

组播路由协议PIM

朱仕耿

www.huawei.com

Author / Email : Zhushigeng 261992 / zhushigeng@huawei.com

Version 1.2 (2016-12-20)



课程目标

- 理解组播路由协议的基本概念。
- 掌握PIM-DM的应用场景及工作机制。
- 掌握PIM-DM的配置。
- 掌握PIM-SM的应用场景及工作机制。
- 掌握PIM-SM的配置。

目录

PIM概述

PIM-DM

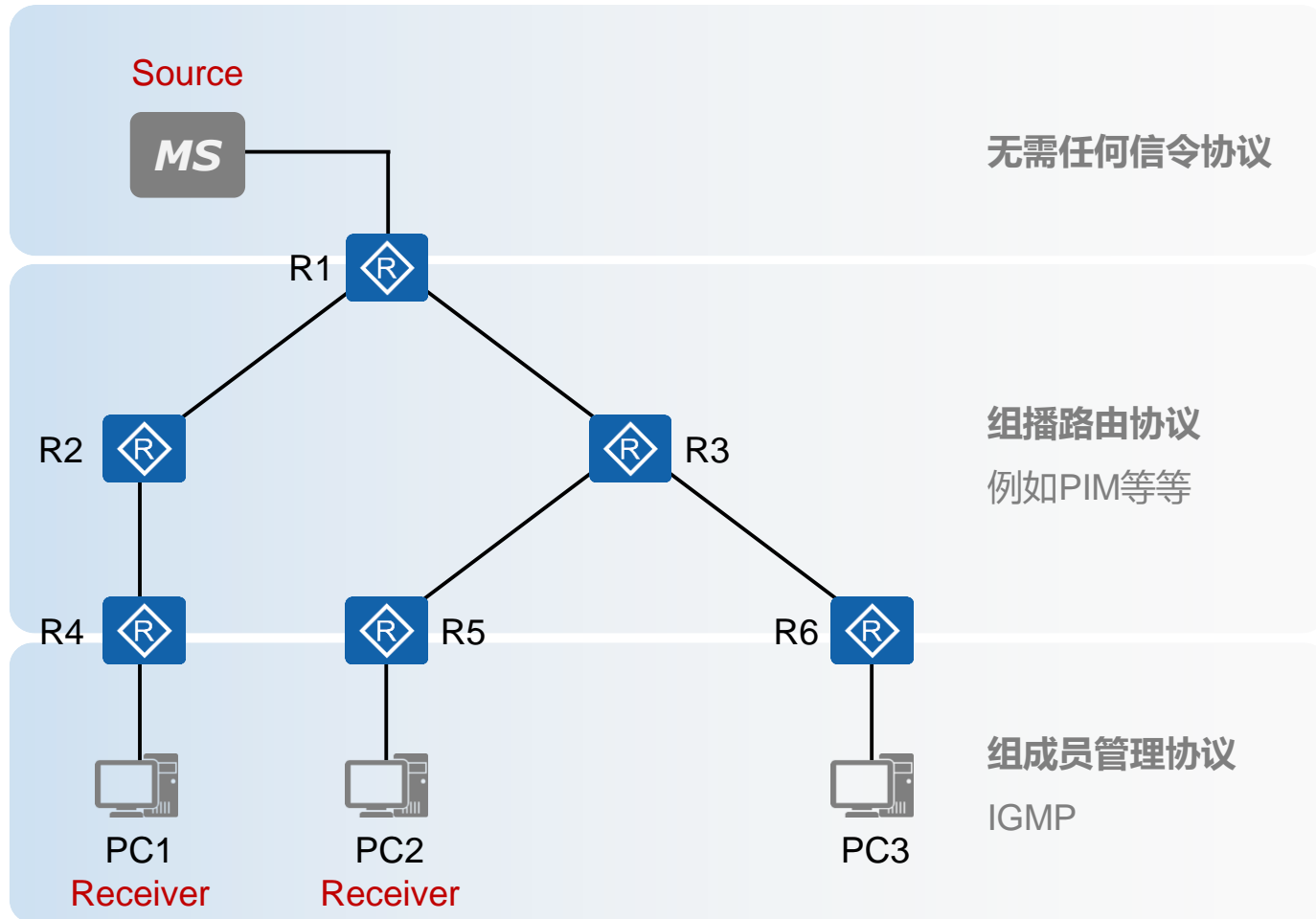
PIM-DM的基础配置

PIM-SM

PIM-SM的基础配置

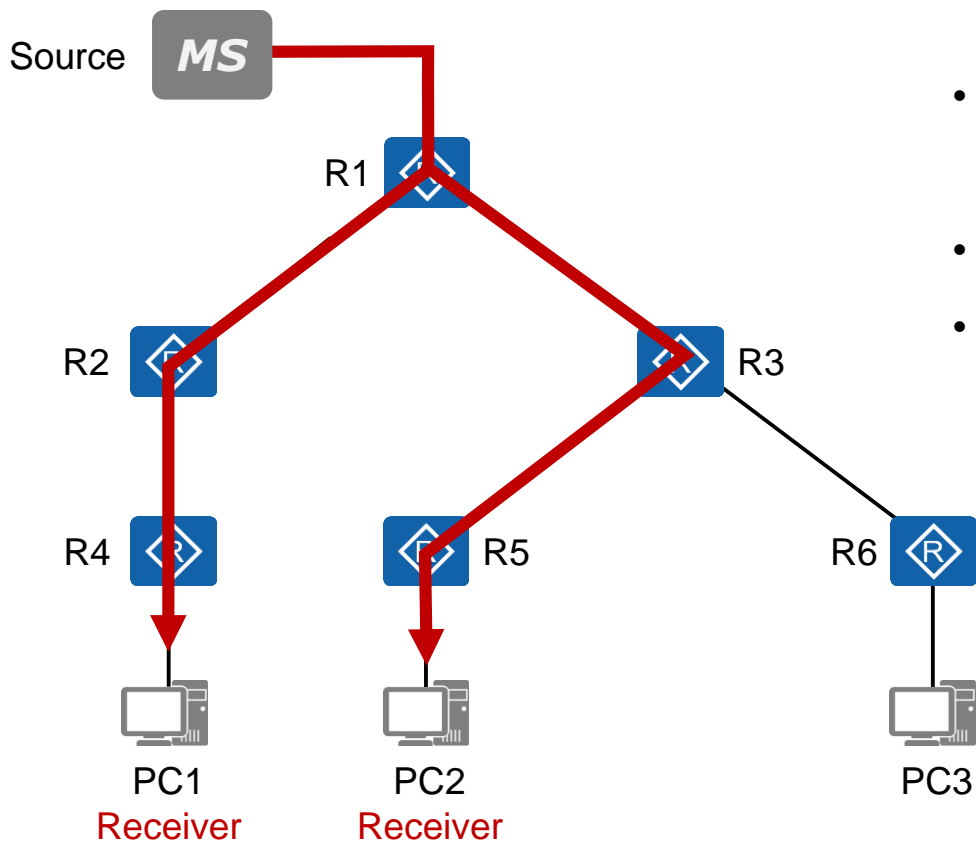


组播路由协议概述



- 组播路由器把数据拷贝并转发给需要该数据或存在组播接收者的网络分支；
- 组播流量的转发路径如何确定？
- 如何确保流量只被转发到正确的分支？
- 如何防止组播流量转发不会出现环路？

组播路由协议的主要功能



- 在接收组播报文时，判断该报文是否在正确的接口上到达，从而确保组播数据转发的无环化。
- 在网络中建立一棵组播分发树（组播流量转发的路径树）。
- 组播分发树体现在每一台组播路由器上便是（S，G）或（*，G）的组播转发表项。

PIM概述

- PIM (Protocol Independent Multicast) 称为协议无关组播。这里的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 (Reverse Path Forwarding , 逆向路径转发) 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主要有三种模式：
 - PIM-DM (PIM-Dense Mode)
 - PIM-SM (PIM-Sparse Mode)
 - 双向PIM (Bidirectional PIM)

PIM-DM密集模式

- PIM-DM使用“推（Push）模式”转发组播报文，一般应用于组播组成员规模相对较小、相对密集的网络。
- 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。当有活跃的组播源出现时，PIM-DM会将组播源发来的组播报文扩散到整个网络的PIM路由器上，再裁剪掉不存在组播报文转发的分支。
- PIM-DM通过周期性的进行“扩散（Flooding）—剪枝（Prune）”，来构建并维护一棵连接组播源和组成员的单向无环SPT（Shortest Path Tree，最短路径树，也称源树）。如果在下一次“扩散-剪枝”进行前，被裁剪掉的分支由于其叶路由器上有新的组成员加入而希望提前恢复转发状态，也可通过嫁接（Graft）机制主动恢复其对组播报文的转发。
- PIM-DM的关键工作机制包括邻居发现、扩散、剪枝、嫁接、断言和状态刷新。其中，扩散、剪枝、嫁接是构建SPT的主要方法。

PIM-SM

- PIM-SM使用“拉（Pull）模式”转发组播报文，一般应用于组播组成员规模相对较大、相对稀疏的网络。基于这样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（共享树）。
- 当网络中出现活跃的组播源（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。
- 在ASM模型中，PIM-SM的关键机制包括邻居发现、DR竞选、RP发现、RPT构建、组播源注册、SPT切换、断言；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。

目录

PIM概述

PIM-DM

PIM-DM的基础配置

PIM-SM

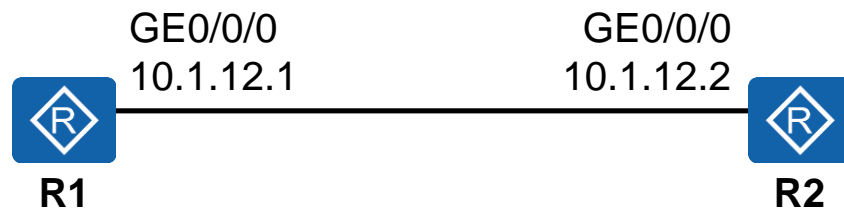
PIM-SM的基础配置

PIM-DM协议报文

报文类型	报文功能
Hello	用于建立及维护PIM邻居关系。
Join/Prune	加入/剪枝报文，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。
Graft	嫁接报文，用于将设备嫁接到组播分发树。
Graft-ACK	嫁接确认报文，用于对邻居发送的Graft报文进行确认。
Assert	断言报文，用于断言机制。

PIM邻居关系

- 在PIM路由器每个使能了PIM协议的接口上，都会对外发送Hello报文。封装Hello报文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报文的作用：发现PIM邻居、协调各项PIM协议报文参数、维持邻居关系。
- 只有邻居关系建立成功后，PIM路由器才能接收其他PIM协议报文，从而创建组播路由表项。



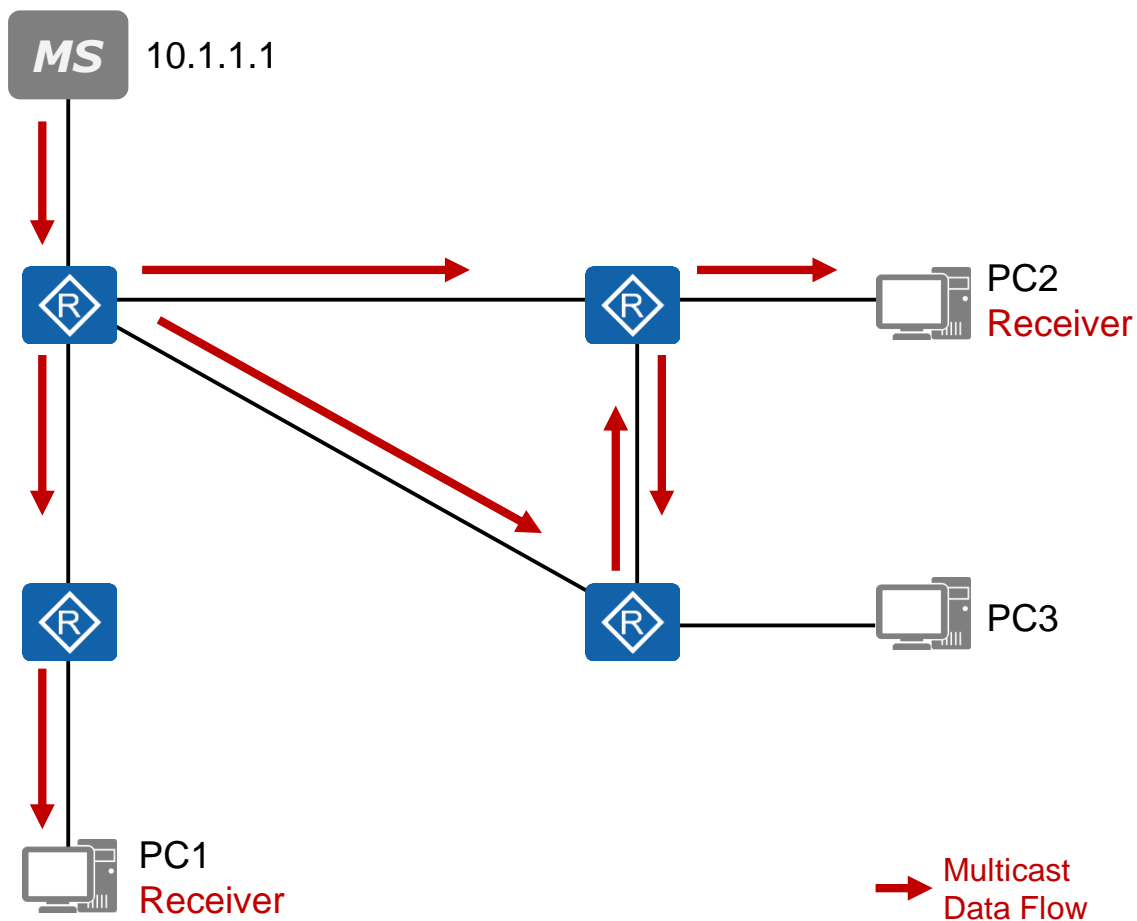
[R1] display pim neighbor

VPN-Instance: public net

Total Number of Neighbors = 1

Neighbor	Interface	Uptime	Expires	Dr-Priority	BFD-Session
10.1.12.2	GE0/0/0	00:02:47	00:01:27	1	N

密集模式协议：初始扩散过程



初始时，组播流量会被扩散到全网各个角落。即使不存在组播组成员的分支也会收到组播流量。每一台组播路由器都会在其组播路由表中创建（S,G）表项，在本例中，假设组播源10.1.1.1向239.1.1.1组播组发送流量，则表项为（10.1.1.1，239.1.1.1）。

密集模式协议：初始扩散过程

- 当PIM-DM网络中出现活跃的组播源之后，组播源发送的组播报文将在全网内扩散。当PIM路由器接收到组播报文，根据单播路由表进行RPF检查通过后，就会在该路由器上创建（S，G）表项，下游接口列表中包括除上游接口之外与所有PIM邻居相连的接口，后续到达的组播报文将从各个下游接口转发出去。
- 最后组播报文扩散到达叶子路由器，会出现以下两种情况：
 - 若与该叶子路由器相连用户网段上存在组成员，则将与该网段相连的接口加入（S，G）表项的下游接口列表中，后续的组播报文会向组成员转发。
 - 若与该叶子路由器相连用户网段上不存在组成员，且不需要向其下游PIM邻居转发组播报文，则执行剪枝机制。

PIM路由表

<Huawei>display pim routing-table

VPN-Instance: public net

Total 0 (*, G) entry; 1 (S, G) entry

(10.1.1.1, 239.1.1.1)

(S,G) 表项, S代表源, G
代表组播组

Protocol: pim-dm, Flag: ACT

UpTime: 00:14:22

Upstream interface: GigabitEthernet0/0/0

上行接口, 也就是RPF接口

Upstream neighbor: 10.1.12.1

RPF prime neighbor: 10.1.12.1

Downstream interface(s) information:

下行接口列表, 也就是组播数
据的出站接口

Total number of downstreams: 1

1: GigabitEthernet0/0/1

Protocol: pim-dm, UpTime: 00:00:15, Expires: never

组播路由表

<Huawei> display multicast routing-table

Multicast routing table of VPN-Instance: public net

Total 1 entry

00001. (10.1.1.1, 239.1.1.1)

Uptime: 00:00:06

Upstream Interface: GigabitEthernet0/0/0

List of 1 downstream interface

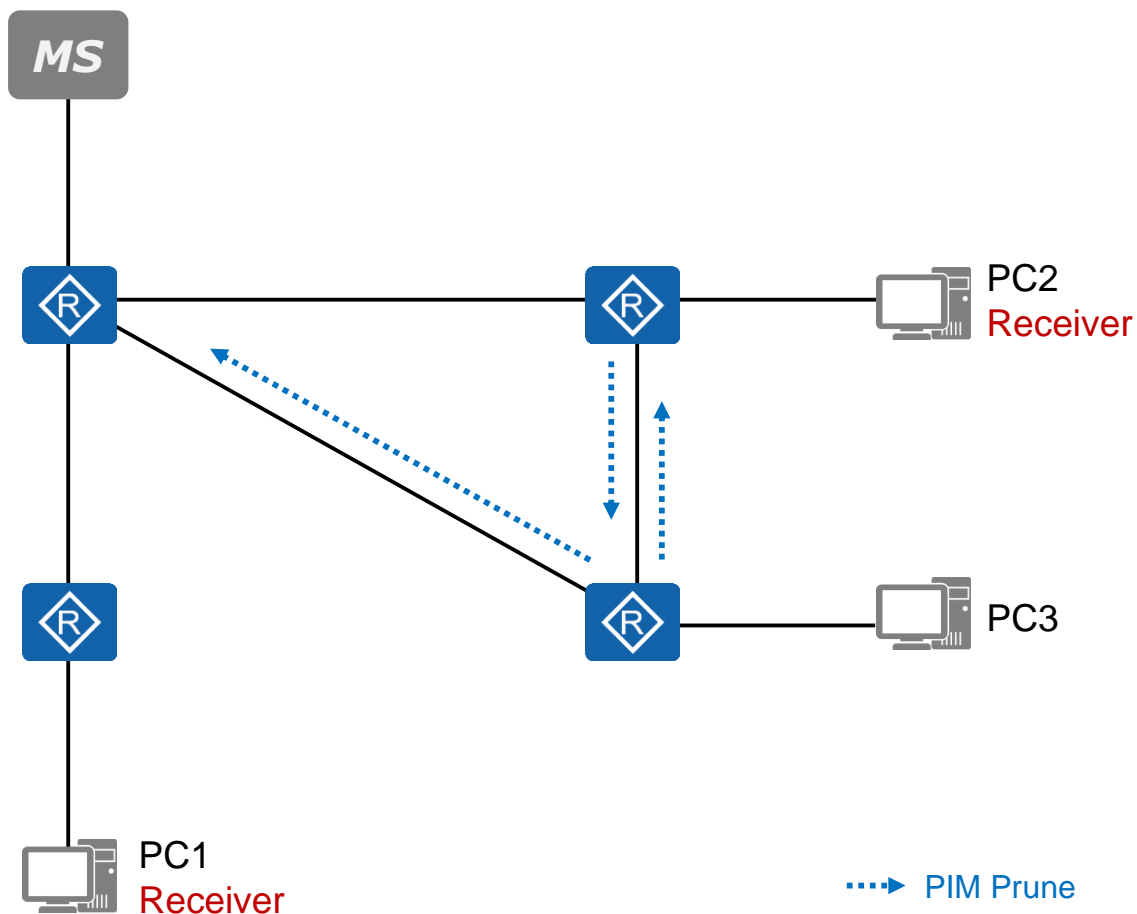
1: GigabitEthernet0/0/1

(S,G) 表项, S代表源, G
代表组播组

上行接口, 也就是RPF接口

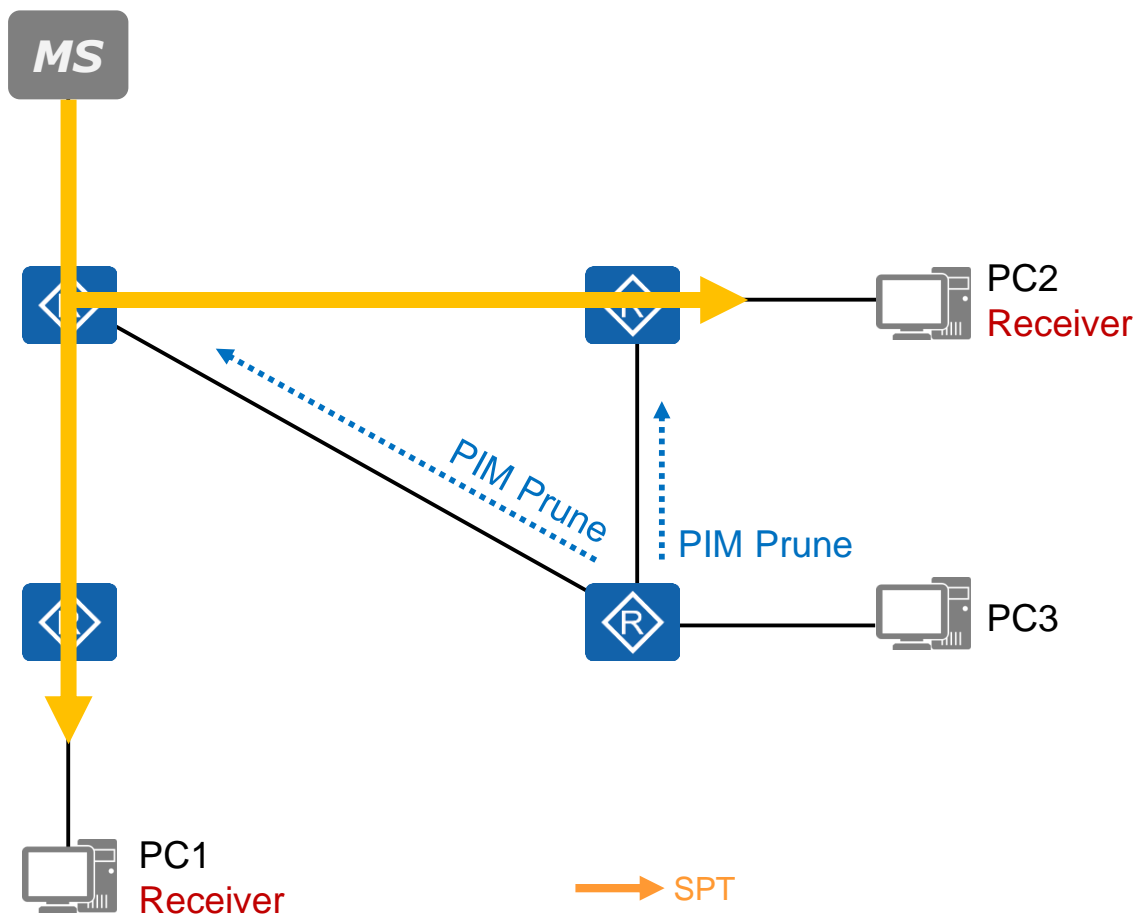
下行接口列表, 也就是组播数
据的出站接口

密集模式协议：剪枝过程



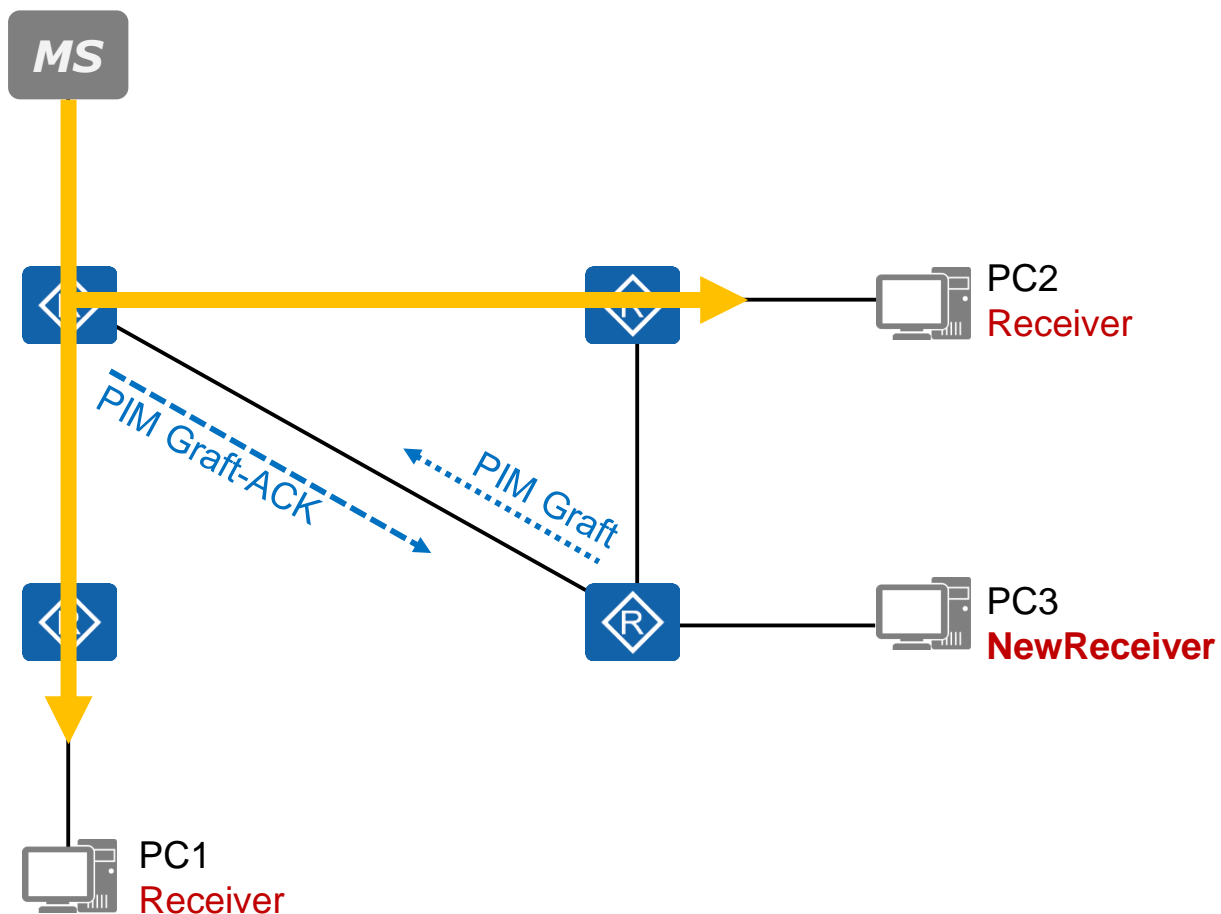
不需要 (10.1.1.1 , 239.1.1.1) 组播流量的分支，使用PIM Prune报文将自己从组播分发树上剪除。上游的路由器收到下游设备发送过来的Prune报文，则将接收该报文的接口从 (10.1.1.1 , 239.1.1.1) 表项的出接口列表中移除。

密集模式协议：新的SPT形成



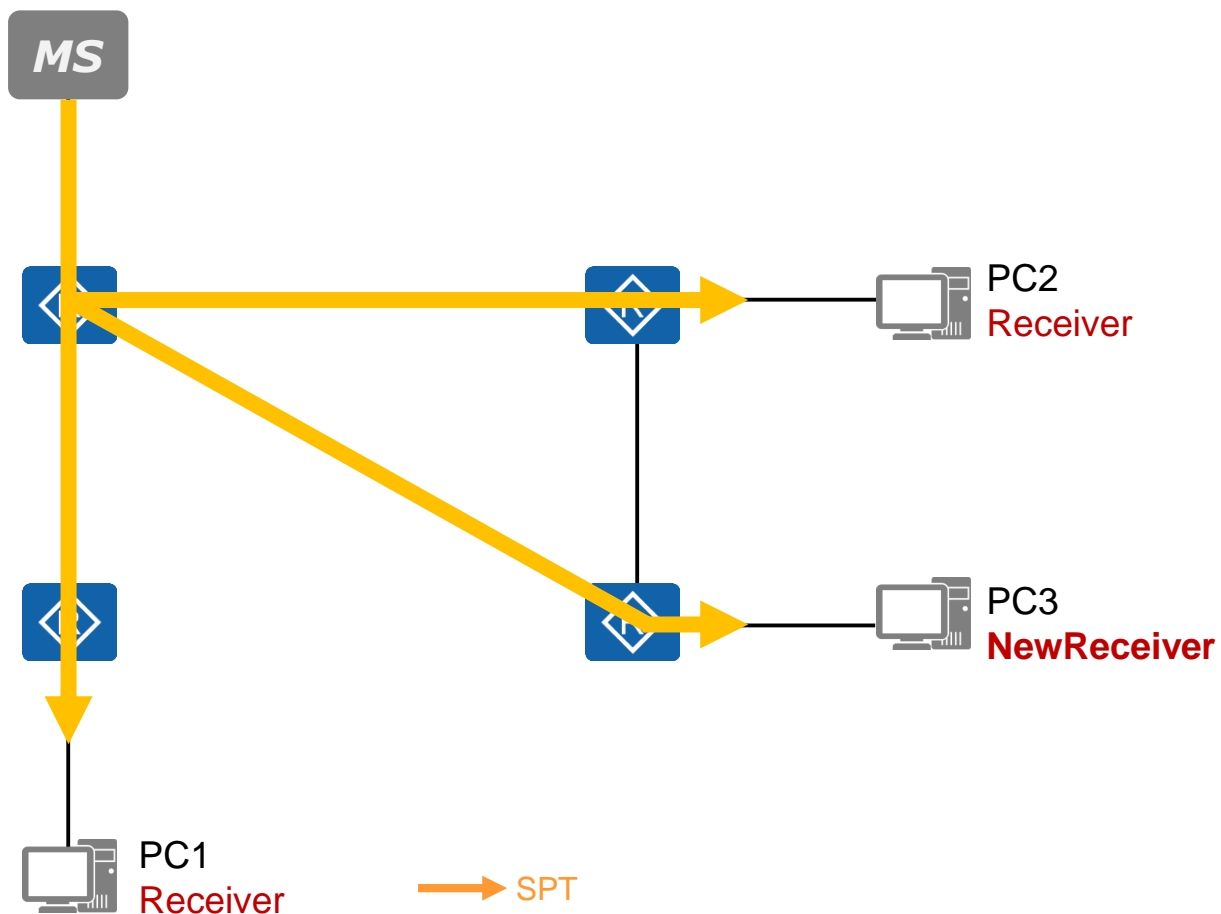
网络收敛完成后，SPT稳定下来，如图所示，组播流量将沿着SPT源源不断地从组播源转发到接收者。
不需要该组播流量的路由器需要周期性地向上游发送Prune报文。

密集模式协议：嫁接（Graft）



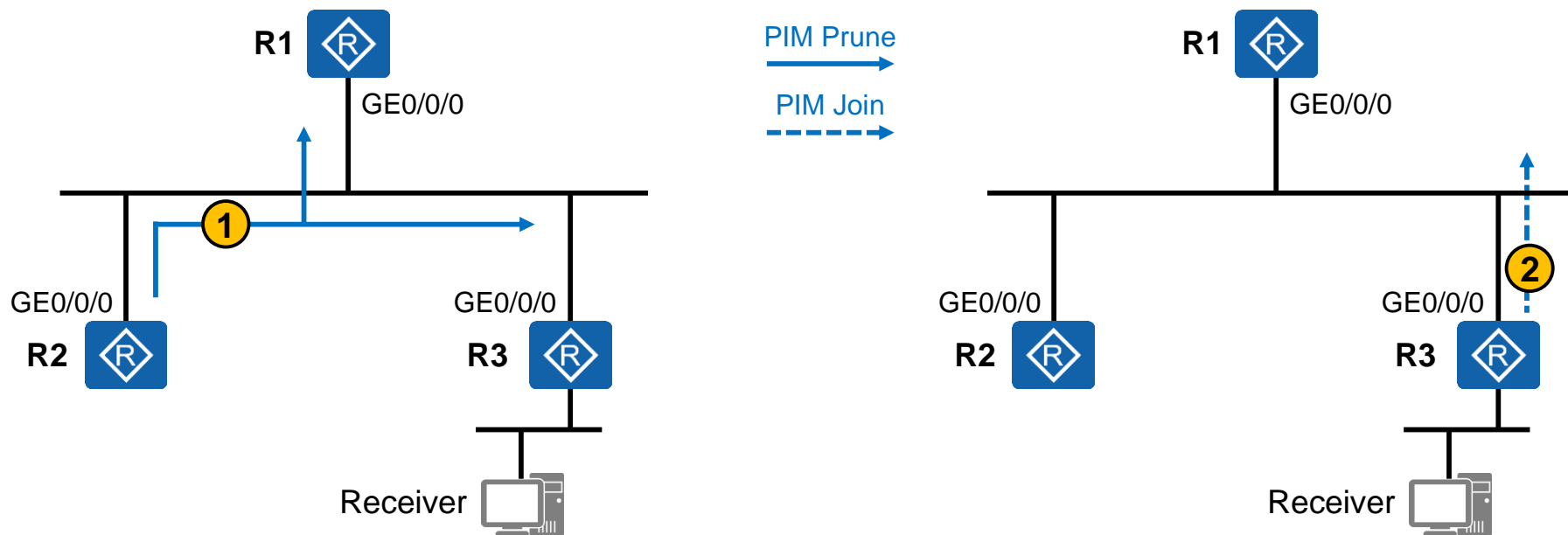
当原先不在SPT上的网络分支出现了接收者，则最后一跳路由器向其上游邻居发送PIM Graft（嫁接）报文，试图将自己嫁接到SPT。

密集模式协议：嫁接（Graft）



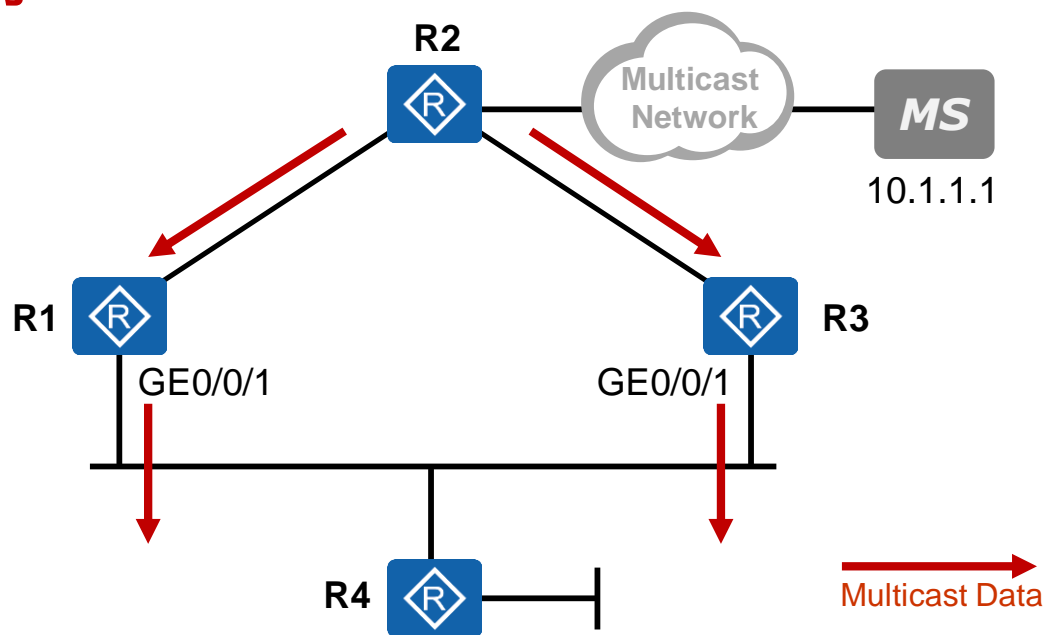
新的SPT形成，组播流量沿着SPT从源转发到Receiver。

剪枝否决机制



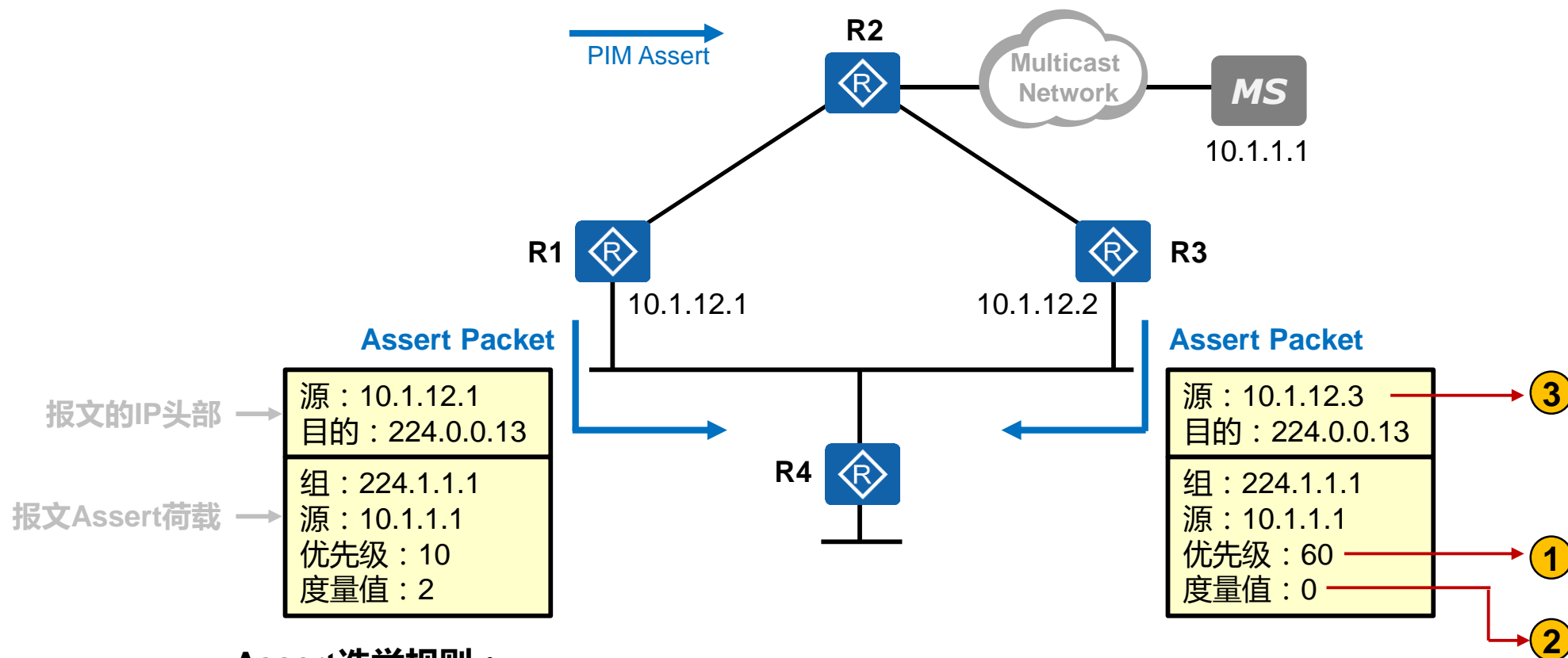
1. R2不再需要 (S,G) 组播流量，因此从上游接口发送Prune报文。R1及R3都将收到该报文。如果R1直接将接收报文的GE0/0/0接口设置为剪枝状态，那么便会导致R3无法接收 (S,G) 组播流量。
2. 因此当R1的GE0/0/0接口上存在多个PIM邻居时，如果它在该接口收到Prune报文，则不会立即将接口置为修剪状态，而是启动剪枝延迟计时器（缺省3s），在这个时间内，如果接口收到Join报文，则会忽略此前接收的Prune报文。
3. R3收到R2发送的Prune报文后立即发送Join报文给R1，用于否决R2的剪枝请求。

断言 (Assert) 机制



- R1、R3都从其RPF接口收到组播数据，如果它们都将组播流量从各自的GE0/0/1接口转发下去，那么R4将收到重复的组播数据。PIM设计了断言机制用于解决这个问题。
- 当一个网段内有多个相连的PIM路由器RPF检查通过向该网段转发组播报文时，则需要通过断言机制来保证只有一个PIM路由器向该网段转发组播报文。PIM路由器在接收到邻居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PIM Assert报文，报文的目的地址为224.0.0.13。其它PIM路由器在接收到Assert报文后，将自身参数与对方报文中携带的参数做比较，进行Assert竞选。

断言 (Assert) 机制



Assert选举规则：

1. 到达组播源 (10.1.1.1) 的单播路由协议优先级较高者 (优先级的值更小者) 获胜。
2. 如果优先级相同，则到组播源的单播路由的开销较小者获胜。
3. 如果以上都相同，则接口IP地址最大者获胜。

目录

PIM概述

PIM-DM

PIM-DM的基础配置

PIM-SM

PIM-SM的基础配置

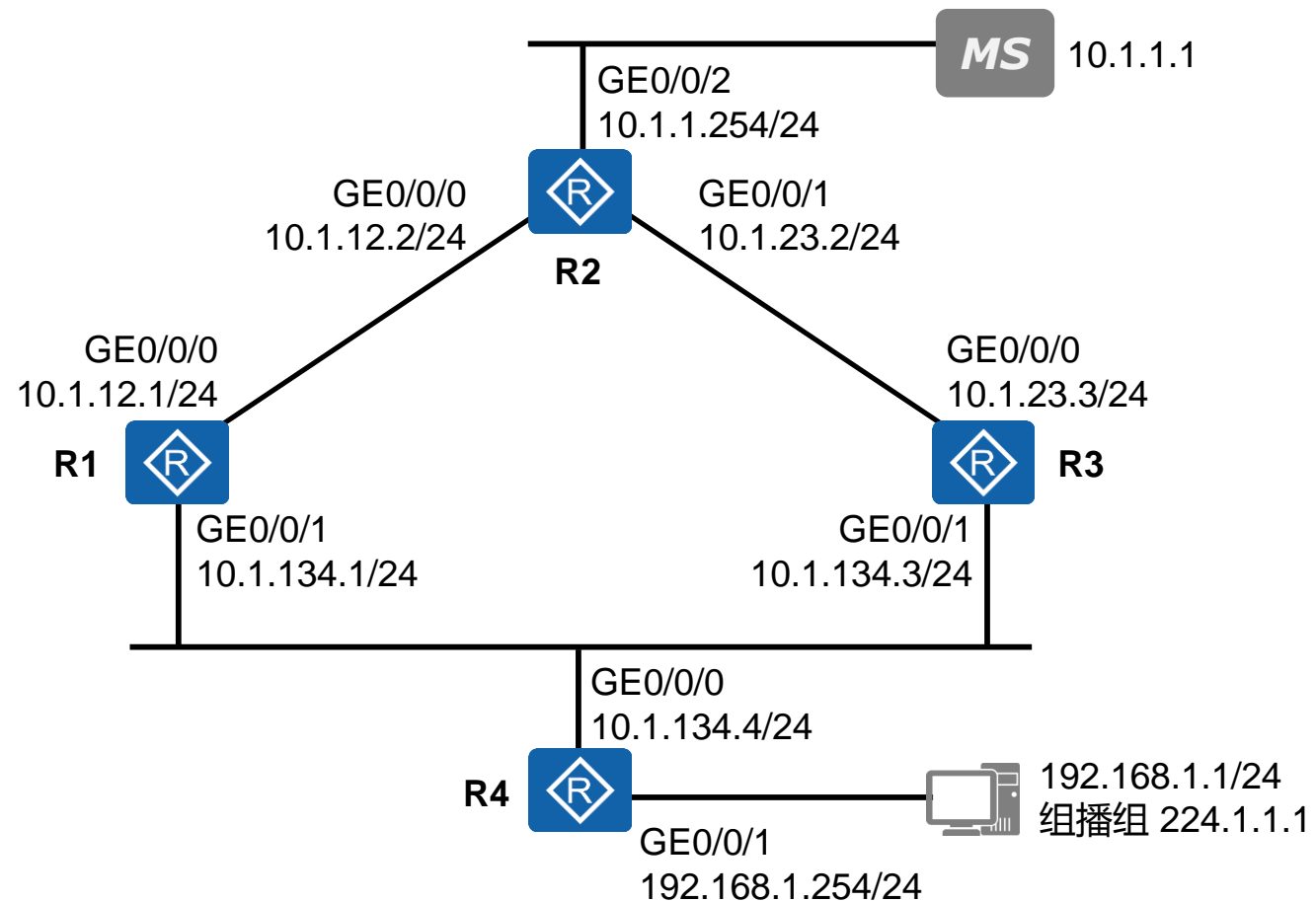
基础命令

- 激活路由器的组播路由功能：
[Huawei] **multicast routing-enable**

- 在接口上激活PIM-DM：
[Huawei] **interface GigabitEthernet 0/0/0**
[Huawei-GigabitEthernet 0/0/0] **pim dm**

- 查看PIM邻居：
[Huawei] **display pim neighbor**

基础实验



基础实验 设备配置

- **R1的关键配置如下（省略接口IP、OSPF的配置）**

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

- **R2的关键配置如下（省略接口IP、OSPF的配置）**

```
[R2] multicast routing-enable
[R2] interface GigabitEthernet0/0/0
[R2-GigabitEthernet0/0/0] pim dm
[R2] interface GigabitEthernet0/0/1
[R2-GigabitEthernet0/0/1] pim dm
[R2] interface GigabitEthernet0/0/2
[R2-GigabitEthernet0/0/2] pim dm
```

- **R3的关键配置如下（省略接口IP、OSPF的配置）**

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

- **R4的关键配置如下（省略接口IP、OSPF的配置）**

```
[R4] multicast routing-enable
[R4] interface GigabitEthernet0/0/0
[R4-GigabitEthernet0/0/0] pim dm
[R4] interface GigabitEthernet0/0/1
[R4-GigabitEthernet0/0/1] igmp enable
```

目录

PIM概述

PIM-DM

PIM-DM的基础配置

PIM-SM

PIM-SM的基础配置

RP (Rendezvous Point)

- **关于RP**

- 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路由器都必须知道RP的地址，类似于一个供求信息的汇聚中心。
- 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。

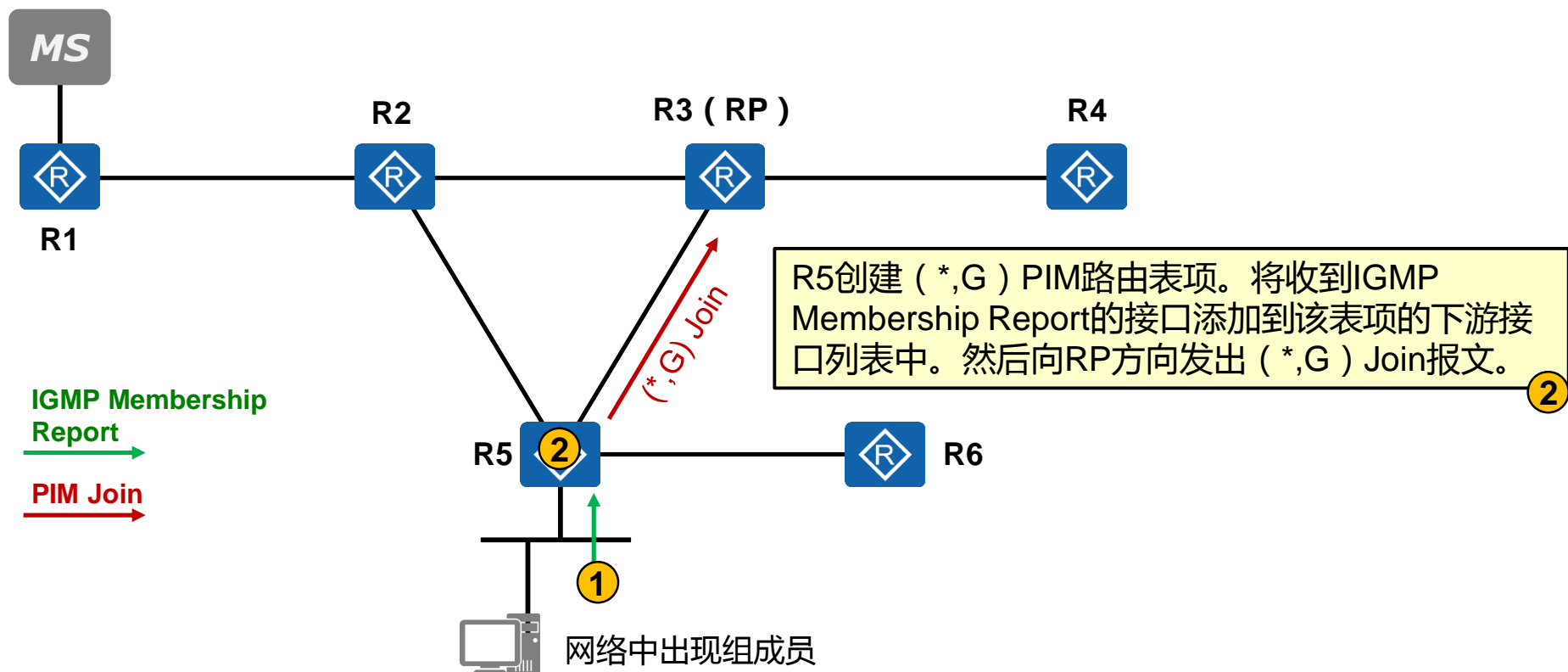
- **RP的发现机制**

- 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：在PIM域内选择几台PIM路由器，配置C-RP (Candidate-RP，候选RP) 来动态竞选出RP。同时，还需要通过配置C-BSR (Candidate-BSR，候选BSR) 选举出BSR，来收集C-RP的通告信息，向PIM-SM域内的所有PIM路由器发布。

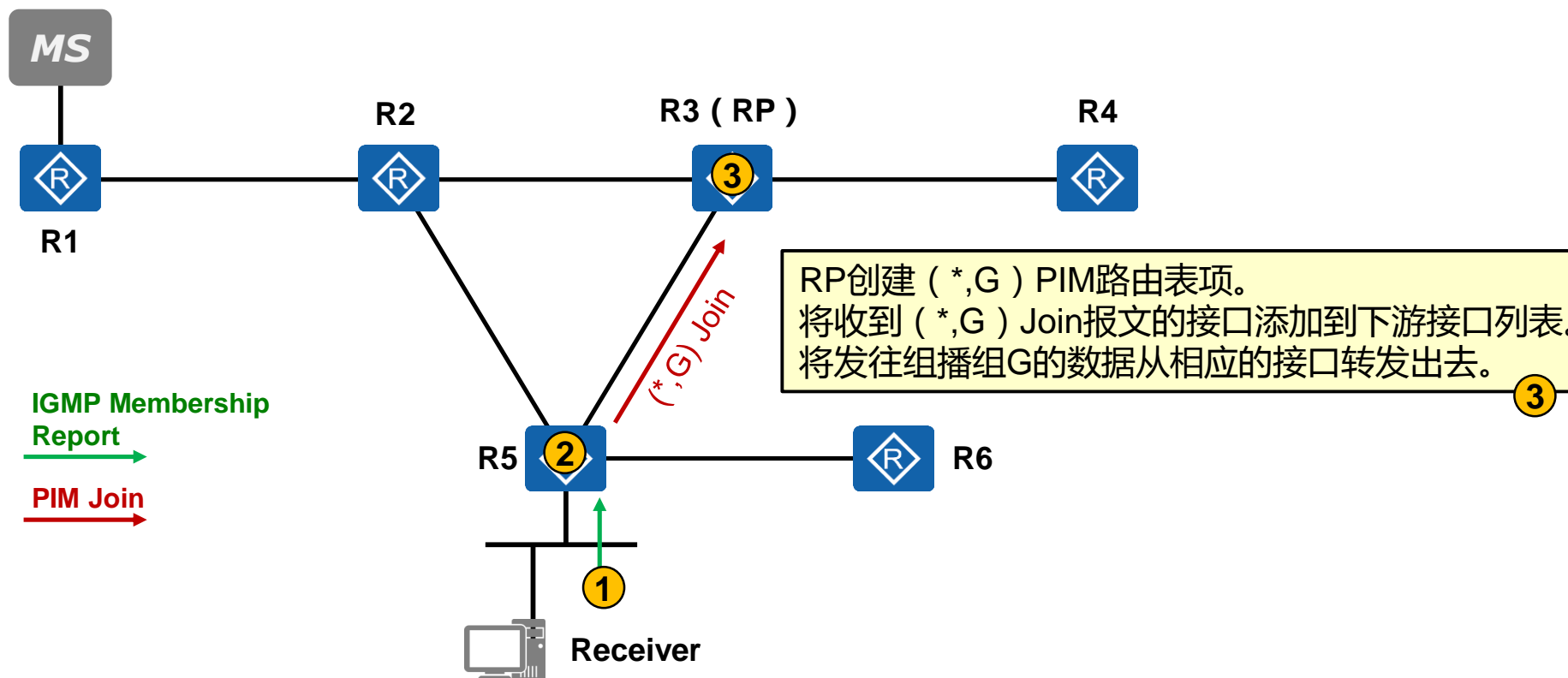
PIM-SM协议报文

报文类型	报文功能
Hello	用于建立及维护PIM邻居关系。
Register	注册报文，用于组播源向RP的注册过程。
Register-Stop	注册停止报文，RP使用该报文通知第一跳路由器停止在注册报文中发送组播流量。
Join/Prune	加入/剪枝报文，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。
Bootstrap	自举报文，用于BSR选举和发送C-RP汇总信息。
Assert	断言报文，用于断言机制。
Candidate-RP-Advertisement	C-RP通告报文，C-RP使用该报文向BSR发送通告。

共享树加入（接收者到RP）

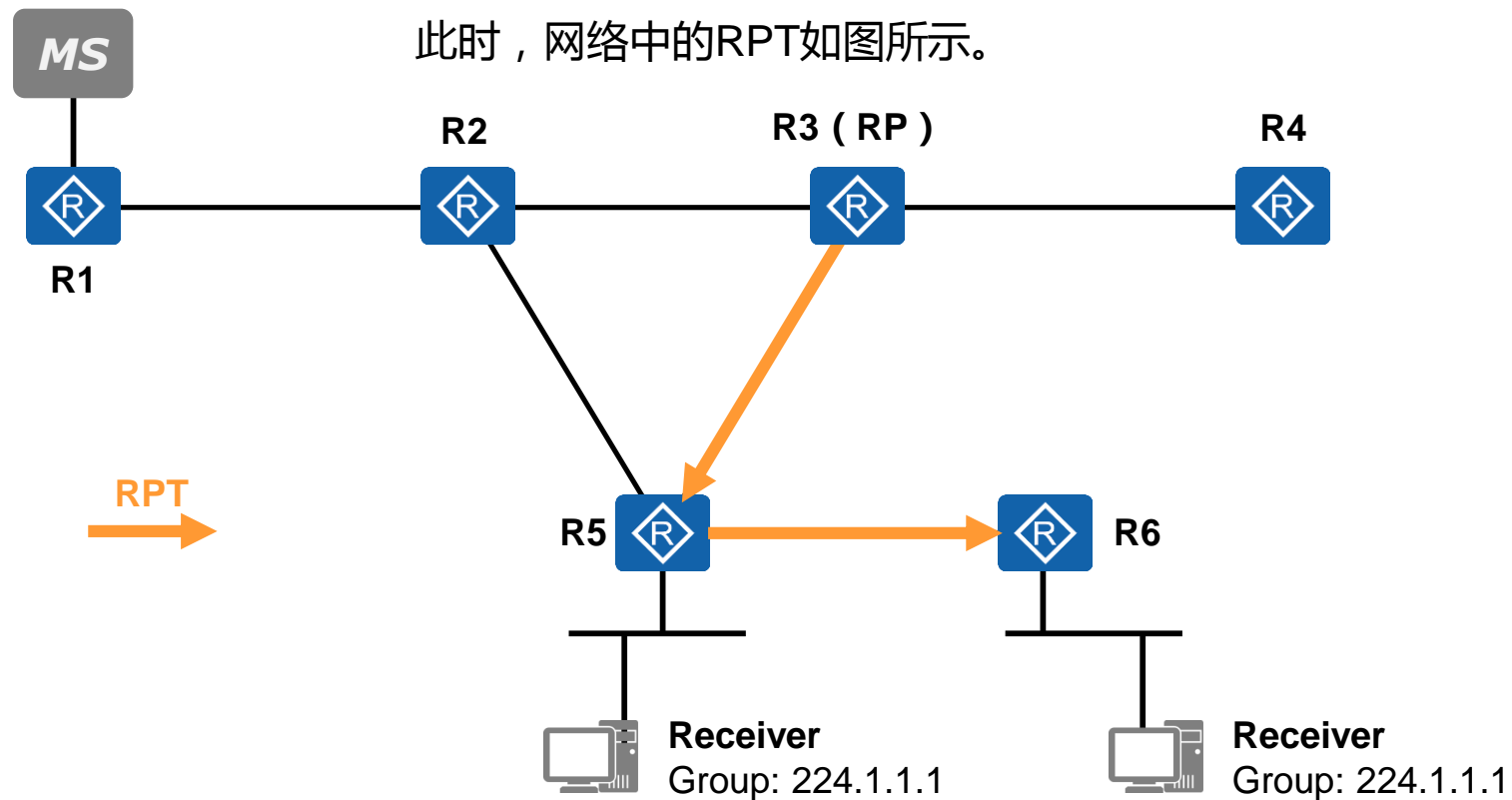


共享树加入 (续)

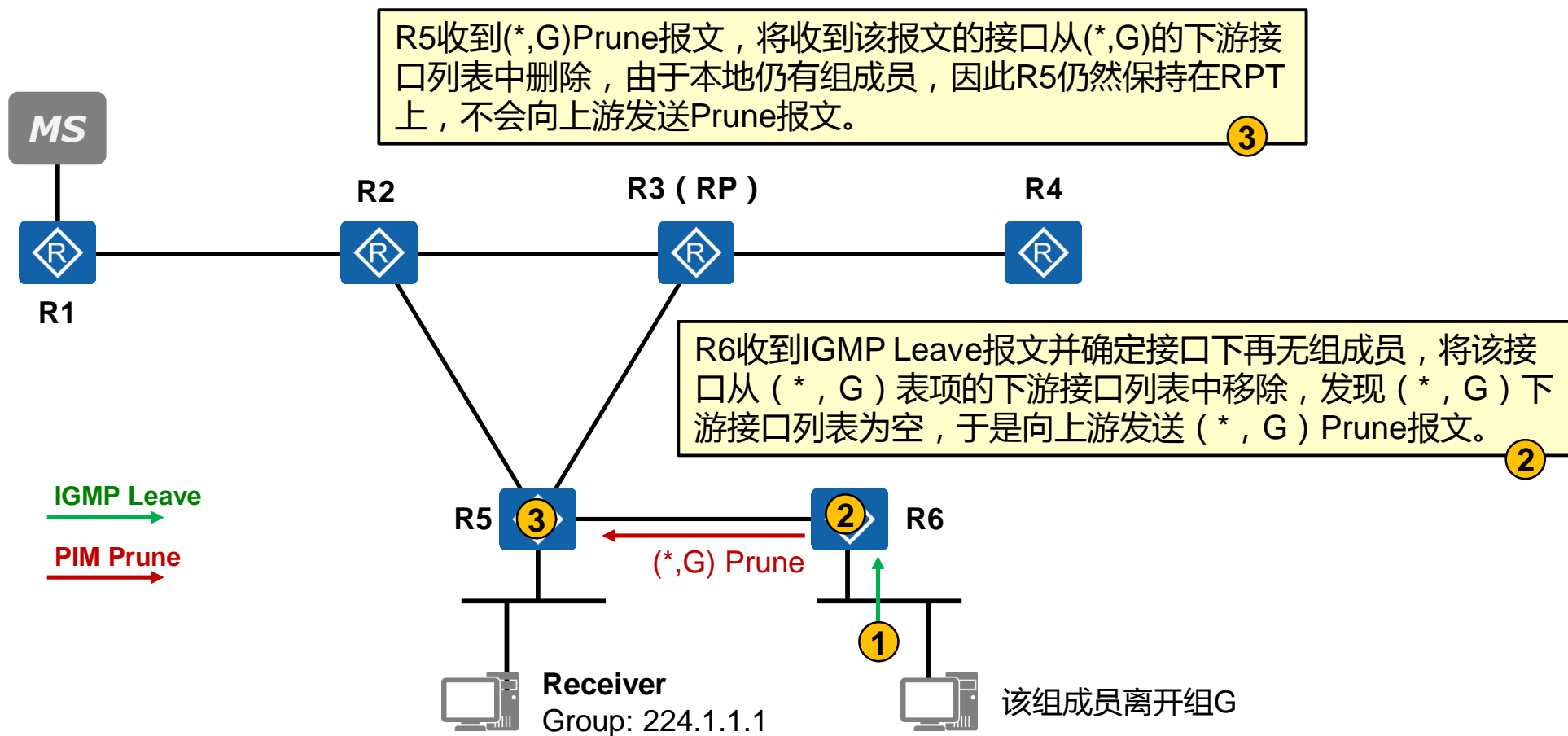


Page 32
HUAWEI TECHNOLOGIES CO., LTD.

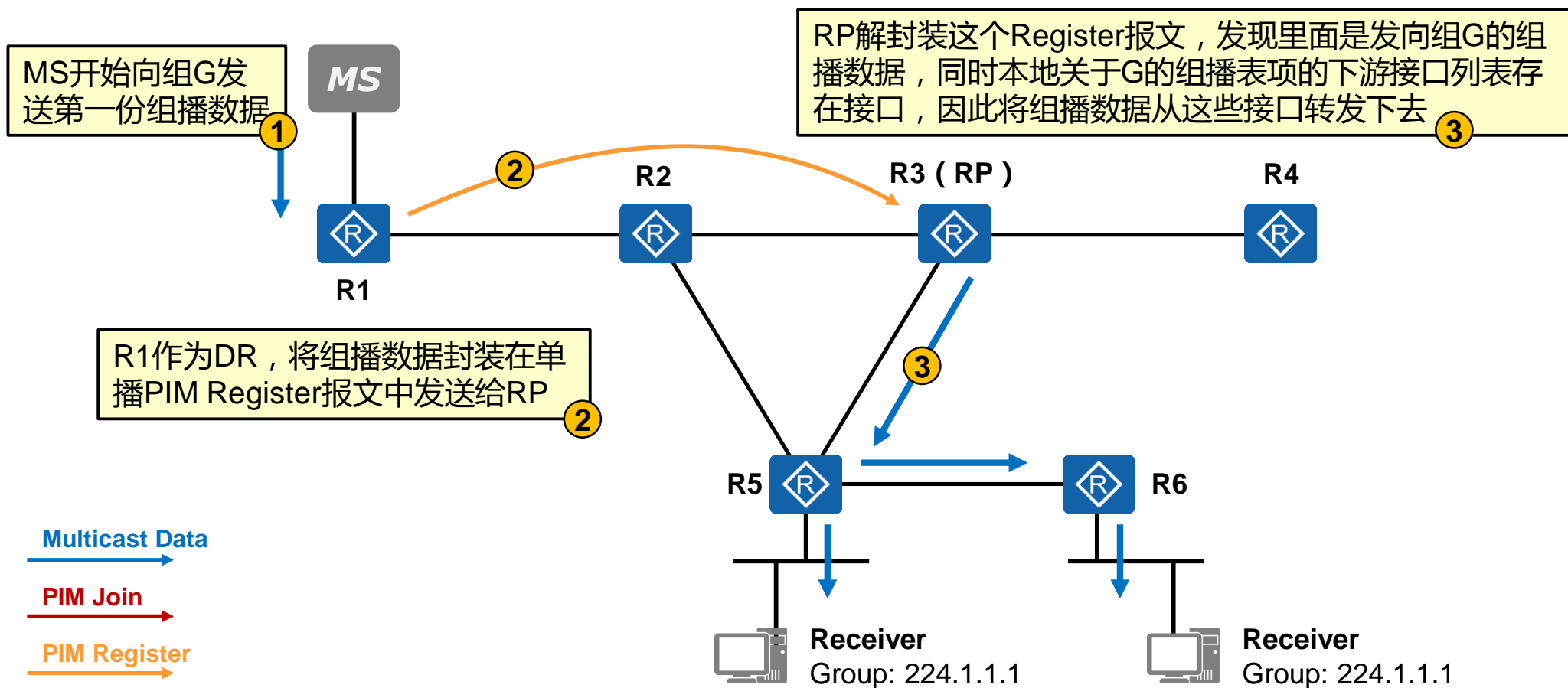
共享树加入（续）



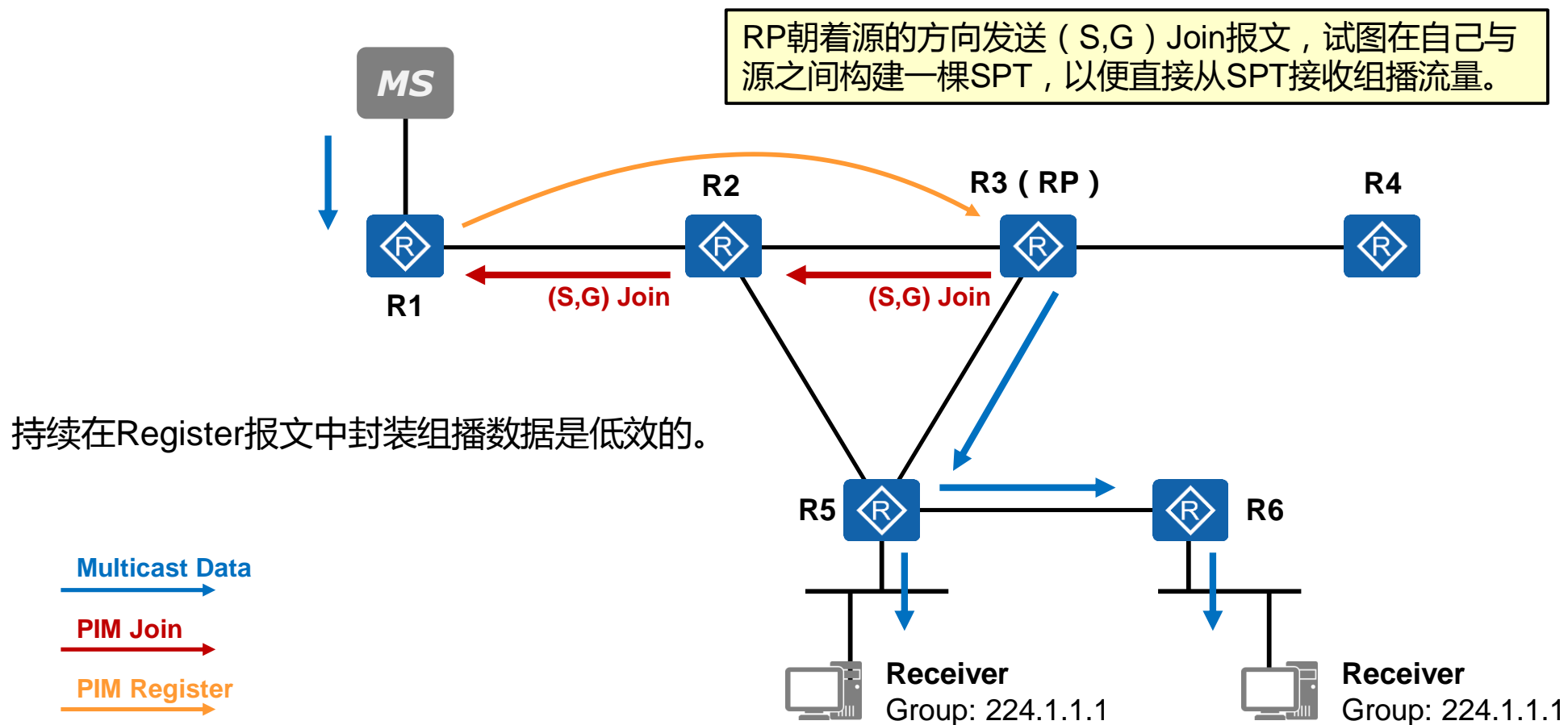
共享树剪枝



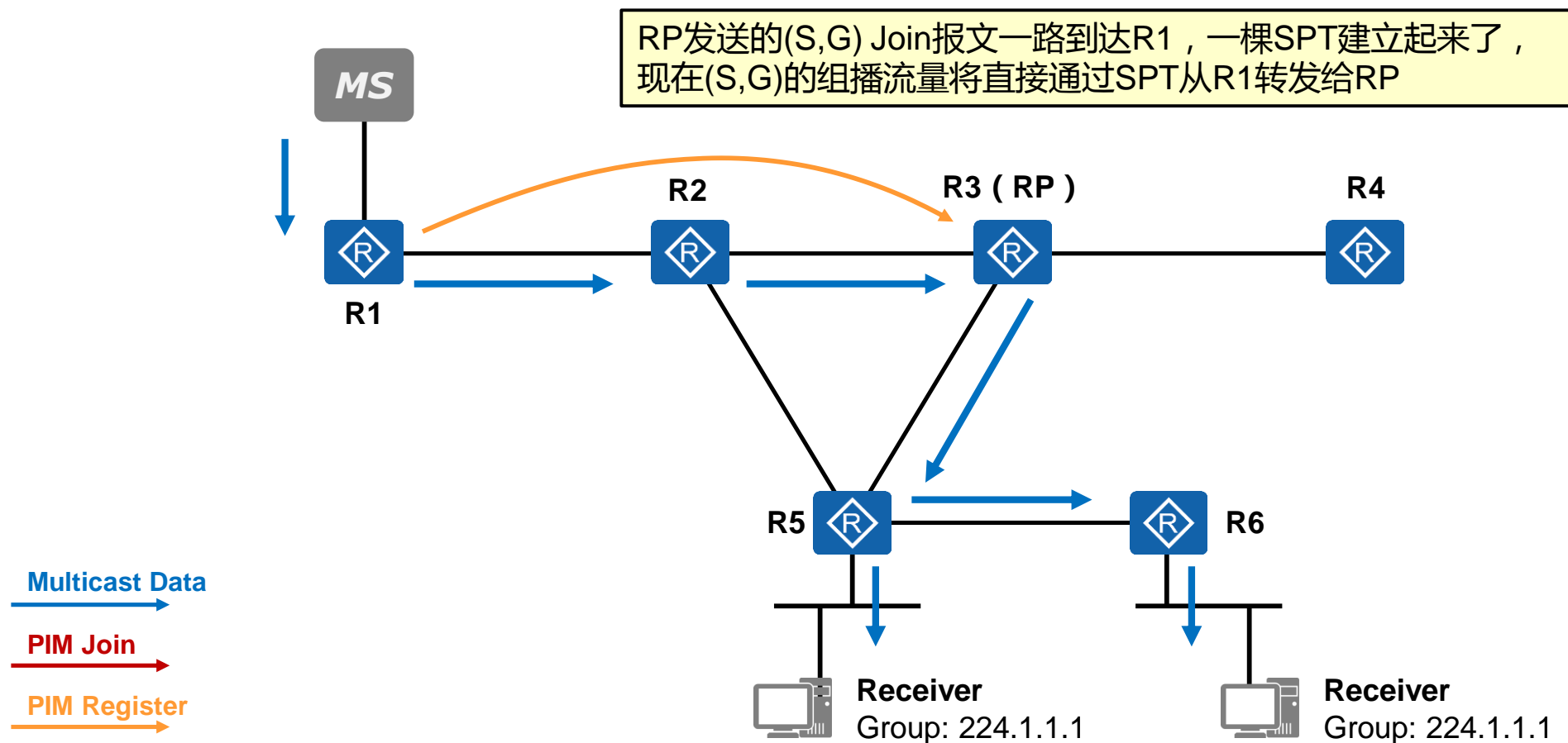
源注册



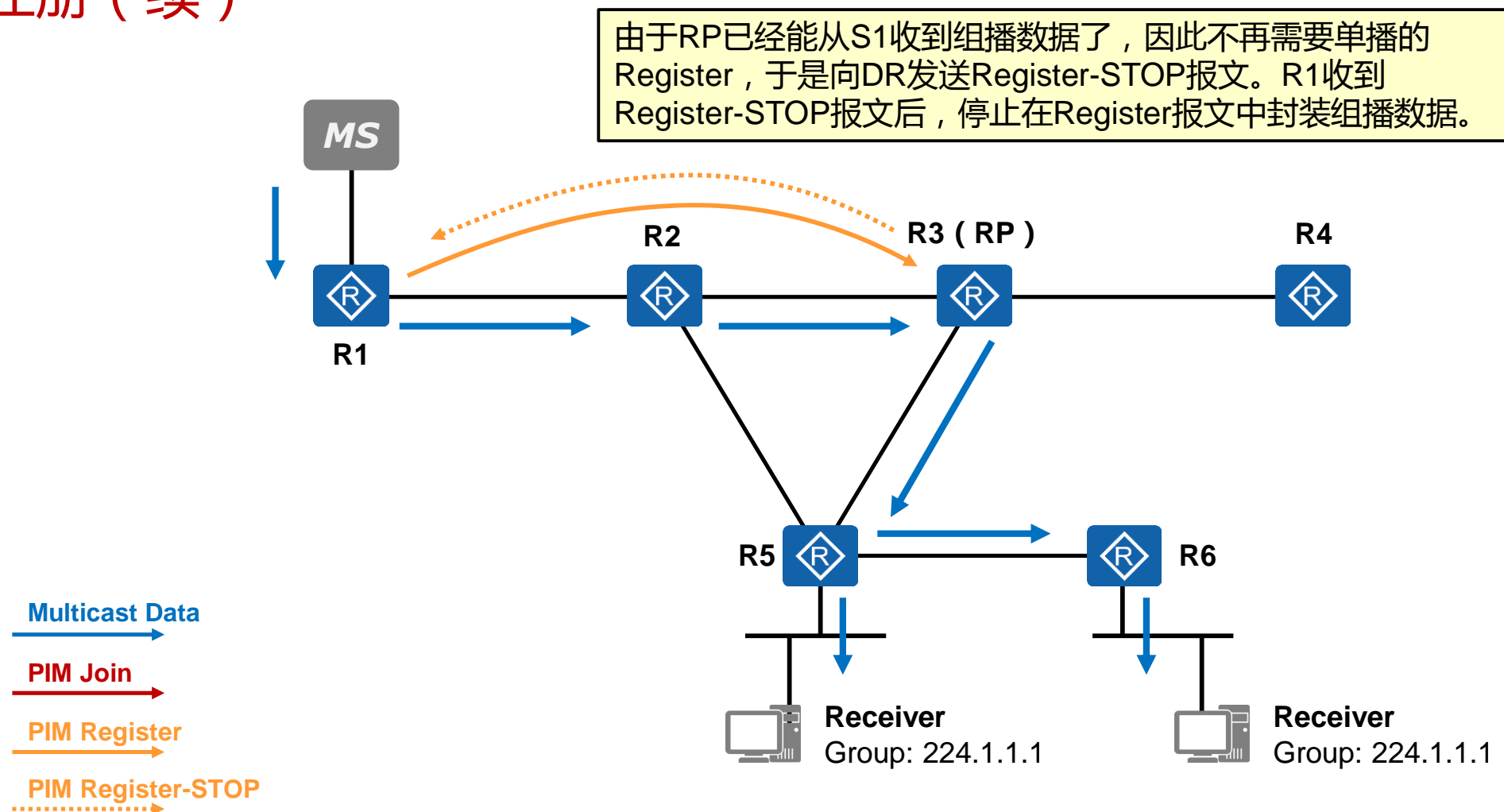
源注册（续）



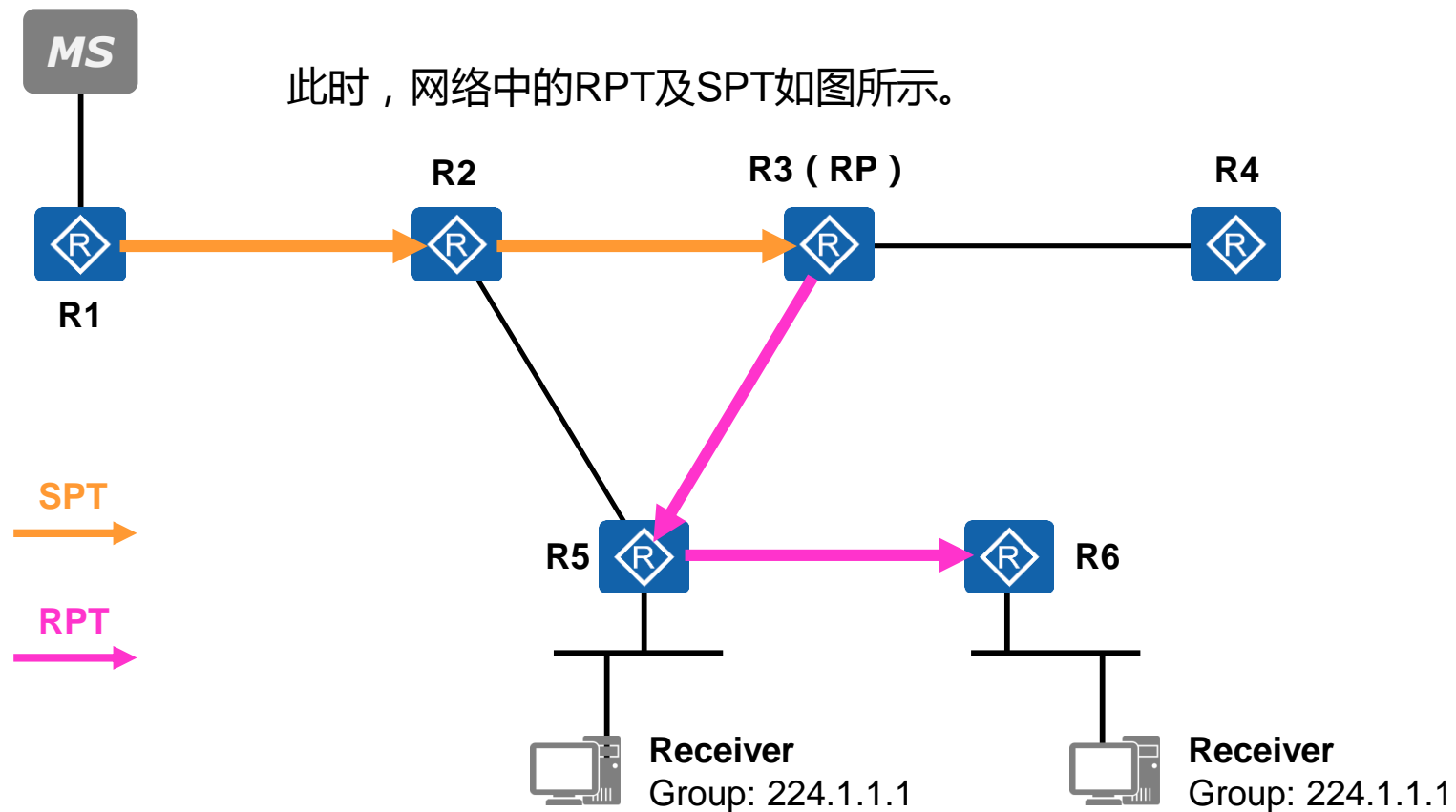
源注册（续）



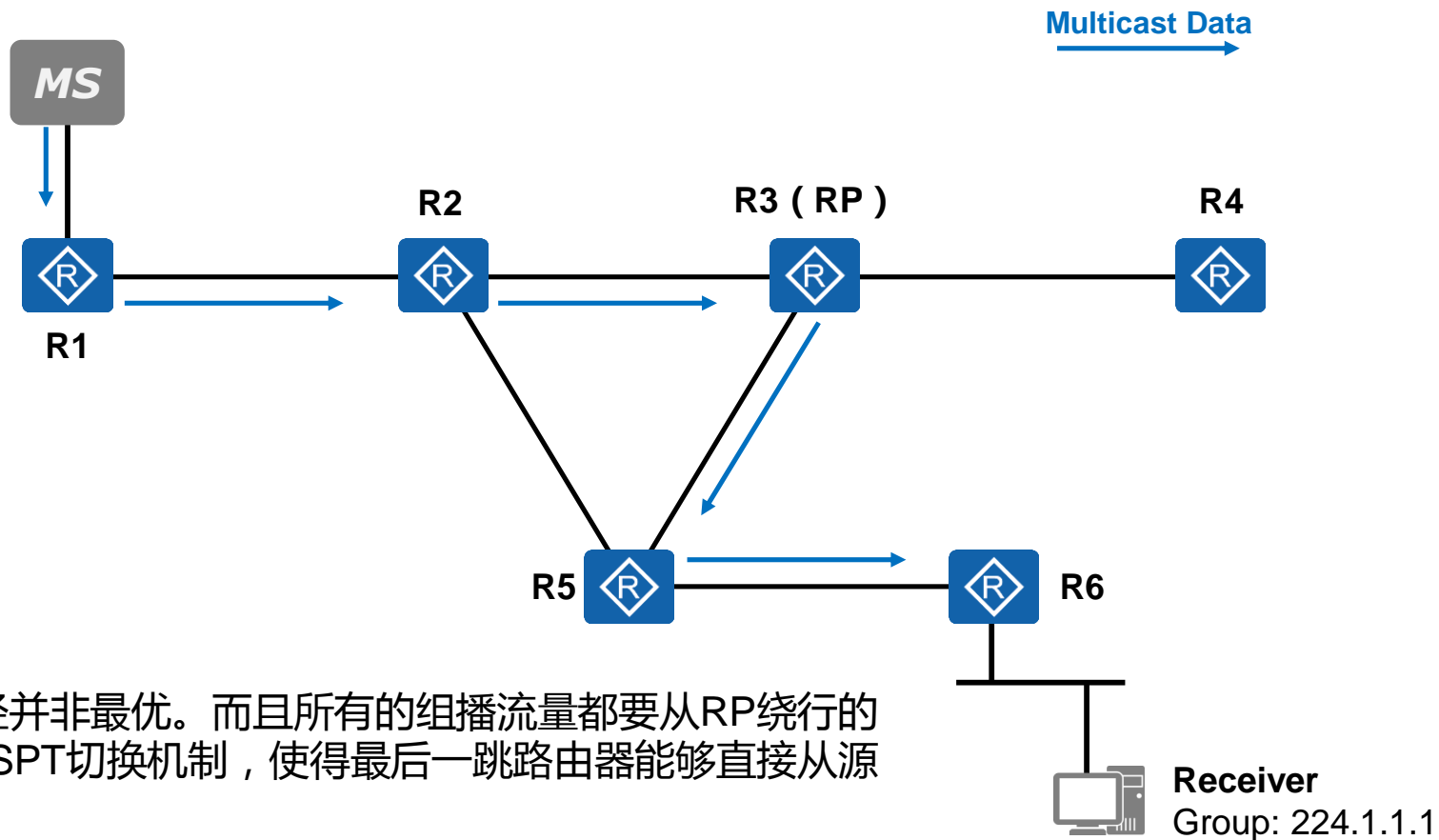
源注册（续）



源注册（续）



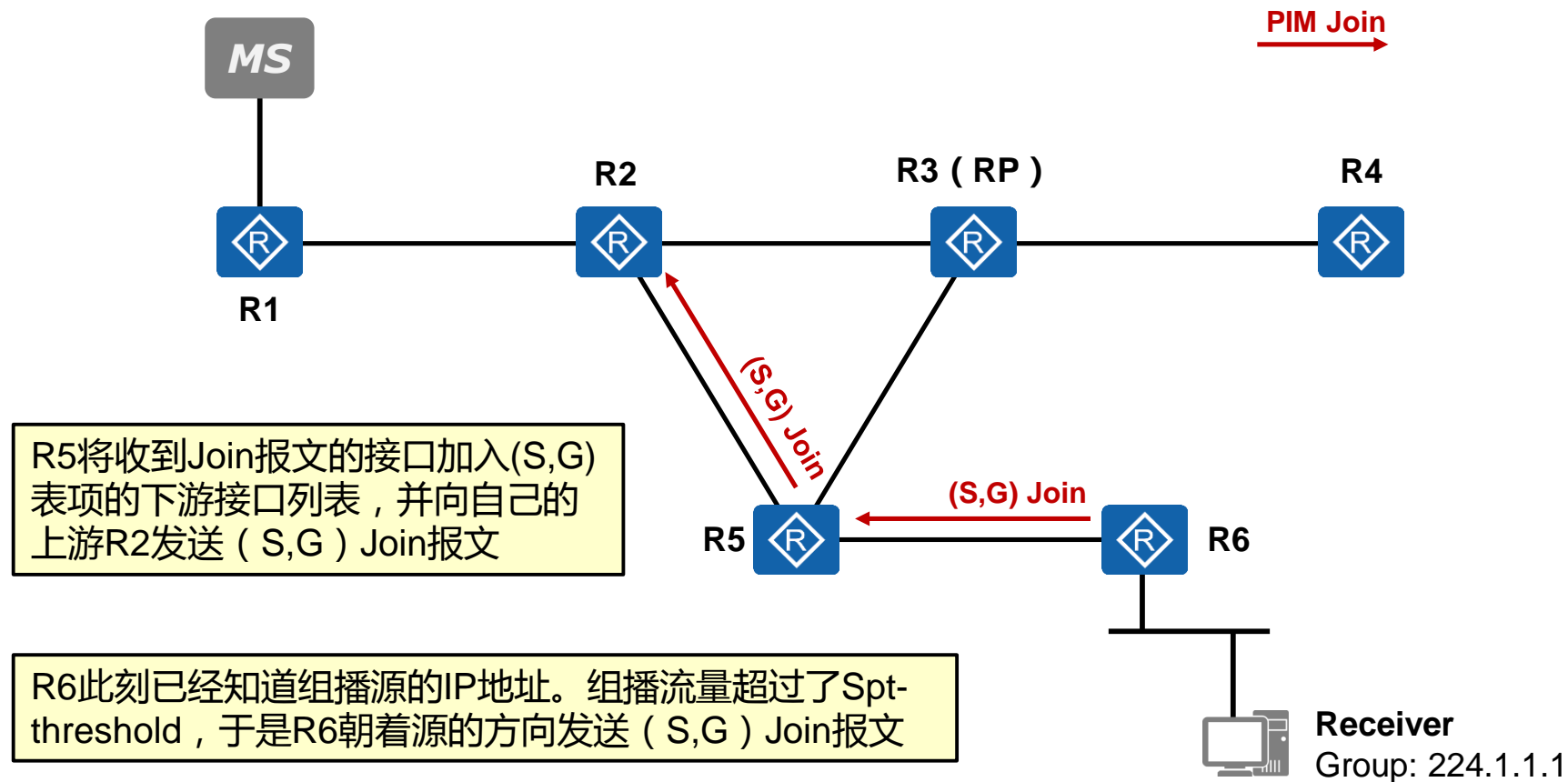
SPT切换过程



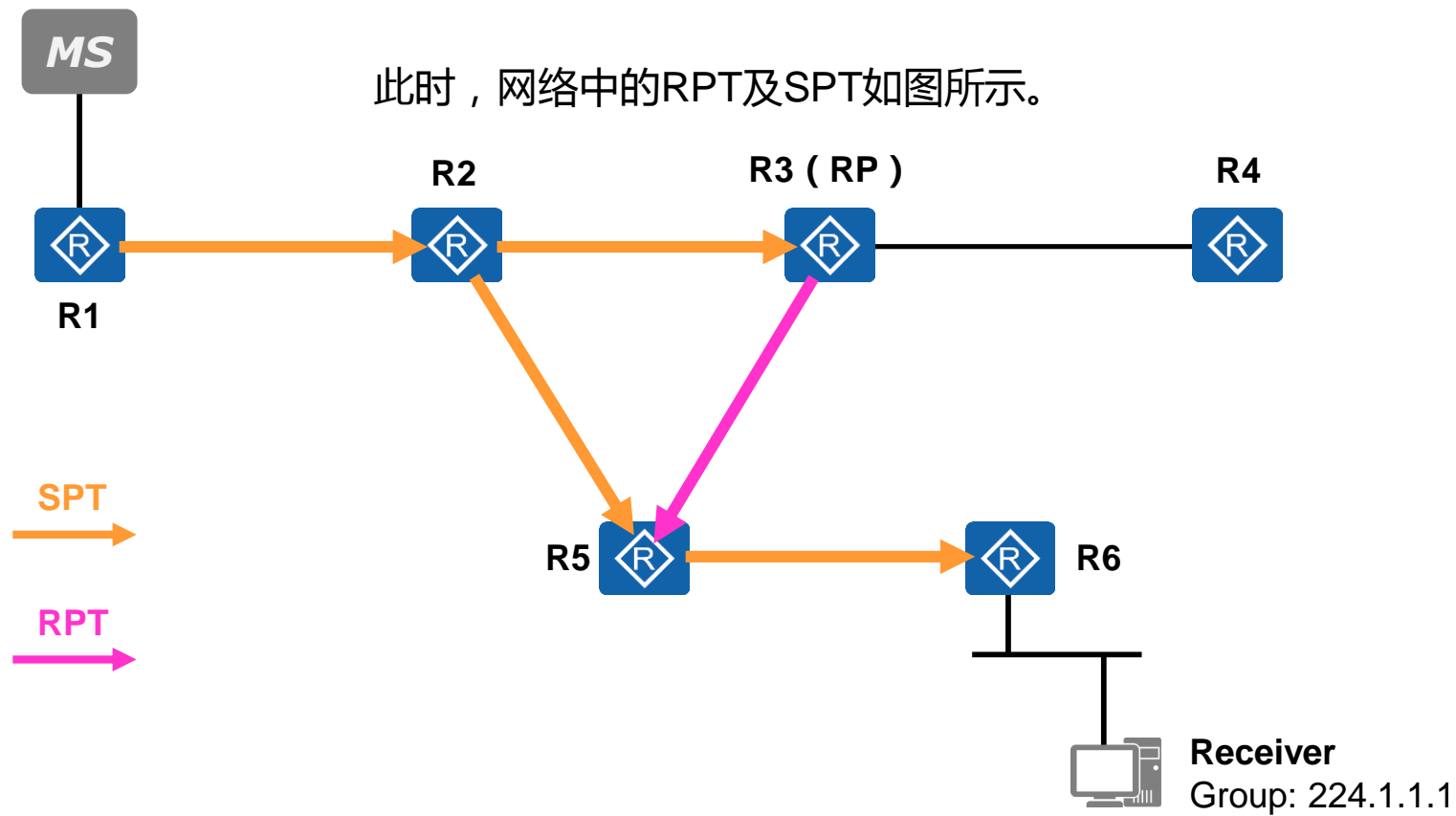
对于R5、R6而言组播流量的传输路径并非最优。而且所有的组播流量都要从RP绕行的话，RP的负担过大。PIM-SM设计了SPT切换机制，使得最后一跳路由器能够直接从源接收组播流量，而不用从RP绕行。

组成员端DR（如图中的R6）周期性检测组播报文的转发速率，一旦发现（S，G）报文的转发速率超过阈值（Spt-threshold），则触发SPT切换。缺省时，DR只要从RPT上接收第一个（S，G）组播报文即开始触发切换。

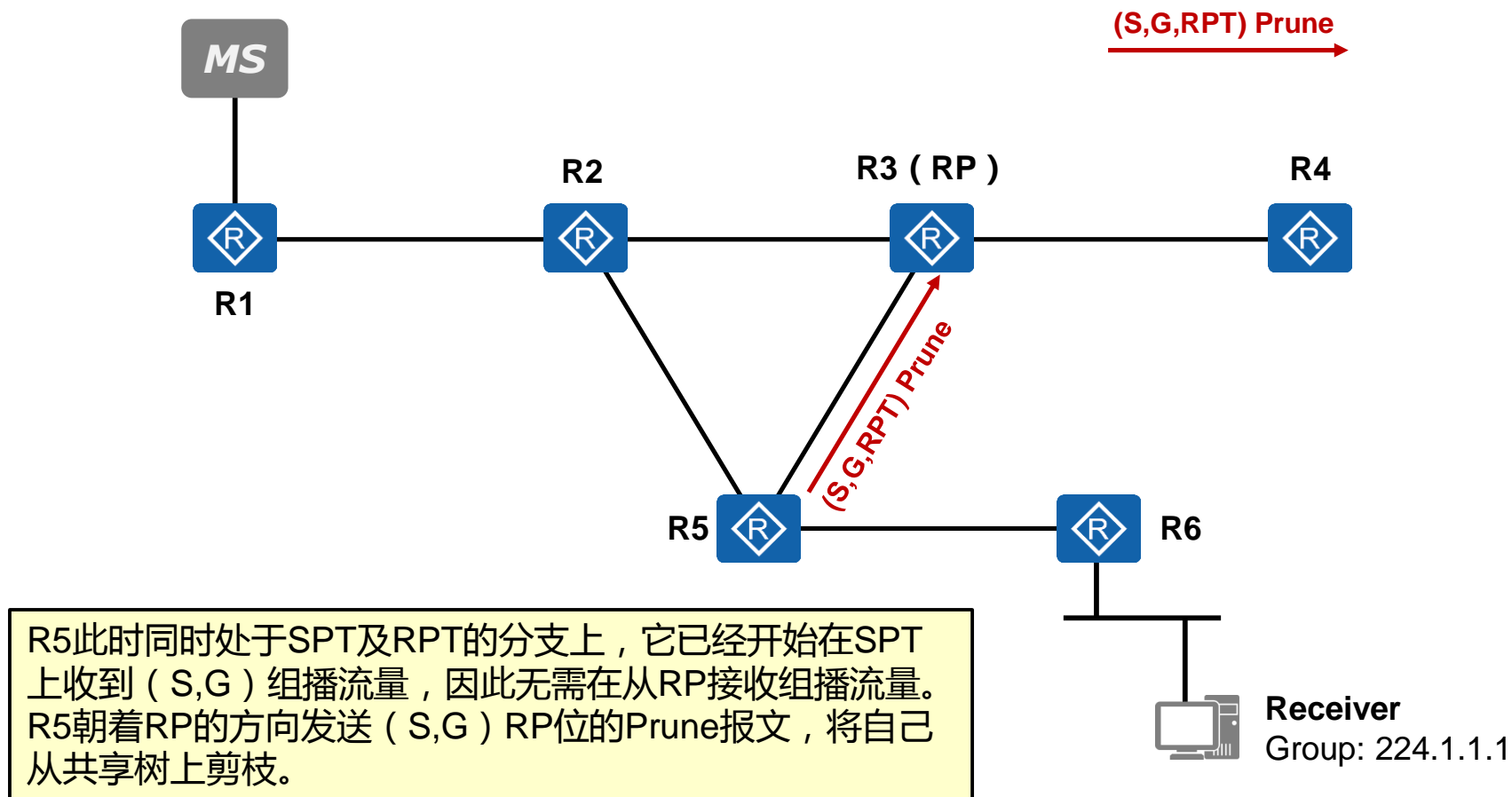
SPT切换过程（续）



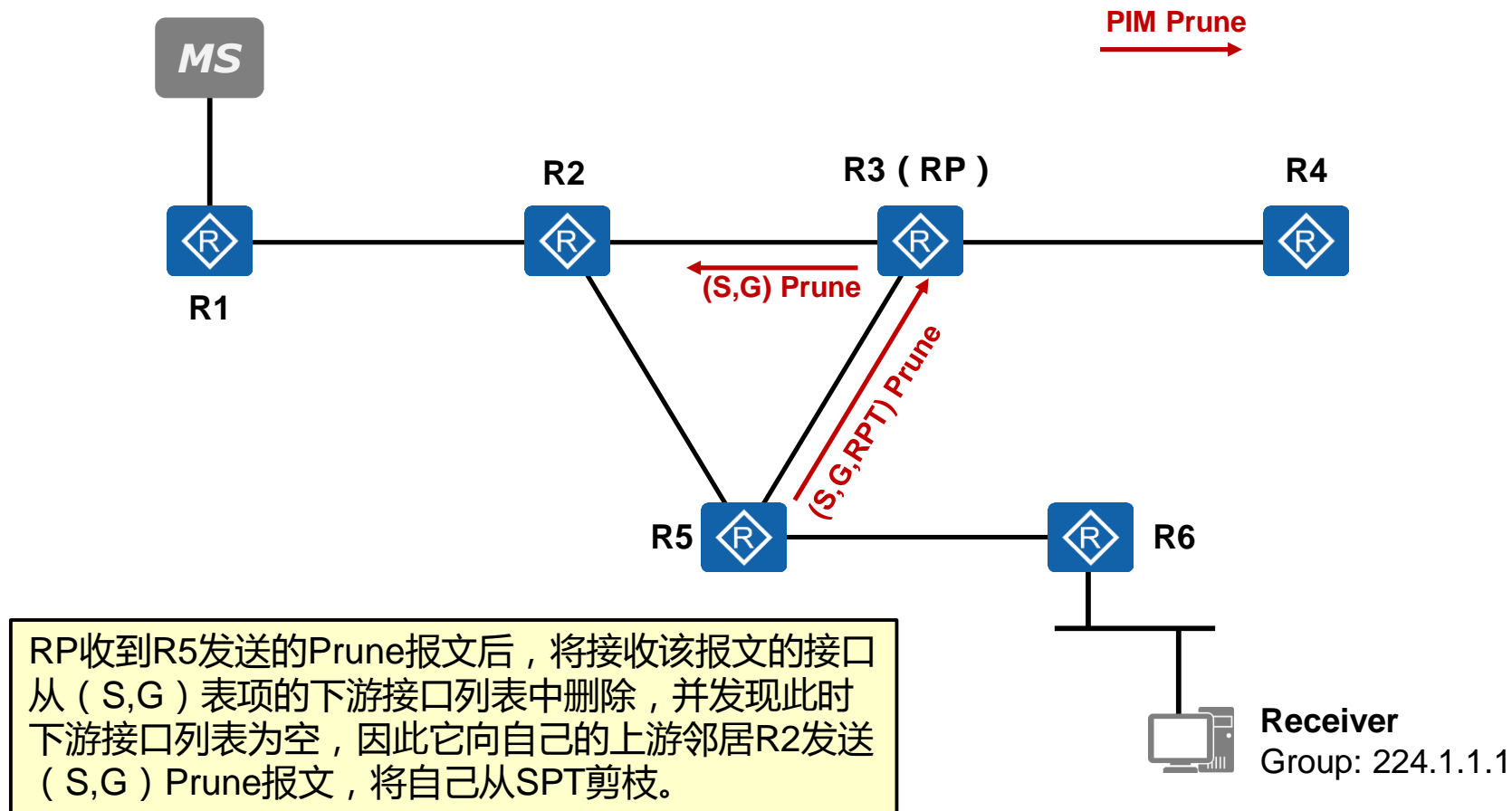
SPT切换过程（续）



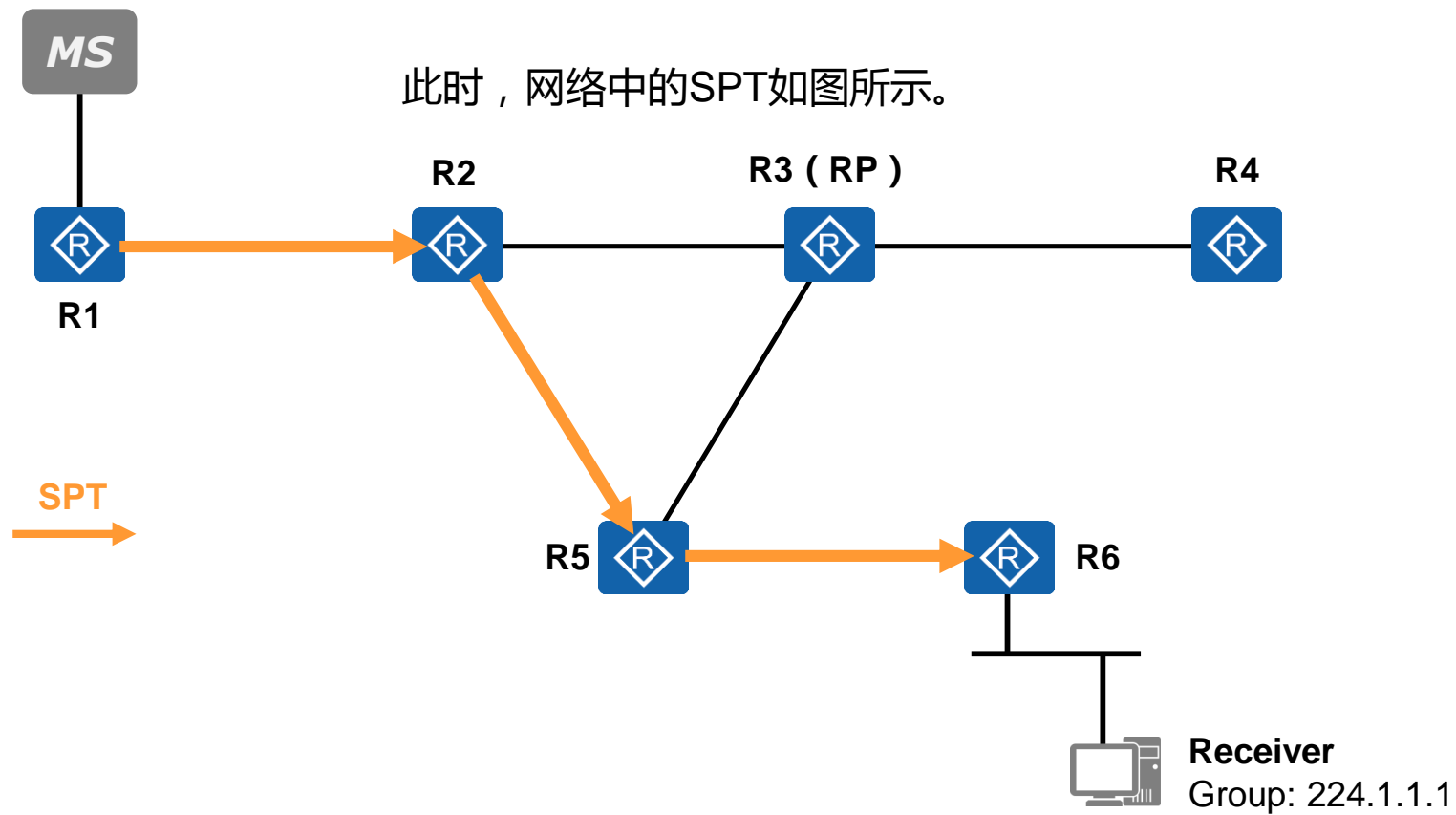
SPT切换过程（续）



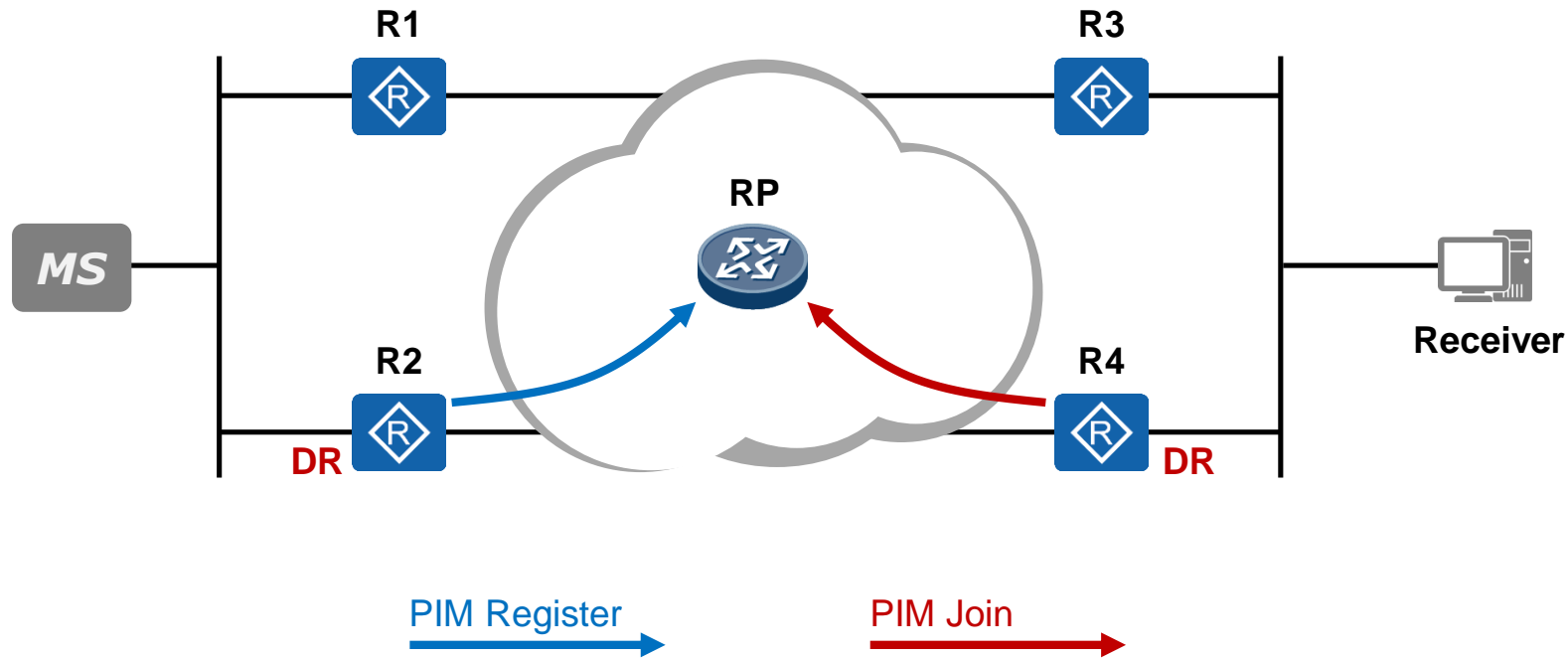
SPT切换过程（续）



SPT切换过程（续）



PIM-SM DR (Designated Router , 指定路由器)



PIM-SM 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报文成为PIM邻居，Hello报文中携带DR优先级和该网段接口地址。PIM路由器将自身条件与对方报文中携带的信息进行比较，选举出唯一的DR来负责源端或组成员端组播报文的收发。竞选规则如下：
 - DR优先级较高者获胜（如果网段中所有PIM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PIM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。
 - 在连接组播源的共享网段，由DR负责向RP发送Register报文。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报文。与组成员相连的DR称为组成员端DR。

目录

PIM概述

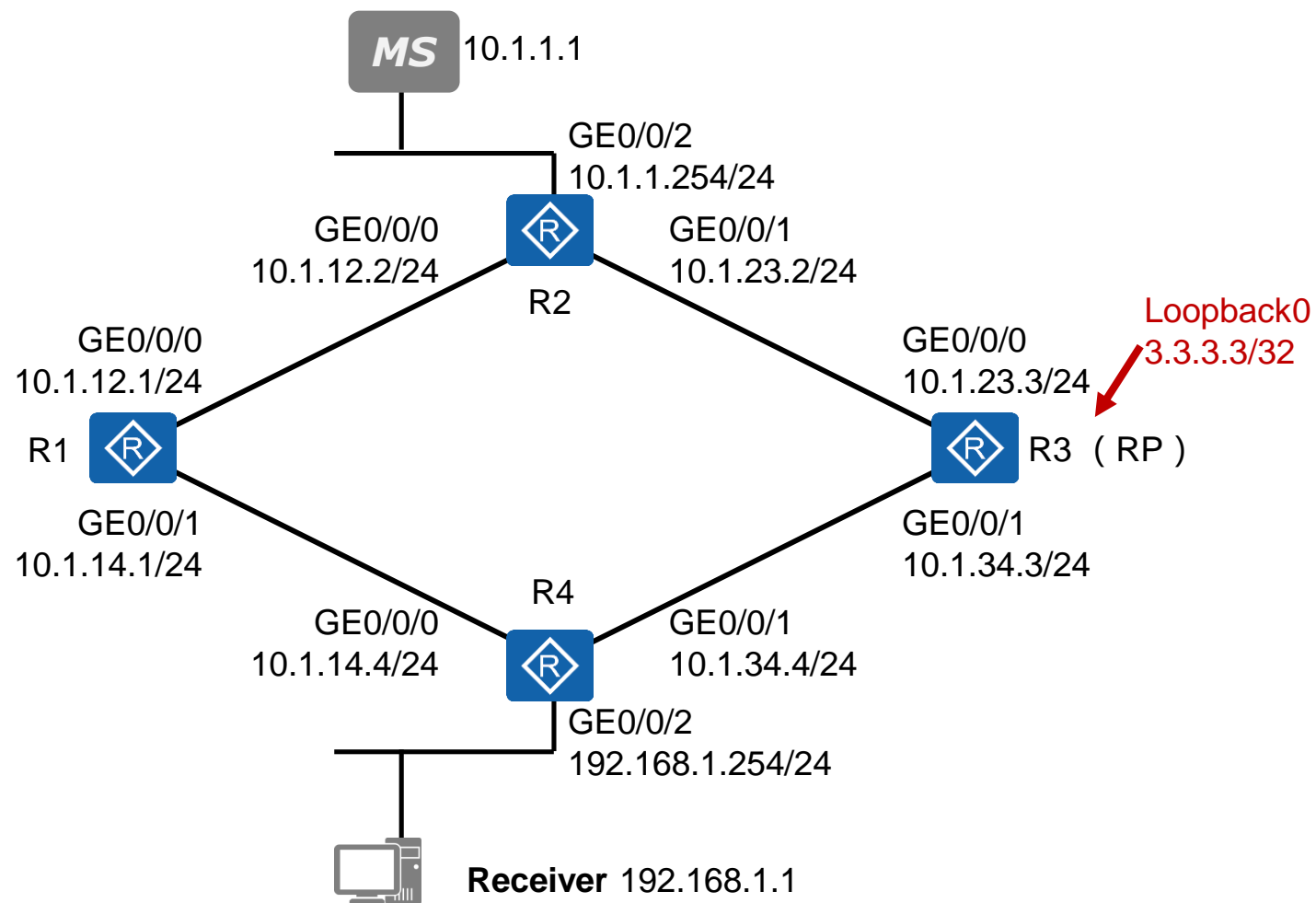
PIM-DM

PIM-DM的基础配置

PIM-SM

PIM-SM的基础配置

PIM-SM基础实验



R1的配置如下：

```
[R1] multicast routing-enable
[R1] interface GigabitEthernet0/0/0
[R1-GigabitEthernet0/0/0] pim sm
[R1] interface GigabitEthernet0/0/1
[R1-GigabitEthernet0/0/1] pim sm
[R1] pim
[R1-pim] static-rp 3.3.3.3
```

```
[R3] multicast routing-enable
[R3] interface GigabitEthernet0/0/0
[R3-GigabitEthernet0/0/0] pim sm
[R3] interface GigabitEthernet0/0/1
[R3-GigabitEthernet0/0/1] pim sm
[R3] interface loopback0
[R3-loopback0] pim sm
[R3] pim
[R3-pim] static-rp 3.3.3.3
```

```
[R2] multicast routing-enable
[R2] interface GigabitEthernet0/0/0
[R2-GigabitEthernet0/0/0] pim sm
[R2] interface GigabitEthernet0/0/1
[R2-GigabitEthernet0/0/1] pim sm
[R2] interface GigabitEthernet0/0/2
[R2-GigabitEthernet0/0/2] pim sm
[R2] pim
[R2-pim] static-rp 3.3.3.3
```

```
[R4] multicast routing-enable
[R4] interface GigabitEthernet0/0/0
[R4-GigabitEthernet0/0/0] pim sm
[R4] interface GigabitEthernet0/0/1
[R4-GigabitEthernet0/0/1] pim sm
[R4] interface GigabitEthernet0/0/2
[R4-GigabitEthernet0/0/2] igmp enable
[R4] pim
[R4-pim] static-rp 3.3.3.3
```

Thank you

www.huawei.com

Copyright ©2014 Huawei Technologies Co.,Ltd. All Rights Reserved.

The information contained in this document is for reference purpose only, and is subject to change or withdrawal according to specific customer requirements and conditions.

©2014 华为技术有限公司 版权所有
本资料仅供参考，不构成任何承诺及保证