

PIM 原理与配置

PIM: Protocol Independent Multicast 协议无关组播
常用版本是 PIMv2，PIM 报文直接封装在 IP 报文中，协议号为 103，组播地址 224.0.0.13。

PIM 网络中存在两种路由表项：(S , G) 路由表项或 (* , G) 路由表项。

S 表示组播源，G 表示组播组，*表示任意。

两个常见的组播路由协议：

PIM-DM (Protocol Independent Multicast Dense Mode) 密集模式

PIM-SM (Protocol Independent Multicast Sparse Mode) 稀疏模式

对于组播而言，有几个数据表是大家需要熟悉和理解。

IGMP 组表 : display igmp group

IGMP 路由表 : display igmp routing-table

组播协议路由表 : display pim routing-table

组播路由表 : display multicast routing-table

组播转发表 : display multicast forwarding-table

=====

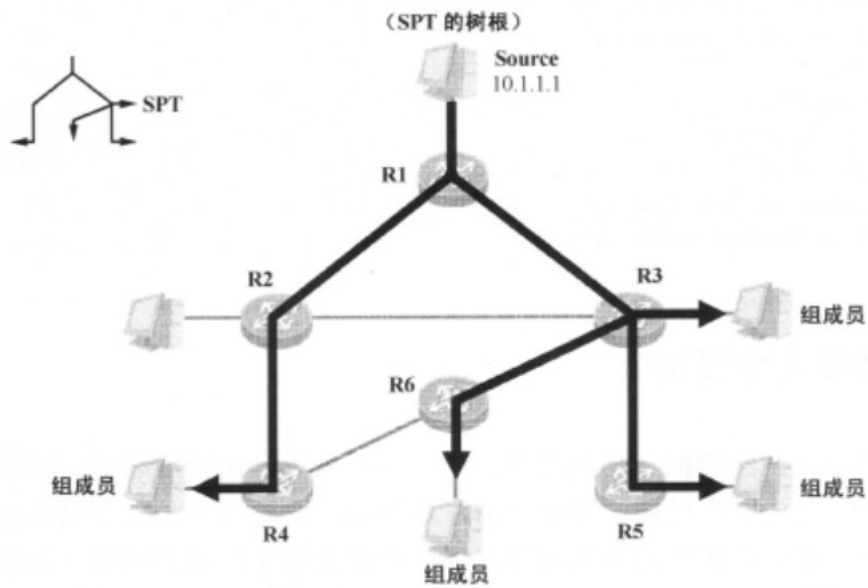


图 12-24 SPT

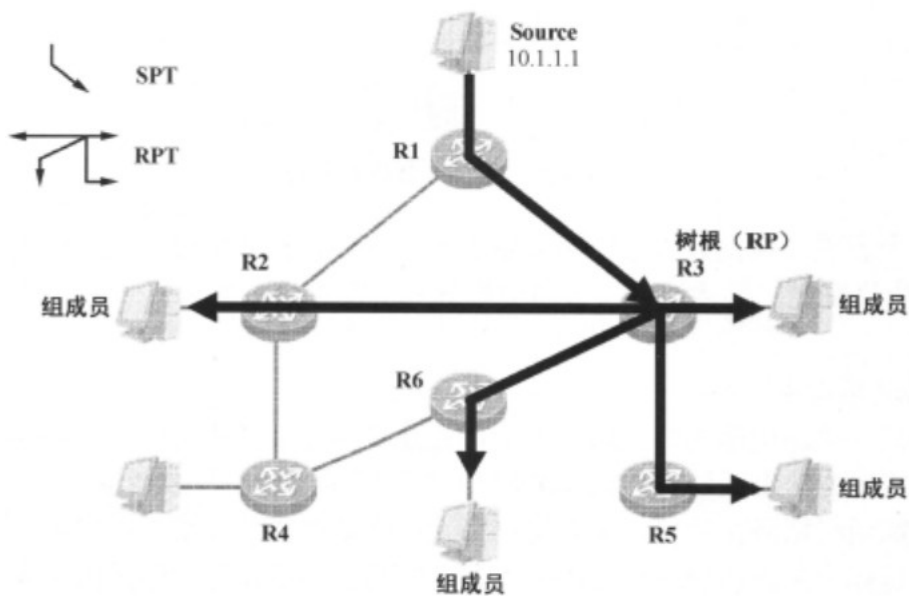


图 12-25 RPT

PIM-DM：采用“推（Push）模式”转发组播报文，（密集模式）

邻居的发现

构建 SPT

扩散与剪枝

状态刷新

嫁接 断言

使用的 5 个 PIMv2 消息

Hello
Join / Prune
Graft
Graft-Ack
Assert

状态刷新

PIM-DM 网络中，设备会周期性的发送状态刷新报文，刷新下游设备启动剪枝定时器的超时时间，使没有组播需求的接口一直处于剪枝状态。执行命令 `state-refresh-interval` 可设置状态刷新报文的发送周期。

缺省情况下，发送 PIM 状态刷新报文的时间间隔是 60 秒。

```
pim
state-refresh-interval 60
```

Hello 消息默认周期是 30 秒，超时时间值为 105 秒。

```
int g0/0/0
pim timer hello interval
pim hello-option holdtime interval
```

DR 选举

DR 在 DM 中并没有太多意义，只是做为 IGMPv1 的查询器，主要应用于 SM 中。

DR 是抢占的，接口 DR 优先级大的路由器将成为该 MA 网络

的 DR，在优先级相同的情况下，接口 IP 地址大的路由器将成为 DR。默认优先级为 1

```
int g0/0/0
pim hello-option dr-priority 9
```

表 12-3 PIM-DM 使用的几种类型报文	
报文类型	报文功能
Hello	用于 PIM 邻居发现、协议参数协商，以及 PIM 邻居关系维护
Join/Prune（加入/剪枝）	加入报文用于加入组播分发树，剪枝报文则用于修剪组播分发树。加入报文和剪枝报文在 PIM 中使用相同的报文格式，只不过报文载荷不同
Graft（嫁接）	用于将设备所在的分支嫁接到组播分发树
Graft-ACK（嫁接确认）	用于确认嫁接报文
Assert（断言）	用于断言机制

1、邻居的发现

在 PIM 域中，路由器通过周期性的向所有 PIM 路由器以组播方式发送 PIM hello 报文，以发现 PIM 邻居，维护各路由器之间的 PIM 邻居关系。发送地址为 224.0.0.13

2、构建 SPT

构建 SPT 的过程就是“扩散---剪枝”（flooding---prune）
扩散：路由器收到组播消息时，把所有端口当作成员端口发送组播消息，沿着路径中的每个路由器，扩散到成员主机连接的路由器，并且在路由器上建立（S,G）组播路由表

剪枝：没有连接成员的路由器，向上游接口发送一个剪枝消息，

当上游接口收到剪枝消息时，则将收到剪枝消息的接口，从(S,G)表项的下游接口中删除。

3、嫁接：(graft)

当路由器收到主机的报告报文，则向上游路由器发送嫁接消息，上游路由器将接收到嫁接消息的接口加入到(S,G)表项的下游接口中，并返回一个嫁接确认，若下游路由器收不到嫁接确认，则重复发送嫁接消息

4、断言 (Assert)

在同一个网段内，存在多个组播路由器，这时需要使用断言机制来选出一台组播路由器。

选举规则：

- 1、优先级高者优胜 到组播源的单播路由协议优先级较小者获胜。
- 2、cost 值小获胜 如果优先级相同，则到组播源的路由协议开销较小者获胜。
- 3、本地接口 IP 大的获胜 如果以上都相同，则连接到接受者 MA 网络接口 IP 地址最大者获胜。

=====

PIM-SM ：使用“拉 (Pull) 模式”转发组播报文。

邻居发现

DR 选举

RP 发现

构建 RPT

组播源注册

RPT 向 SPT 的切换

RP

RP 相当于组播网络中的一个组播数据汇聚中心，而 BSR (BootStrap Router)则相当于组播网络中的管理中心，它管理着动态选举 RP 的整个过程。

BSR 的工作机制并不复杂。首先网络中会选举出一台 BSR，BSR 被选举出来之后，它将向整个网络通告自己的存在。网络中的 PIM-SM 路由器会侦听 BSR 所泛洪的通告并保存 BSR 的相关信息，而 C-RP (Candidate-RP，候选 RP) 则纷纷向 BSR 发送自己的候选通告。BSR 收集所有 C-RP 发送过来的候选通告后，将这些通告加以汇总，然后将汇总的信息向全网进行泛洪。

网络中所有的 PIM-SM 路由器都会收到这个汇总信息，然后各自基于这些信息，采用相同的算法进行计算，最终得到组播组与 RP 的映射关系。由于每台 PIM-SM 路由器所收到的 C-RP 信息集合是一致的，而且基于这些信息采用相同的算法进行计算，因此得出的组播组与 RP 的映射关系必然也是相同的。

C-BSR 竞选 BSR 的规则：

- 1.具有最高优先级的路由器将成为 BSR，默认值为 0
- 2.当优先级相同时，IP 地址较大者将成为 BSR。

C-RP 竞选 RP 的规则：

- 1.C-RP 接口地址掩码最长者获胜。
- 2.C-RP 优先级较高者获胜，默认值为 0
- 3.如果优先级相同，则执行 Hash 函数，计算结果较大者获胜。
- 4.如果以上都相同，则 C-RP 地址较大者获胜。

当希望某个 C-RP 成为 RP 时，可以配置该命令调小该 C-RP 的优先级数值，从而提高 C-RP 的优先级

使用的 7 个 PIMv2 消息：

- Hello
- Register
- Register-Stop
- Join / Prune
- Bootstrap
- Assert
- Candidate-RP-Advertisement

表 12-4 PIM-SM 使用的几种类型报文	
报文类型	报文功能
Hello	用于 PIM 邻居发现、协议参数协商，以及 PIM
Register（注册）	用于实现源的注册过程。这是一种单播报文，组播数据被第一跳路由器封装在单播注册报文
Register-Stop（注册停止）	RP 使用该报文通知第一跳路由器停止通过注册
Join/Prune（加入/剪枝）	加入报文用于加入组播分发树，剪枝则用于修
Bootstrap（自举）	用于 BSR 选举。另外 BSR 也使用该报文（Candidate-RP，候选 RP）的汇总信息
Assert（断言）	用于断言机制
Candidate-RP-Advertisement（候选 RP 通告）	C-RP 使用该报文向 BSR 发送通告，报文中包及优先级等信息

PIM 网络中存在两种路由表项：（S，G）路由表项或（*，G）路由表项。
S 表示组播源，G 表示组播组，*表示任意。

(S , G) 路由表项主要用于在 PIM 网络中建立 SPT。对于 PIM-DM 网络和 PIM-SM 网络适用。

(* , G) 路由表项主要用于在 PIM 网络中建立 RPT。对于 PIM-SM 网络适用。

224.0.0.1 地址 (表示同一网段内所有主机和路由器)

224.0.0.2 地址 (本地网段内的所有组播路由器)

静态组播组

在某些应用场景中，可以在路由器的用户侧接口上配置静态组播组。比如：

- 1.网络中存在稳定的组播组成员，为了实现组播数据的快速、稳定转发，可以在用户侧接口配置静态组播组。
- 2.某网段内没有组播组成员或主机无法发送报告报文，但是又需要将组播数据转发到该网段，可以在接口上配置静态组播组，将组播数据引流到接口。

在接口上配置静态组播组后，路由器就认为此接口网段上一直存在该组播组的成员，从而转发该组的组播数据。该命令配置在与成员主机相连的接口上。可以配置单个组播组或源组，也可以配置批量组播组或源组。

```
interface g0/0/1
igmp static-group 224.1.1.1
```

=====

RPF (Reverse Path Forwarding , 逆向路径转发)

RPF 检查：为了防止组播报文在转发过程中出现重复报文及

环路的情况

RPF 检查的作用有 2 个

- 1.是防止组播路由发送环路
- 2.是防止转发冗余的组播数据报文

RPF 机制确保组播报文在正确的接口到达，只有这些组播报文才会被路由器沿着组播分发树进行转发，如果报文在错误的接口到达，路由器将丢弃这些报文。

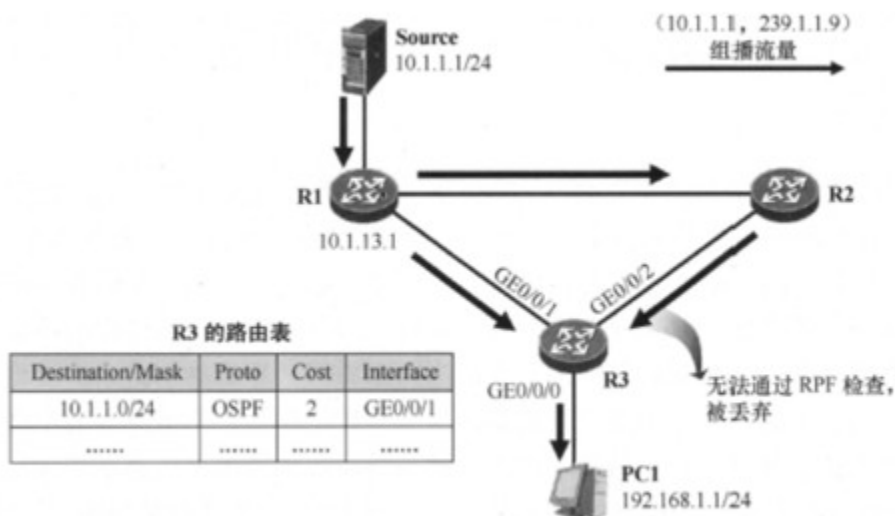


图 12-27 RPF 检查

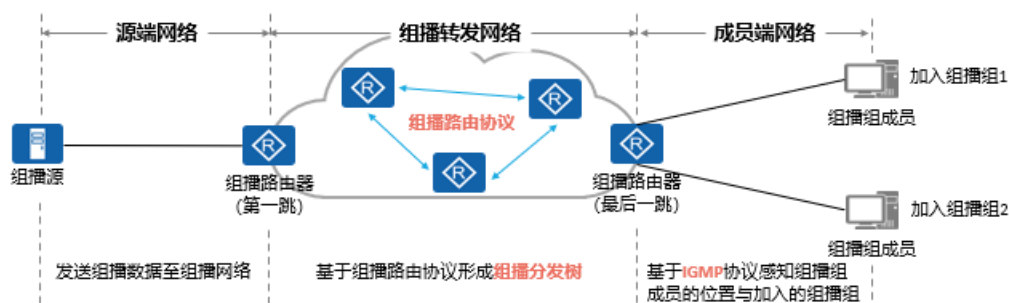
网络设备上的单播路由，MBGP 路由以及组播静态路由都可以作为 RPF 检查的依据。通过情况下，我们采用单播路由进行 RPF 检查。当设备收到一个组播报文时，它将在其单播路由表中查询到达该报文的源 IP 地址的路由，也就是查询到达组播源的单播路由，并检查该单播路由表项的出接口与接收该报文的接口是否一致，如果不一致，则认为报文未通过 RPF 检查并将其丢弃，如果一致，则认为报文通过 RPF 检查并对

其进行转发。

- 组播网络由组播源，组播组成员与组播路由器组成。
- 组播源的主要作用是发送组播数据。
- 组播组成员的主要作用是接收组播数据，因此需要通过 IGMP 让组播网络感知组成员位置与加组信息。
- 组播路由器的主要作用是**将数据从组播源发送到组播组成员**。组播数据转发需要**依赖组播分发树**，因此组播路由器需要通过协议来**构建组播分发树**。
- PIM (Protocol Independent Multicast ，协议无关组播) 协议的主要作用就是**构建组播分发树**。
- 本章主要讲解 PIM 协议的基本概念，PIM-DM 模式的工作原理与 PIM-SM 模式的工作原理。

组播网络基本架构回顾

- 组播网络大体可以分为三个部分：
 - 源端网络：将组播源产生的组播数据发送至组播网络。
 - 组播转发网络：**形成无环的组播转发路径**，该转发路径也被称为**组播分发树 (Multicast Distribution Tree)**。
 - 成员端网络：通过**IGMP**协议，让组播网络**感知组播组成员位置与加入的组播组**。

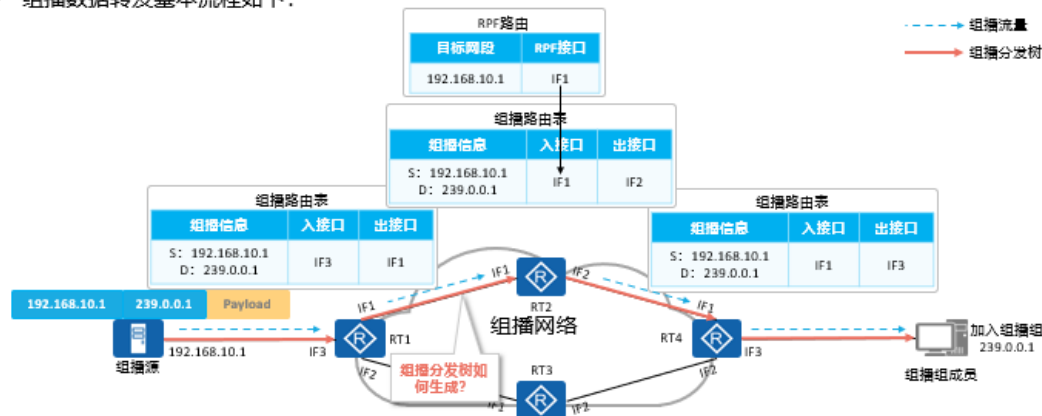


- 成员端网络通过 IGMP 协议可以让组播网络感知组播组成员位置与加入的组播组。
- 组播转发网络形成组播分发树，需要组播路由协议的支持。
- 通过组播路由协议有多种，最常用的是 PIM 协议，也就

是本节课的重点。

组播数据转发流程回顾

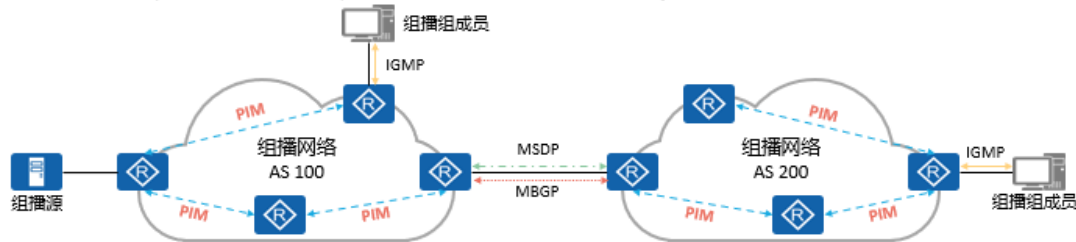
• 组播数据转发基本流程如下：



- 《IGMP 原理与配置》课程介绍了组播网络如何感知组播组成员的位置和所加组播组。
- 本课程将主要介绍如何形成组播分发树。

组播协议介绍回顾

- 组播网络需要基于多种组播协议才能建立转发路径：
 - 工作在成员端网络的主要是IGMP (Internet Group Management Protocol, 因特网组管理协议) 协议, 用于告知组播网络, 组成员的位置与所加组播组。
 - 工作在组播转发网络的协议主要是PIM, MSDP, MBGP。
 - PIM (Protocol Independent Multicast, 协议无关组播) 协议主要作用是生成AS域内的组播分发树。
 - MSDP (Multicast Source Discovery Protocol, 组播源发现协议) 主要作用是帮助生成AS域间的组播分发树。
 - MBGP (Multicast BGP, 组播BGP) 主要作用是帮助跨域组播流进行RPF校验。



- 本课程主要介绍 PIM 协议的工作原理。
- PIM 称为协议无关组播。协议无关指的是与单播路由协议无关，即 PIM 不需要维护专门的单播路由信息。作为组播路由解决方案，它直接利用单播路由表的路由信息，对组播报

文执行 RPF 检查，检查通过后创建组播路由表项，从而转发组播报文。

- 目前在实际网络中，PIM 主要有两种模式：
 - **PIM-DM (PIM-Dense Mode , PIM 密集模式)**。
 - **PIM-SM (PIM-Sparse Mode , PIM 稀疏模式)**，PIM-SM 模式根据组播服务模型又可以分为：
 - **PIM-SM (ASM)**：为任意源组播建立组播分发树。
 - **PIM-SM (SSM)**：为指定源组播建立组播分发树。
-
- PIM (Protocol Independent Multicast) 协议无关组播，目前常用版本是 PIMv2，PIM 报文直接封装在 IP 报文中，协议号为 103，PIMv2 组播地址为 224.0.0.13。
 - 在 PIM 组播域中，以组播组为单位建立从组播源到组成员的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)。
 - 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。

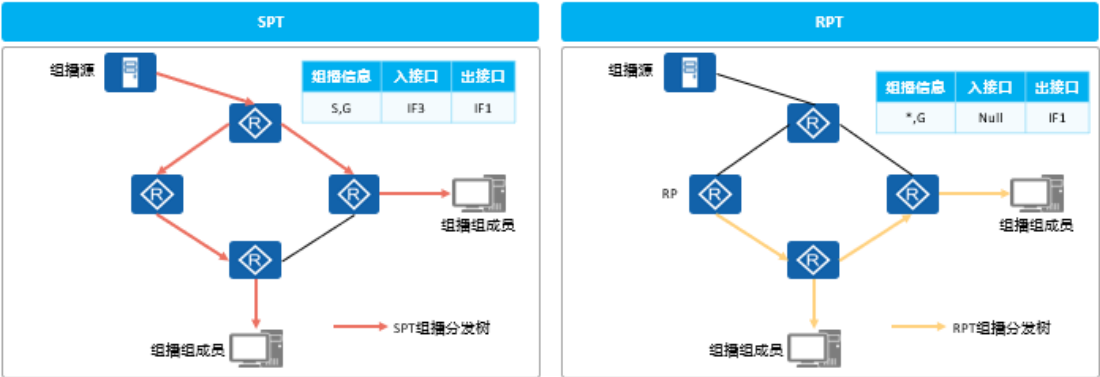
PIM-DM 与 PIM-SM 使用场景

- PIM 形成组播分发树主要有两种模式，即 PIM-DM 模式与 PIM-SM 模式，这两种模式分别用在不同的场景下：
- **PIM-DM 模式**主要用在**组成员较少且相对密集的组播网络中**，该模式建立组播分发树的基本思路是“扩散-剪枝”，即将组播流量全网扩散，然后剪枝没有组成员的路径，最终形成组播分发树。
- **PIM-SM 模式**主要用在**组成员较多且相对稀疏的组播网**

网中，该模式建立组播分发树的基本思路是先收集组成员信息，然后再形成组播分发树。使用 PIM-SM 模式不需要全网泛洪组播，对现网的影响较小，因此现网多使用 PIM-SM 模式。

组播分发树的分类

- 通过PIM形成的组播分发树主要分为以下两种：
 - 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），在PIM-DM与PIM-SM中均有使用。
 - 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），在PIM-SM中使用。



- SPT 又被称为源树，SPT 在 PIM-DM 模式与 PIM-SM 模式的组播网络中均有使用。
- RPT 又被称为共享树，RPT 主要在 PIM-SM 模式的组播网络中被使用。
- RP 的作用将在《PIM-SM》章节讲解。

PIM路由表项

- PIM路由表项即通过PIM协议建立的组播协议路由表项。
- PIM网络中存在两种路由表项：
 - (S, G) 路由表项主要用于在PIM网络中建立SPT。对于PIM-DM网络和PIM-SM网络适用。
 - (*, G) 路由表项主要用于在PIM网络中建立RPT。对于PIM-SM网络适用。

PIM (*, G) 表项

(* , 239.0.0.1)
Protocol: pim-dm, Flag: WC //Flag值的意义见备注
UpTime: 02:07:35
Upstream interface: NULL
Upstream neighbor: NULL
RPF prime neighbor: NULL
Downstream interface(s) information:
Total number of downstreams: 1
1: GigabitEthernet0/0/0
Protocol: igmp, UpTime: 02:07:35, Expires: never

流量触发创建

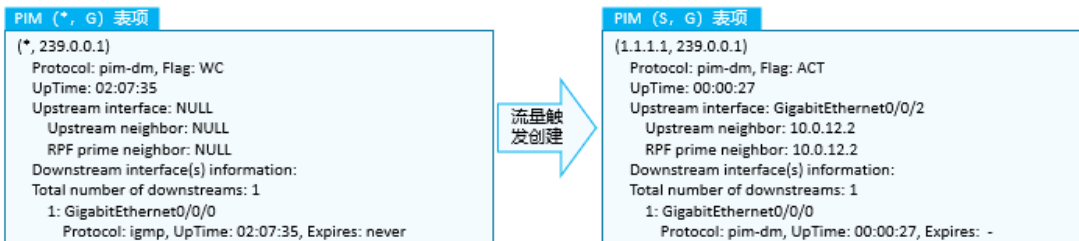
PIM (S, G) 表项

(1.1.1.1, 239.0.0.1)
Protocol: pim-dm, Flag: ACT
UpTime: 00:00:27
Upstream interface: GigabitEthernet0/0/2
Upstream neighbor: 10.0.12.2
RPF prime neighbor: 10.0.12.2
Downstream interface(s) information:
Total number of downstreams: 1
1: GigabitEthernet0/0/0
Protocol: pim-dm, UpTime: 00:00:27, Expires: -

- S 表示组播源，G 表示组播组，*表示任意。
- PIM 路由器上可能同时存在两种路由表项。当收到源地
址为 S，组地址为 G 的组播报文，且 RPF 检查通过的情况下，
按照如下的规则转发：
- 如果存在 (S，G) 路由表项，则由 (S，G) 路由表项
指导报文转发。
- 如果不存在 (S，G) 路由表项，只存在 (*，G) 路由表
项，则先依照 (*，G) 路由表项创建 (S，G) 路由表项，再
由 (S，G) 路由表项指导报文转发。
- Flag 值的意义见下一页。

PIM路由表项

- PIM路由表项即通过PIM协议建立的组播协议路由表项。
- PIM网络中存在两种路由表项：
 - (S，G) 路由表项主要用于在PIM网络中**建立SPT**。对于PIM-DM网络和PIM-SM网络适用。
 - (*，G) 路由表项主要用于在PIM网络中**建立RPT**。对于PIM-SM网络适用。



- PIM 路由表中 Flag 字段的意义如下：



PIM路由表项与组播路由表项

- 在不同的组播路由器上，组播路由表项会基于不同的表项汇总形成。
 - 最后一跳路由器的组播路由表项主要基于PIM路由表项，IGMP组表项和IGMP路由表项汇总形成。
 - 其余组播路由器的组播路由表项主要基于PIM路由表项形成。

基于PIM路由表生成组播路由表

PIM (S, G) 表项

(1.1.1.1, 239.0.0.1)

Protocol: pim-dm, Flag: ACT

UpTime: 00:00:27

Upstream interface: GigabitEthernet1/0/1

Upstream neighbor: 10.0.12.2

RPF prime neighbor: 10.0.12.2

Downstream interface(s) information:

Total number of downstreams: 1

1: GigabitEthernet0/0/0

Protocol: pim-dm, UpTime: 00:00:27, Expires: -

组播路由表项

(1.1.1.1, 239.0.0.1)

Upstream Interface: GigabitEthernet1/0/1

Downstream interfaces

1: GigabitEthernet0/0/0

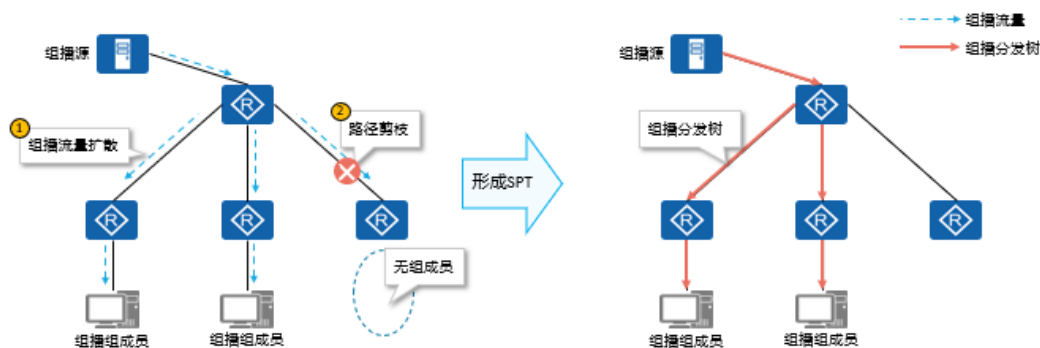
- 组播路由表项只能基于PIM (S, G) 路由表项形成。PIM (*, G) 路由表项缺少入接口信息无法形成组播路由表。

- 最后一跳路由器的组播表项形成可以参考《IGMP 原理与配置》课程。



PIM-DM基本概念

- PIM-DM主要用在组成员较少且相对密集的网络中，通过“扩散-剪枝”的方式形成组播转发树（SPT）。
- PIM-DM在形成SPT的过程中，除了扩散（Flooding），剪枝（Prune）机制外，还会涉及邻居发现（Neighbor Discovery），嫁接（Graft），断言（Assert）和状态刷新（State Refresh）机制。





PIM-DM协议报文

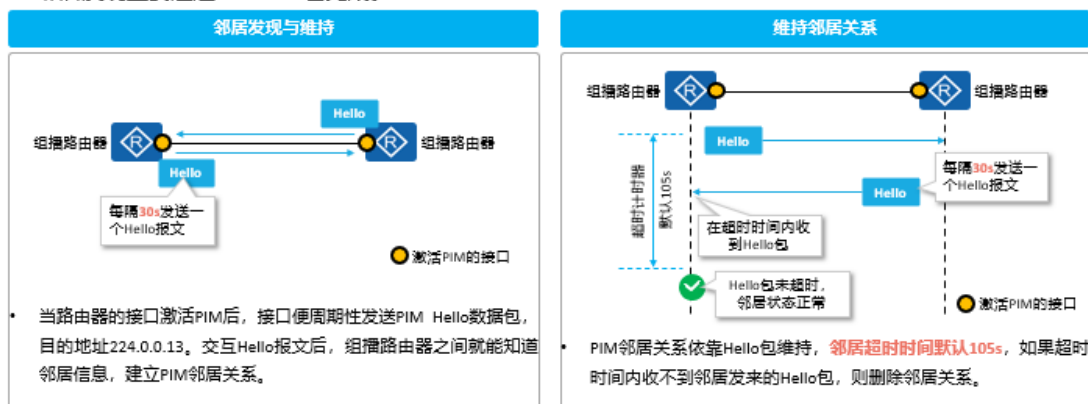
- PIM协议报文直接采用IP封装，**目的地址224.0.0.13**，**IP协议号103**。
- PIM-DM与PIM-SM使用的协议报文类型有所不同。
- PIM-DM使用报文主要是以下几类：

报文类型	报文功能
Hello	用于PIM邻居发现，协议参数协商，PIM邻居关系维护等
Join/Prune（加入/剪枝）	加入报文用于加入组播分发树，剪枝报文则用于修建组播分发树。加入及剪枝报文在PIM中使用相同的报文格式，只不过报文载荷中的字段内容有所不同
Graft（嫁接）	用于将设备所在的分支嫁接到组播分发树
Graft-ACK（嫁接确认）	用于对邻居发送的Graft报文进行确认
Assert（断言）	用于断言机制



邻居发现

- **组播转发路径只能在PIM邻居之间建立**，因此邻居发现是形成组播分发树的先决条件。
- 邻居发现主要通过PIM Hello包完成。

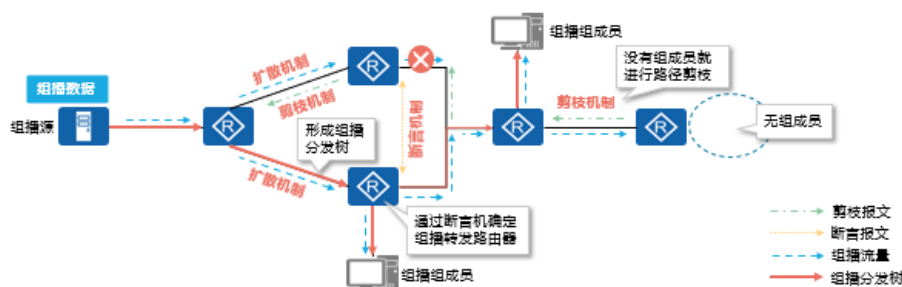


- Hello 报文中携带多项 PIM 协议报文参数，主要用于 PIM 邻居之间 PIM 协议报文的控制。具体如下：
- DR_Priority：表示各路由器接口竞选 DR 的优先级，优先级越高越容易获胜。
- Holdtime：表示保持邻居为可达状态的超时时间。如果在超时时间内没有收到 PIM 邻居发送的 Hello 报文，路由器则认为邻居不可达。
- LAN_Delay：表示共享网段内传输 Prune 报文的延迟时间。

- Neighbor-Tracking：表示邻居跟踪功能。
- Override-Interval：表示 Hello 报文中携带的否决剪枝的时间间隔。

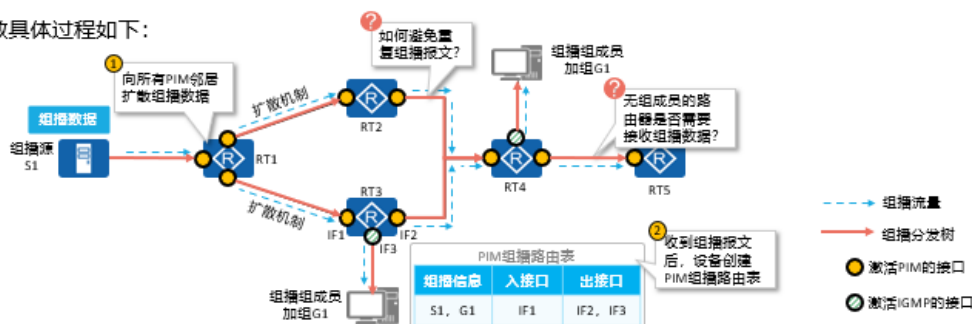
首次形成组播分发树

- PIM-DM模式首次形成组播分发树主要依赖扩散机制、剪枝机制、断言机制与DR选举机制。
 - 扩散机制：组播数据包向所有的PIM邻居泛洪，同时组播路由器产生组播路由表项。
 - 断言机制：当组播转发过程中存在多路访问网络，则需要选举出一个组播转发路由器，避免重复组播报文。
 - 剪枝机制：如果组播路由器下没有组成员，则将源到该组播路由器的组播转发路径剪枝。



扩散机制

- 组播源发送的组播报文会在全网内扩散。当PIM路由器接收到组播报文，先进行RPF检查，通过后会在该路由器上创建 (S, G) 表项，之后会向所有PIM邻居发送。
- PIM-DM形成的 (S, G) 表项有老化时间（默认210s），如果老化时间超时前没有收到新的组播报文，则删除 (S, G) 表项。
- 扩散具体过程如下：

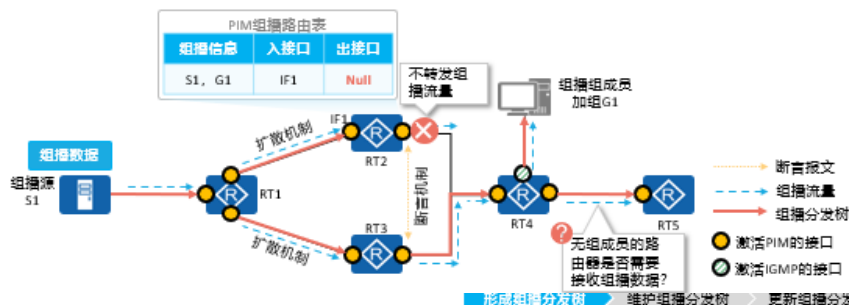


- 扩散机制会周期性（默认 180s）全网扩散组播数据，周期性扩散的主要目的是探测是否有新成员加组，但是由于全网扩散组播数据会浪费大量带宽，所以现在的组播网络一般使用“状态刷新机制”加上“嫁接机制”来实现周期性全网扩散感知新成员加组的目的。



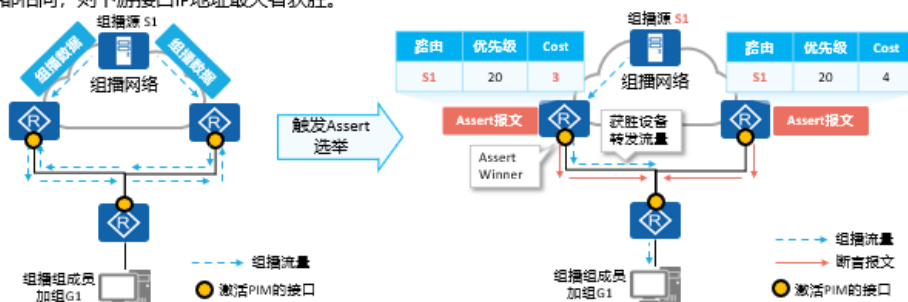
断言机制

- 当一个网段内有多个相连的PIM路由器向该网段转发组播报文时，需要通过断言机制（Assert）来保证**只有一个PIM路由器向该网段转发组播报文**。
- 通过断言机制的选举规则将决定组播路由器的转发行为：
 - 获胜一方的下游接口称为**Assert Winner**，将负责后续对该网段组播报文的**转发**。
 - 落败一方的下游接口称为**Assert Loser**，后续**不会**对该网段**转发**组播报文，PIM路由器也会将其从（S，G）表项**下游接口列表中删除**。



断言机制选举规则

- PIM路由器在**接收到**邻居路由器发送的**相同组播报文**后，会向该网段**发送断言（Assert）报文**，进行Assert选举。Assert报文内会携带到组播源的单播路由前缀，路由优先级与开销。选举规则如下：
 - 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。

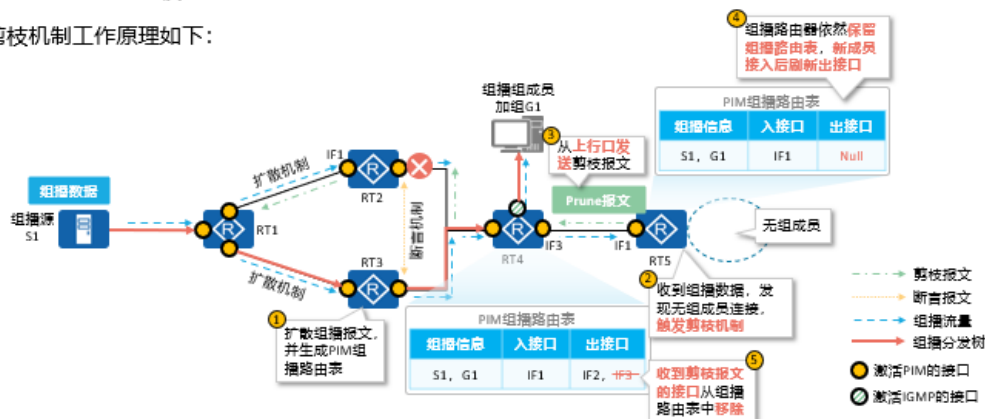


- Assert 选举失败的设备会抑制转发，并将这种抑制转发的状态保持一段时间，这段时间就被称为 Assert 保持时间，默认 180s。
- Assert 保持时间超时后，竞选失败的设备会恢复转发从而触发新一轮竞选。



剪枝机制

- 对于没有组成员连接的组播路由器，组播网络无需再将组播流量继续放往该设备。通过剪枝机制，组播网络可以将此类路径剪枝。
- 剪枝机制工作原理如下：

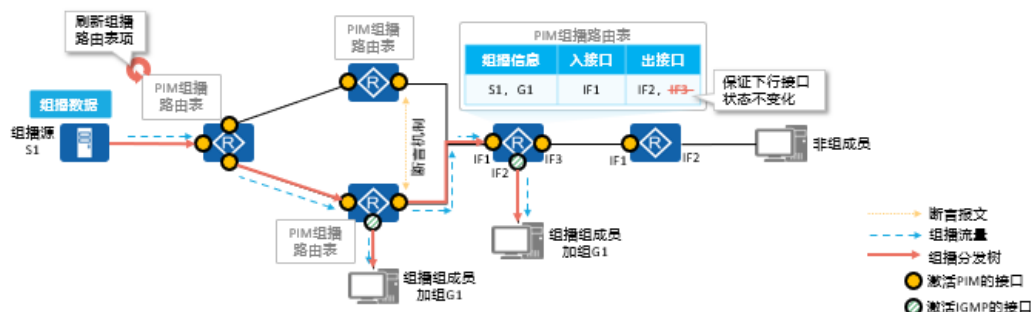


- 路由器为被裁剪的下游接口启动一个剪枝定时器（默认 210s），定时器超时后接口恢复转发。组播报文重新在全网范围内扩散，新加入的组成员可以接收到组播报文。随后，下游不存在组成员的叶子路由器将向上发起剪枝操作。通过这种周期性的扩散-剪枝，PIM-DM 周期性的刷新 SPT。
- 当下游接口被剪枝后：
- 如果下游叶子路由器有组成员加入，并且希望在下次“扩散-剪枝”前就恢复组播报文转发，则执行嫁接动作。
- 如果下游叶子路由器一直没有组成员加入，希望该接口保持抑制转发状态，则执行状态刷新动作。



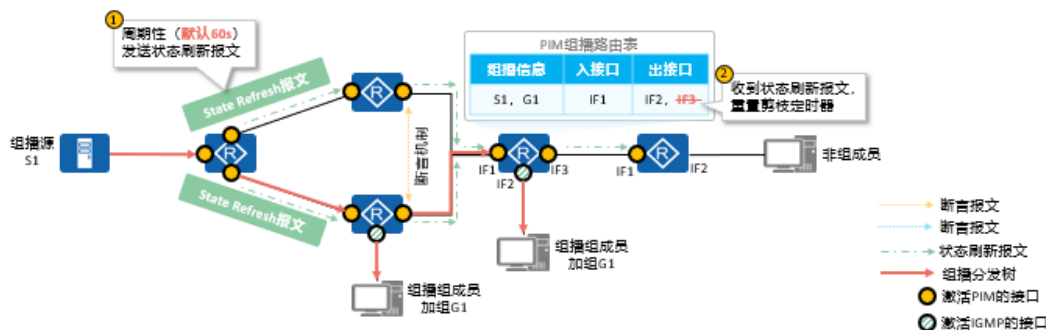
维护组播分发树

- 组播分发树形成后不会一直存在，也不会一直不变。
- 在PIM邻居关系稳定，组成员没有变化的情况下，维护组播分发树一般有两种方式：
 - 持续发送组播报文，保证组播路由表项能一直存在。
 - 发送状态刷新报文，保证组播路由表项的下行接口状态不发生变化。



状态刷新机制

- 在PIM-DM网络中，为了避免被裁剪的接口因为“剪枝定时器”超时而恢复转发，离组播源最近的第一跳路由器会周期性地触发State Refresh报文在全网内扩散。
- 收到State Refresh报文的PIM路由器会刷新剪枝定时器的状态。被裁剪接口的下游叶子路由器如果一直没有组成员加入，该接口将一直处于抑制转发状态。



新成员加入

- 当有新成员加入组播组后，组播网络需要**更新组播分发树**，才能将组播数据发往组成员。PIM-DM模式在使用“扩散-剪枝”的方式建立组播分发树后，通过**状态刷新机制**，使**下行接口一旦被抑制就无法自动恢复**。
- 因此需要一些机制来更新组播分发树，一般PIM-DM模式更新组播分发树的方法有两种：
- 等待组播路由表超时后，全网重新泛洪。该方法不可控，

在现网中无法实现

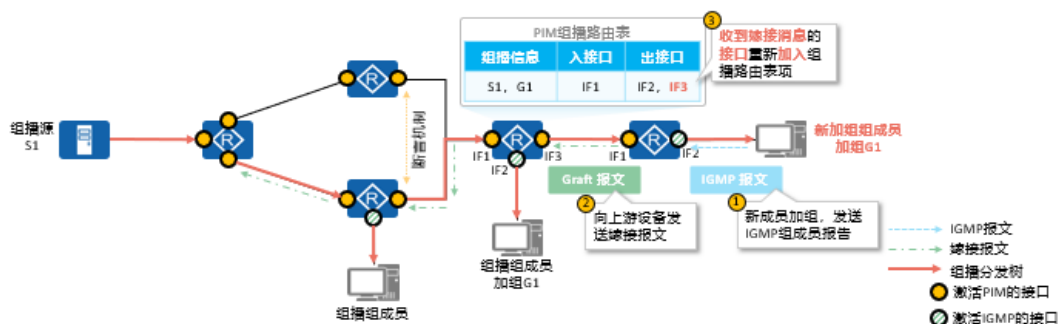
- 使用嫁接 (Graft) 机制，当新成员加组后，**主动反向建立组播分发路径**。现网中一般使用嫁接机制来实现新成员加组。

形成组播分发树 > 维护组播分发树 > 更新组播分发树



嫁接机制

- PIM-DM通过嫁接机制，使有新组成员加入的网段快速得到组播报文。
- 叶子路由器通过IGMP了解到与其相连的用户网段上，组播组G有新的组成员加入。随后叶子路由器会**基于本地的组播路由表**向上游发送Graft报文，请求上游路由器**恢复相应出接口**转发，将其添加在 (S, G) 表项下游接口列表中。



PIM-DM的基本配置

- 使能路由器的组播路由功能

```
[Huawei] multicast routing-enable
```

- 在接口上使能PIM-DM

```
[Huawei - GigabitEthernet1/0/0] pim dm
```

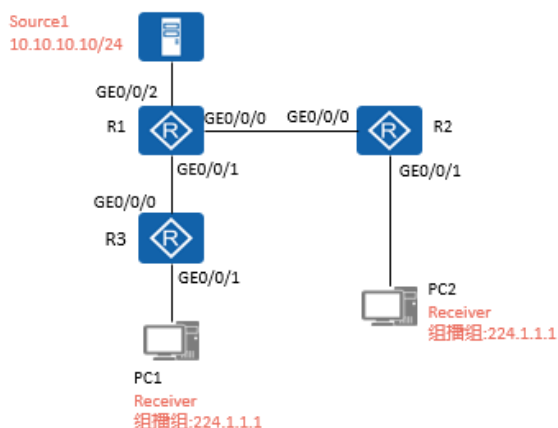
- 查看PIM邻居参数

```
<Huawei> display pim neighbor
```

- 查看PIM路由表参数

```
<Huawei> display pim routing-table
```

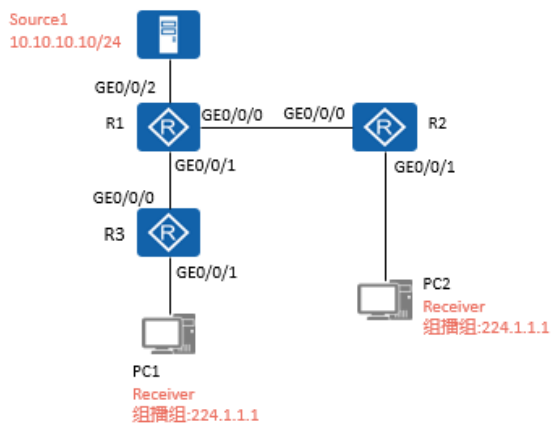
PIM-DM基础实验



实验要求:

- 通过PIM-DM协议让PC1、PC2可以接收到组播源的数据包。

PIM-DM的配置 (1)



(省略接口IP、OSPF的配置)

R1的配置如下:

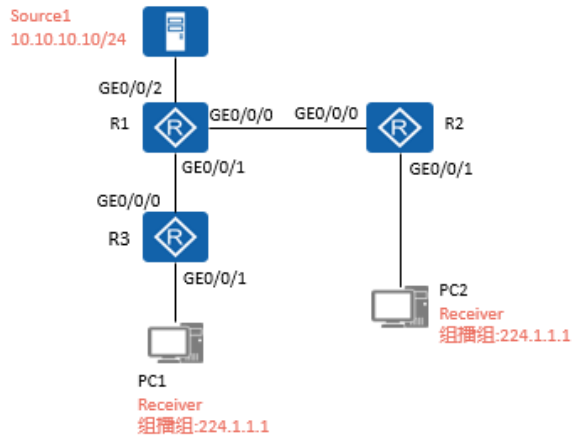
```
[R1]multicast routing-enable
[R1]interface g0/0/2
[R1-GigabitEthernet0/0/2]pim dm
[R1-GigabitEthernet0/0/2]interface g0/0/0
[R1-GigabitEthernet0/0/0]pim dm
[R1-GigabitEthernet0/0/0]interface g0/0/1
[R1-GigabitEthernet0/0/1]pim dm
```

R2的配置如下:

```
[R2]multicast routing-enable
[R2]interface g0/0/0
[R2-GigabitEthernet0/0/0]pim dm
[R2-GigabitEthernet0/0/0]interface g0/0/1
[R2-GigabitEthernet0/0/1]igmp enable
```



PIM-DM的配置 (2)



R3的配置如下:

```
[R3]multicast routing-enable
[R3]interface g0/0/0
[R3-GigabitEthernet0/0/0]pim dm
[R3-GigabitEthernet0/0/0]interface g0/0/1
[R3-GigabitEthernet0/0/1]igmp enable
```

(省略接口IP、OSPF的配置)



查看参数

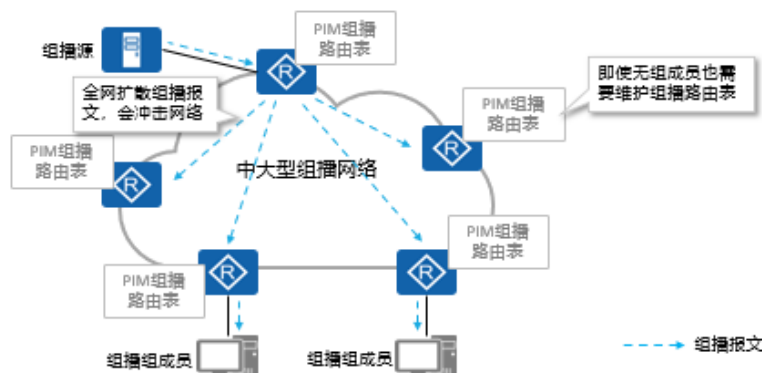
```
<R1>display pim routing-table
VPN-Instance: public net
Total 0 (*, G) entry; 1 (S, G) entry

(10.10.10.10, 224.1.1.1)
Protocol: pim-dm, Flag: LOC ACT
UpTime: 00:00:34
Upstream interface: GigabitEthernet0/0/2
Upstream neighbor: NULL
RPF prime neighbor: NULL
Downstream interface(s) information:
Total number of downstreams: 2
1: GigabitEthernet0/0/0
Protocol: pim-dm, UpTime: 00:00:34, Expires: never
2: GigabitEthernet0/0/1
Protocol: pim-dm, UpTime: 00:00:34, Expires: never
```



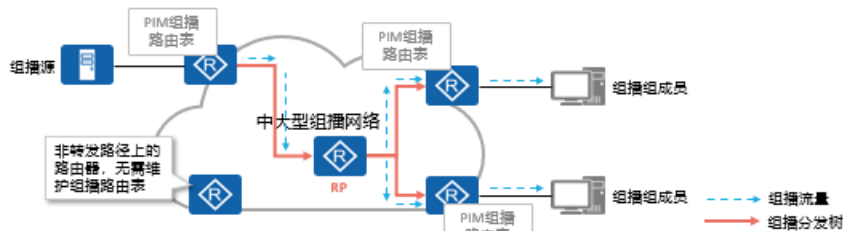
PIM-DM的局限性

- 中大型组播网络中由于网络较大，如果依然使用PIM-DM会遇到很多问题：
 - 使用“扩散-剪枝”方式需要全网扩散组播报文，对于网络有一定冲击。
 - 所有组播路由器均需要维护组播路由表，即使该组播路由器无需转发组播数据。
 - 对于组成员较为稀疏的组播网络，使用“扩散-剪枝”形成组播分发树的效率不高。



PIM-SM(ASM)介绍

- PIM-DM模型使用“扩散-剪枝”形成组播分发树的原因是：组播网络中大部分组播路由器无法得知组成员的位置。
- PIM-SM (ASM) 模型形成组播分发树的方法是：
 - 将组成员的位置事先告知某台组播路由器 (Rendezvous Point, RP)，形成RPT (RP Tree)。
 - 组播源在发送组播数据时，组播网络先将组播数据发送至RP，然后由RP再将组播数据转发给组成员。
 - 对于部分次优的组播转发路径，PIM-SM (ASM) 能自动优化为最优路径 (SPT)。



- 通过 PIM-SM (ASM) 模式形成组播分发树有如下好处：
- 只有组播转发路径上的组播路由器需要维护组播路由表。
- 通过 RP 可以让所有组播路由器获知组成员的位置。
- 避免“扩散-剪枝”机制，提高组播分发树的形成效率。



PIM-SM(ASM)协议报文

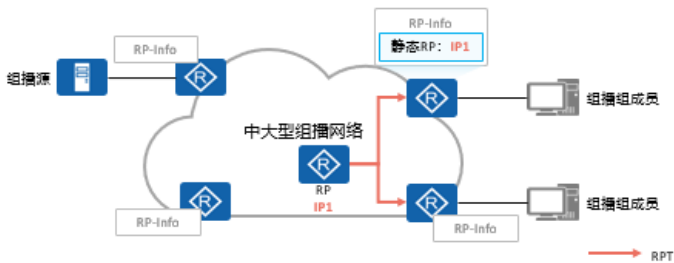
- PIM协议报文直接采用IP封装，目的地址224.0.0.13，IP协议号103。
- PIM-SM使用报文主要是以下几类：

报文类型	报文功能
Hello	用于PIM邻居发现，协议参数协商，PIM邻居关系维护等
Register（注册）	用于事先源的注册过程。这是一种单播报文，在源的注册过程中，组播数据被第一跳路由器封装在单播注册报文中发往RP
Register-Stop（注册停止）	RP使用该报文通知第一跳路由器停止通过注册报文发送组播流量
Join/Prune（加入/剪枝）	加入报文用于加入组播分发树，剪枝则用于修剪组播分发树
Assert（断言）	用于断言机制
Bootstrap（自举）	用于BSR选举。另外BSR也使用该报文向网络中扩散C-RP（Candidate-RP，候选RP）的汇总信息
Candidate-RP-Advertisement（候选RP通告）	C-RP使用该报文向BSR发送通告，报文中包含该C-RP的IP地址及优先级等信息



RP介绍

- 汇聚点RP（Rendezvous Point）为网络中一台重要的PIM路由器，用于处理源端DR注册信息及组成员加入请求，网络中的**所有PIM路由器都必须知道RP的地址**，类似于一个供求信息的汇聚中心。
- 目前可以通过以下方式配置RP：
 - **静态RP**：在网络中的所有PIM路由器上配置相同的RP地址，静态指定RP的位置。
 - **动态RP**：通过选举机制在多个C-RP（Candidate-RP，候选RP）之间选举出RP。

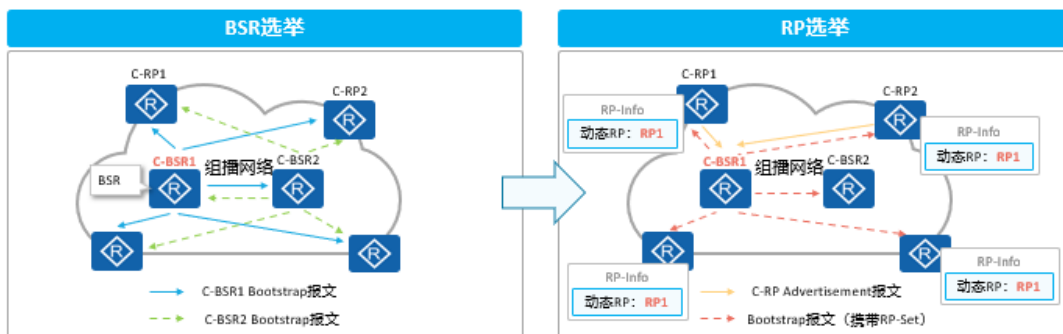


- 静态 RP 或者是动态 RP 在设置时均可以指定该 RP 为哪些组播组提供服务。



动态选举RP

- 动态选举RP会涉及两类角色C-BSR（Candidate-Bootstrap Router）与C-RP（Candidate-RP）：
 - C-BSR通过竞选能选举出一个唯一的BSR。
 - BSR的作用是收集C-RP的信息并形成RP-Set信息，BSR通过PIM报文将RP-Set信息扩散给所有PIM路由器。
 - PIM路由器收到RP-Set消息后，根据RP选举规则选举出合适的RP。

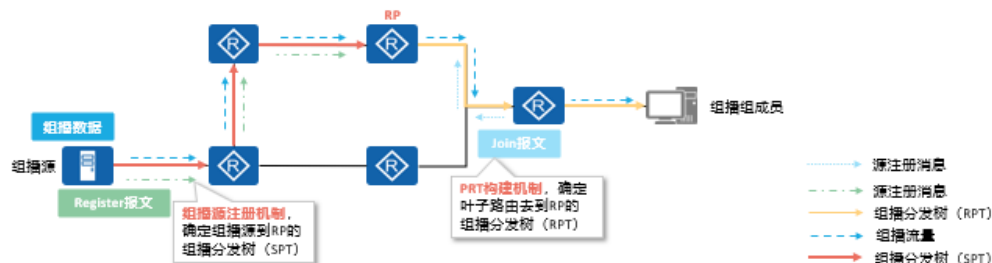


- BSR 竞选规则如下：
 - 优先级较高者获胜（优先级数值越大优先级越高）。
 - 如果优先级相同，IP 地址较大者获胜。
- RP 竞选规则如下：
 - 与用户加入的组地址匹配的 C-RP 服务的组范围掩码最长者获胜。
 - 如果以上比较结果相同，则 C-RP 优先级较高者获胜（优先级数值越小优先级越高）。
 - 如果以上比较结果都相同，则执行 Hash 函数，计算结果较大者获胜。
 - 如果以上比较结果都相同，则 C-RP 的 IP 地址较大者获胜。



首次形成组播分发树

- PIM-SM (ASM) 模式首次形成组播分发树主要依赖RPT构建机制, 组播源注册机制与DR选举机制。
 - RPT构建机制: 组播叶子路由器主动建立到RP的组播分发树 (RPT)
 - 组播源注册机制: 通过该机制形成组播源到RP的组播分发树 (SPT)
 - DR选举机制: DR负责源端或组成员端组播报文的收发, 避免重复组播报文, 同时成员端DR还负责发送Join加组消息。

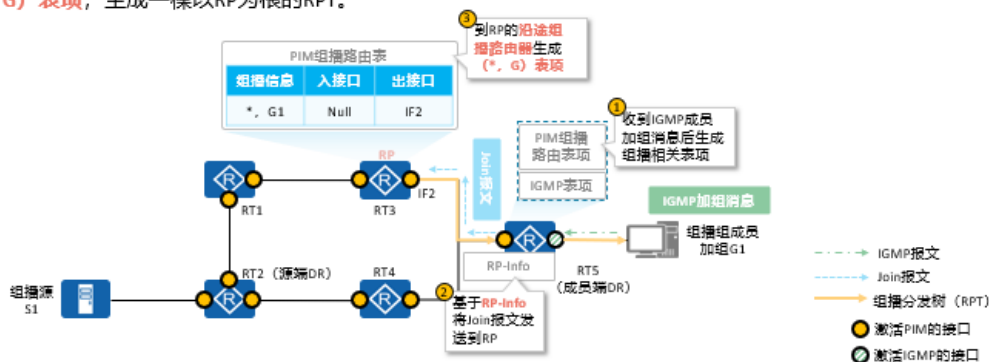


- DR 选举机制与 PIM-DM 中的 DR 选举机制一致, 不再赘述。



RPT构建

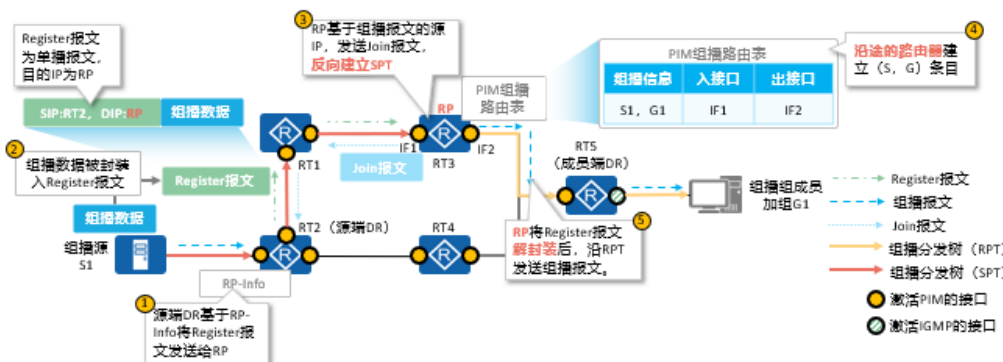
- RPT (RP Tree) 是一棵以RP为根, 以存在组成员关系的PIM路由器为叶子的组播分发树。
- 当网络中出现组成员 (形成IGMP表项) 时, 组成员端DR向RP发送Join报文, 在通向RP的路径上逐跳创建 (*, G) 表项, 生成一棵以RP为根的RPT。





组播源注册机制 – 形成SPT

- PIM-SM (ASM) 模型中，源端DR到RP的组播分发树无法使用Join报文创建，因此需要**组播源注册机制**帮助形成源端DR到RP的组播分发树 (SPT)。
- 形成SPT需要基于Register报文与Join报文，具体过程如下：

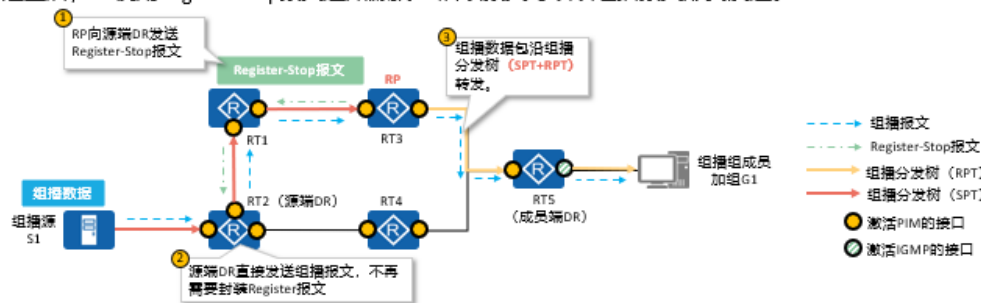


- 组播源与源端 DR 之间没有 IGMP 协议，无法通过 IGMP 生成 PIM (*, G) 表项，进而无法发送 Join 消息形成组播分发树。



组播源注册机制 – 转发组播数据

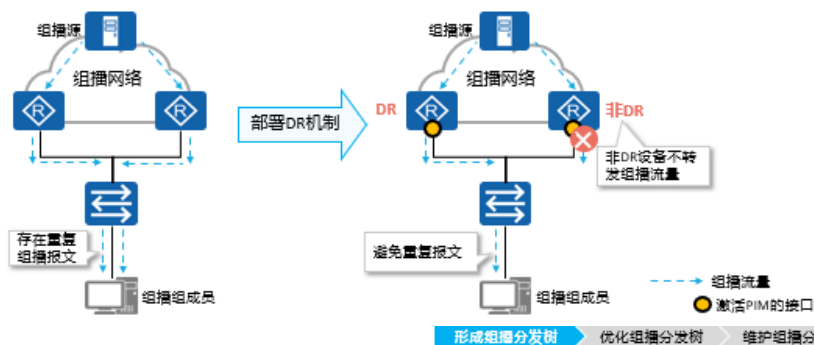
- 组播源信息注册到RP后，就形成了组播源到RP的SPT，但源端DR此时仍然会将组播数据包封装入Register报文，该方式会造成一些问题：
 - 源端DR最初发送的是单播Register报文，但是该方式会**加重源端DR与RP的工作量**。
 - 源端DR形成到RP的SPT后，**会同时发送单播Register报文和组播报文**，造成重复组播包的问题。
- SPT建立后，RP使用Register-Stop报文通知源端DR后续报文可以以组播报文形式发送。





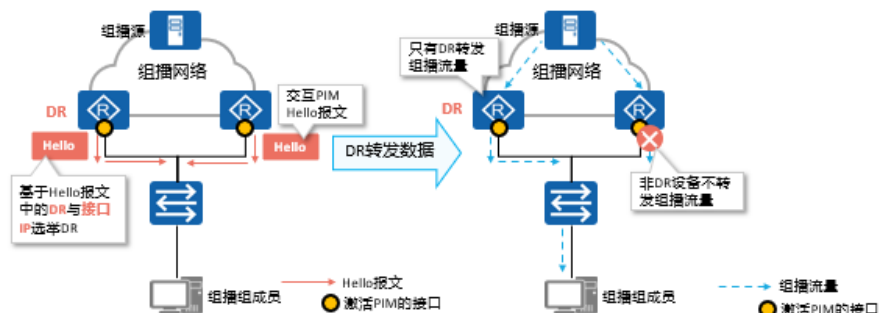
源/末端网络中的重复组播报文

- 在源端网络或者成员端网络中，有可能有多台组播路由器转发组播流量，从而造成重复组播报文的问题。
- PIM DR (Designated Router) 是源端网络或者成员端网络的唯一组播转发者，由于不存在别的组播转发路由器就避免了重复组播报文的问题。



PIM DR选举

- PIM DR的选举：
 - 在PIM-SM (ASM) 中各路由器通过比较Hello消息上携带的优先级和IP地址，为多路访问网络选举指定路由器DR。
 - 接口DR优先级高的路由器将成为该MA网络的DR，在优先级相同的情况下，接口IP地址大的路由器将成为DR。
 - 当DR出现故障后，邻居路由器之间会重新选举DR。

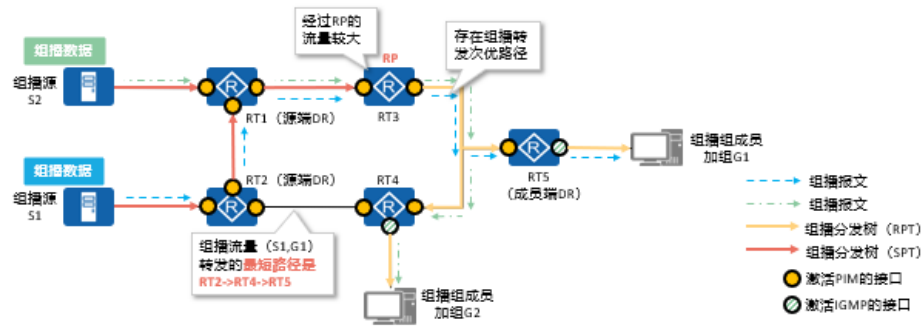


- DR 优先级默认为1，数值越大越优。
- 对于成员端网络，如果有多台组播路由器，则组播路由器的下行接口需要同时开启 IGMP 与 PIM。
- DR 还可充当 IGMPv1 的查询器。



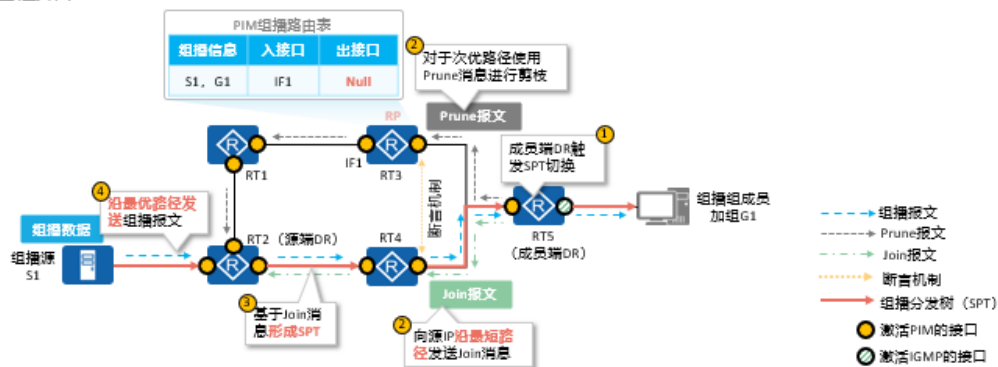
RPT次优路径问题

- 在PIM-SM网络中，一个组播组只对应一个RP。因此组播数据最初都会发往RP，由RP进行转发，这会导致两个问题：
 - 过大的组播流量会对RP形成巨大的负担。
 - 组播转发路径有可能是次优路径。



SPT切换机制

- 当数据发送至RP后，RP会沿RPT将数据发送给成员端DR。为了解决RPT潜在的次优路径问题，成员端DR会基于组播数据包中的源IP，反向建立从成员端DR到源的SPT。
- 具体过程如下：

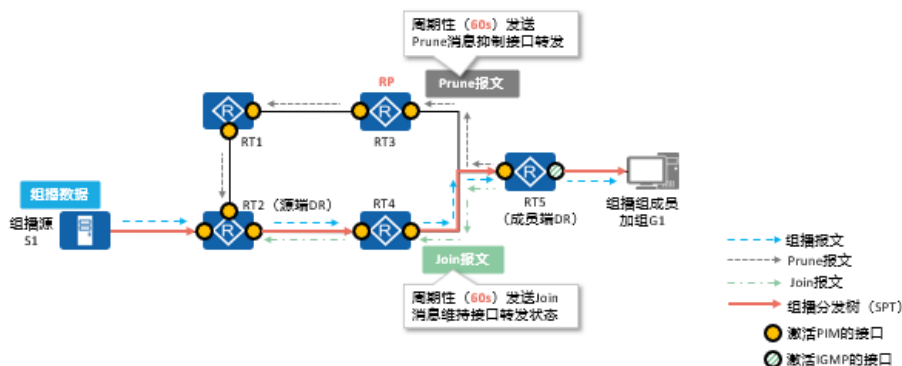


- 设备沿最短路径发送 Join 消息，该最短路径基于 RPF 选举规则决定。设备将 Join 消息从 RPF 选举得出的上行接口发出。
- 多路访问网络在 SPT 切换的过程中可能会存在重复报文，需要利用断言机制快速选定下行接口。
- SPT 切换的触发条件



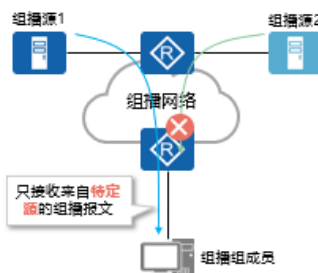
维护组播分发树

- 当组播分发树（SPT或RPT）稳定后，成员端DR会周期性发送Join/Prune报文，用于维护组播分发树。
- 如果组播在一段时间后（默认210s）没有流量则SPT树会消失，成员端DR恢复到RP的RPT树。



SSM概念回顾

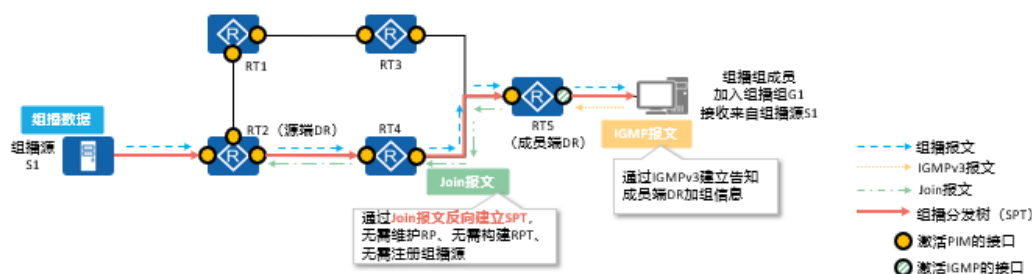
- SSM模型针对特定源和组的绑定数据流提供服务，接收者主机在加入组播组时，可以**指定只接收**哪些源的数据或**指定拒绝接收**来自哪些源的数据。加入组播组以后，主机**只会收到指定源**发送到该组的数据。
- SSM模型对组地址不再要求全网唯一，只需要每个组播源保持唯一。这里的“唯一”指的是同一个源上不同的组播应用必须使用不同的SSM地址来区分。不同的源之间可以使用相同的组地址，因为SSM模型中针对每一个（源，组）信息都会生成表项。这样一方面节省了组播组地址，另一方面也不会造成网络拥塞。





PIM-SM(SSM)基本概述

- 由于SSM提前定义了组播的源地址，所以PIM-SM（SSM）可以在成员端DR上基于组播源地址直接反向建立SPT。
- PIM-SM（SSM）**无需维护RP、无需构建RPT、无需注册组播源**，可以直接在组播源与组成员之间建立SPT。
- 在PIM-SM（SSM）模型中，关键机制包括邻居发现、DR竞选、构建SPT。

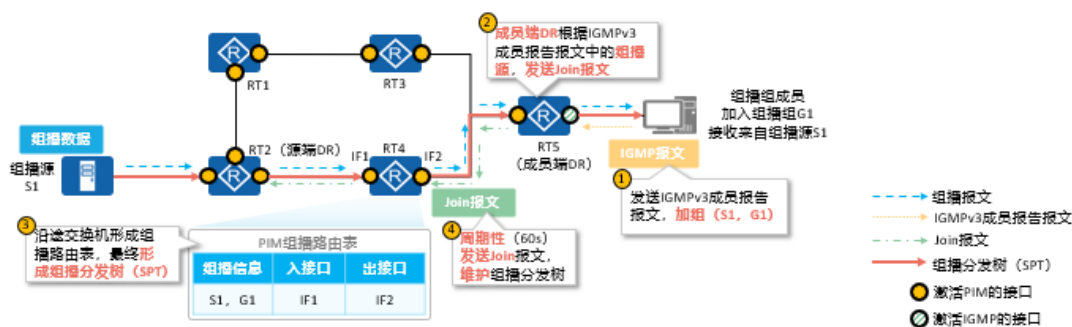


- PIM-SM (SSM) 无需 Assert 机制



组播分发树形成与维护

- PIM-SM（SSM）模型构建组播分发树的形成主要依赖IGMPv3报文与Join报文。
- PIM-SM（SSM）模型形成的**组播分发树会一直存在**，不会因为组播流量而消失。



- DR 与邻居发现机制与 PIM-DM 模式一样，此处不再赘述。



PIM模型比较

- 单击此处输入文字

协议	模型分类	适用场景	工作机制
PIM-DM	ASM模型	适合规模较小、组播组成员相对比较密集的局域网	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
PIM-SM	ASM模型	适合网络中的组成员相对比较稀疏，分布广泛的大型网络	采用接收者主动加入的方式建立组播分发树，需要维护RP、构建RPT、注册组播源
	SSM模型	适合网络中的用户预先知道组播源的位置，直接向指定的组播源请求组播数据的场景	直接在组播源与组成员之间建立SPT，无需维护RP、构建RPT、注册组播源



PIM-SM的基本配置

1. 使能路由器的组播路由功能

```
[Huawei] multicast routing-enable
```

2. 在接口上使能PIM-SM

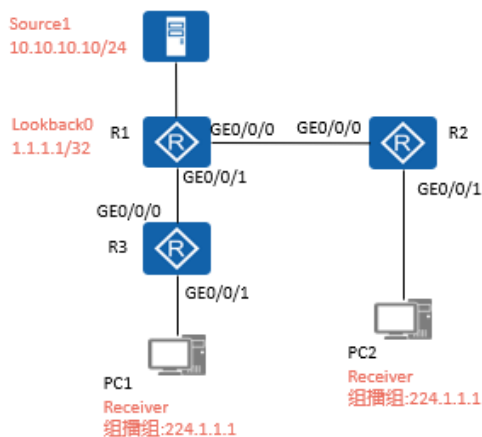
```
[Huawei - GigabitEthernet1/0/0] pim sm
```

3. 配置静态RP

```
[Huawei -pim] static-rp rp-address
```



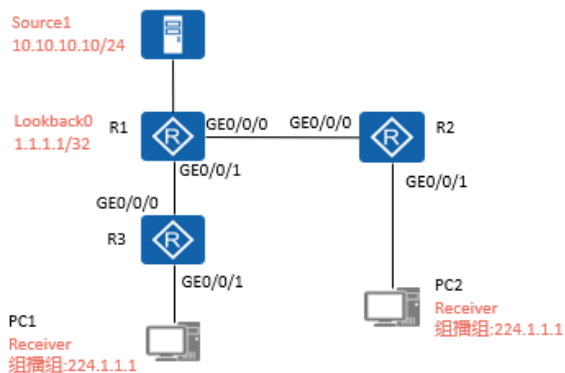
PIM-SM基础实验



- 实验要求:
- 通过PIM-SM协议让PC1、PC2可以接收到组播源的数据包。



PIM-SM的配置 (1)



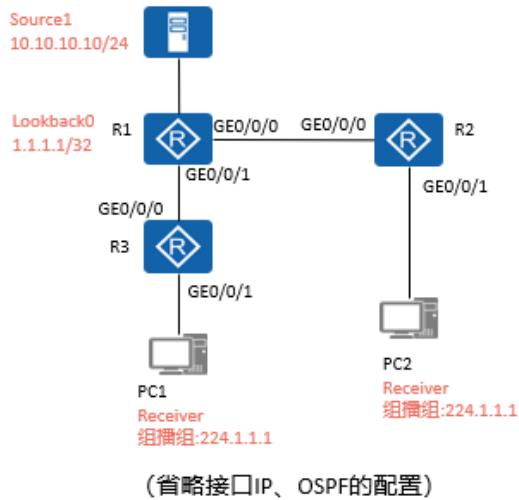
R1的配置如下:

```
[R1]multicast routing-enable
[R1]interface g0/0/2
[R1-GigabitEthernet0/0/2]pim sm
[R1-GigabitEthernet0/0/2]interface g0/0/0
[R1-GigabitEthernet0/0/0]pim sm
[R1-GigabitEthernet0/0/0]interface g0/0/1
[R1-GigabitEthernet0/0/1]pim sm
[R1] pim
[R1-pim] static-rp 1.1.1.1
```

(省略接口IP、OSPF的配置)



PIM-SM的配置 (2)



R2的配置如下:

```
[R2]multicast routing-enable
[R2]interface g0/0/0
[R2-GigabitEthernet0/0/0]pim sm
[R2-GigabitEthernet0/0/0]interface g0/0/1
[R2-GigabitEthernet0/0/1]igmp enable
[R2] pim
[R2-pim] static-rp 1.1.1.1
```

R3的配置如下:

```
[R3]multicast routing-enable
[R3]interface g0/0/0
[R3-GigabitEthernet0/0/0]pim sm
[R3-GigabitEthernet0/0/0]interface g0/0/1
[R3-GigabitEthernet0/0/1]igmp enable
[R3] pim
[R3-pim] static-rp 1.1.1.1
```

思考题 :

- (单选题) PIM 组播报文的目的 IP 地址是 ?
- 224.0.0.2
- 224.0.0.1
- 224.0.0.5
- 224.0.0.13
- (多选题) 组播防止重复报文的机制有 ?
- RPF 机制
- Assert 选举机制
- DR 选举机制
- IGMP 查询者选举机制

答案 :

- D
- A , B , C
-