 PIM-SSM 组播地址范围
- **ssm-policy** *basic-acl6-number*

## 查看IPv6组播路由信息

```
[RouterA] display pim ipv6 routing-table
Vpn-instance: public net
Total 1 (S, G) entry

 (2001::5, FF3E::1)
    Protocol: pim-ssm, Flag: LOC
    UpTime: 00:00:11
    Upstream interface: GigabitEthernet1/0/0
        Upstream neighbor: FE80::200:AFF:FE01:10D
        RPF prime neighbor: FE80::200:AFF:FE01:10D
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: Pos2/0/0
            Protocol: pim-ssm, UpTime: 00:00:11, Expires: 00:03:19
```

- 如果 Host A 需要接收组播源 S（2001::5）发给组 G（FF3E::1）的信息，Router B 建立到源的 SPT。SPT 路径上的 Router A 和 Router B 生成（S,G）项，SPT 路径外的 Router D 不存在（S,G）项。

## 组播路由管理简介（IPv6）

- 为了实现组播路由转发路径的控制与维护，组播路由管理提供了一系列特性。主要分为RPF（Reverse Path Forwarding）和组播负载分担。

| 特性 | 功能 |
| --- | --- |
| RPF检查 | 用于保证组播数据沿正确的转发路径进行传输。 |
| 组播负载分担 | 用于选取不同的等价路由进行组播数据转发分流组播数据。 |

- 组播路由和转发与单播路由和转发类似，首先每个组播路由协议都各自建立并维护了一张协议路由表。各组播路由协议的组播路由信息经过综合形成一个总的组播路由表（Multicast Routing-Table）。最后，路由器根据组播路由和转发策略，从组播路由表中选出最优的组播路由，并下发到组播转发表（Multicast Forwarding-Table），直接用于控制组播数据的转发。

- 通过组播转发表，整个网络建立了一条以组播源为根，组成员为叶子的一点到多点的转发路径。为了实现转发路径的控制与维护，组播路由管理提供了一系列特性。

## IPv6组播协议路由表

- 组播协议路由表是运行各种组播路由协议时由各个协议自己维护的表项，是组播路由和转发的基础。PIM路由表项信息如下：

```
<HUAWEI> display pim ipv6 routing-table
VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (FC00::2, FFE3::1)
    Protocol: pim-sm, Flag: SPT LOC ACT
    UpTime: 00:04:24
    Upstream interface: Vlanif20
        Upstream neighbor: FE80::A01:100:1
        RPF prime neighbor: FE80::A01:100:1
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: Vlanif10
            Protocol: pim-sm, UpTime: 00:04:24, Expires: 00:02:47
```

- (FC00::2, FFE3::1) (S, G)表项。
- Protocol: pim-sm 协议类型。第一个 Protocol 表示生成表项的协议类型，第二个 Protocol 表示生成下游接口的协议类型。
- Flag: SPT LOC ACT PIM 路由表项的标志。
- UpTime: 00:04:24 存在时间。第一个 UpTime 表示表项已存在的时间，第二个 UpTime 表示下游接口已存在的时间。
- Upstream interface: Vlanif20 上游接口。
- Upstream neighbor: FE80::A01:100:1 上游邻居。NULL 表示不存在上游邻居。
- RPF prime neighbor: FE80::A01:100:1 RPF 邻居。NULL 表示不存在 RPF 邻居。
- Downstream interface(s) information: 下游接口信息。
- Total number of downstreams: 1 下游接口数量。

- Expires: 00:02:47 下游接口老化时间。

## IPv6组播路由表

- 组播路由表是组播路由管理模块生成的路由表。如果组播路由管理支持多种组播协议，那这里应该能看到多种协议生成的优选出的路由信息。

```
<HUAWEI> display multicast ipv6 routing-table
IPv6 multicast routing table
 Total 1 entry

 00001. (FC00::2, FFE3::1)
     Uptime: 00:00:14
     Upstream Interface: Vlanif10
     List of 1 downstream interface
       1: Vlanif20
```

- 00001. (FC00::2, FFE3::1) 第 00001 号表项，是(S, G)形式。
- Uptime: 00:00:14 组播路由表项更新时间。
- Upstream Interface: Vlanif10 上游接口。
- List of 1 downstream interface 下游接口列表。

## 组播转发表

- 组播转发表是路由管理模块依据组播路由表信息生成的用于指导组播数据实际转发的表项，通常称为MFIB。这张表与单播中FIB表的功能是一样的，用于指导组播数据转发。

```
<HUAWEI> display multicast ipv6 forwarding-table
IPv6 Multicast Forwarding Table
Total 2 entries, 2 matched

00001. (FC00:1::3, FF1E::1)
    MID: 10, Flags: ACT
    Uptime: 02:54:43, Timeout in: 00:02:27
    Incoming interface: Vlanif10
    List of 1 outgoing interfaces:
      1: LoopBack0
        Activetime: 00:23:15
    Matched 0 packets(0 bytes), Wrong If 0 packets
    Forwarded 0 packets(0 bytes)
```

真正指导组播数据转发的是组播转发表，转发表项中概括性记录了报文转发的统计信息。

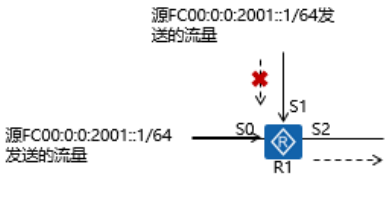
- 00001. (FC00:1::3, FF1E::1)　第 00001 号表项，是(S, G)形式。
- MID: 10　组播转发表项在 MFIB 表中的唯一标识，用于快速检索组播转发表。

- Flags: ACT 组播转发表项的标志。
- UpTime: 02:54:43 组播转发表项已存在的时间。
- Timeout in: 00:03:26　组播转发表项超时时间。
- Incoming interface: Vlanif10 表项入接口。
- List of 1 outgoing interfaces　表项出接口列表。
- Activetime: 00:23:15 出接口已存在时间。
- Matched 38264 packets(1071392 bytes)　匹配该表项的报文数目。
- Wrong If 0 packets　从错误接口进入的报文数目。
- Forwarded 38264 packets(1071392 bytes) 已转发的报文数目。

## RPF检查

- RPF检查
  - 确保组播流量沿正确路径转发，避免环路
  - 基于单播路由表
  - 检查过程：
    - 路由器确认组播报文是从自身连接到组播源的接口上收到的，才进行转发，否则丢弃

| IPv6路由表 | |
|---|---|
| 网段 | 接口 |
| FC00:0:0:2001::1/64 | S0 |
| FC00:0:0:3001::1/64 | S1 |
| FC00:0:0:4001::1/64 | S2 |




- RPF 检查原理
- 路由器收到一份组播报文后，会根据报文的源地址通过单播路由表查找到达“报文源”的路由，查看到“报文源”的路由表项的出接口是否与收到组播报文的入接口一致。如果一致，则认为该组播报文从正确的接口到达，从而保证了整个转发路径的正确性和唯一性。这个过程就被称为 RPF 检查。
- 如果这几条等价路由都是来自同一张路由表项，则选取下一跳地址最大的路由作为 RPF 路由。

- RPF 检验可以基于单播路由、MBGP 路由和组播静态路由。他们之间的优先顺序为组播静态路由、MBGP 路由、单播路由。

- 拓扑描述
- 来自组播源 FC00:0:0:2001::1/64 的组播流从 S1 口到达路由器，路由器检查路由表，发现可以转发该组播流的端口为 S0，RPF 检查失败。因此达到 S1 口的数据流被丢弃。
- 来自组播源 FC00:0:0:2001::1/64 的组播流从 S0 口达到路由器，检查路由表发现入接口与接收该组播流的接口 S0 一致，RPF 检查成功。因此组播流将被正确的转发。
- 组播路由协议通过已有的单播路由信息来确定上、下游邻居设备，创建组播路由表项。运用 RPF 检查机制，来确保组播数据流能够沿组播分发树（路径）正确的传输，同时可以避免转发路径上环路的产生。
- 在实际组播数据转发过程中，如果对每一份接收到的组播数据报文都通过单播路由表进行 RPF 检查，会给路由器带来很大负担。因此，路由器在收到一份来自源 S 发往组 G 的组播数据报文之后，首先会在组播转发表中查找有无相应的（S，G）组播转发表项：
- 如果不存在（S，G）转发表项，则对该报文执行 RPF 检查，将检查到的 RPF 接口作为入接口，创建组播路由表项，下发到组播转发表中。其中，对 RPF 检查结果的处理方式为：如果检查通过，表明接收接口为 RPF 接口，向转发表项的所有出接口转发；如果检查失败，表明报文来源路径错误，丢弃该报文。
- 如果存在（S，G）转发表项，并且接收该报文的接口与转发表项的入接口一致，则向所有的出接口转发该报文。
- 如果存在（S，G）转发表项，但是接收该报文的接口与转发表项的入接口不一致，则对此报文进行 RPF 检查。对 RP

F 检查结果的处理方式为：

- 若 RPF 检查选取出的 RPF 接口与转发表项的入接口一致，则说明（S，G）表项正确，报文来源路径错误，将其丢弃。
- 若 RPF 检查选取出的 RPF 接口与转发表项的入接口不符，则说明（S，G）表项已过时，于是把表项中的入接口更新为 RPF 接口。然后再根据 RPF 检查规则进行判断：如果接收该报文的接口正是其 RPF 接口，则向转发表项的所有出接口转发该报文，否则将其丢弃。

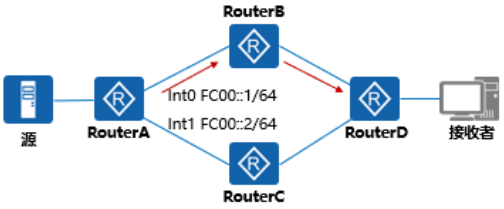
## 组播负载分担

- “负载分担”与“负载均衡”是不同的概念。
- “负载分担”是指如果发往某一目的地的数据流存在多条等价的转发路径，就将数据在这多条路径上转发，达到分流的目的。在进行数据转发时，每一条路径上转发的数据流量并不一定相同，转发流量多少需要根据负载分担方式来决定。
- “负载均衡”属于“负载分担”的一种特殊形式，不仅将数据流在这多条路径上转发，并且每条路径转发等量的数据流量。

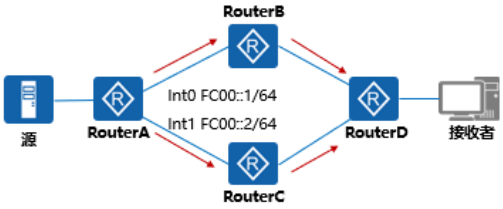
- 缺省情况下，组播报文转发过程中如果存在多条等价的最优转发路径，按照 RPF 检查对等价路由的处理规则，只会从 IGP 路由表中选取出下一跳地址最大的路由作为 RPF 路由。

## 组播负载分担前后对比示意图

- 组播负载分担是指如果存在多条等价的最优转发路径时，不按照RPF检查规则来选取下一跳地址最大的路由，而是根据配置的组播负载分担方式将组播流在这多条路径上进行分流转发。



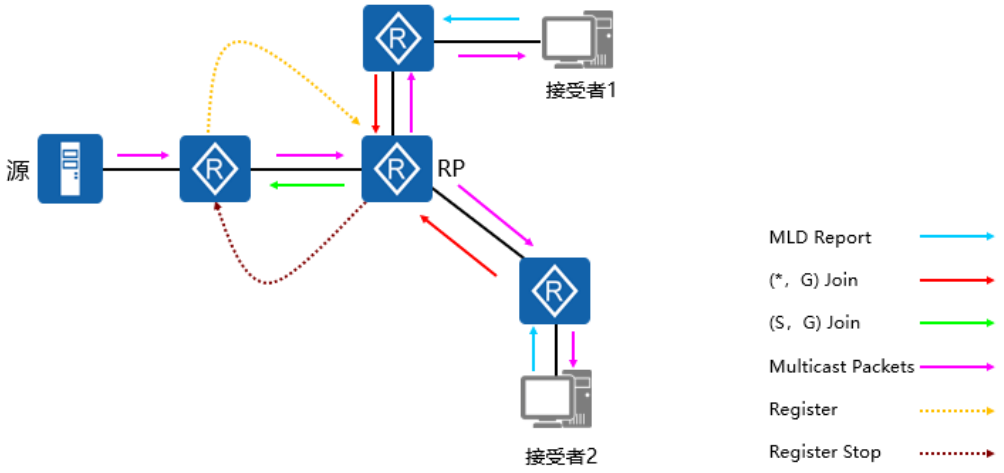

配置负载分担前




配置负载分担后

- 组播源 Source 向组播组 G 发送组播流，路由器 RouterA 和 RouterD 之间运行某种 IGP 协议（如 OSPF），RouterA→RouterB→RouterD 和 RouterA→RouterC→RouterD 是 2 条等价转发路径。
- 缺省情况下，根据 RPF 检查规则，组播流会从 Int1 端口转发，因为 Int1 的 IP 地址比 Int0 地址大。配置组播负载分担之后，就不会根据下一跳地址来选取转发路径，RouterA→RouterB→RouterD 和 RouterA→RouterC→RouterD 都会转发组播流。

## 域内组播:PIM SM + MLD



- 如图所示，域内路由器运行 PIM SM，和接收者相临的接口运行 MLD v1，在 IPv6 组播中 MLDv1 协议等同于 IPv4 组播中的 IGMPv2，用于获取组播组成员信息并通知上层协议。
- 区域内所有路由器通过 RP 的静态配置、BSR、或者自动发现方式得到 RP 信息。
- 和 IPv6 接收者相连的倒数第一跳路由器收到接收者发送的 MLD report 报文，沿 RPF 邻居向上游发送(*，G)加入消息，直到 RP 收到(*，G)加入消息，沿途路由器都创建(*，G)项，生成以 RP 为根的共享树。
- 组播源发出组播数据，第一跳路由器向 RP 发送 PIM 注册消息，RP 收到后回应注册停止消息。
- RP 向通过 RPF 邻居向第一跳路由器发(S，G)加入消息，沿途路由器创建(S，G)项，生成以第一跳路由器为根的源路径树。
- 组播数据沿源路径树到达 RP，并沿(*，G)转发，沿途路由器生成(S，G)项，组播数据到达接收者。

## 本章总结

- PIM-SM协议机制
- IPv6 PIM-SM配置
- IPv6 PIM-SSM工作原理
- 组播路由管理
- IPv6组播典型应用

# 思考题

1. IPv6 PIM-SM和IPv4 PIM-SM有哪些不同?
2. IPv6 PIM-SSM的工作机制是怎样的?

- IPv6 PIM-SM 和 IPv4 PIM-SM 有哪些不同？
- 地址不同，协议机制完全一样。
- IPv6 PIM-SSM 的工作机制是怎样的？
- IPv6 PIM-SSM 的实现可以概括成邻居发现、DR 选举和 SPT 生成：
- 邻居发现和 DR 选举过程与 IPv6 PIM-SM 中的描述相同，是通过在路由器间发送 Hello 消息来实现的；
- 由于 PIM-SSM 也使用 PIM-SM 协议，路由器生成 RPT 还是生成 SPT 的判决取决于组播地址是否在 SSM 组地址范围内。