PIM 协议原理与配置

IGMP（Internet Group Management Protocol,因特网组管理协议）

PIM: Protocol Independent Multicast　协议无关组播

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

==========

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
嫁接
断言

使用的 5 个 PIMv2 消息
Hello
Join / Prune
Graft
Graft-Ack
Assert

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
2、cost 值小获胜
3、本地接口 IP 大的获胜

==========

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
主播源注册
RPT 向 SPT 的切换
断言

RP 是 PIM-SM 的核心设备，所有的组播信息通过 RP 转发，RP 一般在路由器上静态指定。一般 PIM-SM 的域规模非常大，RP 的压力巨大，因此可以在 SM 中选举多个 C-RP（候选 RP），通过自动选举机制选举 RP，使不同的 RP 服务不同的组播组。
此时需要选举 BSR（bootstrap router 自举路由器），BSR 动态映射组播组与 RP 的关系，C-BSR 是 BSR 的备份。

使用的 7 个 PIMv2 消息：
Hello
Bootstrap
Candidate-RP-Advertisement
Join / Prune
Assert
Register
Register-Stop

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。

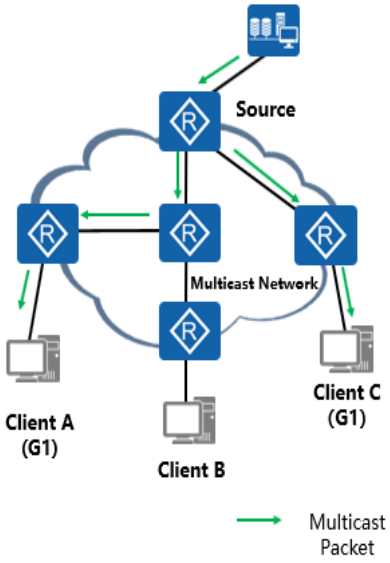
（S，G）路由表项主要用于在 PIM 网络中建立 SPT。对于 PIM-DM 网络和 PIM-SM 网络适用。
（*，G）路由表项主要用于在 PIM 网络中建立 RPT。对于 PIM-SM 网络适用。
224.0.0.1 地址（表示同一网段内所有主机和路由器）
224.0.0.2 地址（本地网段内的所有组播路由器）
RPF（Reverse Path Forwarding，逆向路径转发）。

前言

- 组播报文发送给一组特定的接收者，这些接收者可能分布在网络中的任意位置。为了实现组播报文正确、高效地转发，组播路由器需要建立和维护组播路由表项。
- 随着多个组播路由协议的开发与应用，人们渐渐感觉到，如果像单播路由一样通过多种路由算法动态生成组播路由，会带来不同路由协议间在互相引入时操作繁琐的问题。
- PIM（Protocol Independent Multicast）直接利用单播路由表的路由信息进行组播报文RPF检查，创建组播路由表项，转发组播报文。

路由器如何转发组播报文

- 路由器需要依据哪些信息进行转发：
  - 各接口所在网段有无潜在接收者。
  - 接收者需要接收哪些组的数据。
- 人工配置上述信息存在一些问题：
  - 实时性差。
  - 灵活性差。
  - 工作量大、易出错。



- 在单播报文的转发机制中，路由器依据单播报文的目的IP地址，查找单播路由表进行转发。其中，单播路由表可以通过静态配置或者动态路由协议来学习路由。
- 在组播中，接收者可能存在于全网中的任意位置，所以如果静态配置组播路由的话，存在实时性差、灵活性差以及工作量大容易出错的问题。
- 为了正确、高效的转发组播数据报文，路由器之间则需要运行组播路由协议。

PIM-DM基本概述

- 采用“推（Push）模式”转发组播报文。
- PIM-DM的关键任务：
  - 建立SPT（Shortest Path Tree，最短路径树）。
- PIM-DM的工作机制：
  - 邻居发现。
  - 扩散与剪枝。
  - 状态刷新。
  - 嫁接。
  - 断言。

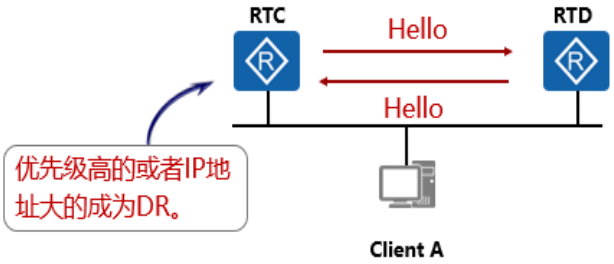
- PIM（Protocol Independent Multicast）协议无关组播，目前常用版本是 PIMv2，PIM 报文直接封装在 IP 报文中，协议号为 103，PIMv2 组播地址为 224.0.0.13。
- 在 PIM 组播域中，以组播组为单位建立从组播源到组成员的点到多点的组播转发路径。由于组播转发路径呈现树型结构，也称为组播分发树（MDT，Multicast Distribution Tree）。
- 组播分发树的特点：
- 无论网络中的组成员有多少，每条链路上相同的组播数据最多只有一份。
- 被传递的组播数据在距离组播源尽可能远的分叉路口才开始复制和分发。
- PIM 有两种模式：
- PIM-DM（Protocol Independent Multicast – Dense Mode）。
- PIM-SM（Protocol Independent Multicast – Sparse Mode）。
- 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。

- 其设计思想是：
- 首先将组播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。
- PIM-DM 的关键工作机制包括邻居发现、扩散与剪枝、状态刷新、嫁接和断言。

- 使用Hello机制发现邻居：

- 选举DR：

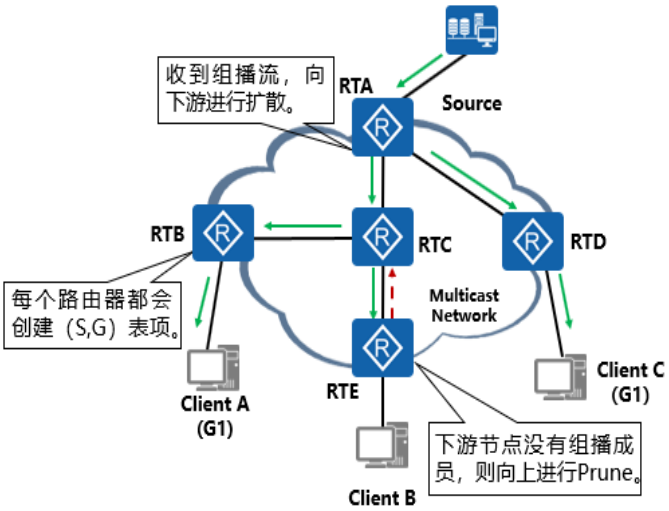


- 在 PIM-DM 网络中，路由器周期性发送 Hello 消息来发现、建立并维护邻居关系。
- pim timer hello interval，在接口视图下配置发送 Hello 消息的时间间隔。 Hello 消息默认周期是 30 秒。
- pim hello-option holdtime interval，在接口视图下配置 Hello 消息超时时间值。默认情况超时时间值为 105 秒。
- DR 的选举：
- 在 PIM-DM 中各路由器通过比较 Hello 消息上携带的优先级和 IP 地址，为多路访问网络选举指定路由器 DR。

- DR 充当 IGMPv1 的查询器。
- 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。
- 当 DR 出现故障后，邻居路由器之间会重新选举 DR。

PIM-DM构建SPT

- 扩散过程。
- RPF检查。
- 剪枝过程。



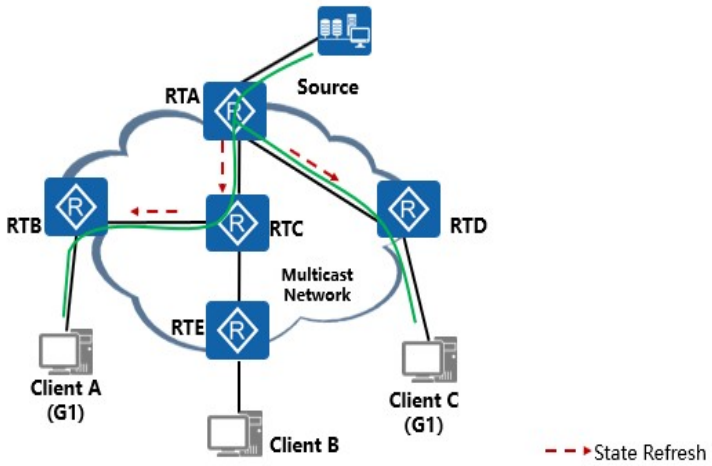
- 扩散过程：PIM-DM 假设网络中所有主机都准备接收组播数据，当某组播源开始向组播组 G 发送数据时，具体过程如下：
- 路由器接收到组播报文时会进行 RPF 检查。
- 如果 RPF 检查通过，则创建（S，G）表项，然后将数据向所有下游 PIM-DM 节点转发，这个过程称为扩散（Flooding）。
- 如果 RPF 检查没有通过，则将报文丢弃。
- RPF 检查： 为了防止组播报文在转发过程中出现重复报文及环路的情况，路由器必须执行 RPF 检查。
- 所谓 RPF 检查，就是指路由器通过查找去往组播源的路由来判断所收到的组播报文是否来自于“正确的”上游接口。某一路由器去往某一组播源的路由所对应的出接口称为该路由器

上关于该组播源的 RPF 接口。一台路由器从某一接口收到一个组播报文后，如果发现该接口不是相应组播源的 RPF 接口，就意味着 RPF 检查失败，所收到的组播报文将被丢弃。

- 剪枝过程：当下游有没有组播成员，扩散组播报文会导致带宽资源的浪费。为避免带宽的浪费 PIM-DM 使用剪枝机制。
- 当下游节点没有组播组成员，则路由器向上游节点发 Prune 消息，通知上游节点不用再转发数据到该分支。上游节点收到 Prune 消息后，就将相应的接口从其组播转发表项（S，G）对应的输出发送列表中删除。剪枝过程继续直到 PIM-DM 中仅剩下了必要的分支，这就建立了一个以组播源为根的 SPT。
- 各个被剪枝的节点同时提供超时机制，当剪枝超时时重新开始扩散—剪枝过程。剪枝状态超时计时器的默认值为 210 秒。

状态刷新

- 周期性地刷新剪枝状态。



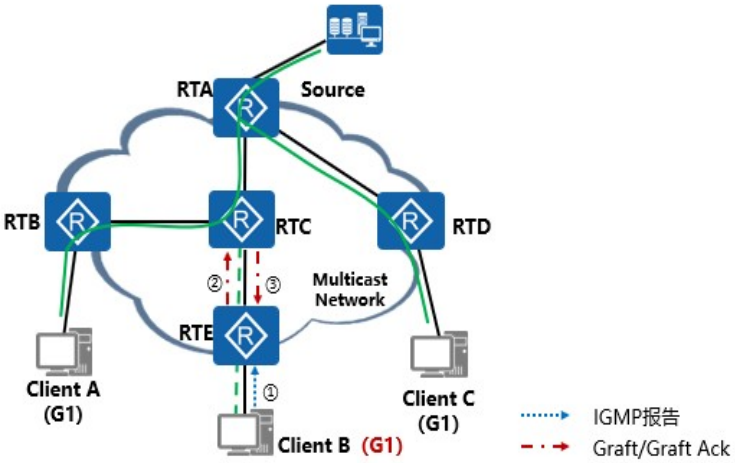
- PIM DM 协议采用状态刷新特性解决周期性“扩散-剪枝”带来的问题：离组播源最近的第一跳 RTA 周期性触发 State R

efresh 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。

- 状态刷新使得 RTE 不再周期性的收到组播数据，但是当 Client B 加入 G1 组之后，如果一直是剪枝状态，Client B 无法收到组播数据。
- 上述问题将如何解决？

Graft机制

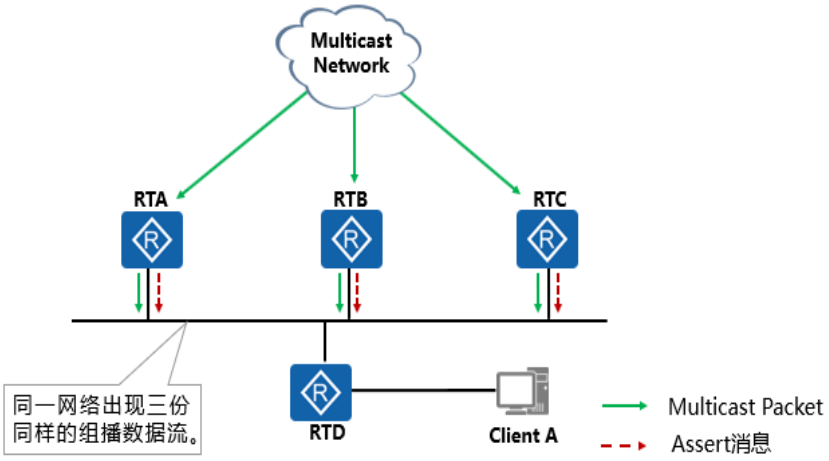
- 新的组成员加入组播组后，快速得到组播报文。



- 如图所示，当 Client B 发送组播组 G1 的 IGMP Report 报文请求组播数据后。RTE 收到 Client B 的 IGMP Report 报文，说明 RTE 具有转发组播数据需求，则立即向上游路由器 RTC 发送 Graft 消息，请求上游路由器恢复对应出接口的转发。RTC 收到 Graft 消息后，向 RTE 回复 Graft Ack 并将连接 RTE 的出接口恢复为转发状态。

Assert机制

- 避免重复组播报文。

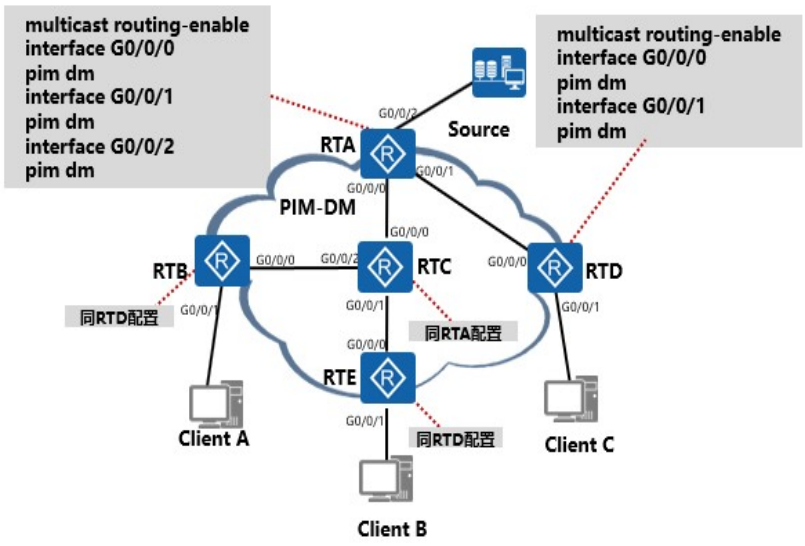


- 如图所示，RTA、RTB、RTC 均从上游接口收到组播报文并通过了 RPF 检查，三台路由器的下游接口连接在同一网段。RTA、RTB、RTC 都向该网段发送组播报文，三份重复的组播报文浪费带宽资源。
- 为避免重复的组播报文浪费带宽资源，PIM 路由器在接收到邻居路由器发送的相同组播报文后，会以组播的方式向本网段的所有 PIM 路由器发送 Assert 消息，其中目的地址为 224.0.0.13。其它 PIM 路由器在接收到 Assert 消息后，将自身参数与对方报文中携带的参数做比较，进行 Assert 竞选。竞选规则如下：
- 到组播源的单播路由协议优先级较小者获胜。
- 如果优先级相同，则到组播源的路由协议开销较小者获胜。
- 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。
- 根据 Assert 竞选结果，路由器将执行不同的操作：
- 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。
- Assert 竞选结束后，该网段上只存在一个下游接口，只传输一份组播报文。
- 所有 Assert Loser 可以周期性地恢复组播报文转发，从而引发周期性的 Assert 机制。

PIM-DM配置实现



PIM-DM配置验证

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

(192.168.0.1, 239.255.255.250)
    Protocol: pim-dm, Flag: ACT
    UpTime: 00:00:09
    Upstream interface: GigabitEthernet0/0/0
        Upstream neighbor: 10.1.14.1
        RPF prime neighbor: 10.1.14.1
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: GigabitEthernet0/0/1
            Protocol: pim-dm, UpTime: 00:00:09, Expires:  -
```

```
<RTD>display pim neighbor
VPN-Instance: public net
Total Number of Neighbors = 1

Neighbor        Interface    Uptime   Expires  Dr-Priority  BFD-Session
10.1.14.1       GE0/0/0      00:12:19 00:01:16      1          N
```

PIM-DM的局限性

- PIM-DM适用于组播成员分布较为密集的园区网络。
- PIM-DM的局限性：
  - 在组播成员分布较为稀疏的网络中，组播流量的周期性扩散会给网络带来较大负担。

- PIM-DM 适用于组播成员分布较为密集的园区网络。
- 在组播成员分布相对较为稀疏的大规模网络中（Internet），组播流量的周期性扩散/剪枝将给网络带来极大的负担。
- 对于 PIM-DM 的局限性，PIM-SM 可以提供相对更加有效的解决方案。

PIM-SM基本概述

- 使用“拉（Pull）模式”转发组播报文。
- 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。

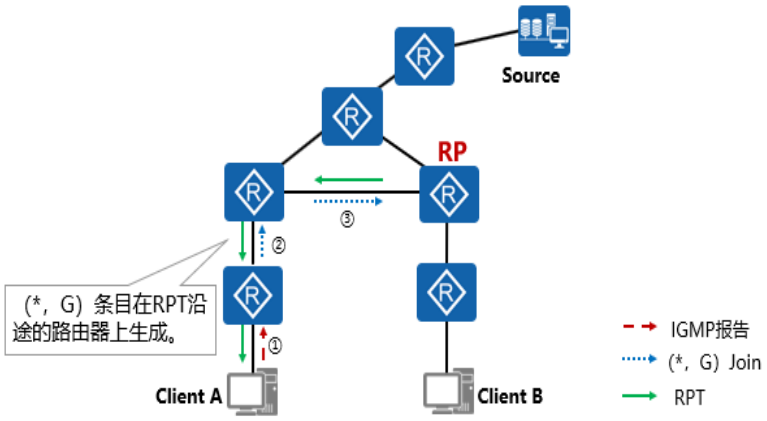
- 相对于 PIM-DM 的“推（Push）模式”，PIM-SM 使用“拉（Pull）模式”转发组播报文。PIM-SM 假设网络中的组成员分布非常稀疏，几乎所有网段均不存在组成员，直到某网段出现组成员时，才构建组播路由，向该网段转发组播数据。一般应用于组播组成员规模相对较大、相对稀疏的网络。
- 基于这一种稀疏的网络模型，它的实现方法是：
- 在网络中维护一台重要的 PIM 路由器：汇聚点 RP（Rendezvous Point），可以为随时出现的组成员或组播源服务。网络中所有 PIM 路由器都知道 RP 的位置。
- 当网络中出现组成员（用户主机通过 IGMP 加入某组播组 G）时，最后一跳路由器向 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。

- RP 的作用：
- RP 是 PIM-SM 域中的核心路由器，担当 RPT 树根节点。
- 共享树里所有组播流量都要经过 RP 转发给接收者。
- 用户通过配置命令限制 RP 所提供服务的组播组范围。
- RP 可以静态指定也可动态选举：
- 静态指定是指由管理员在每台 PIM-SM 路由器上进行配置，使得每台路由器获知 RP 的位置。
- 动态选举是指通过专用协议在若干台 C-RP（Candidate-RP）中选举产生。管理员需要开启选举协议并配置若干台 PIM-SM 路由器成为 C-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。

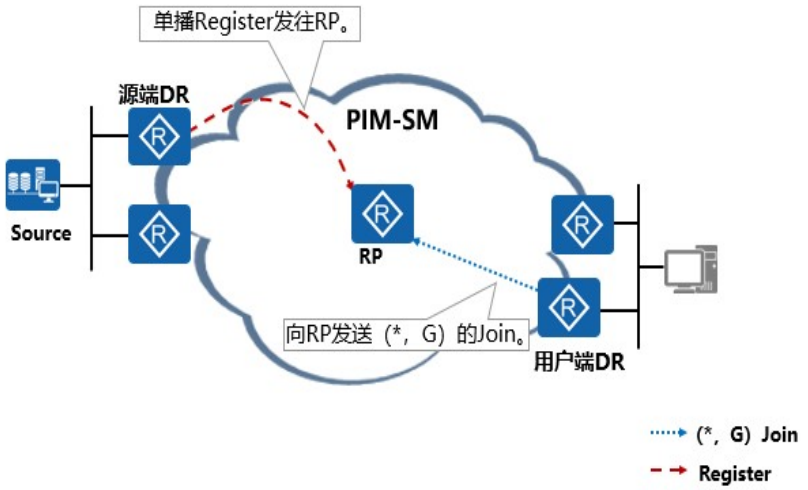
RPT及其建立过程



- 思考：如果连接Client A的路由器有两台，这两台路由器都会向RP发送（*，G）Join消息吗?

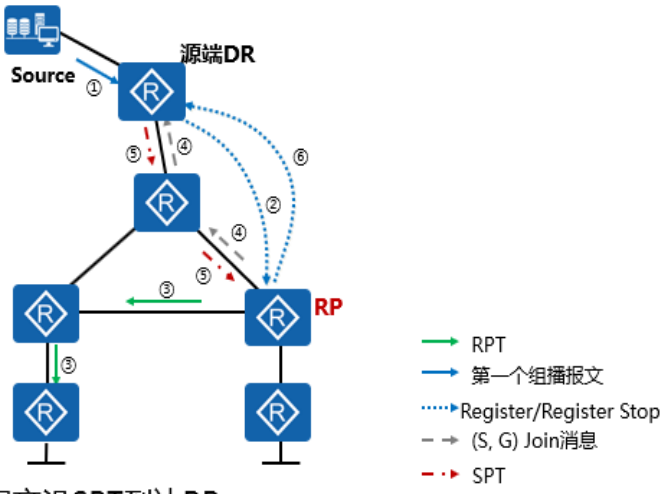
- RPT 的建立过程：
- 主机加入某个组播组时，发送 IGMP 成员通告。
- 最后一跳路由器向 RP 发送（*，G）Join 消息。
- （*，G）Join 消息到达 RP 的过程中，沿途各路由器都会生成相应的（*，G）组播转发条目。
- RPT 实现了组播数据按需转发的目的，减少了数据泛洪对网络带宽的占用。

组播接收者侧DR与组播源侧DR



- 运行 PIM-SM 的网络，都会进行 DR（Designated Router）的选举。其中有两种 DR 分别称为接收者侧 DR 和组播源侧 DR。
- 组播接收者侧 DR：与组播组成员相连的 DR，负责向 RP 发送（*，G）的 Join 加入消息。
- 组播源侧 DR：与组播源相连的 DR，负责向 RP 发送单播的 Register 消息。
- PIM-SM 中 DR 的选举原则与 PIM-DM 相同。

SPT的建立过程



- SPT建好之后，组播报文沿SPT到达RP。

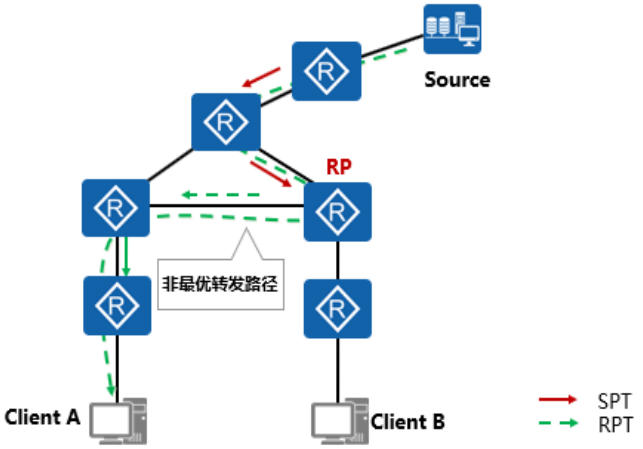
- 如图所示，在 PIM-SM 网络中，任何一个新出现的组播源都必须首先在 RP 处“注册”，继而才能将组播报文传输到组成员。具体过程如下：
- 组播源向组播组发送第一个组播报文。
- 源端 DR 将该组播报文封装成 Register 报文并以单播方式发送给相应的 RP。
- RP 收到注册消息后，一方面从 Register 消息中提取出组播报文，并将该组播报文沿 RPT 分支转发给接收者。
- 另一方面，RP 向源端 DR 发送(S，G)Join 消息，沿途路由器上都会生成相应(S，G)表项。从而建立了一颗由组播源至 RP 的 SPT 树。
- SPT 树建立后，组播源发出的组播报文沿该 SPT 转发至 RP。
- RP 沿 SPT 收到该组播报文后，向源端 DR 单播发送 Register-stop 消息。

(*，G) 与 (S，G) 条目关系

模式	类型	使用场景
PIM-DM	(S，G)	第一跳路由器到最后一跳路由器的SPT。
PIM-SM	(*，G)	RP到最后一跳路由器的RPT。
	(S，G)	源端DR到RP的SPT。
	(S，G)	Switchover之后，从第一跳路由器到最后一跳路由器的SPT。

PIM-SM的转发树

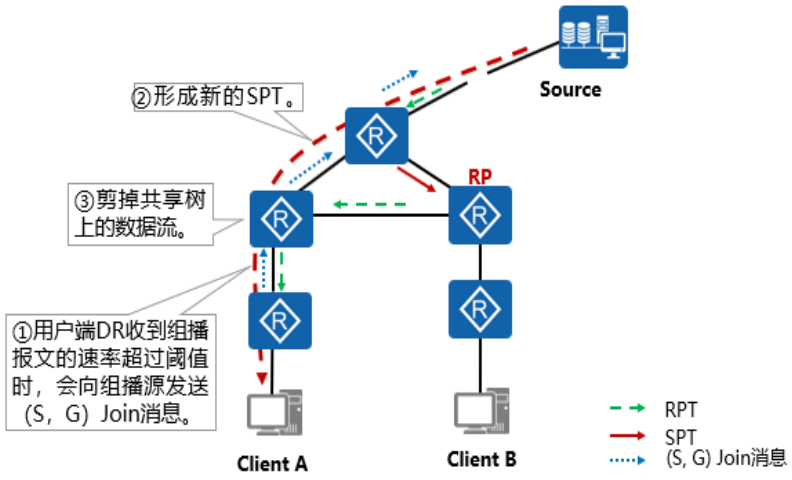
- SPT和RPT构成组播报文的转发路径，存在哪些问题?



- PIM-SM 同时包含了 SPT 和 RPT。通常情况下，组播源发出的组播报文会沿 SPT 到达 RP，然后从 RP 沿 RPT 到达接收者。
- 在这种情况下，从组播源到接收者的路径不一定是最优的，并且 RP 的工作负担非常大。为此，我们可以启用 RPT

向 SPT 进行的切换机制。

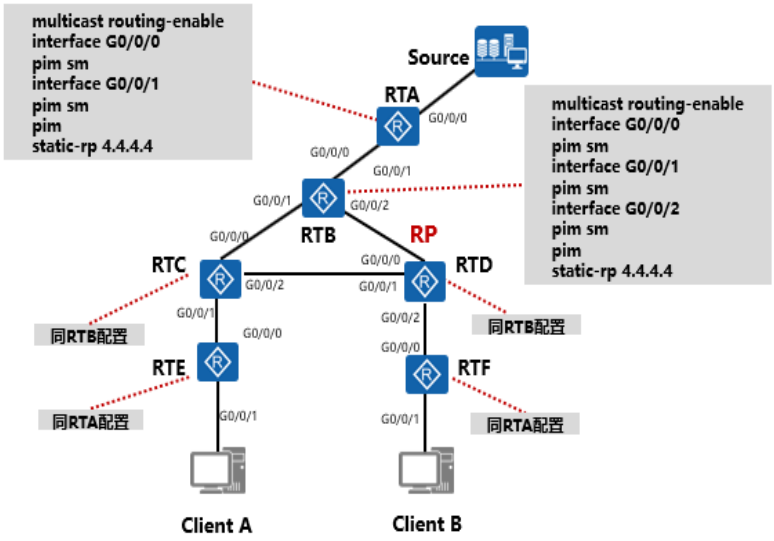
Switchover机制



- PIM-SM 通过指定一个利用带宽的 SPT 阈值可以实现 RPT 到 SPT 的切换。
- 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。

- 通过 RPT 树到 SPT 树的切换，PIM-SM 能够以比 PIM-DM 更精确的方式建立 SPT 转发树。

PIM-SM配置实现



PIM-SM配置验证

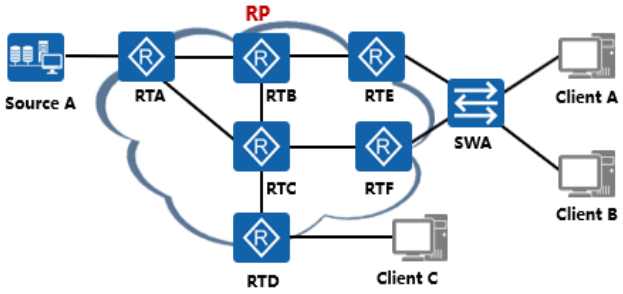
```
<RTF>display pim routing-table
VPN-Instance: public net
Total 1 (*, G) entry; 0 (S, G) entry

(*, 224.1.1.1)
    RP: 4.4.4.4
    Protocol: pim-sm, Flag: WC
    UpTime: 00:00:20
    Upstream interface: GigabitEthernet0/0/0
        Upstream neighbor: 10.1.46.4
        RPF prime neighbor: 10.1.46.4
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: GigabitEthernet0/0/1
            Protocol: igmp, UpTime: 00:00:20, Expires: -
```

```
<RTB>display pim neighbor
VPN-Instance: public net
Total Number of Neighbors = 3

Neighbor        Interface       Uptime   Expires  Dr-Priority  BFD-Session
10.1.12.1       GE0/0/0         00:04:08 00:01:27 1            N
10.1.23.3       GE0/0/1         00:01:29 00:01:16 1            N
10.1.24.4       GE0/0/2         00:03:19 00:01:25 1            N
```

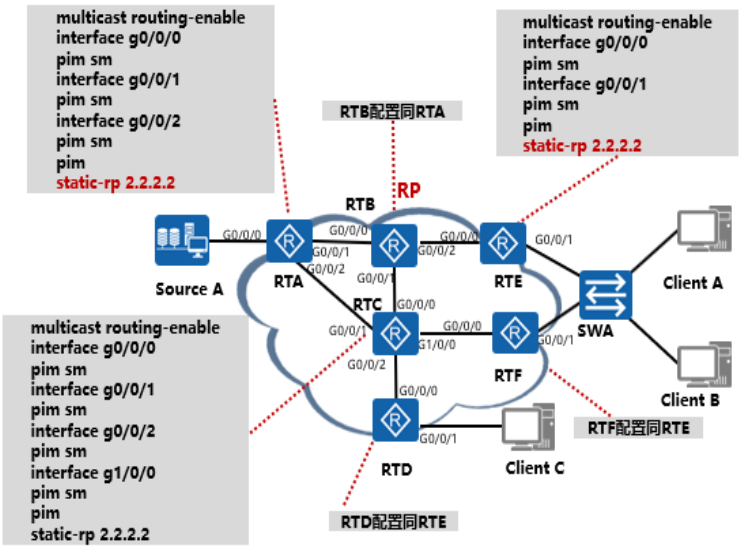
组播综合实验



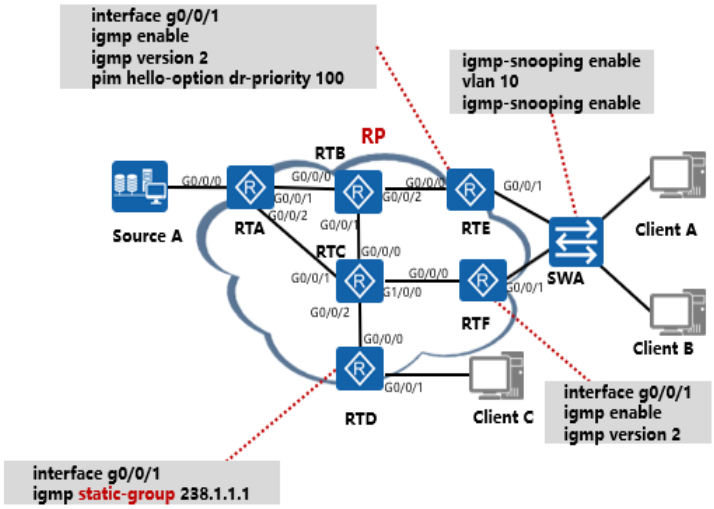
- 如图所示，要求在网络中部署PIM-SM协议且以静态方式指定RTB为RP。
- 用户侧网络配置IGMPv2协议，同时需要尽可能地降低用户侧网络资源消耗，提高安全性。
- RTE和RTF连接接收者，要求在RTE和RP之间建立RPT。
- RTD连接重要的用户网络，当用户加入组播组238.1.1.1组后，需要马上就能收到组播数据。

- 需求分析：
- 使能组播路由功能是配置 PIM-SM 的前提，首先在路由器上使能组播路由功能，其次在路由器接口上使能 PIM-SM 功能，最后在 PIM 视图下静态配置 RTB 为 RP。
- 在与用户侧相连的路由器接口上使能 IGMPv2。为了降低资源消耗、提高安全性，需要在 SWA 上使能 IGMP Snooping 功能，使交换机进行有效、安全的组播帧转发。
- 用户端 DR 负责向 RP 建立 RPT。根据 DR 的选举规则，需要把 RTE 接口的 DR 优先级设置为大于 1 的值（DR 优先级默认为 1）。
- 也就是说 RTD 上需要具有 238.1.1.1 的组播转发表项，RTD 在收到 IGMP Report 报文后，立即转发组播报文。可通过在 RTD 接口下配置静态加入 238.1.1.1 命令实现。

PIM-SM配置实现 (1)



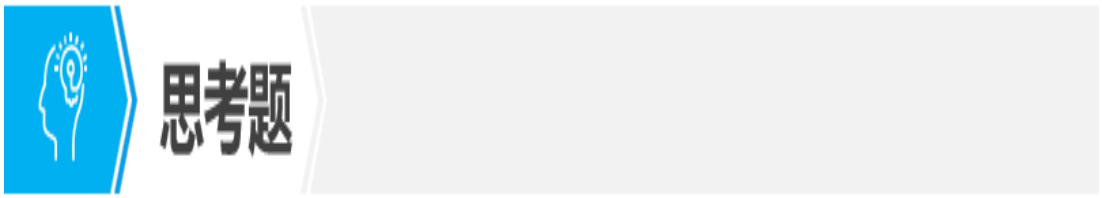
PIM-SM配置实现 (2)



- 配置静态组播组：
- 在某些特殊的应用场景中，比如：网络中存在稳定的组播组成员；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。
- 为了实现组播数据的快速、稳定转发，或者将组播数据

引流到接口，可以在组播路由器的用户侧接口上配置静态组播组。在接口上配置静态组播组后，组播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。

- 当成员主机无法解析组播 ping 报文并作出回应时，可以在组播路由器的用户侧接口上配置组播 ping 功能。这样，接口除了正常接收组播数据之外，还可以对收到的组播 ping 报文作出回应，从而使定位问题更加灵活、方便。



- 答案：组播分发树是指从组播源到接收者之间形成的一个单向无环数据传输路径。组播分发树有两类：SPT 和 RPT。
- 答案：Assert 可以避免在共享网络（如 Ethernet）中相同报文的重复发送。通过 Assert 机制在共享网络中来选定一个唯一的转发者。其他落选路由器则剪掉对应的接口以禁止转发信息。
- 答案：错误，与组播源相连的 DR 负责向 RP 发送单播的 Register 消息。
-