

Multicast

目录

概述.....	2
组播的三个组成部分.....	3
组播地址.....	4
组成员机制.....	9
组播协议.....	14
PIM	16
组播树.....	16
组播反向路径转发.....	16
PIM模式.....	17
PIM-SM RP	20
PIM DR	24
PIM前转器.....	25
PIM-DM数据包.....	26
PIM-SM数据包.....	29
RP的确立.....	31
Pim sparse-dense-mode.....	33
PIM Dense Mode Fallback.....	34
共享树切换到源树.....	35
PIM-SM 之NBMA Mode.....	35
配置组播.....	37
配置PIM-DM	37
配置PIM-SM.....	58
PIM-SM的NBMA Mode.....	85
Source Specific Multicast (SSM)	91
MSDP (Multicast Source Discovery Protocol).....	94
MSDP概述.....	96
MSDP RPF检测.....	99
MSDP RPF检测详细规则.....	101
Default MSDP Peer	103
MSDP Mesh Group.....	105
MSDP SA Filter	106
PIM-SM域边界.....	107

组播流RPF检测详细规则.....	109
Anycast RP.....	111
配置一对一PIM-SM域的MSDP实验.....	113
1. 配置初始网络环境.....	115
2. 配置PIM-SM.....	129
3. 配置GRE Tunnel隧道.....	139
4. 配置MSDP.....	151
5. 通过单播路由表解决PIM-SM域间组播通信问题.....	159
6. 通过静态组播路由表解决PIM-SM域间组播通信问题.....	163
7. 通过MBGP解决PIM-SM域间组播通信问题.....	167
配置3个PIM-SM域全互联的MSDP实验.....	193
1. 配置初始网络环境.....	194
2. 配置PIM-SM.....	215
3. 配置MSDP.....	227
4. 配置MBGP帮助MSDP的SA通过RPF检测.....	234
5. 测试default MSDP peer帮助MSDP的SA通过RPF检测.....	242
6. 测试MSDP Mesh Group帮助MSDP的SA通过RPF检测.....	248
配置Anycast RP实验.....	258
1. 配置初始网络环境.....	260
2. 配置PIM-SM.....	283
3. 配置MSDP.....	296
4. 测试Anycast RP冗余性.....	310
IPv6 Multicast.....	319

概述

在当前的IP网络中，某台主机将数据包发向另一台主机时，就需要在数据包的目标IP位置写上那台主机的IP地址，再将数据包发出去，这个数据包发出去后，只有那台主机才能收到并且打开，而其它主机是不能收到和打开的。如果还想发送数据包给别的主机，就需要为数据包重新写上别的主机的IP地址，然后发出去。要将数据包发给几台主机，就需要为每个独立的数据包写上相应的目标IP地址。一个数据包包含一个特定的目标IP地址，并且这个数据包只能由相应的某台主机能够接收并且查看，这样的数据包称为单播（Unicast）。当要将同一份数据发送给多台主机时，如果使用单播的传送方式，那么需要发给几台主机，就需要重新封装几次数据包，并且将每份数据包单独发送给每台主机。

当使用广播来发送数据包时，目标 IP 为广播地址的数据将被网络中的每台主机接收并查看，但广播是不能被路由器转发的。

在网络上，当需要将一份同样的数据发送给多台主机时，如数字电视、视频会议等应用，这样的数据有多种传送方式，如下面两种：

1.单播

在使用单播的情况下，需要为每个接收者重复发送单播，如果接收者数目过多，那么数据源就需要多次发送而承受巨大的压力，并且低速的 WAN 链路也会成为潜在的瓶颈，如果数据对时延比较敏感，还会造成延迟。

2.广播

在使用广播的情况下，数据源只需要将同一份数据发送一次，但是负担却转移到了网络中的其他主机，因为不管想不想接收这个数据，都必须接收；并且广播是不可跨越路由器的，如果接收者在远程网络，将会造成数据丢失的情况。

3.组播

从上面的结论中可以看出，当需要将一份同样的数据发送给多台主机时，虽然使用单播可以跨越路由器，但是需要将同一份数据发送多次，不切实际；而使用广播只需要发送一次数据，但是却让网络中每个人都必须接收数据，并且数据不能穿越路由器，造成远程网络收不到数据，所以也不可行。考虑到这些因素，便开发出了一种新的数据传输方式，这样的传输方式结合了单播和广播的优势，即将一份数据发出去后，这样的数据可以同时被多台主机接收，并且数据可以穿越路由器，从而被路由到远程网络，这样的数据就是组播（Multicast），因此，组播数据发出去后，可以只被一组特定的主机接收，而不想接收的主机，是收不到的，组播还可以被路由器转发到远程网络，前提是路由器必须开启组播功能。在组播中，想要接收组播的主机，被称为组员，或组成员。

组播的三个组成部分

1.组播地址（能被组播识别的地址集）

2.组成员机制（主机加入和退出组的机制）

3.组播路由协议（路由器有效传送组播到各个网络的组成员，且不会过度消耗网络资源的路由协议）

当需要将一份同样的数据发送给多台主机时，在使用组播的情况下，就需要将需要接收数据包的主机标识出来，要区别于不接收的主机，只有想要接收的主机，才能收到相应的组播数据，这时就需要为组播数据包写上特定的 IP 地址，被写上组播地址的数据包，只能被特定的组成员接收，所以要将组播正常的发送到组成员，就必须为组播数据包写上组播地址，当网络中有多种组播数据时，每种数据应该写上不同的组播地址。

当写上了组播地址的数据包在网络中传送时，这样的数据包只应该被特定组的组成员接收，只有属于同一个组的成员，才能接收该组的数据包，所以必须确认哪些主机是组成员，哪些主机不是。要想接收组播，主机要做的事就是加入特定的组，特定的组，就是由组播地址来区分的。当主机加入了某个组之后，便能收到该组的数据，而当主机不想接收组播时，就应该退出相应的组，这样就可以停止组播的接收和转发，要完成这一切，就需要一种主机加入和退出组的机制，要区别哪些是组成员，哪些不是。

如果某个组的成员分布在不同的网络中，那么就需要路由器转发组播，才能保证远程主机能够收到组播数据。要让路由器为组播数据提供转发，就需要让路由器拥有像单播路由表一样的转发表，依据路由表来决定数据包该从哪个接口发出去。要让路由器依据路由表来转发组播，就需要有组播路由表，而路由器的组播路由表，就需要靠特定的组播路由协议来收集组播路由表，组播路由表指导路由器如何将组播正确转发到组成员。

下面分别来详细介绍组播地址、组成员机制以及组播路由协议的工作过程：

组播地址

组播 IP 地址

在我们区分一个 IP 地址是哪类地址时，只需要看第一个字节便能得出结果。

CCIE LAB认证经验分享千人群：539730342

在 IP 地址一个字节的 8 个 bit 中，相应 bit 位为 1 时，便得到相应的值，每个 bit 位的取值分别如下：

128 64 32 16 8 4 2 1

A 类 IP 地址第一字节的第 1 个 bit 总是为 0,后面 7 个 bit 可以随意设置，所以 A 类 IP 地址第一个字节的取值范围为：

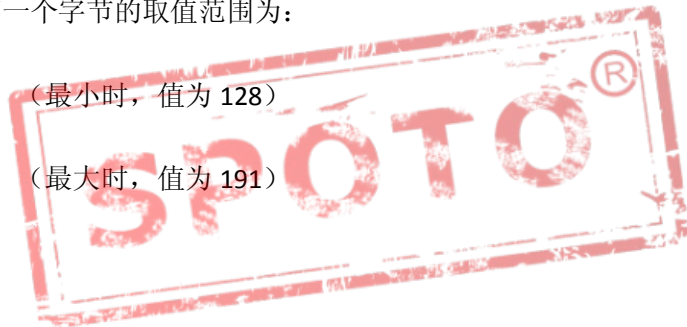
00000000 （最小时，值为 0）

01111111 （最大时，值为 127）

B 类 IP 地址第一字节的前面 2 个 bit 总是为 10,后面 6 个 bit 可以随意设置，所以 B 类 IP 地址第一个字节的取值范围为：

10000000 （最小时，值为 128）

10111111 （最大时，值为 191）



C 类 IP 地址第一字节的前面 3 个 bit 总是为 110,后面 5 个 bit 可以随意设置，所以 C 类 IP 地址第一个字节的取值范围为：

11000000 （最小时，值为 192）

11011111 （最大时，值为 223）

而 D 类 IP 地址第一字节的前面 4 个 bit 总是为 1110,后面 4 个 bit 可以随意设置，所以 D 类 IP 地址第一个字节的取值范围为：

11100000 （最小时，值为 224）

11101111 （最大时，值为 239）

CCIE LAB认证经验分享千人群：539730342

组播地址采用 **D类** IP 地址表示，因为 IP 地址共 32 bit，D 类地址前 4 bit 总是为 1110,所以

组播地址取值范围为 224.0.0.0——239.255.255.255。

组播地址可以是永久的，也可以是临时的，“永久”是指一个组地址被永久分配给某个协议，临时的可以自由定义。

部分永久的组播地址：

224.0.0.1 子网中的所有系统和主机

224.0.0.2 子网中的所有路由器

224.0.0.4 DVMRP 路由器

224.0.0.5 所有 OSPF 路由器

224.0.0.6 OSPF 指定路由器

224.0.0.9 RIP-2 路由器

224.0.0.10 EIGRP 路由器

224.0.0.13 PIM 路由器

224.0.0.15 CBT 路由器

224.0.1.39 Cisco-RP-Announce

224.0.1.40 Cisco-RP-Discovery

组播地址除了分为永久地址和临时地址之外，还可以细分，我们需要知道的细分地址为：

保留给网络协议的地址范围：

224.0.0.0 —— 224.0.0.255

全局地址范围：

224.0.1.0 —— 238.255.255.255

限制私有地址范围：

239.0.0.0 —— 239.255.255.255

组播二层地址：

一个 IP 数据包要在网络中传送，必须依照 OSI 七层模型由上至下封装，比如先封装 TCP 或 UDP 端口号，然后封装 IP 地址，如果是组播数据包，那么目标 IP 为组播地址，最后再封装数据链路层地址，如果介质是以太网，那么就需要封装 MAC 地址。

当一个数据包为组播数据时，那么这个数据包将被多台主机接收，所以数据包的 MAC 地址不能为某台主机的真实 MAC 地址，这时就需要根据组播 IP 地址，来封装一个拥有对应关系的组播 MAC 地址，因为这个组播 MAC 是与组播 IP 地址对应的，且不是主机的真实 MAC 地址，所以能够被多台主机接收到。

组播的 MAC 地址和组播 IP 地址拥有对应关系，也就是说组播 MAC 地址是根据组播 IP 地址计算得到的。一个 MAC 地址为 48 bit，使用十六进制表示，组播 MAC 地址的前面 24 bit 固定为 01 00 5E，第 25 bit 固定为 0，而剩下的 23 bit 则使用组播 IP 地址后 23 bit 填充。如下图表示：



从上图中可以看出，组播 MAC 地址前面 24 bit 固定为 01 00 5E，第 25 bit 固定为 0，而后 23 bit 由组播 IP 地址的后面 23 bit 映射过来。

一个组播 IP 地址为 32 bit，而前面 4 个 bit 固定为 1110，后面 23 bit 被映射到了组播 MAC 地址，所以中间还有多余的 5 个 bit 可以随意设置，正因为这 5 bit 可以随意取值，最终造成有 2 的 5 次方共 32 个 IP 地址被映射到同一 MAC 地址。如下面出现的结果：

组播 IP 地址为：

224.0.0.1 后 23 bit 为 00 00 01

映射出组播 MAC 地址为 01 00 5E 00 00 01

225.0.0.1 后 23 bit 为 00 00 01

映射出组播 MAC 地址为 01 00 5E 00 00 01

226.0.0.1 后 23 bit 为 00 00 01

映射出组播 MAC 地址为 01 00 5E 00 00 01

等等.....。

由上可以看出，将会有多个组播 IP 地址映射到同一个组播 MAC 地址，数量为 32 个，但是这并不会影响到组播的通信，因为即使组播的 MAC 地址相同，但是组播 IP 地址可能是不同的，所以数据量不会被混淆，即使组播 IP 地址是相同的，不同数据流也会使用不同的上层端口号，最终也能将不同数据流区分开来。

组成员机制

要将组播数据准确发送给组成员，必须先确定哪些网络的哪些主机是组成员，只有先确定了成员的位置，才能正确转发组播。当组成员不再需要接收组播的时候，就应该停止向成员发送组播，要确定组成员不再需要接收组播，就必须在成员退出时明确通告发送者。要确定组成员，有两种方式：查询和报告。

查询，就是一台路由器向网络中发出查询消息，查询是否要主机要加入组，如果有主机应答，那么路由器就可以请求上游路由器把组播流量前转到这个子网中，如果没有主机应答，则请求上游路由器停止向其前转组播流量。

报告，主机也可以不必等待路由器的查询，可以主动向路由器请求加入某个组，退出时也要向路由器发送退出消息，让路由器停止向其前转组播流量。

在网络中，要确定组成员，需要使用一种协议，这种协议就是 IGMP (Internet Group Management Protocol)因特网组管理协议，IGMP 运行在路由器和主机之间，因为当组播发送者和组成员在不同网络时，需要路由器为组播数据提供转发，那么路由器就必须确认自己直连的网络中是否存在组成员，可以使用查询和报告来发现组成员，IGMP 就可以完成这样的工作。

IGMP 共有 3 个版本，version 1，version 2，version 3，而思科路由器接口默认使用 version 2，下面分别来介绍 version 1 和 version 2，而 version 3 暂不介绍。

IGMP version 1

因为 IGMP 是运行在路由器和主机之间的，所以当 IGMP 分别运行在路由器和

主机上时具有不同的功能。

IGMP v1 路由器:

运行了 IGMP 的路由器目的是确定哪些主机是组成员，主要靠发送 queries（查询）来确定，当路由器发送 queries 查询主机时，使用目的地址为 224.0.0.1，当有主机回复时，便认为网络中存在组成员，则将组播发送到该网络，并且 queries 每一分钟发送一次，主机每次都回复，如果连续三个 queries 没有回复，即三分钟，那么路由器便认为网络中的组成员已经离开，也就停止向网络中发送组播数据了。

IGMP v1 主机:

运行了 IGMP 的主机发送 reports 来告诉路由器自己是组成员，reports 同时也可以用来回复路由器的 queries。Reports 的目标地址是需要加入的组的地址，所以其它路由器和主机都能收到此数据包。一个网络中，因为有一个组成员时，路由器需要向该网络转发组播数据，组播是发送一次，有多个组成员时，路由器还是只向该网络发送一次组播数据，所以当在一个网络中有多台主机需要成为组成员时，并不需要每台主机都向路由器发送 reports，一个网络中只需要一台主机向路由器发送 reports 即可，其它主机全部不发，因为一台发送了之后，其它要接收组播的主机也是能正常接收到的，也为了避免所有组员都发 reports 而充斥整个网络。

当网络中的组成员离开后，并没有特定的机制来通知路由器组员的离开，唯一的方法就是路由器在 3 分钟后没有收到组成员对 queries 的回复，也就认为网络中的组成员已经离开，便停止向网络中发送组播数据了。而其它主机还需要继续接收组播时，就需要再次向路由器发送 reports 要宣告自己的存在。

IGMP version 2

IGMP version 2 与 version 1 具有相同的功能，具有相同的数据包，但也有不同的数据包，下面分别来看 IGMP version 2 运行在路由器和主机上的数据包：

IGMP v2 主机:

IGMP 主机发送三种数据包来向 IGMP 路由器通告自己的状态：

1. Membership Report

2. Version 1 Membership Report

3. Leave Group

Membership Report

Membership Report 是主机发向路由器用来加入组的数据包，当路由器从接口上收到主机的 Membership Report，便认为该网络上有组成员，就开始将组播转发到该网络。主机同样使用 Membership Report 来回复路由器的查询。Membership Report 的目的地址是组地址，所以除了路由器之外，其它组员也能听到，一个网络中只会有一台主机向路由器发送 Membership Report 来通告网络中存在组员。

Version 1 Membership Report

Version 1 Membership Report 是兼容 IGMP v1 时发送的。



Leave Group

在 IGMP v1 中，并没有特定的机制来通知路由器组员的离开，唯一的方法就是路由器在 3 分钟后没有收到组成员对 queries 的回复，也就认为网络中的组成员已经离开。但是在 IGMP v2 中，组成员离开时，需要向路由器发送 Leave Group 来通告自己的离开。但是只有向路由器发送过 Membership Report 的组成员离开时才需要发送 Leave Group，其它组员离开时，也是没有任何动作的。Leave Group 目的地址为 224.0.0.2，也就是说组成员离开时，只有路由器知道，其它组员是不知道的。

IGMP v2 路由器：

IGMP 路由器靠查询消息来确认组成员的存在，运行 IGMP V2 的路由器使用 2 种查询消息：

1. General Query
2. Group-Specific Query

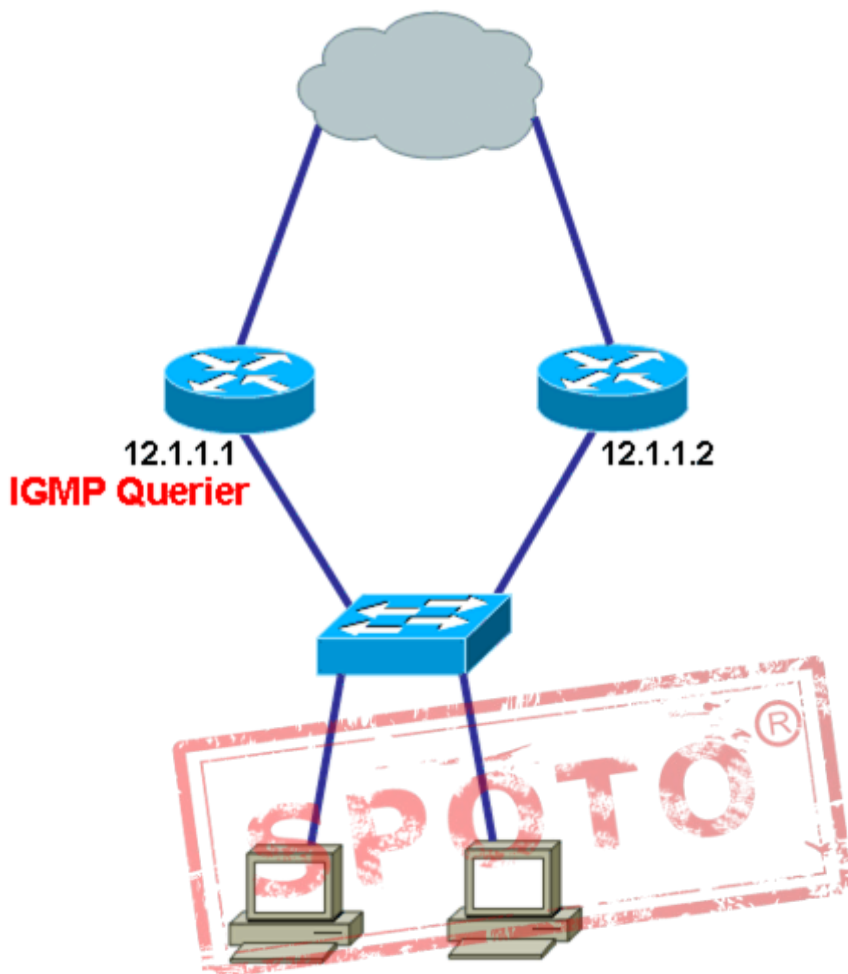
General Query

General Query 的作用和 IGMP v1 中 queries 的作用是一样的，当路由器发送 General Query 查询主机时，使用目的地址为 224.0.0.1，当有主机回复时，便认为网络中存在组成员，则将组播发送到该网络，并且 General Query 每分钟发送一次，主机每次都回复，如果连续三个 General Query 没有回复，即三分钟，那么路由器便认为网络中的组成员已经离开，也就停止向网络中发送组播数据了。

Group-Specific Query

在 IGMP v2 中，当网络中的组成员离开之后，会向路由器发送 Leave Group 来通告自己的离开，而只有向路由器发送过 Membership Report 的组成员离开时才需要发送 Leave Group，其它组员离开时，也是没有任何动作的，因此当网络中的组成员离开后，路由器并不知道网络中是否还有其它组员，因为其它组员没有通告过自己的存在。所以如果路由器马上停止向子网转发组播的话，可能会造成某些还存在的组员无法接收组播。最终当路由器收到组成员离开时发来的 Leave Group 后，并不能马上停止组播的转发，还要向网络中发送 Group-Specific Query，目的地址为组的地址，用来查询网络中是否还有别的组员存在，如果有组员回复了，那么组播流就不会中断。为了防止包丢失而误认为没有组员，路由器每 1 秒分别发两个 Group-Specific Query。

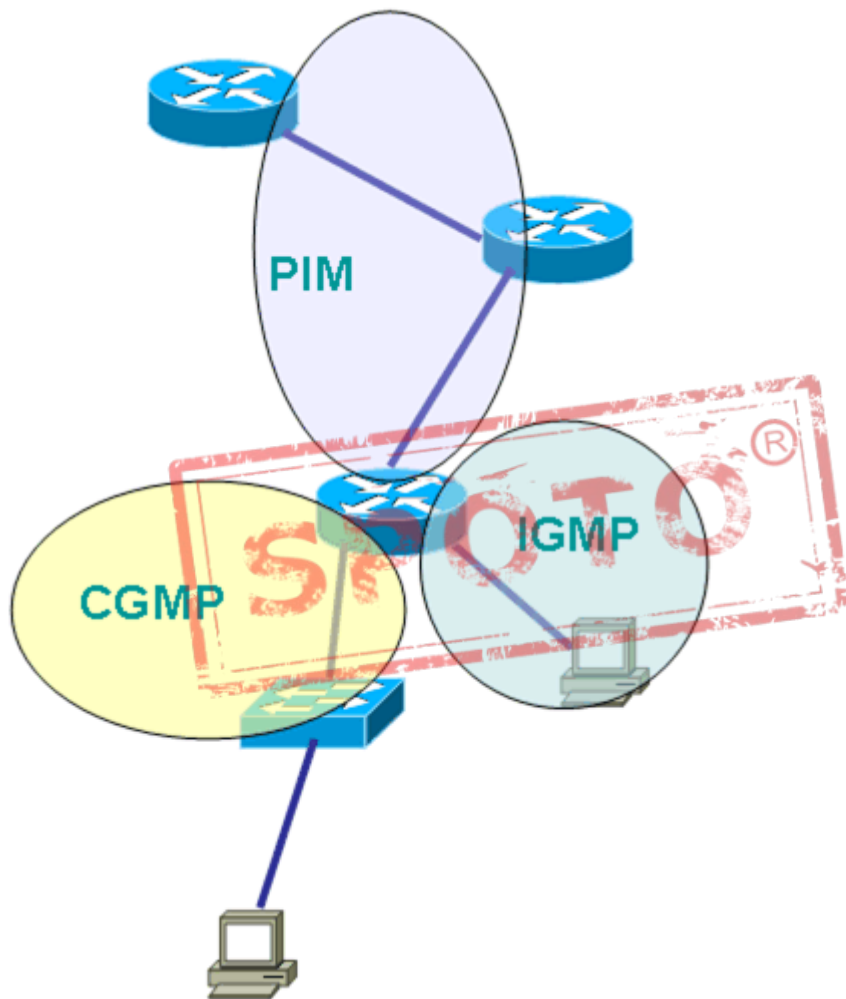
IGMP v2 Querier:



由上图可以看出，当有两台或更多路由器连接到同一个网络时，路由器要确认网络中是否有组成员，发送 **General Query** 就可以确认组员的存在，而当主机在使用 **Membership Report** 回复路由器的 **General Query** 时，因为 **Membership Report** 的目的地址是组地址，所以网络中所有路由器都能收到 **Membership Report**。如果网络中的每一台路由器都发送 **General Query** 来查询组员的存在，那么每台路由器得到的结果都是一样的，这样也就没必要让网络中的每台路由器都进行组成员的查询。这时，就从网络的多台路由器中选举出唯一的一台路由器来查询组成员，这台路由器就是 **IGMP querier**（IGMP 查询器），只要 **IGMP** 查询器查询一次，所有路由器都能得到相同的结果。网络中拥有最低 IP 地址的路由器将被选举为 **IGMP** 查询器，因此上图中被选为 **IGMP** 查询器的路由器为 12.1.1.1。当路由器正常启动 **IGMP v2** 后，就开始发送目标地址为 224.0.0.1 查询数据包，每台路由器收到后，查看数据包的源 IP 地址，最后得到结果谁是 **IGMP** 查询器，并且之后每一分钟发送一次查询数据包，如果路由器在两倍查询时间内没有收到查询器的查询数据包，便认为 **IGMP** 查询器已经失效，然后重新选举 IP 地址最小的为查询器。**IGMP** 查询器在 **IGMP v1** 环境下是没有的。

组播协议

要让组播正常运行起来，需要一些协议的协同工作，如下图所示，下面分别来介绍各协议的功能：



IGMP

因为组播数据只会朝着有组员的地方发去，所以要让想接收组播的主机正常收到组播数据，那主机就需要让发送者知道它的存在，当发送者知道了组成员的存在，就会向组成员发送组播数据。主机要宣告自己想要接收组播，可以发送一些数据包，而这些数据包，正是主机通过 IGMP 协议来发送的。当路由器想要知道网络中是否存在组成员时，同样也是发送 IGMP 数据包来查询的。在网络中确认组成员的存在，这就是 IGMP 协议的功能，可以看出，IGMP 是主机与路由器之间运行的组播协议。

CGMP

因为当路由器获知网络中存在组成员之后，就会将组播向该网络中转发。又因为网络中无论是一个组员，还是许多个组员，对于路由器来说，都得向网络中发送一份组播数据，如果网络中是用交换机连接了多台主机时，交换机收到组播后，会像转发广播一样，将组播发送给每一台主机。这样一来，如果只是部分主机是组成员的话，那么交换机的这种行为将导致组播转发的不精确，所以不能让交换机像处理广播那样处理广播，而必须将组播像处理单播一样，只发给需要接收的组成员。要实现这样的功能，就必须让交换机知道哪些主机才是组员，从而根据组员的 MAC 地址来转发组播。主机和交换机之间是没有协议的，却和路由器之间有协议，要准确的了解主机的 MAC 地址情况，就通过路由器得知，路由器再将此信息发送给交换机，最终使交换机能够像处理单播一样根据目标 MAC 地址来转发组播。这时就需要在路由器和交换机之间运行一种协议，那就是 CGMP (Cisco Group Management Protocol)，在 CCIE R&S 考试中，CGMP 并不是考点，所以此篇中将不会对 CGMP 作出详细介绍。



PIM

只有当组播发送者知道组成员的存在，才会向组成员的方向发送组播。在某些情况下，发送者和组成员可能在远程网络，那么这样一来，发送者发出的数据，必须经过路由器才能到达组成员的网络，所以要使组播数据准确地被转发到组成员的网络，就必须让中间的路由器也知道组成员网络的位置所在。两个不同网络的主机使用单播通信时，数据可以被中间路由器准确地转发，是因为路由器的单播路由表中能够找到目标网络的位置，如果要想让路由器也能像转发单播数据一样，将组播根据路由表来精确地转发到目的地，那就需要让路由器拥有像单播路由表一样的组播路由表，从而让路由器在收到组播时，就像查单播一样，去查组播该从什么样的接口被发出去才能到达目标网络。要让路由器生成一张功能完全的组播路由表，就需要在路由器之间运行一种协议，这种协议可以让组播源和目的之间的路由表生成单播表一样地生成组播表，最后路由器根据这张组播路由表来完成组播的转发。这个协议就是 PIM (Protocol Independent Multicast)。其实要让路由器知道目标组员的位置，完全可以依靠单播来找到组员，所以只要组播的源和目的之间单播是通的，那么组播路由表就能建立，而不用管单播运行的是动态路由协议还是静态路由协议，但是前提是 PIM 必须依靠单播路由表才能生成。

PIM

当组播的源和目的在远程网络时，需要依靠路由器转发组播数据，要让路由器像转发单播一样地转发组播，就必须生成一张组播路由表，而 PIM 协议正是用来帮助路由器生成组播路由表的协议，只有当路由器都拥有完整的组播路由表时，远程网络之间的组播才能正常通信。因为如果路由器不能对组播数据负责而随意转发组播，这将带来不可想象的结果，所以路由器必须像依靠单播路由表精确转发单播一样去转发组播，下面来详细介绍组播的重点知识，PIM 组播路由协议：

组播树

因为在需要将一份数据同时发给多个接收者时，而开发了组播技术，所以组播的发送者通常面临着要将数据发向多个接收者，并且这些接收者可以分布在任意网络的任意位置。如果接收者在远程网络，那么就需要路由器提供组播转发，所以为了保证接收者能够正常收到组播，就必须让路由器知道自己该将组播从什么样的接口发出去，当组播到达下一跳路由器后，下一跳路由器同样也必须知道该将组播从什么样的接口发出去，即使接收者不是与自己直连的，只有这样让路由器之间协同工作，都能够记住组播的出口，最终在发送者与接收者之间形成一条连线，这样才能完成组播的转发。当多个网络存在接收者时，那么这样的连线就会有多条，组播发送者到接收者之间的这些转发线路，被称为组播转发树，而组播发送者就好比是组播树的树根，组播总是从根发向接收者。很容易想象得出，从发送者到接收者之间的路由器，都是在组播树上的，因为这些路由器在中间提供组播转发，那么如果一台根本没有与接收者相连的路由器，与组播树是没有任何关系的。要完成从发送者到接收者之间的组播转发，组播树上的路由器都应该记住组播的出口，每台中间路由器都记住出口之后，最终便形成了组播树，而要记住组播的出口信息，这就是组播路由表的工作。

组播反向路径转发

因为组播经常会有多个接收者分布在不同的网络，所以当路由器在转发组播时，也通常需要将同一份数据从多个接口发出去，如果别的路由器不小心或者由于各种

原因再将组播发送回来，对于这样已经出现环路的数据，如果路由器收到后，再次将组播转发出去，那么只会形成组播风暴，最终危害整个网络。由此可以看出，转发组播的路由器，必须拥有发现环路、避免环路的能力。要实现这样的功能，所有的组播路由器就必须只将数据往接收者的方向转发，而绝不能往发送者的方向转发，因为向发送者的方向转发，就等于是将数据往回发，那就是引起路由环路的原因。让组播路由器只将数据往接收者的方向转发，而不能往发送者的方向转发，这样的机制被称为组播反向路径转发（mRPF），这也是组播路由器必须遵循的机制。组播路由器在收到组播数据后，都要对数据进行 RPF 检测，只有从源的方向发来的数据才能被转发，从其它接口过来的数据被认为是无效的。

因为组播路由器在转发数据时，必须遵循组播反向路径转发（mRPF）机制，所以路由器除了要记住数据从什么接口发出去是接收者之外，还要记住组播发送者是在哪个接口，而数据是绝不能发向发送者的。由此可以看出，组播路由器在正常工作时，组播转发表中必须同时记住接收者的接口和发送者的接口，其它不相关的接口，可以不用记住，也可以记为剪除状态，被剪除的接口是没有组播流量的。

组播路由表记住了接收者的接口后，就能正常将数据发给组成员，而记住了发送者的接口后，就可以避免路由器将数据往回发而产生路由环路了。组播路由器上朝发送者方向的接口被称为 RPF 接口，这个接口是在单播路由表中指向组播源地址的出口，因为从 RPF 接口发出去的数据就等于将数据发向发送者，所以组播只能发往除 RPF 接口之外的其它接口，并且收到的组播数据如果不是从 RPF 接口进来的，全部都将被丢弃。

PIM 模式

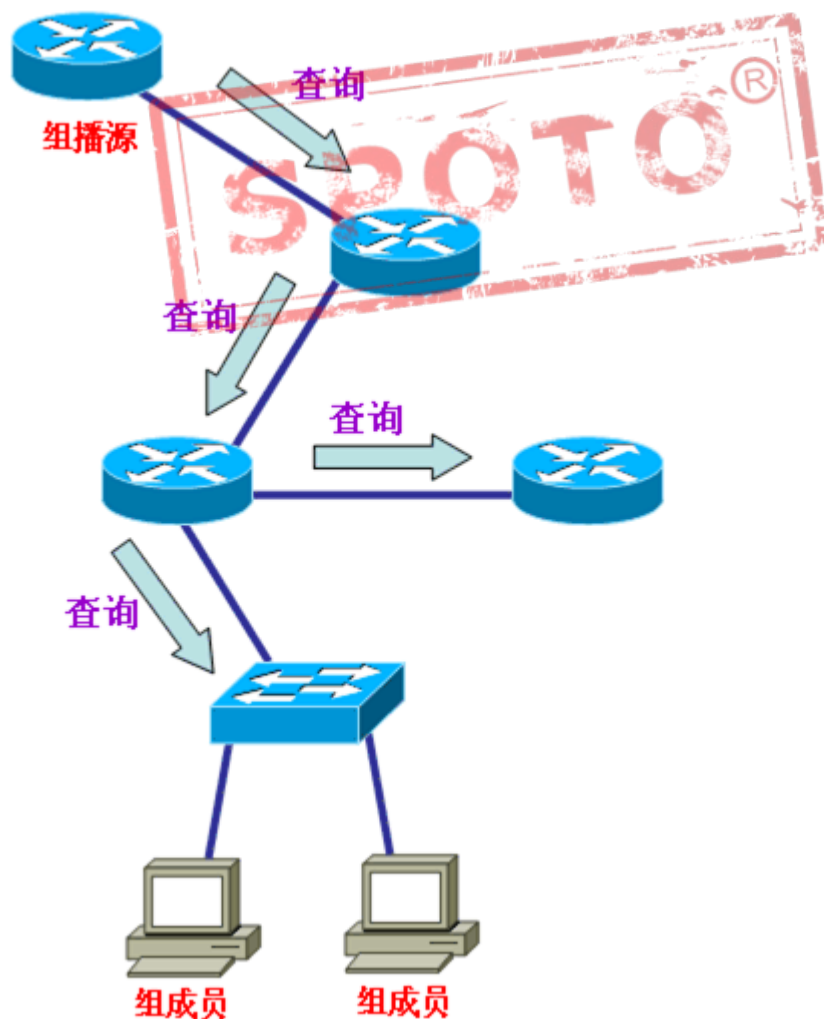
在组播正常通信之前，组播路由器必须知道哪个接口是通往发送者的，即 RPF 接口，组播数据只能从 RPF 接口进来。组播路由器还必须知道哪些接口是通往接收者的，组播将从这些接口被转发出去。从组播源到接收者之间的所有路由器都记住了 RPF 接口和组播出口的组合，被称为组播树，组播树用来指导路由器如何转发组播，而组播树就是记录在组播路由表中的，因此，有了正确的组播转发表后，组播就能够正常通信了。

要让组播路由器生成组播路由表，就是 PIM 协议的工作，PIM 在组播路由器之间运行，通过路由器之间的协商，从而获得组播路由表，构建出组播树。PIM 要为路由器学习组播路由表从而建立组播树，有两种不同的方式，这两种不同方式在 PIM

中分两种模式来运行，为 PIM-DM（密集模式）和 PIM-SM（稀疏模式），下面来详细介绍 PIM 的两种运行模式：

PIM-DM（密集模式）

组播树是用来指导路由器如何正确转发组播的，它的相关信息全部都是记录在组播路由表中的，PIM 用来帮助路由器生成组播路由表。要形成组播树，路由器需要知道哪些接口出去能够到达接收者，并记录下来，然后再记录到发送者的 RPF 口。要让路由器知道哪些接口存在接收者，有两种方式，第一种方法是接收者主动向路由器报告，第二种方法是路由器主动向网络中发出查询，而 PIM-DM 模式中，采用的方法为路由器主动向网络查询是否有接收者，如下图所示：



在上图显示的 PIM-DM 工作模式中，组播源会向所有 PIM 邻居发出查询，查询数据包中包含组的地址，下一跳 PIM 邻居还会继续向它的邻居发出查询数据包，这些查询数据包会在所有 PIM 邻居之间传递。如果查询数据包发向一个连接了组成员的网络，这时路由器收到组成员的报告之后，就会向自己上一跳邻居（RPF 接口方向的邻居）发送加入组的消息，以宣布自己要接收组播，从而将组播转发到组成员。在这些过程中，如果某些 PIM 路由器根本没有与组成员相连，那么它将会向自己的上一跳邻居发送剪除消息，以宣布自己不需要接收组播，最终组播从源发出后，只会沿着组播树被发到连接了组成员的网络，而其它不相关的网络是不会有组播流量的。

以上的信息，路由器都会记录在组播路由表中，这个路由表是不同与单播路由表的，组播路由表拥有自己的格式，因为网络中可能存在多个组，所以路由器在记录时，应该记录下组地址，然后就是该组需要从哪些接口被发出去，最后还由于路由器需要对组播数据进行 RPF 检测，所以还必须记录与组对应的组播源地址，也就是记录发送组播数据的源 IP 地址。例如网络中有一个组，地址为：224.1.1.1，发送者为 100.1.1.1，那么就记录为 (100.1.1.1, 224.1.1.1)，除了记录下组地址和源 IP 地址之外，路由器还要记录的就是组播的出口，收到的组播就会从这些出口发出去。PIM-DM 模式中这样记录组播的方式被称为 (S,G)，其中 S 就是组播源地址，G 就是组地址，而出口则会被标为 forwarding。

如果只有一个组 224.1.1.1，而发送者除了 100.1.1.1 之外，还有 100.1.1.2 和 100.1.1.3，那么依照 (S, G) 的记录方式，就需要同时记录 (100.1.1.1, 224.1.1.1)，(100.1.1.2, 224.1.1.1)，(100.1.1.3, 224.1.1.1)，也就是说 (S,G) 的记录方式，会因为组源地址的增加而增加记录条目。

可以看出，对于一个组，PIM-DM 模式中，路由器需要记录组地址，源地址，出口信息，除了这些之外，比如路由器上有 5 个接口有 PIM 邻居，其中只有一个出口，则这个接口被标为 forwarding，再去掉 1 个 RPF 接口，那么还剩 3 个 PIM 接口是不需要接收组播的，对于不需要接收组播的 PIM 接口，PIM-DM 模式照样会将其记录在路由表中，但被标为 pruning。因为 PIM-DM 这样的记录方式，可以看出当路由器上只有少量的 PIM 接口与组成员相连时，PIM-DM 会消耗更多的资源去记录一些没有组成员的不相关的接口，而这样的事情是可以避免的。

在 PIM-DM 模式下，组播发送源将数据发给组播路由器，然后路由器依照组播路由表朝着接收者的方向转发，这样的路径，是依靠单播路由表计算出来的最短路径，也就是说从发送者到接收者之间的路径，总是最短的，所以 PIM-DM 模式建立起来的组播树，如 (S, G) 记录的组播树被称为最短路径树 shortest-path tree (SPT)，因为 SPT 中记录每个组播发送者的源地址，故又被称为源树。

PIM-SM（稀疏模式）

在 PIM-DM 模式中，组播树的生成是靠路由器发送查询消息来建立的，在组播路由表中记录下组地址，发送者的源 IP 地址，同时还需要记录下所有 PIM 接口状态，而不管这些接口是连接着组成员的，还是没有。

而 PIM-SM 模式的工作过程和 PIM-DM 模式是不同的，在建立组播树时，PIM-SM 并不会让路由器发送查询数据包去查询组成员，而组成员的发现是靠组成员自己主动向路由器发送报告数据包，当一台路由器从接口上收到组成员的报告之后，就会向自己的上一跳邻居发送加入消息，以通告自己需要接收组播，如果上一跳邻居还不是组播发送者，那么上一跳邻居会继续再向上一跳邻居发送加入消息，直到组播源收到加入消息为止，通过这样的方式，就可以建立组播源到组成员之间的组播树。

PIM-SM 模式和 PIM-DM 模式除了在发现组成员的方式上不同之外，记录路由表的方式也不一样，例如网络中有一个组，地址为：224.1.1.1，发送者为 100.1.1.1，PIM-SM 模式记录为 (*, 224.1.1.1)，可以看出，PIM-SM 模式并不关心组播的源地址，而统统将源地址使用星号*来表示，这样一来，PIM-SM 为一个组只记录一个条目，而不管这个组有多少个发送者。这样的记录方式称为(*, G)，其中*就是组播源地址，G 就是组地址。如果有 10 个组，每个组有 5 个发送者，使用 (S, G) 的记录方式，需要记录 $10 \times 5 = 50$ 条，而使用 (*, G) 的记录方式，则只需要 10 条，所以使用 PIM-SM 模式，可以大大缩减组播路由表的空间，从而大大节省系统资源。

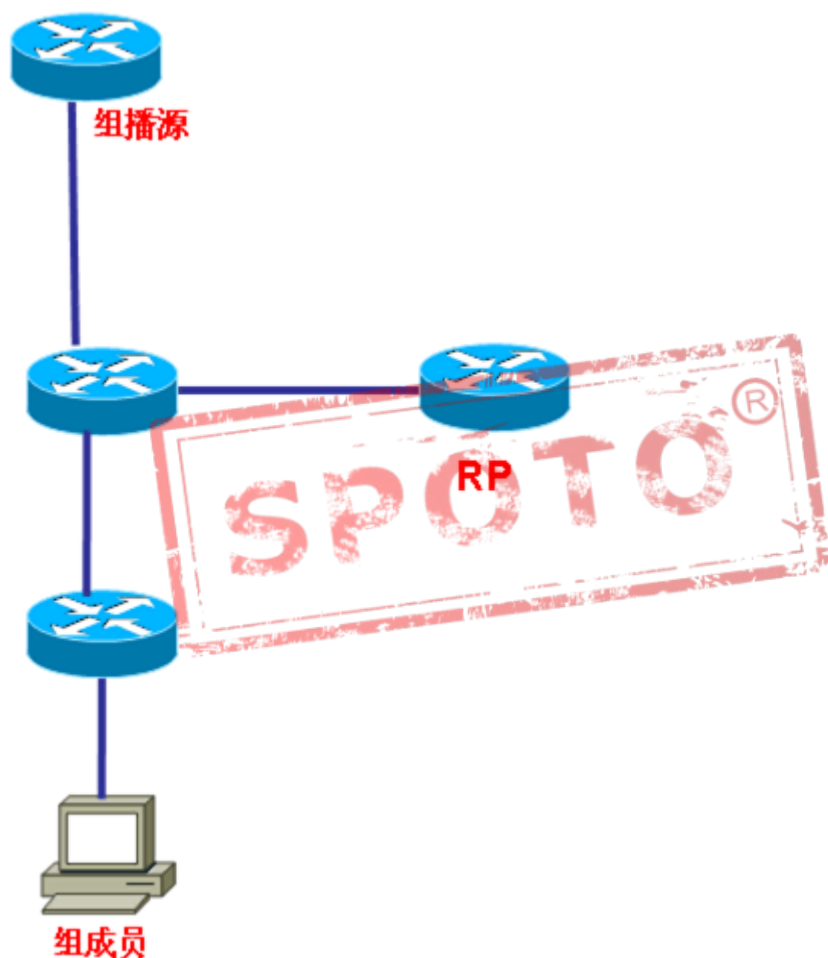
PIM-SM 模式不仅需要记录(*, G)信息，也和 PIM-DM 模式一样需要记录在该组中，哪些接口是出口，从此接口将数据发给接收者。但是与 PIM-DM 模式不同的是，PIM-SM 模式只记录连接着接收者的接口，其它没有接收者，不需要接收组播的接口是不会被记录的。比如路由器上有 5 个接口有 PIM 邻居，其中只有一个出口，再去掉 1 个 RPF 接口，还剩 3 个 PIM 接口是不需要接收组播的，在 PIM-DM 模式中，会记录下一个 RPF 口和一个 forwarding 状态的出口，以及三个不需要接收组播的 pruning 状态的接口，而在 PIM-SM 模式中，只会记录一个 RPF 口和一个 forwarding 状态的出口，其它的都不作记录，因此，PIM-SM 模式的接口记录方式会比 PIM-DM 模式更省资源。

PIM-SM RP

在 PIM-SM 模式中，由于记录组播信息采用(*, G)的方式，而并不关心组播源地址，因此造成路由器不知道组播发送者的 IP 地址是什么，也就无法完成 RPF 反向路径检

测。在这种情况下，PIM-SM 在网络中选出一个组播会聚点，即 Rendezvous Point (RP)，RP 就是组播网络的核心，发送者统一将组播数据发送到 RP，然后 RP 再将数据发到接收者，也就是说接收者收到的数据，都是由 RP 转发过来的，路由器也就认为 RP 的地址，就是组播源的 IP 地址。

在 PIM-SM 模式中，通过 RP 转发数据，如下图所示：

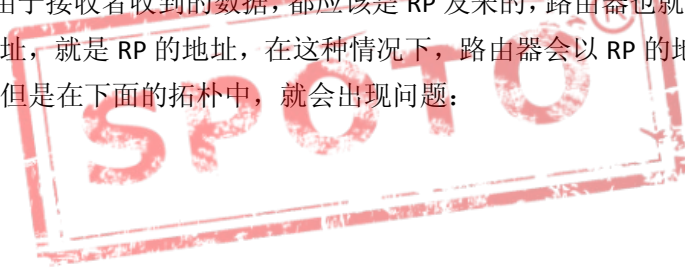


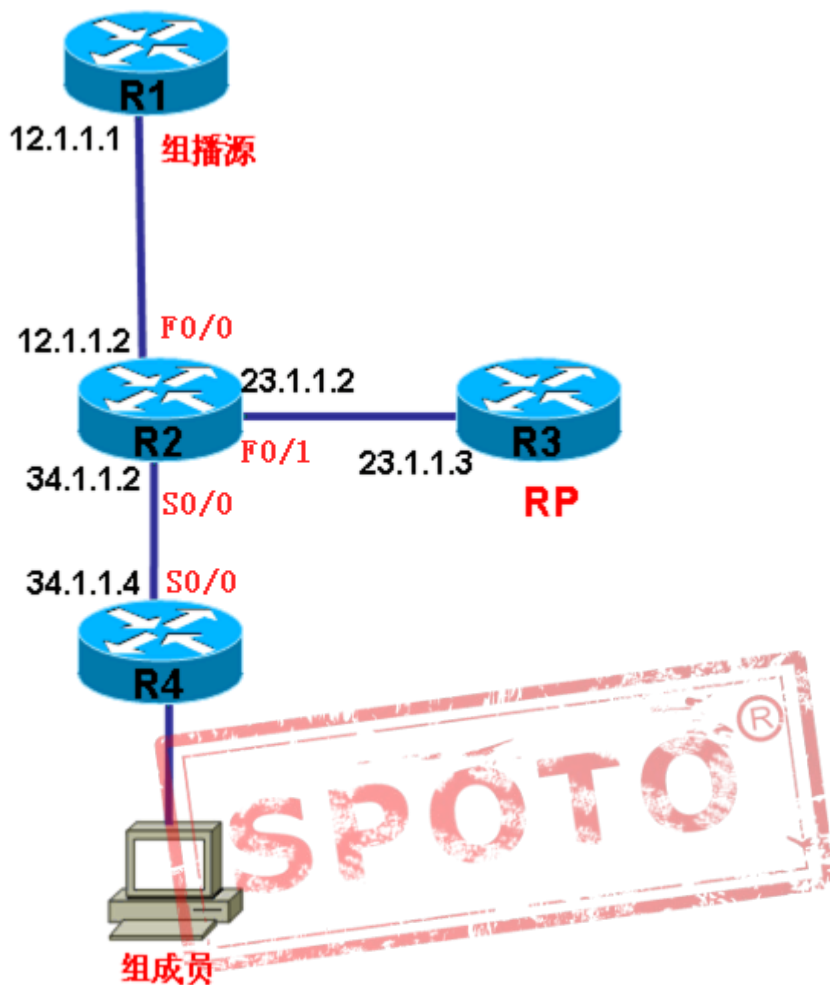
因为数据总是先发送到 RP，然后 RP 再发给组成员，如果 RP 出现了故障，也就意味着组播无法正常通信。组播中的一个组，只有一个 RP 是处于转发状态的，有时 RP 路由器的位置可能不是离每个组成员都是最近的，并且数据从源到 RP 然后再到组成员的路径也可能会很远，所以 RP 的位置，会导致源到组成员的数据包走的不是最优路径，也会在大流量的情况下，成为组播通信的瓶颈。但是在这种情况下，PIM-SM 有自己的解决方法，就是在正常组播通信时，路径会切换到源到组成员最短的路径上，也就是挑选网络中的最优路径来转发，和源树的路径方式相同，这个规则可以随意更改，定义流量超过多少时可被启用，思科路由器默认收到第一个包

后即启用。

因为 RP 在 PIM-SM 中被当作核心，而(*, G)的记录方式中并不知道组播源的地址，大家都会认为 RP 就是组播源，最后 PIM-SM 中的组播树，即(*, G)形式的树被称为共享树（RPT）。

PIM-SM 模式中的 RP，可以手工指定，也可以使用动态协议来选择。如果使用手工指定，当使用中的 RP 出现故障后，也必须手工更改，才能恢复通信。而使用动态协议，则可以配置多个备份 RP，但活动 RP 只有一个，当活动中的 RP 出现故障后，协议会重新选择备用 RP 替换为当前的活动 RP，能够实现冗余功能。

在 PIM-SM 中，因为(*, G)形式的记录中不知道组播源的地址，也就无法完成 RPF 检测，但是又由于接收者收到的数据，都应该是 RP 发来的，路由器也就认为 PIM-SM 中的组播源地址，就是 RP 的地址，在这种情况下，路由器会以 RP 的地址为源地址做 RPF 检测，但是在下面的拓扑中，就会出现问题：



如上图所示，当所有路由器都认为 RP 地址 23.1.1.3 就是组播源地址时，这样去做 RPF 检测，在 R4 上检测时，因为 R4 的路由表中会显示从接口 S0/0 出去可以到达 23.1.1.3，所以组播从 R4 的 S0/0 被发进来时，则 RPF 检测可以通过，所以在 R4 上没有任何问题。

当在 R2 上做 RPF 检测时，因为 R2 的路由表中显示到达 23.1.1.3 应该从接口 F0/1 出去，所以 R2 只会接收从接口 F0/1 发进来的组播，从其它接口发进来的，都会被认为是环路而被丢弃。因此在真正的源 12.1.1.1 将组播从 R2 的 F0/0 发进来时，R2 会因为 RPF 检测失败，从而丢弃所有的组播数据，这样一来，组播也就无法通信了。

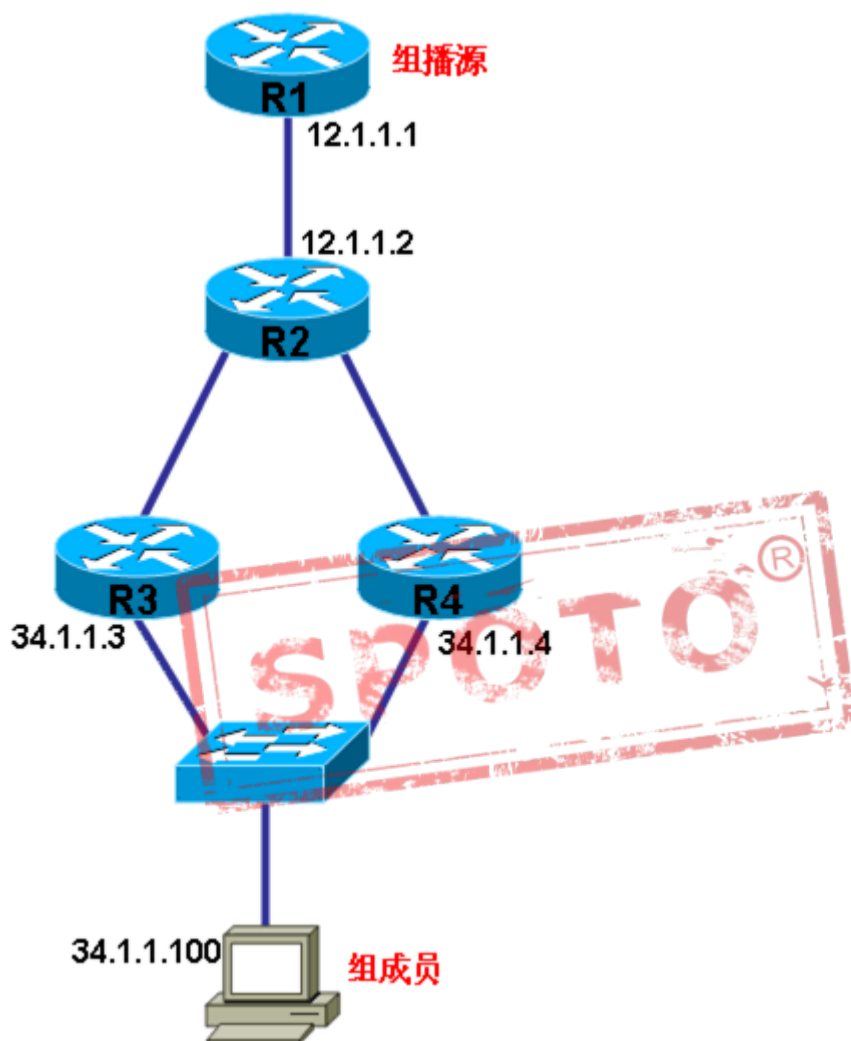
PIM DR

因为 IGMP V1 不选举 IGMP 查询器，如果有多台路由器连接同一个多路访问网段时，就必须选举 PIM Designated Router （DR）来避免重复查询，这个 DR 的选举，是由 PIM 进行的，选择 IP 地址高的为 DR。PIM DR 路由器在 IGMP v1 中充当 IGMP 查询器的功能，来向主机发送 IGMP 查询。

在 PIM-SM 中，组播源没有机制用来宣告自己的存在，又因为 PIM 路由器都认为 RP 才是组播源，而当直接的组播源向 RP 发送组播时，会导致某些路由器 RPF 检测失败，因此在 PIM 网络中，真正的源需要向 RP 发送注册消息，以宣告自己的存在，而发送注册消息在多路访问的网络中则是由 DR 来代为完成的。真正的源向 RP 发出第一个组播包时，DR 将此包封装在单播中发向 RP，这称为注册，发送的注册消息会从 DR 到 RP 之间建立一条源树，也就是（S，G）的记录，这样，在源到 RP 之间创建的源树就可以帮助避免 RPF 检测失败，当 RP 和真正的源之间创建（S，G）条目之后，就会通知 DR 停止以单播发送，从而转回发送真正的组播。因为由于误认为 RP 是源而导致 RPF 检测失败的情况，只有在 RP 到源这段上游网络中的路由器才会存在，而 RP 到组成员的这段网络中是不会出现这种情况的。由此可以想象得出，当源所在的网络需要向 RP 发送数据包时，应该将 DR 的位置选为最靠近 RP 的路由器，否则也会因为 RPF 检测失败而导致组播不通。

选择 DR 的数据包为 30 秒一次，105 秒保持时间，如果过了保持时间没有收到 DR 的数据包，则会重新选举 IP 地址最高的为 DR。

PIM 前转器



从上图中可以看出，当组播源 **R1** 将数据发向组成员时，数据到达 **R2** 之后，可以分别从 **R3** 和 **R4** 两条路径到达组成员。要知道组播通常是应用在视频和语音环境下的，假如 **R2** 选择将数据包以两条路径负载均衡的方式发往组成员，当 **R3** 和 **R4** 两条路径之间的延时不相同时，数据包到达组成员就会出现明显的先后顺序，这在普通数据传输时是没有问题，但是如果当数据是视频时，那么就可能出现视频画面先后不一致的情况，又如果当传输的是语音时，对方在报数字，比如 **R2** 向组成员发送 123456，在采用负载均衡传输时，将 1 发向 **R3**，将 2 发向 **R4**，再将 3 发向 **R3**，再将 4 发向 **R4**，再将 5 发向 **R3**，再将 6 发向 **R4**，当 **R3** 链路上的速度比 **R3** 快时，数据 135 就会先到达组成员，246 就会后到达，最终导致 **R2** 发出的数据 123456，

在组成员收到时，就可能变成了 135246。由于这些原因，组播负载均衡可能导致数据传输错误，所以当组播源到组成员之间有多条路径可走时，是一定要选出其中唯一的一条路径的，如上图，就 R2 就必须选出要么从 R3 走，要么从 R4 走，而绝不能负载均衡，两个下一跳中被 R2 选中的路由器，称为 PIM 前转器，选 PIM 前转器的规则是，比较哪台路由器的路由表中，到组播源的路由的 AD 值最小，如果 AD 值相同（如使用同一种路由协议），再比较谁到源的 metric 值最小，如果 metric 值还相同，则选择 IP 地址大的。在上图中，如果 R3 和 R4 到组播源 R1 的路由条目中是使用同一个路由协议学到的，则比较 metric 值，如果 metric 值相同，则比较 R3 和 R4 的 IP 地址，这个 IP 地址是同网段相互建 PIM 邻居所使用的 IP 地址，因为 34.1.1.4 比 34.1.1.3 要大，所以前转器为 R4，因此 R2 会选择从 R4 到达组成员，并且把到 R3 的出口剪除。

PIM-DM 数据包

在 PIM-DM 模式中，所使用的源树，会产生 (S, G) 条目，但是也会创建 (*, G) 的条目，创建 (*, G) 是为了作为 (S, G) 的“父”数据结构，所有 PIM 接口都会加入输出接口，而输入接口总是空的。

在运行 PIM-DM 时，路由器会发送 5 种数据包：

1. Hello
2. Join/prune
3. Graft
4. Graft-Ack
5. Assert

1. Hello

PIM 路由器用来发现邻居，会在启动了 PIM 的接口上周期性地发送 Hello，默认 30 秒发送一个，有个保持时间，如果过了保持时间没有收到邻居的消息便认为邻居丢失，保持时间为 Hello 的 3.5 倍（105 秒）。PIM 邻居为 0.0.0.0 的，表示邻居就是

组播源。

2. Join/prune

PIM-DM 启动后，组播源开始向所有邻居发送查询消息，邻居会再向自己的邻居发送查询消息，如果有路由器连接着组成员，那么就会向上一跳邻居发送 join 数据包，用来通告自己需要接收组播，从而将自己接入组播树。其它没有连接组成员或不需要接收组播的路由器则向上一跳邻居发送 prune 数据包，用来通告自己不需要接收组播，状态被标为 prune，而不是 forward。被标为剪除状态，是方便被剪除的路由器在必要时把自己接入树中。

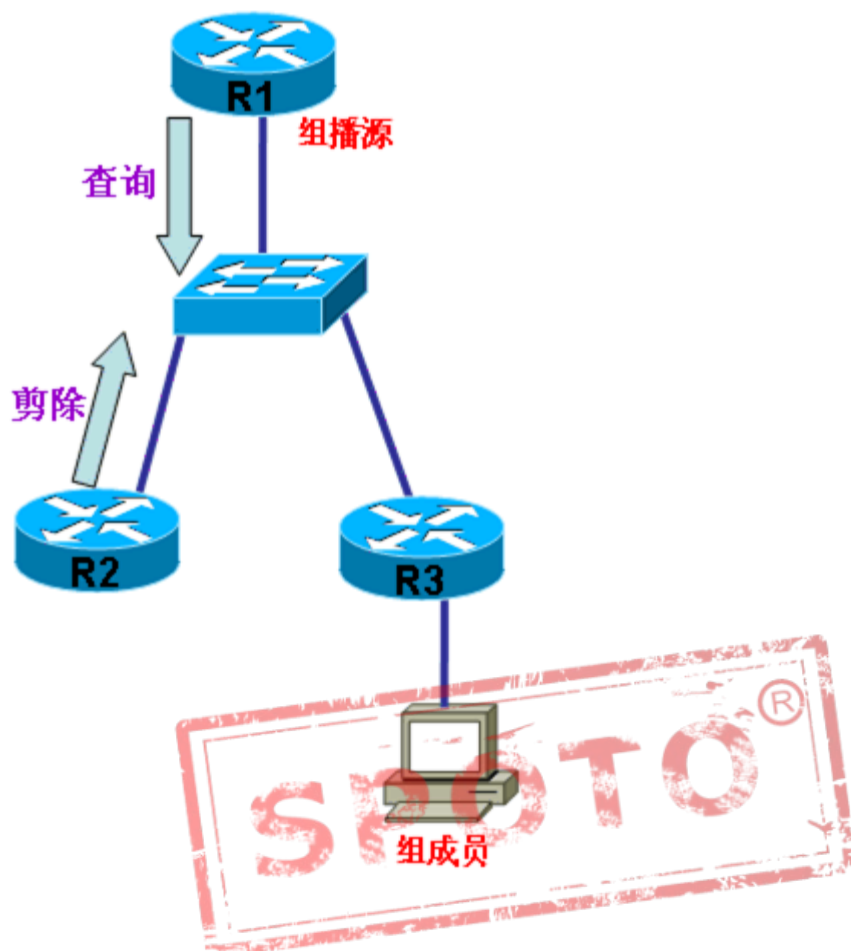
3. Graft

在路由器向下一跳邻居发送查询数据包，查询是否有组成员存在时，邻居必须发送 join 数据包来通告自己需要接收组播，然后不需要接收的，则发送 prune 将自己剪除，如果剪除后的路由器后来需要再接收组播，是不能发送 join 数据包的，而必须发送 graft 来将自己接入组播树。当被剪除的路由器向上一跳邻居发送 graft 数据包后，上一跳会将邻居接口标为组播的出口，即 forward 状态。如果上一跳邻居也是被剪除的，那么也会再向上一跳邻居发送 graft。

4. Graft-Ack

当被剪除的路由器向上一跳邻居发送 graft 数据包后，上一跳除了会将邻居接口标为组播的出口之外，还会向下一跳邻居发送一个 Graft-Ack。

Prune 消息覆盖



当一台路由器收到 PIM 查询后, 如果自己不需要接收组播, 或者没有连接组成员, 就会向上游发送 Prune 消息将自己剪除。如上图, 因为 R1 有两个 PIM 邻居, 为 R2 和 R3, 当 R1 向网络中发出查询后, 因为 R2 不需要接收组播, 所以会发送 prune 消息将自己剪除, 在正常情况下, R1 是应该将出口剪除的, 但是可以看出, R3 连有组成员, 是需要接收组播的, 如果 R1 因为收到 R2 的剪除消息而将出口剪除后, 那么 R3 发出的 join 消息是不能再收到组播的, 而必须发送 graft, 因为这样, 就会导致 R3 在收到查询后发送 join 却不能收到组播, 要避免这样的事情发生, PIM 路由器向网络中发出查询后, 即使收到剪除消息, 也不会马上将接口剪除, 而会启动一个 3 秒的计时器, 如果在 3 秒内, 收到 join 后, 就会覆盖前面的剪除消息, 就不会将接口变成剪除状态, 这被称为 Prune 消息覆盖。

5. Assert

Assert 是用来选举 PIM 前转器的, 当从源到组成员有多条路径可走时, 则选择其中唯一的一条路径, 被选中的路由器称为 PIM 前转器。

Assert 消息中包括源和组地址，以及到源的单播路由 AD 值和 metric 值，先选 AD 值低的，其次是 Metric，最后是最高 IP 地址，落选的把自己的出口剪除。

PIM-SM 数据包

PIM-SM 模式因为组播源要把数据包先发送到 RP，然后 RP 再将数据发到组成员，这样的共享树可能出现走远路的情况，所以有时不如 PIM-DM 中的源树。但是又因为 PIM-SM 可以从共享树切换到源树，也就是同时支持共享树和源树，而且在记录组播路由表时，更节省资源，所以被多数选用。

在运行 PIM-DM 时，路由器会发送 5 种数据包：

PIM-SM 采用了 7 种 PIMv2 消息：

1. Hello
2. Bootstarp
3. Candidate-RP-Advertisement
4. Join/Prune
5. Assert
6. Register
7. Register-Stop



1. Hello

PIM-SM 中的 Hello 同 PIM-DM 中的 Hello 功能相同，是用于 PIM 邻居的建立和维护。

2. Bootstrap

Bootstrap 用于 RP 的选举，稍微将详细介绍。

3. Candidate-RP-Advertisement

Candidate-RP-Advertisement 用于 RP 的选举，稍微将详细介绍。

4. Join/Prune

PIM-SM 中的 Join/Prune 同 PIM-DM 中的 Join/Prune 功能相同，是用于路由器加入组播树或将自己从中剪除。

5. Assert

PIM-SM 中的 Assert 同 PIM-DM 中的 Assert 功能相同，用于选举 PIM 前转路由器。

6. Register

DR 在功能，在前面已经提到过，就是在 PIM-SM 中，因为 PIM 路由器都认为 RP 才是组播源，而当直接的组播源向 RP 发送组播时，会导致中间路由器 RPF 检测失败，因此在 PIM-SM 网络中，真正的源需要向 RP 发送组播，就应该建立一条 (S, G) 的记录来避免 RPF 检测失败。当真正的源向 RP 发出第一个组播包时，DR 将此包封装在单播中发向 RP，这称为注册，也就是 Register 消息，发送的注册消息会从 DR 到 RP 之间建立一条源树，也就是 (S, G) 的记录，这样，在源到 RP 之间创建的源树就可以帮助避免 RPF 检测失败，当 RP 和真正的源之间创建 (S, G) 条目之后，就会通知 DR 停止以单播发送，从而转回发送真正的组播。

7. Register-Stop

因为 PIM-SM 中真正的源需要向 RP 发送 Register 消息来注册，从而建立 (S, G) 条目，将组播正常发向 RP，当 RP 和真正的源之间创建 (S, G) 条目之后，就会发送 Register-Stop 通知对方停止发送单播，从而转回发送真正的组播。

最后可以看出，在 PIM-DM 中，不仅会建立 (S, G) 的记录，也会建立 (*, G) 的记录，而在 PIM-SM 中，除了建立 (*, G) 的记录之外，同样也有 (S, G) 的记录，组播树建立后，会周期性发送 Join/Prune，60 秒一个，保持时间为 3 倍，即 180 秒，也就是说在 180 秒之后没有正常的组播流量，那么这些组播树是会被清除的。

RP 的确立

在 PIM-SM 中，组播源必须将数据发送到 RP，再由 RP 转发给组成员，当网络中没有 RP 时，组播是不正常的，因为要保证组播的正常通信，首先要让所有 PIM 路由器知道 RP 的地址，在共享树建立之前，必须确立 RP，可以单独为某个组配置一个 RP，也可以为多个组配置一个 RP，要确立 RP，有三种方法，分别是：

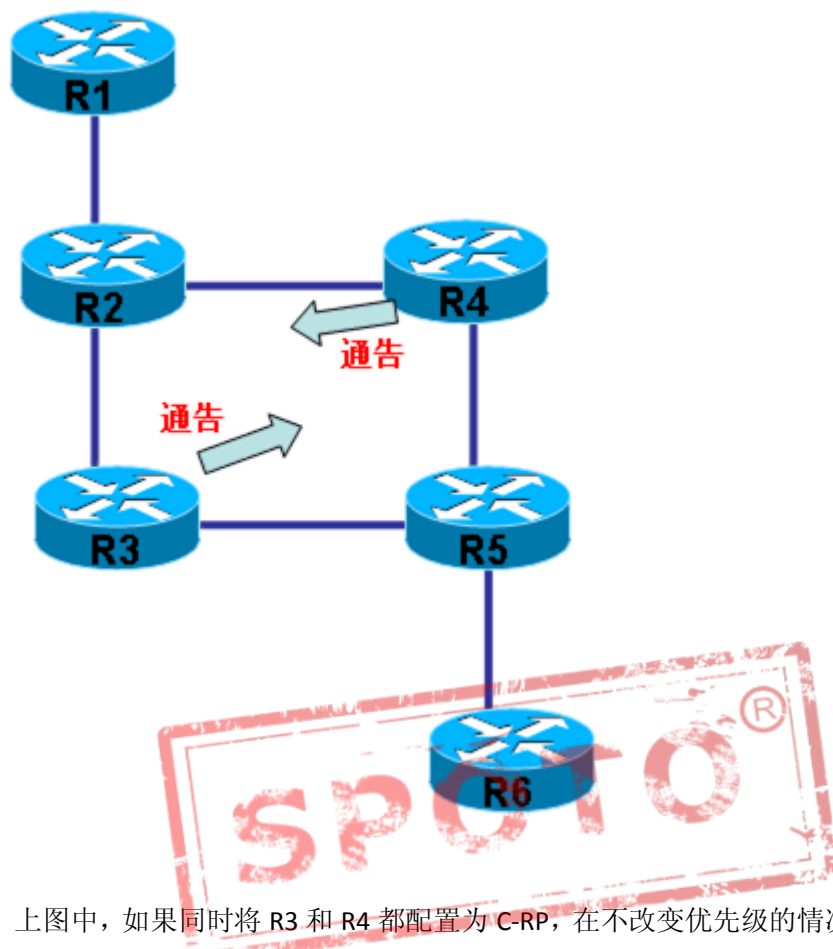
1. 手工静态配置
2. Bootstrap Router (BSR)来通告
3. 自动 RP (Auto-RP，思科独有协议)

1. 手工静态配置

一个网络中，PIM-SM 的 RP 可以手工静态为每台 PIM 路由器配置，RP 也需要配置，因为要让每台路由器都知道谁是 RP，但是这种方式不能提供冗余功能的 RP，当配置的 RP 失效后，必须手工更改，否则组播将不能通信。

2. Bootstrap Router (BSR)来通告

RP 的确立，除了手工为每台路由器静态配置之外，还可以使用协议自动选择，使用协议的好处在于，可以在网络中配置多个 RP，起到备份的作用，当正在使用的 RP 失效后，协议可以立即重新选择其它路由器成为活动 RP。自举协议 BSR 的工作方式为在网络中配置多个 RP，称为候选 RP (C-RP)，但只有一个 RP 是正在使用的活动 RP，在多个 C-RP 中，要成为活动 RP，选举规则为优先级最高的成为活动 RP，优先级默认为 0，范围 0 到 255，数字越小，优先级越高，如果优先级都相同的情况下，再比较 IP 地址高的，所以每个 C-RP 都会有一个 IP 地址用来表示自己的身份，也通过此 IP 来竞选 RP。如下图所示：



上图中，如果同时将 R3 和 R4 都配置为 C-RP，在不改变优先级的情况下，谁的 IP 地址最大，谁就是活动 RP，但是如果因为某些原因，R3 发出的 RP 竞选消息只被 R1 和 R2 收到，而 R4 发出的 RP 竞选消息只被 R5 和 R6 收到，那么 R1 和 R2 都会认为 R3 是 RP，而 R5 和 R6 都会认为 R4 是 RP，那么这样一来，网络中路由器得到的 RP 信息就变的 不一致，从而导致组播故障，所以一个网络中，谁才是活动的 RP，并不能让每台路由器自己去计算结果，因为可能会出现大家计算出不同的结果。要解决这个问题，方法就是在网络中选举出一个 RP 裁判，称为 BSR，而所有 C-RP 将自己竞选 RP 的消息统一发送到 BSR，是通过单播发向 BSR 的，最后由 BSR 从收到的竞选消息中，选择出活动 RP，再将活动 RP 的地址统一发给网络的每台路由器，这样就能保证每台路由器得到的消息都是统一的，每台路由器获知的 RP 地址都是统一的。而 C-RP 发出的竞选消息是 Candidate-RP-Advertisement。而 BSR 是通过路由器发送 Bootstarp 选出的，每一台候选 BSR(C-BSR 都)有一个 0 到 255 范围的优先级，默认为 0，优先级相同的情况下，IP 地址最高的为 BSR。Candidate-RP-Advertisement 和 Bootstarp 都是 60 秒发送一次。

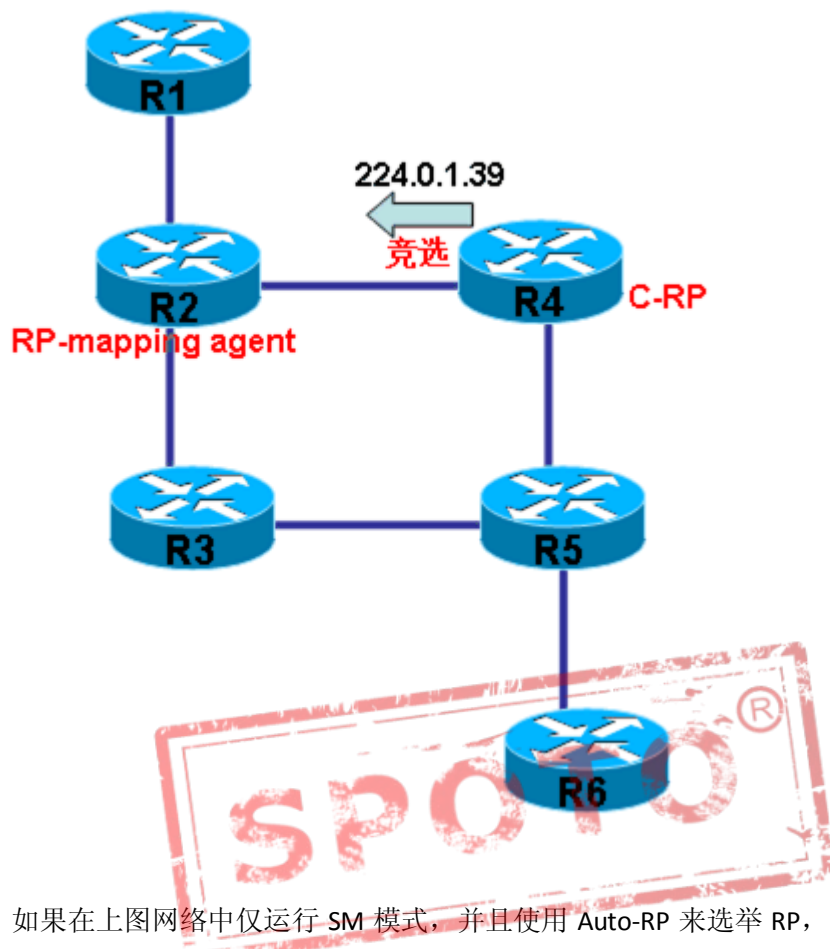
3. 自动 RP (Auto-RP, 思科独有协议)

Auto-RP 的工作过程与 BSR 相同，同样是在网络中配置多个候选 RP 和 RP 裁判，候选 RP 向 RP 裁判发送竞选消息(RP-Announce),最后 RP 裁判从候选 RP 的竞选消息中选出 IP 地址最高的为活动 RP，然后发送 RP-Discovery 通告给每台路由器，不同的是，竞选活动 RP 只根据 IP 地址大小，没有优先级之分。在 Auto-RP 中，侯选 RP 仍然被称为 C-RP，但 RP 裁判被称为映射代理（RP-mapping agent）。所有的 C-RP 向映射代理发送竞选消息，使用目的地址为 224.0.1.39，每 60 秒发送一次，而映射代理从众多 C-RP 中选出活动 RP 后，以目的地址为 224.0.1.40 发给每台路由器，也是每 60 秒发送一次。

Pim sparse-dense-mode

组播路由器在运行 PIM 时，可以运行在 SM 模式下，也可以运行在 DM 模式下，当运行在 SM 模式下时，必须有 RP，否则网络不通，而运行 DM 时，不需要 RP 组播就能通信。但 PIM 路由器可以同时运行两种模式，即 sparse-dense-mode，当同时运行这两种模式时，如果一个组有 RP 时，则使用共享树，但是当 RP 失效时，则可以使用最短路径树来保证组播的通信。

在下图的网络情况下：



如果在上图网络中仅运行 SM 模式，并且使用 Auto-RP 来选举 RP，当配置 R4 为 C-RP，配置 R2 为 RP-mapping agent 时，R4 必须发送 Announce 来竞选 RP，当 R2 收到 R4 的 Announce 后，才能得到最后的结果。配置 SM 模式时，在没有 RP 的情况下，组播是不通的，所以 C-RP 要发送 Announce 将 RP 选出来，但是 C-RP 在发送 Announce 时使用的目的地为 224.0.1.39，这也就意味着 C-RP 发出的 Announce 是到达不了远程 RP-mapping agent 的，因为没有 RP，所以最终导致没有 RP 的情况下，组播就不通，组播不通就选不出 RP，结果是组播永远通不了，要解决这个问题，就是要在 RP 没选出来的情况下就要让组播能够通信，这种情况下，就可以将 PIM 模式配置为 sparse-dense-mode，因为这种混合模式下，在没有 RP 时，可以走 SPT 从而连通组播，这样目的地为 224.0.1.39 的 Announce 也就能够正常发送到 RP-mapping agent，就能选出 RP 了。

PIM Dense Mode Fallback

当将 PIM 路由器配置成 sparse-dense-mode 时，在 RP 不可用的情况下，将从 RPT 切换到 SPT，这种特性被称为 Dense mode fallback，默认是开启的，但是如果接口只配了 PIM-SM，那么就不可能往 SPT 切换的，也就必须拥有 RP 才能通信。在配置

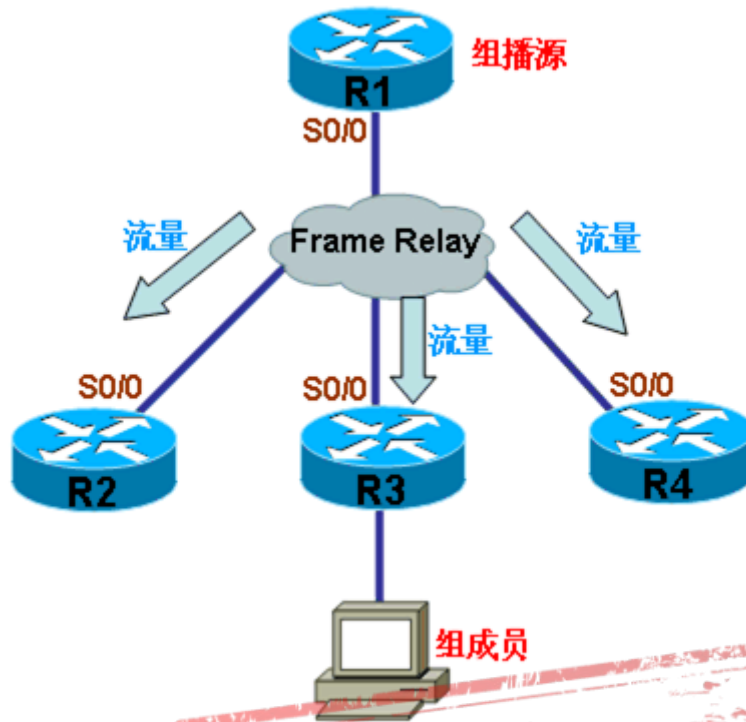
sparse-dense-mode 后，如果想阻止从 RPT 切换回 SPT，可关闭 Dense mode fallback 功能，命令为全局模式下输入 `no ip pim dm-fallback`。

共享树切换到源树

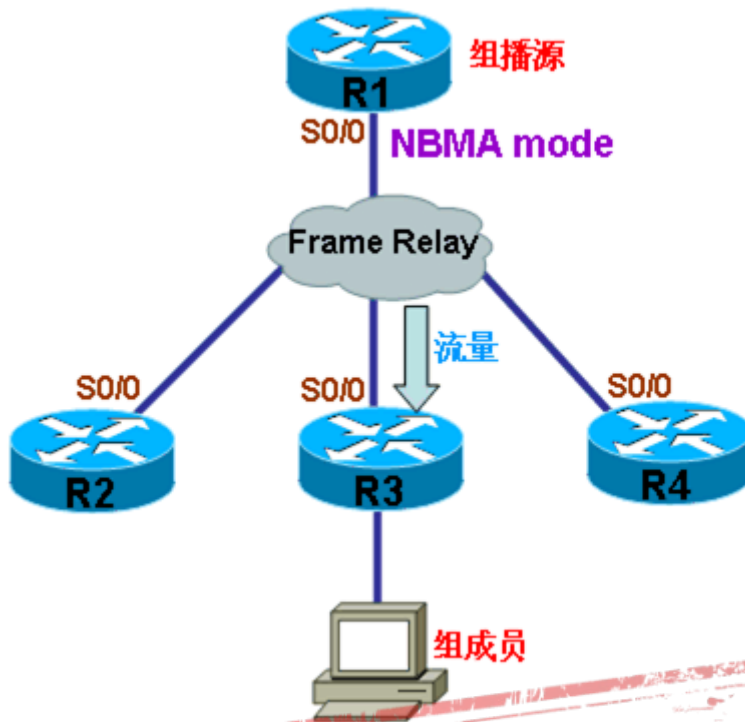
在 PIM-SM 中，组播源会先将数据发到 RP，然后 RP 再发给组成员，数据从源到 RP 再到组成员，这样的路径在很多时候可能不是网络中的最优路径，所以为了从源到组成员之间走最短路径，而不希望走次优路径，这就需要使用最短路径树，即 SPT。在路由器收到组播流量后，在源到组成员之间动态地创建最短路径树，从而替代共享树，这种行为在 IOS 中是默认开启的，但也可以单独为某个组开启，也可以关闭此特性。IOS 默认收到第一个组播包后，就从共享树切换到最短路径树，可以手工定义在共享树中的流量超过多少 kbps 后开始切换，流量低下去之后，60 秒会再切换回去。通过在全局模式下配置命令 `ip pim spt-threshold`，比如希望在流量超过 4Kbps 后切换到 SPT，命令为 `ip pim spt-threshold 4`。而命令 `ip pim spt-threshold infinity` 表示永远使用共享树而不切换到最短路径树。

PIM-SM 之 NBMA Mode

因为帧中继中没有内置的处理广播和组播的能力，所以在收到广播和组播时，会在所有配了关键字 `broadcast` 的 PVC 上转发，这样当帧中继接口对端有多个邻居，而只有部分邻居要接收组播时，那么组播的被当作广播转发的行为，将影响不需要接收组播的路由器，如下图：



在上图中，当 R1 通过帧中继主接口连接多个对端时，在收到组播后，将从帧中继主接口发送给每一位配置了关键字 `broadcast` 的邻居，但并不是所有邻居都有组成员，而只有 R3 需要接收组播，这样就会影响到 R2 的 R4。因此需要让组播路由器在帧中继环境下，只将组播转发给要接收组播的路由器，可以通过配置子接口或者将接口模式在 PIM 下改为 NBMA Mode，子接口类似多个物理接口，会建立多个邻居，所以会根据每个邻居情况进行组播转发，而配置 NBMA Mode 后，组播在主接口上会跟踪 PIM-SM 模式的 join 消息，以记录需要接收组播的路由器地址，之后就只会将组播转发到相应的地址中。如下图：



通过将 R1 的帧中继主接口配置成 NBMA Mode 后,R1 将在接口上跟踪 join 消息的源地址,所以可以得知只有 R3 需要接收组播,在需要转发组播时,只会转发给 R3。NBMA Mode 只能用在 PIM-SM 环境下,在 PIM-DM 下是不可用的。

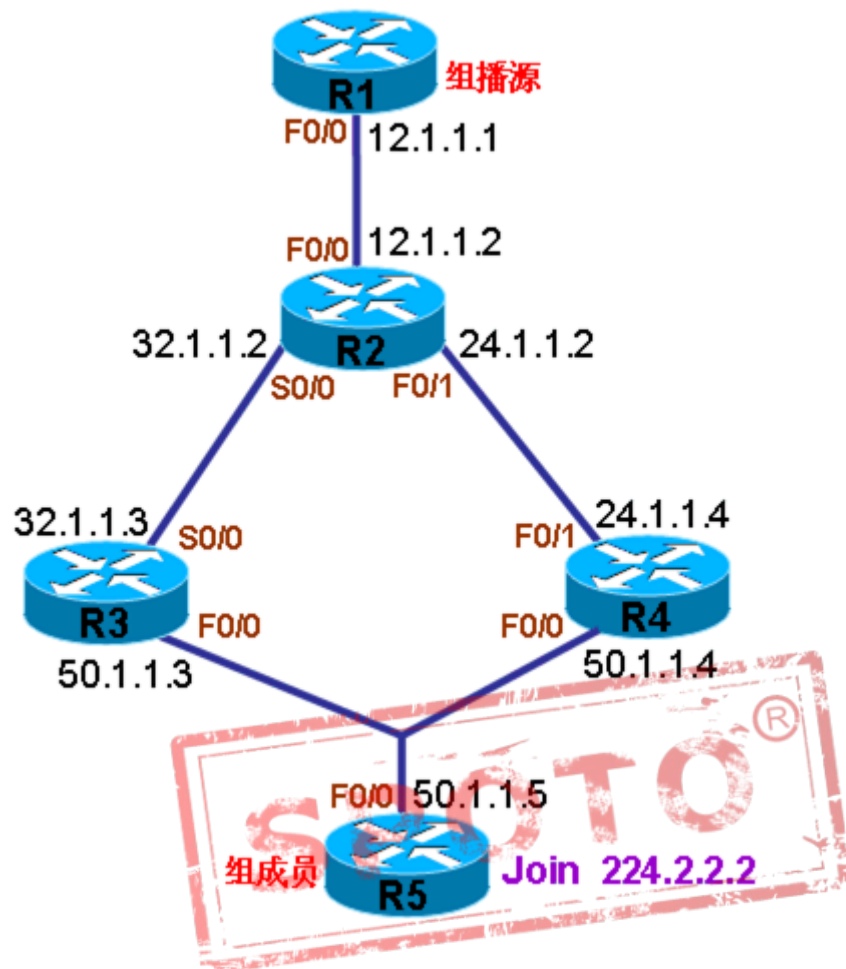
配置组播

默认情况下, Cisco 路由器的组播路由功能是关闭的,所以在配置组播之前,需要在路由器上全局输入 `ip multicast-routing` 激活组播路由功能。

在配置组播之前,必须保证全网的单播路由是通的。

配置 PIM-DM

说明: 以下图为例,其中 R1 为组播源, R5 为组成员,接收发往组 224.2.2.2 的数据。



1 配置单播（此步略）

说明：全网配置 OSPF 来完成单播通信。

2 开启组播路由功能

（1）在每台路由器上开启组播路由功能

R1:

```
R1(config)#ip multicast-routing
```

R2:

R2(config)#ip multicast-routing

R3:

R3(config)#ip multicast-routing

R4:

R4(config)#ip multicast-routing

R5:

R5(config)#ip multicast-routing



3 在接口上开启 PIM Dense-Mode

说明：在全网所有路由器的所有接口上开启 PIM，从而建立 PIM 邻居，要在所有接口开启，是因为要让 PIM 自己决定该在什么接口上转发组播，这样可以避免 RPF 检测失败，因为有可能只开部分接口，将导致接口不符合 RPF 接口而组播失败。

(1) 在所有路由器的每个接口上开启 PIM：（只举例一个接口，其它接口配置相同）

R1(config)#int f0/0

R1(config-if)# ip pim dense-mode

4 查看 PIM 邻居

说明：相邻的两台路由器，接口上开启 PIM 后，将成为 PIM 邻居

(1) 查看 R1 的 PIM 邻居

CCIE LAB认证经验分享千人群：539730342

```
r1#show ip pim neighbor
```

PIM Neighbor Table

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
12.1.1.2	FastEthernet0/0	00:01:08/00:01:35	v2	1 / DR S

```
r1#
```

说明：R1 可以看到的 PIM 邻居有 R2,结果正常。因为双方的 DR 优先级相同，而 R2 的 IP 地址大于 R1，所以 R1 看到 R2 为 DR。

(2) 查看 R2 的 PIM 邻居

```
r2#show ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
12.1.1.1	FastEthernet0/0	00:00:55/00:01:19	v2	1 / S
32.1.1.3	Serial0/0	00:01:04/00:01:40	v2	1 / DR S
24.1.1.4	FastEthernet0/1	00:00:52/00:01:22	v2	1 / DR S

```
r2#
```

说明：R2 可以看到的 PIM 邻居有 R1，R3，R4，结果正常。

(3) 查看 R3 的 PIM 邻居

```
r3#show ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
32.1.1.2	Serial0/0	00:01:28/00:01:44	v2	1 / S
50.1.1.5	FastEthernet0/0	00:01:11/00:01:31	v2	1 / DR S
50.1.1.4	FastEthernet0/0	00:01:19/00:01:32	v2	1 / S

r3#

说明：R3 可以看到的 PIM 邻居有 R2，R4，R5，结果正常。

(4) 查看 R4 的 PIM 邻居

```
r4#show ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
50.1.1.3	FastEthernet0/0	00:01:28/00:01:15	v2	1 / S

50.1.1.5 FastEthernet0/0 00:01:28/00:01:44 v2 1 / DR S

24.1.1.2 FastEthernet0/1 00:01:33/00:01:40 v2 1 / S

r4#

说明：R4 可以看到的 PIM 邻居有 R2，R3，R5，结果正常。

(5) 查看 R5 的 PIM 邻居

r5#show ip pim neighbor

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
50.1.1.3	FastEthernet0/0	00:01:20/00:01:24	v2	1 / S
50.1.1.4	FastEthernet0/0	00:01:49/00:01:24	v2	1 / S

r5#

说明：R4 可以看到的 PIM 邻居有 R3，R4，结果正常。

5 将 R5 加入组 224.2.2.2

说明：组成员要加入组，使用的协议为 IGMP，而 Cisco 路由器开启 PIM 后，接口上自动开启 IGMP，所以 R5 在加入组时，无需额外开启 IGMP 而直接加入组 224.2.2.2

(1) 配置 R5 加入组 224.2.2.2:

r5(config)#interface f0/0

```
r5(config-if)#ip igmp join-group 224.2.2.2
```

6 测试组播通信

说明：在没有组播数据之前，也就无法确认组播源的位置，在没有组播源的情况下，组播树是不能建立的。在测试 R1 向组 224.2.2.2 发送组播流量时，只需要 ping 224.2.2.2，只要数据包的目标地址为组播地址，就是使用组播方式来传递的。在 R1 上 ping 组播，那么 R1 就是组播源。

(1) 从组播源 R1 上 ping 224.2.2.2

```
r1#ping 224.2.2.2
```

Type escape sequence to abort.

Sending 1, 100-byte ICMP Echos to 224.2.2.2, timeout is 2 seconds:

Reply to request 0 from 50.1.1.5, 8 ms

```
r1#
```

说明：从回包中看出，已经收到组成员 50.1.1.5 的回包，说明组播成功被转发到组成员 R5。

7 查看 IGMP 查询器

说明：因为 R3 和 R4 同时连接着同网段的组成员 R5，所以为了避免重复查询，会选举出一个 IGMP 查询器，网络中 IP 地址最小的路由器将成为查询器。

(1) 查看网络中的 IGMP 查询器

```
r5#show ip igmp interface f0/0
```

CCIE LAB认证经验分享千人群：539730342

FastEthernet0/0 is up, line protocol is up

Internet address is 50.1.1.5/24

IGMP is enabled on interface

Current IGMP host version is 2

Current IGMP router version is 2

IGMP query interval is 60 seconds

IGMP querier timeout is 120 seconds

IGMP max query response time is 10 seconds

Last member query count is 2

Last member query response interval is 1000 ms

Inbound IGMP access group is not set

IGMP activity: 2 joins, 0 leaves

Multicast routing is enabled on interface

Multicast TTL threshold is 0

Multicast designated router (DR) is 50.1.1.5 (this system)

IGMP querying router is 50.1.1.3

Multicast groups joined by this system (number of users):

224.0.1.40(1) 224.2.2.2(1)

r5#

说明：从结果中可以看出，R3 在网络中拥有最小地址 50.1.1.3，所以被选为 IGMP 查询器。

8 查看 PIM 前转器

说明：因为从组播源 R1 到组成员 R5 的路径中，有两条可选择，即经过 R3 或者经过 R4，被选择的路由器就是 PIM 前转器，选择的规则为比较双方到组播源 R1 的单播 AD 值，其次是 metric 值，最后是 IP 地址最大的。

(1) 查看 R3 和 R4 到组播源的单播路由表

R3:

```
r3#show ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

50.0.0.0/24 is subnetted, 1 subnets

C 50.1.1.0 is directly connected, FastEthernet0/0

32.0.0.0/24 is subnetted, 1 subnets

C 32.1.1.0 is directly connected, Serial0/0

24.0.0.0/24 is subnetted, 1 subnets

CCIE LAB认证经验分享千人群：539730342

O 24.1.1.0 [110/2] via 50.1.1.4, 00:13:41, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/3] via 50.1.1.4, 00:13:41, FastEthernet0/0

r3#

R4:

r4#show ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

50.0.0.0/24 is subnetted, 1 subnets

C 50.1.1.0 is directly connected, FastEthernet0/0

32.0.0.0/24 is subnetted, 1 subnets

O 32.1.1.0 [110/65] via 50.1.1.3, 00:14:35, FastEthernet0/0

[110/65] via 24.1.1.2, 00:14:35, FastEthernet0/1

24.0.0.0/24 is subnetted, 1 subnets

C 24.1.1.0 is directly connected, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/2] via 24.1.1.2, 00:14:35, FastEthernet0/1

r4#

说明：在 AD 值相同的情况下，因为 R4 到组播源的 metric 值为 2，而 R3 到源的 metric 值为 3，所以 R4 将被选为 PIM 前转器，最后结果将在组树路由表中体现出来。

9 分析组播树

说明：在组播路由表的输出中，将尽量只保留有用的信息给大家，其它信息将被抑制。

(1) 查看 R1 的组播树情况

说明：查看组播树，即查看组播路由表

r1#show ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report, Z - Multicast Tunnel

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.0.1.40), 00:07:14/00:02:15, RP 0.0.0.0, flags: DCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Dense, 00:07:14/00:00:00

r1#

说明：在组播源上，不会有任何组的记录信息。



(2) 查看 R2 的组播树

说明：在 R2 上，即可看出 R3 和 R4 谁才是 PIM 前转器，前转器将成为组的出口。

r2#show ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

CCIE LAB认证经验分享千人群：539730342

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.2.2.2), 00:06:18/stopped, RP 0.0.0.0, flags: D

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:06:18/00:00:00

Serial0/0, Forward/Dense, 00:06:18/00:00:00

FastEthernet0/0, Forward/Dense, 00:06:20/00:00:00

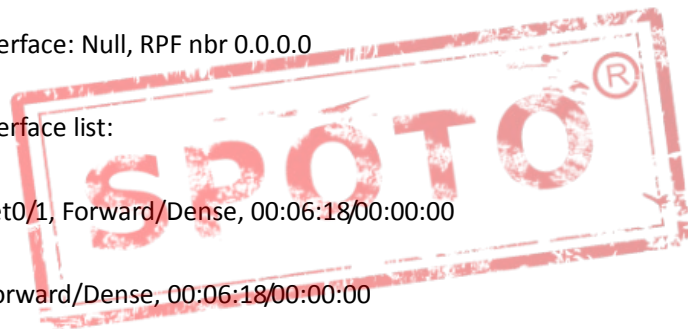
(12.1.1.1, 224.2.2.2), 00:01:57/00:02:56, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 0.0.0.0

Outgoing interface list:

Serial0/0, Prune/Dense, 00:01:53/00:01:06, A

FastEthernet0/1, Forward/Dense, 00:01:57/00:00:00



(*, 224.0.1.40), 00:12:45/00:02:50, RP 0.0.0.0, flags: DCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:12:12/00:00:00

Serial0/0, Forward/Dense, 00:12:25/00:00:00

FastEthernet0/0, Forward/Dense, 00:12:45/00:00:00

r2#

说明：可以看出，PIM-DM 模式中同样会有（*, G）的记录，而所有正常的 PIM 接口都为输出接口，输入接口为空。在（S, G）中可以看出，进口为 F0/0，所以此接口就是组播源所在的接口，因此就是 RPF 接口，而 PIM 邻居为 0.0.0.0，表示对方邻居就是组播源。而在出口中，可以看到有两个出口，分别为 S0/0 和 F0/1，但是由于 R4 被选为 PIM 前转器，所以连 R4 的接口 F0/0 的状态就是 Forward 状态，而 S0/0 则被剪除了。

（3）查看 R4 的组播树

r4#show ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

CCIE LAB认证经验分享千人群：539730342

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.2.2.2), 00:16:36/stopped, RP 0.0.0.0, flags: DC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:16:36/00:00:00

FastEthernet0/0, Forward/Dense, 00:16:36/00:00:00

(12.1.1.1, 224.2.2.2), 00:00:12/00:02:52, flags: T

Incoming interface: FastEthernet0/1, RPF nbr 24.1.1.2

Outgoing interface list:

FastEthernet0/0, Forward/Dense, 00:00:12/00:00:00

(* , 224.0.1.40), 00:19:08/00:02:01, RP 0.0.0.0, flags: DCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:19:04/00:00:00



CCIE LAB认证经验分享千人群：539730342

FastEthernet0/0, Forward/Dense, 00:19:08/00:00:00

r4#

说明：从输出中看出，（*，G）的内容同 R3，而（S，G）中显示 R4 的 F0/1 为进口，而连着组成员的接口 F0/0 被标为 Forward，因此结果正常。

（4）查看 R2 的组播树

r3#show ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.2.2.2), 00:17:44/stopped, RP 0.0.0.0, flags: DC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Dense, 00:17:44/00:00:00

Serial0/0, Forward/Dense, 00:17:44/00:00:00

(12.1.1.1, 224.2.2.2), 00:01:19/00:01:46, flags: PT

Incoming interface: FastEthernet0/0, RPF nbr 50.1.1.4

Outgoing interface list:

Serial0/0, Prune/Dense, 00:01:20/00:01:39

(*, 224.0.1.40), 00:20:25/00:02:57, RP 0.0.0.0, flags: DCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Dense, 00:20:15/00:00:00

Serial0/0, Forward/Dense, 00:20:25/00:00:00

r3#

说明：(*, G) 的内容同其它路由器，但是(S,G)中显示没有出口是 Forward，因此 R3 不提供组播转发，在网络中属正常。

10 改变 PIM 前转器

说明：因为 R3 和 R4 到组播源 R1 的单播路由表中，R4 拥有最小 metric 值，所以被选为 PIM 前转器，所有从组播源到组成员的流量均从 R4 被转发，要想控制从 R3 被

转发，则可通过调整 R3 到组播源的单播 metric 值，只要比 R4 到组播源的 metric 值小就行。

(1) 改变 R3 接口 S0/0 的 cost 值

```
r3(config)#int s0/0
```

```
r3(config-if)#ip ospf cost 1
```

(2) 查看 R3 到组播源 R1 的 metric 值

```
r3#show ip route 12.1.1.0
```

Routing entry for 12.1.1.0/24

Known via "ospf 2", distance 110, metric 2, type intra area

Last update from 32.1.1.2 on Serial0/0, 00:01:37 ago

Routing Descriptor Blocks:

* 32.1.1.2, from 2.2.2.2, 00:01:37 ago, via Serial0/0

Route metric is 2, traffic share count is 1

```
r3#
```

说明： R3 到组播源 R1 的 metric 值为 2，同 R4 到组播源 R1 的 metric 值相同。

(3) 查看 PIM 前转器结果：

```
r2#sh ip mroute
```

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

CCIE LAB认证经验分享千人群：539730342

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.2.2.2), 00:13:59/stopped, RP 0.0.0.0, flags: D

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:13:59/00:00:00

Serial0/0, Forward/Dense, 00:13:59/00:00:00

FastEthernet0/0, Forward/Dense, 00:14:00/00:00:00

(12.1.1.1, 224.2.2.2), 00:00:59/00:02:54, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 0.0.0.0

Outgoing interface list:

Serial0/0, Prune/Dense, 00:01:00/00:01:59

FastEthernet0/1, Forward/Dense, 00:01:00/00:00:00

(*, 224.0.1.40), 00:33:34/00:02:34, RP 0.0.0.0, flags: DCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:33:02/00:00:00

Serial0/0, Forward/Dense, 00:33:16/00:00:00

FastEthernet0/0, Forward/Dense, 00:33:36/00:00:00

r2#

说明：因为 R3 和 R4 到组播源 R1 的 metric 值相同，所以由最大 IP 地址的路由器成为 PIM 前转器，而 R3 和 R4 建 PIM 邻居的地址分别为 50.1.1.3 和 50.1.1.4，因为 R4 的地址比 R3 大，所以前转器仍然是路由器 R4。

(4) 改大 R3 的 IP 地址

说明：只要将 R3 和 R4 建 PIM 邻居的地址改大后，方可成为 PIM 前转器

r3(config)#int f0/0

r3(config-if)#ip address 50.1.1.10 255.255.255.0

(5) 查看 PIM 前转器结果：

r2#sh ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

CCIE LAB认证经验分享千人群：539730342

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

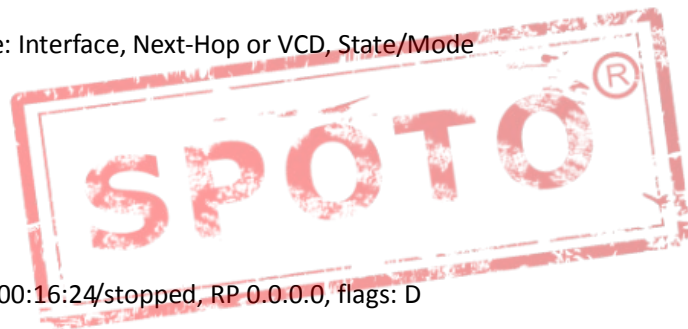
Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode



(* , 224.2.2.2), 00:16:24/stopped, RP 0.0.0.0, flags: D

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:16:24/00:00:00

Serial0/0, Forward/Dense, 00:16:24/00:00:00

FastEthernet0/0, Forward/Dense, 00:16:24/00:00:00

(12.1.1.1, 224.2.2.2), 00:03:23/00:00:31, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 0.0.0.0

Outgoing interface list:

Serial0/0, Forward/Dense, 00:01:43/00:00:00

FastEthernet0/1, Prune/Dense, 00:00:25/00:02:34

FastEthernet0/1, Forward/Dense, 00:35:38/00:00:00

(*, 224.0.1.40), 00:35:56/00:02:39, RP 0.0.0.0, flags: DCL

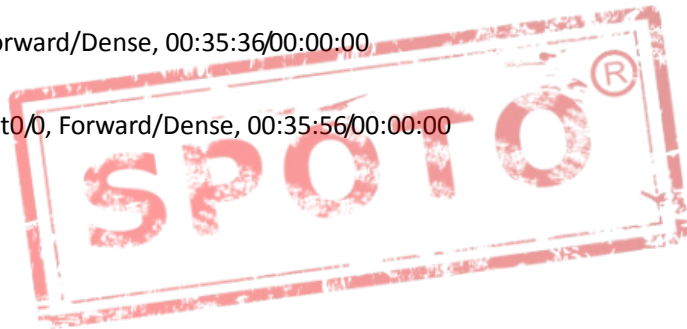
Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Dense, 00:35:23/00:00:00

Serial0/0, Forward/Dense, 00:35:36/00:00:00

FastEthernet0/0, Forward/Dense, 00:35:56/00:00:00



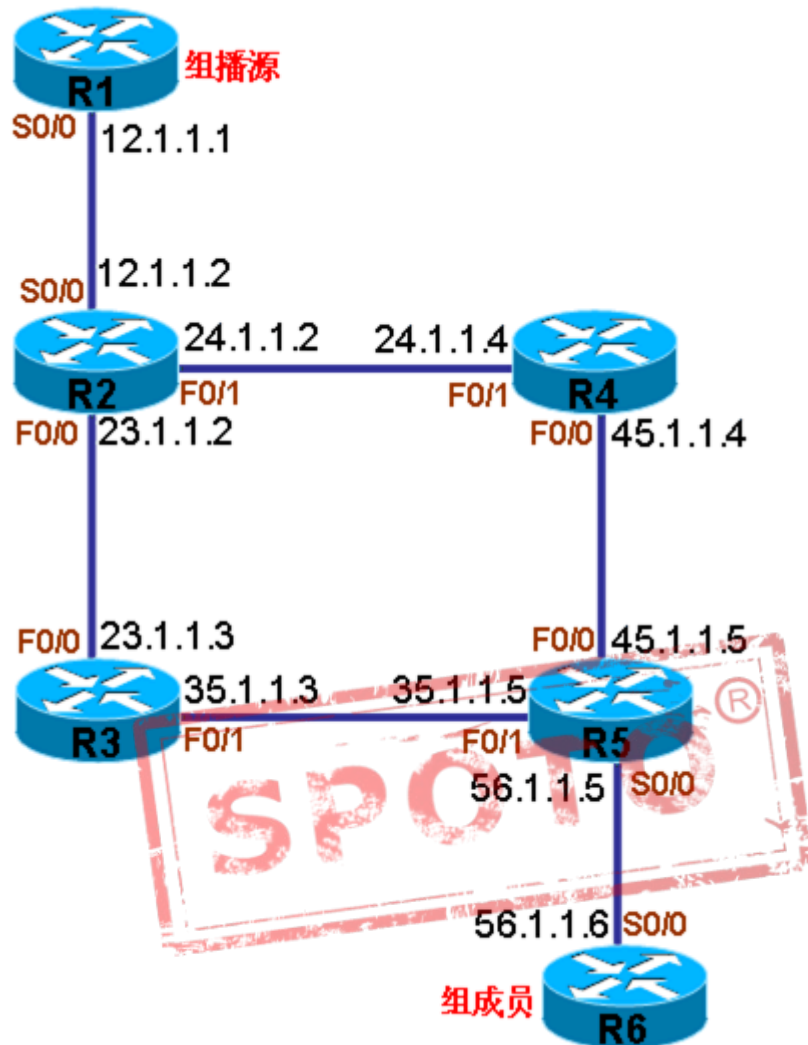
r2#

说明：因为 R3 的地址已经改为 50.1.1.10，而 R4 的地址为 50.1.1.4，所以在 AD 和 metric 值相同的情况下，R3 的 IP 地址最大，所以被选为 PIM 前转器。

配置 PIM-SM

说明：在下图中，R1 为组播源，R6 为组成员，接收发往组 224.6.6.6 的数据。

每台路由器均配有 loopback 地址，分别为 R1：1.1.1.1，R2：2.2.2.2，R3：3.3.3.3，R4：4.4.4.4，R5：5.5.5.5，R6：6.6.6.6，因为配置 PIM-SM 模式，所以网络中必须存在 RP。



1 配置单播（此步略）

说明：全网配置 OSPF 来完成单播通信。

2 开启组播路由功能

（1）在每台路由器上开启组播路由功能

R1:

```
R1(config)#ip multicast-routing
```

R2:

```
R2(config)#ip multicast-routing
```

R3:

```
R3(config)#ip multicast-routing
```

R4:

```
R4(config)#ip multicast-routing
```

R5:

```
R5(config)#ip multicast-routing
```

R6:

```
R6(config)#ip multicast-routing
```



3 在接口上开启 PIM Sparse-Mode，并建立 PIM 邻居。

说明：在全网所有路由器的所有接口上开启 PIM，从而建立 PIM 邻居，要在所有接口开启，是因为要让 PIM 自己决定该在什么接口上转发组播，这样可以避免 RPF 检测失败，因为有可能只开部分接口，将导致接口不符合 RPF 接口而组播失败。

(1) 在所有路由器的每个接口上开启 PIM：（只举例一个接口，其它接口配置相同）

```
R1(config)#int f0/0
```

CCIE LAB认证经验分享千人群：539730342

```
R1(config-if)# ip pim sparse-mode
```

4 将 R6 加入组 224.6.6.6

(1) 配置 R6 加入组 224.6.6.6

```
r6(config)#int s0/0
```

```
r6(config-if)#ip igmp join-group 224.6.6.6
```

5 静态配置 R4 为 RP

说明：将 R4 的 loopback 地址 4.4.4.4 配置为 RP，静态配置 RP 时，RP 地址的接口可以不用开启 PIM。RP 的配置方法有多种，此处首先采用手工静态配置方法，并且需要在每一台路由器上手工配置静态 RP。

(1) 在 R1 上静态配置 R4 为 RP （每台路由器的配置同 R1）

```
R1(config)#ip pim rp-address 4.4.4.4
```

说明：配置后面不跟 ACL 限制，默认为所有组的 RP。

(2) 查看 PIM 路由器的 RP 情况 （所有路由器的结果都应该相同）

```
r1#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

Group(s): 224.0.0.0/4, Static

RP: 4.4.4.4 (?)

```
r1#
```


说明：R1 已经正常看到所有组的 RP 为 4.4.4.4。

6 查看 RP 到组成员的组播树情况

说明：因为 PIM-SM 模式中，组成员会主动加入组，所以 RP 到组成员之间在没有源的情况下，也会形成组播树。

（1）在 RP：R4 上查看组播路由表：

```
r4#sh ip mroute
```

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 00:02:16/00:03:07, RP 4.4.4.4, flags: S

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

CCIE LAB认证经验分享千人群：539730342

FastEthernet0/0, Forward/Sparse, 00:02:16/00:03:07

(*, 224.0.1.40), 00:06:13/00:03:06, RP 4.4.4.4, flags: SJCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:01:35/00:02:53

FastEthernet0/0, Forward/Sparse, 00:02:22/00:03:05

r4#

说明：可以看到，RP 到组成员方向的接口已经被标为 Forward 状态，所以 RP 在有组播流量的情况下，是会向组成员发送的。

(2) 查看 R5 上的组播路由表：

r5#sh ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 00:05:36/00:02:45, RP 4.4.4.4, flags: SJC

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:02:41/00:02:45

(* , 224.0.1.40), 00:06:26/00:02:47, RP 4.4.4.4, flags: SJCL

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:01:55/00:02:33

Serial0/0, Forward/Sparse, 00:02:42/00:02:45

r5#

说明：R5 上连 RP 的接口 F0/0 就是组播的进口，而连组成员 R6 的接口 S0/0 已经是出口，状态为 Forward，所以正常。

注：从 PIM-SM 的组播路由表中看出，只记录 Forward 状态的出口，被剪除的出口是不做记录的，而 PIM-DM 会将所有出口记录下来，其中包括被剪除的接口。

6 测试组播通信

CCIE LAB认证经验分享千人群：539730342

(1) 从组播源 R1 上 ping 224.6.6.6

```
r1#ping 224.6.6.6
```

Type escape sequence to abort.

Sending 1, 100-byte ICMP Echos to 224.6.6.6, timeout is 2 seconds:

Reply to request 0 from 56.1.1.6, 128 ms

```
r1#
```

说明：从回包中看出，已经收到组成员 56.1.1.6 的回包，说明组播成功被转发到组成员 R6。



7 查看组播树的情况：

(1) 查看 R1 的组播树情况

```
r1#show ip mroute
```

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report, Z - Multicast Tunnel

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.0.1.40), 00:10:04/00:02:52, RP 4.4.4.4, flags: SJPL

Incoming interface: Serial0/0, RPF nbr 12.1.1.2

Outgoing interface list: Null

r1#

说明：在组播源上，不会有任何组的记录信息。



(2) 查看 R2 的组播树

r2#sh ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

CCIE LAB认证经验分享千人群：539730342

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 00:02:05/stopped, RP 4.4.4.4, flags: SPF

Incoming interface: FastEthernet0/1, RPF nbr 24.1.1.4

Outgoing interface list: Null

(12.1.1.1, 224.6.6.6), 00:02:05/00:03:24, flags: FT

Incoming interface: Serial0/0, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:02:05/00:03:22

(* , 224.0.1.40), 00:10:33/00:02:25, RP 4.4.4.4, flags: SJCL

Incoming interface: FastEthernet0/1, RPF nbr 24.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:10:33/00:01:59

r2#

说明：可以看出，（*，G）中说明了 RP 是 4.4.4.4，但是组播源到 RP 之间会创建（S，G）的记录，并且通往 RP 的路径正常。RPF 邻居为 0.0.0.0，也正常。

(3) 再次查看 RP 上的组播树

r4#sh ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 00:07:39/00:02:39, RP 4.4.4.4, flags: S

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:07:39/00:02:39

(12.1.1.1, 224.6.6.6), 00:03:35/00:02:01, flags: T

Incoming interface: FastEthernet0/1, RPF nbr 24.1.1.2

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:03:35/00:02:50

(*, 224.0.1.40), 00:11:37/00:02:58, RP 4.4.4.4, flags: SJCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:07:00/00:03:22

FastEthernet0/0, Forward/Sparse, 00:07:45/00:02:38

r4#

说明：RP 上也有了（S，G）的记录，并且接口状态都是正常。



（4）再次查看 R5 上的组播树

r5#sh ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

CCIE LAB认证经验分享千人群：539730342

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 00:11:55/00:03:21, RP 4.4.4.4, flags: SJC

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:09:00/00:03:21

(12.1.1.1, 224.6.6.6), 00:04:55/00:02:11, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:04:55/00:03:21

(* , 224.0.1.40), 00:12:47/00:03:21, RP 4.4.4.4, flags: SJCL

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:08:16/00:03:08

Serial0/0, Forward/Sparse, 00:09:01/00:03:21



r5#

说明：同样也有（S，G）的记录，方面后面从 RPT 切换到 SPT，正常。

（5）再次查看 R3 上的组播树

r3#sh ip mroute

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.0.1.40), 00:15:23/00:02:22, RP 4.4.4.4, flags: SJCL

Incoming interface: FastEthernet0/1, RPF nbr 35.1.1.5

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:15:23/00:02:22

r3#

说明：因为 R3 不需要转发组播，所以没有任何记录。

注：IGMP 查询器，PIM 前转器不再查看和控制，方法同 PIM-DM 模式。

配置 Auto-RP

说明：以上配置为静态 RP 的方式，下面配置 Auto-RP

1 配置 C-RP

说明：配置为 RP 的接口必须开启 PIM。

(1) 配置 R3 为 C-RP:

r3(config)#access-list 24 permit 224.5.5.5

r3(config)#access-list 24 permit 224.6.6.6

r3 (config)#int loopback 0

r3 (config-if)# ip pim sparse-mode

r3(config)#ip pim send-rp-announce loopback 0 scope 10 group-list 24

说明：配置 R3 的 loopback 0 为 RP 地址，并且使用 ACL 限制只做组 224.5.5.5 和 224.6.6.6 的 RP。

(2) 配置 R4 为 C-RP:

```
R4(config)#access-list 24 permit 224.5.5.5
```

```
R4(config)#access-list 24 permit 224.6.6.6
```

```
R4 (config)#int loopback 0
```

```
R4 (config-if)# ip pim sparse-mode
```

```
R4(config)#ip pim send-rp-announce loopback 0 scope 10 group-list 24
```

说明：配置 R4 的 loopback 0 为 RP 地址，并且使用 ACL 限制只做组 224.5.5.5 和 224.6.6.6 的 RP。

2 配置 RP-mapping agent

(1) 配置 R1 为 RP-mapping agent

```
R1 (config)#int loopback 0
```

```
R1 (config-if)# ip pim sparse-mode
```

```
R1(config)#ip pim send-rp-discovery loopback 0 scope 16
```

说明：配置 R1 的 loopback 0 为 RP-mapping agent。

3 查看 RP 信息

说明：所有 C-RP 的竞选消息都发送到 224.0.1.39，由 RP-mapping agent 接收后将选为 RP 的地址公布给所有 PIM 路由器，要让 RP-mapping agent 成功收到 C-RP 的竞选消息，就必须能够接收组播到 224.0.1.39 的数据。

(1) 查看 RP-mapping agent 的组加入情况

```
r1#show ip igmp interface
```

Serial0/0 is up, line protocol is up

Internet address is 12.1.1.1/24

IGMP is enabled on interface

Current IGMP host version is 2

Current IGMP router version is 2

IGMP query interval is 60 seconds

IGMP querier timeout is 120 seconds

IGMP max query response time is 10 seconds

Last member query count is 2

Last member query response interval is 1000 ms

Inbound IGMP access group is not set

IGMP activity: 2 joins, 0 leaves

Multicast routing is enabled on interface

Multicast TTL threshold is 0

Multicast designated router (DR) is 12.1.1.2

IGMP querying router is 12.1.1.1 (this system)

Multicast groups joined by this system (number of users):

224.0.1.39(1)

Loopback0 is up, line protocol is up

CCIE LAB认证经验分享千人群：539730342

Internet address is 1.1.1.1/24

IGMP is enabled on interface

Current IGMP host version is 2

Current IGMP router version is 2

IGMP query interval is 60 seconds

IGMP querier timeout is 120 seconds

IGMP max query response time is 10 seconds

Last member query count is 2

Last member query response interval is 1000 ms

Inbound IGMP access group is not set

IGMP activity: 2 joins, 0 leaves

Multicast routing is enabled on interface

Multicast TTL threshold is 0

Multicast designated router (DR) is 1.1.1.1 (this system)

IGMP querying router is 1.1.1.1 (this system)

Multicast groups joined by this system (number of users):

224.0.1.40(1) 224.0.1.39(1)

r1#

说明：可以看出，配置成为 RP-mapping agent 后，所有的接口都已加入了 224.0.1.39，但是这并不能保证 RP-mapping agent 就一定能够收到 C-RP 的竞选消息，因为 PIM-SM 模式下，没有 RP 时，组播是不通的。

4 解决 RP-mapping agent 接收 C-RP 竞选消息

说明：因为组播不通，所以当 RP-mapping agent 和 C-RP 不是同一台路由器时，无法接收 C-RP 的竞选消息，可以通过在路由器上配置 `Autorp listener` 来解决，网络中最好所有路由器都配置。

(1) 配置 `Autorp listener`（此处只举例配置 R1，所有路由器配置相同）

```
R1 (config)#ip pim autorp listener
```

注：解决方法还有就是将 PIM 的模式都改为 `sparse-dense-mode`，此模式配置省略。

5 查看 RP 情况

(1) 查看 R1 的 RP 情况：

```
r1#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

This system is an RP-mapping agent (Loopback0)

Group(s) 224.5.5.5/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:05:49, expires: 00:02:11

RP 3.3.3.3 (?), v2v1

Info source: 3.3.3.3 (?), via Auto-RP

Uptime: 00:05:14, expires: 00:02:43



Group(s) 224.6.6.6/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:05:49, expires: 00:02:10

RP 3.3.3.3 (?), v2v1

Info source: 3.3.3.3 (?), via Auto-RP

Uptime: 00:05:14, expires: 00:02:45

r1#

说明：在 RP-mapping agent 上，同时看到所有信息，因为 R4 的地址最大，所以被选举为两个组的 RP。

(2) 查看 R5 的 RP 情况：

r5#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.5.5.5/32

RP 4.4.4.4 (?), v2v1

Info source: 1.1.1.1 (?), elected via Auto-RP

Uptime: 00:01:25, expires: 00:02:31

Group(s) 224.6.6.6/32

RP 4.4.4.4 (?), v2v1

Info source: 1.1.1.1 (?), elected via Auto-RP

Uptime: 00:01:25, expires: 00:02:32

r5#

说明：R5 正常学习到 RP 信息，其它路由器都应该正常学习到 RP 信息。

6 测试组播通信

(1) 测试 R1 到各组的通信情况：

r1#ping 224.5.5.5

Type escape sequence to abort.

Sending 1, 100-byte ICMP Echos to 224.5.5.5, timeout is 2 seconds:

Reply to request 0 from 56.1.1.6, 129 ms

Reply to request 0 from 56.1.1.6, 165 ms

r1#ping 224.6.6.6

Type escape sequence to abort.

Sending 1, 100-byte ICMP Echos to 224.6.6.6, timeout is 2 seconds:

Reply to request 0 from 56.1.1.6, 124 ms

Reply to request 0 from 56.1.1.6, 152 ms

r1#

说明：R1 到组 224.5.5.5 和 224.6.6.6 通信正常。

7 限制 RP 竞选消息

说明：在 C-RP 上可以控制只竞选特定组的 RP，而在 RP-mapping agent 上，可以配置只接收某个 RP 竞选某特定组的消息，那么其它消息都将被过滤。

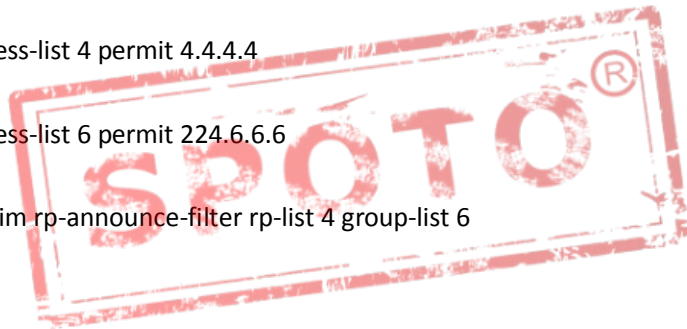
（1）配置 RP-mapping agent 限制 R4 的竞选消息

说明：配置 RP-mapping agent 只接收 R4 竞选 224.6.6.6 的消息，那么 R4 也就不可能竞选组 224.5.5.5 的 RP 了。

r1(config)#access-list 4 permit 4.4.4.4

r1(config)#access-list 6 permit 224.6.6.6

r1(config)#ip pim rp-announce-filter rp-list 4 group-list 6



8 查看 RP 情况

（1）在 R1 上查看 RP 信息：

r1#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP-mapping agent (Loopback0)

Group(s) 224.5.5.5/32

RP 3.3.3.3 (?), v2v1

Info source: 3.3.3.3 (?), elected via Auto-RP

CCIE LAB认证经验分享千人群：539730342

Uptime: 00:02:42, expires: 00:02:17

Group(s) 224.6.6.6/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:08:18, expires: 00:02:40

RP 3.3.3.3 (?), v2v1

Info source: 3.3.3.3 (?), via Auto-RP

Uptime: 00:07:43, expires: 00:02:17

r1#

说明：因为 R4 竞选组 224.5.5.5 的消息被过滤了，所以组 224.5.5.5 的 RP 只有 R3。

(2) 在 R2 上查看 RP 信息：

r2#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.5.5.5/32

RP 3.3.3.3 (?), v2v1

Info source: 1.1.1.1 (?), elected via Auto-RP

Uptime: 00:00:59, expires: 00:02:57

Group(s) 224.6.6.6/32

RP 4.4.4.4 (?), v2v1

CCIE LAB认证经验分享千人群：539730342

Info source: 1.1.1.1 (?), elected via Auto-RP

Uptime: 00:20:01, expires: 00:02:56

r2#

说明：R2 上的 RP 信息和预料中相同。

配置 BSR

说明：在 BSR 中，在不修改 IP 地址的情况下，可以通过修改优先级来控制竞选。

1 配置 C-RP

(1) 配置 R3 为 C-RP，优先级为 10（数字越大，优先级越高）

R3(config)#access-list 24 permit 224.5.5.5

R3(config)#access-list 24 permit 224.6.6.6

R3(config)#int loopback 0

R3 (config-if)# ip pim sparse-mode

R3(config)#ip pim rp-candidate loopback 0 group-list 24 priority 10

(2) 配置 R4 为 C-RP，优先级默认为 0

R4(config)#access-list 24 permit 224.5.5.5

R4(config)#access-list 24 permit 224.6.6.6

```
R4(config)#int loopback 0
```

```
R4 (config-if)# ip pim sparse-mode
```

```
R4(config)#ip pim rp-candidate loopback 0 group-list 24
```

2 配置 C-BSR

(1) 配置 R1 为 C-BSR

```
R1(config)#int loopback 0
```

```
R1 (config-if)# ip pim sparse-mode
```

```
R1(config)#ip pim bsr-candidate loopback 0
```



3 查看 RP 情况

(1) 查看 BSR 情况

```
r1#sh ip pim bsr-router
```

PIMv2 Bootstrap information

This system is the Bootstrap Router (BSR)

BSR address: 1.1.1.1 (?)

Uptime: 00:01:45, BSR Priority: 0, Hash mask length: 0

CCIE LAB认证经验分享千人群：539730342

Next bootstrap message in 00:00:14

r1#

说明：因为 C-BSR 只有 R1，所以被选为 BSR 的也是 R1。

(2) 查看 RP 结果：

r1#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is the Bootstrap Router (v2)

Group(s) 224.5.5.5/32

RP 3.3.3.3 (?), v2

Info source: 23.1.1.3 (?), via bootstrap, priority 10

Uptime: 00:13:25, expires: 00:02:08

RP 4.4.4.4 (?), v2

Info source: 24.1.1.4 (?), via bootstrap, priority 20

Uptime: 00:09:43, expires: 00:02:00

Group(s) 224.6.6.6/32

RP 3.3.3.3 (?), v2

Info source: 23.1.1.3 (?), via bootstrap, priority 10

Uptime: 00:13:25, expires: 00:02:07

RP 4.4.4.4 (?), v2



Info source: 24.1.1.4 (?), via bootstrap, priority 20

Uptime: 00:12:43, expires: 00:01:58

r1#

说明：通过查看 RP，无法看出 BSR 选择了谁成为 RP。

(3) 查看其它路由器的 RP 结果：

r2#sh ip pim rp mapping in-use

PIM Group-to-RP Mappings

Group(s) 224.5.5.5/32

RP 3.3.3.3 (?), v2

Info source: 1.1.1.1 (?), via bootstrap, priority 10, holdtime 98

Uptime: 00:14:10, expires: 00:01:19

RP 4.4.4.4 (?), v2

Info source: 1.1.1.1 (?), via bootstrap, priority 20, holdtime 91

Uptime: 00:10:28, expires: 00:01:14

Group(s) 224.6.6.6/32

RP 3.3.3.3 (?), v2

Info source: 1.1.1.1 (?), via bootstrap, priority 10, holdtime 98

Uptime: 00:14:10, expires: 00:01:20

RP 4.4.4.4 (?), v2



Info source: 1.1.1.1 (?), via bootstrap, priority 20, holdtime 89

Uptime: 00:13:27, expires: 00:01:12

Dynamic (Auto-RP or BSR) RPs in cache that are in use:

Group(s): 224.6.6.6/32, RP: 3.3.3.3, expires: 00:00:51

Group(s): 224.5.5.5/32, RP: 3.3.3.3, expires: 00:00:53

r2#

说明：在 R2 上看出，因为 C-RP 的数字越 小，优先级越高，所以 RP 为 3.3.3.3。

PIM-SM 的 NBMA Mode

说明：因为帧中继中没有内置的处理广播和组播的能力，所以在收到广播和组播时，会在所有配了关键字 **broadcast** 的 PVC 上转发，这样当帧中继接口对端有多个邻居，而只有部分邻居要接收组播时，那么组播的被当作广播转发的行为，将影响不需要接收组播的路由器，因此需要配置 **NBMA** 模式来跟踪要接收组播的邻居，从而只是将组播发送给要接收的邻居。**NBMA** 模式只支持 **PIM-SM** 模式，因为是跟踪 **PIM-SM** 模式的 join 消息。帧中继下的解决方法除了 **NBMA Mode** 之外，还有配置子接口，子接口类似多个物理接口，会建立多个邻居，此处不对子接口作过多解释。

没配之前：

1.查看 R5 没有配置 NBMA Mode 的组播树

r5#sh ip mroute

IP Multicast Routing Table

CCIE LAB认证经验分享千人群：539730342

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 01:53:44/00:03:13, RP 3.3.3.3, flags: SCF

Incoming interface: FastEthernet0/1, RPF nbr 35.1.1.3

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:01:16/00:03:13

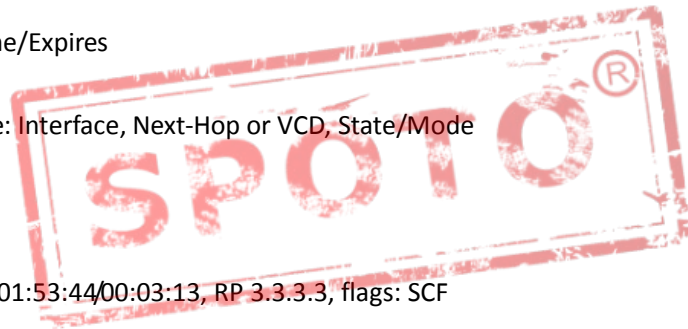
Serial0/0, Prune/Sparse, 00:01:17/00:00:00

(1.1.1.1, 224.6.6.6), 00:00:48/00:03:24, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:00:49/00:03:12



CCIE LAB认证经验分享千人群：539730342

(12.1.1.1, 224.6.6.6), 00:00:49/00:03:23, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:00:49/00:03:12

(* , 224.5.5.5), 01:16:35/00:03:00, RP 3.3.3.3, flags: SC

Incoming interface: FastEthernet0/1, RPF nbr 35.1.1.3

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:00:30/00:02:59

Serial0/0, Prune/Sparse, 00:01:21/00:00:00



(1.1.1.1, 224.5.5.5), 00:02:04/00:02:41, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, Forward/Sparse, 00:01:05/00:03:22

Serial0/0, Prune/Sparse, 00:01:22/00:00:00

(12.1.1.1, 224.5.5.5), 00:02:04/00:02:41, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

CCIE LAB认证经验分享千人群：539730342

Serial0/0, Forward/Sparse, 00:01:05/00:03:24

Serial0/0, Prune/Sparse, 00:01:22/00:00:00

(*, 224.0.1.40), 00:55:08/00:02:59, RP 0.0.0.0, flags: DCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Loopback0, Forward/Sparse, 00:55:09/00:00:00

Serial0/0, Forward/Sparse, 00:55:09/00:00:00

FastEthernet0/1, Forward/Sparse, 00:55:10/00:00:00

FastEthernet0/0, Forward/Sparse, 00:55:10/00:00:00

r5#



说明：从结果中看出，因为没有配置 NBMA Mode，所以 R5 将收到的组播从 S0/0 的每一条 PVC 发出去。（虽然本例中只配了一条 PVC）

2.配置 R5 连组成员 R6 的帧中继接口 S0/0 为 NBMA Mode(最好双方都配置)

```
r5(config)#int s0/0
```

```
r5(config-if)#ip pim nbma-mode
```

3.查看 R5 配置 NBMA Mode 之后的组播树

```
r5#sh ip mroute
```

IP Multicast Routing Table

CCIE LAB认证经验分享千人群：539730342

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.6.6.6), 01:54:34/00:03:22, RP 3.3.3.3, flags: SCF

Incoming interface: FastEthernet0/1, RPF nbr 35.1.1.3

Outgoing interface list:

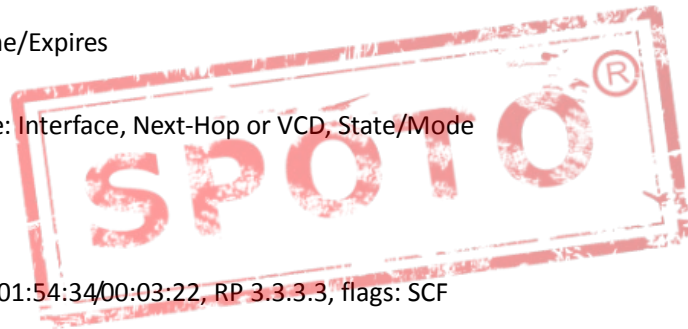
Serial0/0, 56.1.1.6, Forward/Sparse, 00:00:07/00:03:22

(1.1.1.1, 224.6.6.6), 00:01:38/00:02:34, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, 56.1.1.6, Forward/Sparse, 00:00:07/00:03:22



CCIE LAB认证经验分享千人群：539730342

(12.1.1.1, 224.6.6.6), 00:01:39/00:02:33, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, 56.1.1.6, Forward/Sparse, 00:00:09/00:03:20

(* , 224.5.5.5), 01:17:25/00:03:09, RP 3.3.3.3, flags: SC

Incoming interface: FastEthernet0/1, RPF nbr 35.1.1.3

Outgoing interface list:

Serial0/0, 56.1.1.6, Forward/Sparse, 00:00:20/00:03:09

(1.1.1.1, 224.5.5.5), 00:02:56/00:03:20, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, 56.1.1.6, Forward/Sparse, 00:00:26/00:03:29

(12.1.1.1, 224.5.5.5), 00:03:02/00:03:13, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 45.1.1.4

Outgoing interface list:

Serial0/0, 56.1.1.6, Forward/Sparse, 00:00:32/00:03:25

(* , 224.0.1.40), 00:56:07/00:02:55, RP 0.0.0.0, flags: DCL



Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Loopback0, Forward/Sparse, 00:56:07/00:00:00

Serial0/0, Forward/Sparse, 00:56:11/00:00:00

FastEthernet0/1, Forward/Sparse, 00:56:11/00:00:00

FastEthernet0/0, Forward/Sparse, 00:56:11/00:00:00

r5#

说明：R5 在帧中继接口上配置了 NBMA Mode，所以看到组播树中，明确显示组播将会发给特定的邻居（56.1.1.6），由于在组播树中记录了对端邻居的地址，所以在接口对端有多台路由器时，组播不会发给不想接收组播的路由器。

Source Specific Multicast (SSM)

概述

在组播树中，被记录为（*, G）的组条目表示，对于一个特定的组，任何主机都可以向该组发起组播流量，也就是说，一个组可以拥有多个组播源，任何组播源都可以发送组播流量，这样的组播被称为任意源组播 Any Source Multicast (ASM)。

因为一个组一般表示一个应用，如果网络中两个应用使用了同一个组地址，这样就会造成组成员将两个不同应用的流量误当作同一个应用来处理，就会造成数据的混乱或错误，所以当两个应用不小心使用了同一个组地址，这样会给应用带来问题。

如果一个组成员想要接收某个组的组播流量，可以通过 IGMP 向路由器报告，例如 IGMP ver1, IGMP ver2，报告中指出了组成员想要接受的组地址，当路由器收到 IGMP 报告之后，就会将发往相应组地址的流量转发到组成员。

正因为路由器会将任何组播源发到同一个组的流量转发给相同组成员，所以可能会造成多个应用使用同一个组地址时，不能只将组成员想要接收的流量发到组成员。如果要实现只将特定组播源发来的流量转发给相应的组成员，那么这样的组播机制，被称为特定源组播 **Source Specific Multicast(SSM)**。

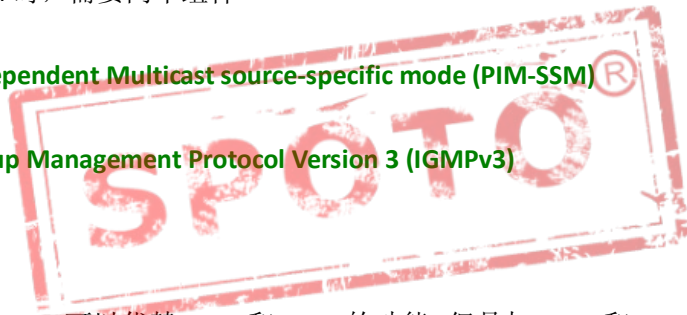
因为特定源组播（SSM）只将特定的组播源发来的流量，而不是任何源发来的流量转发给组成员，所以组成员在向路由器报告自己想要接受的组播流量时，除了明确指出组地址之外，还必须指出组播源地址，而这样的 IGMP 报告，需要 IGMP ver3 来支持。

并且可以想象，如果多个应用程序在同一个源，那么就要多个组，但如果多个应用在不同源，那么组地址就可以相同，也可以不同，因为 SSM 可以根据源地址区分出不同的应用程序。

在运行 SSM 时，需要两个组件

- **Protocol Independent Multicast source-specific mode (PIM-SSM)**

- **Internet Group Management Protocol Version 3 (IGMPv3)**



其中 IGMP ver3 可以代替 ver 1 和 ver 2 的功能，但是与 ver 1 和 ver 2 不同之处在于，IGMP ver 3 支持对源地址的过滤，IGMP ver 3 在报告中，会明确指出想要接收的组播源地址。

SSM 基于(S, G) 传输，但 SSM 也可以和其它组播树共存，只要配好自己的组地址范围即可，为 SSM 保留地址范围是：

232.0.0.0 - 232.255.255.255 （232.0.0.0/8）

但思科 IOS 可任意配地址范围。

虽然如此，要先有 PIM SM，才能有 SSM，但 SSM 也可以独立存在。

如果已经有了 PIM-SM，那么只有最后一跳路由器需要开启 SSM 即可，也就是

说只需要直接连接着组成员，直接接收组成员 IGMP 数据包的路由器需要开启 SSM。

注：SSM 不需要 RP，当最后一跳路由器开启 SSM 后，正常的 PIM-SM 就失去意义了。

配置 SSM

说明：只在直接连接着组成员，直接接收组成员 IGMP 数据包的路由器 上开启 SSM

1. 全局开启组播

(1) 全局开启组播路由功能

```
Router(config)#ip multicast-routing
```

2. 配置 SSM 组地址范围

(1) 配置默认的 SSM 组地址范围

```
Router(config)#ip pim ssm default
```

说明：配置此命令后，默认的 SSM 组地址范围为：232.0.0.0 – 232.255.255.255

(2) 配置 SSM 组地址为 232.1.1.1

```
Router(config)#access-list 1 permit 232.1.1.1
```

```
Router(config)#ip pim ssm range 1
```

说明：ACL 1 所匹配的地址即为 SSM 的组地址范围。

3. 在接口开启 SSM

(1) 在接口开启 PIM

```
Router(config-if)#ip pim sparse-mode
```

说明：必须在接口开启 sparse-mode 或 sparse-dense-mode

(2) 在接口开启 IGMP ver 3

```
Router(config-if)#ip igmp version 3
```

说明：默认为 IGMP ver 2.

附：SSM 查看命令

```
Router# show ip igmp groups
```

```
Router# show ip mroute
```

MSDP (Multicast Source Discovery Protocol)



PIM 概述

通过路由器组建的网络需要通信，前提是所有路由器上都必须拥有正确的路由表，对于单播是这样子，而如果需要使用组播通信；同样道理，所有路由器上都必须拥有正确的组播路由表，准确的说，是必须拥有正确的组播树，组播树从组播源到组成员（组播接收者）之间建立，组播源发出的数据沿着组播树流向组成员，从而保证组播通信正常。换句话说就是，如果没有组播树，或组播树不正常，那么组播是不能通信的。

组播路由器需要靠组播路由协议的配合来建立组播树，目前最为流行的组播路由协议为 Protocol Independent Multicast (PIM)，PIM 可工作在两种模式下，分别为 PIM dense mode (PIM-DM)和 PIM sparse mode (PIM-SM)，PIM-DM 就是一个纯傻瓜式的工作方式，它就像咱们平时见到的傻瓜式交换机一样，交换机打开之后不需要做任何配置就能用，并且谁都会用，PIM-DM 只要在相应的路由器的相应接口上开启之后，就算是大功告成，组播源只要一发数据，组播树铁定就能正确建立，组成员也一定能够收到数据，相当快捷方便；但遗憾的是 PIM-DM 虽然简洁，但缺乏控制力度，就相当于傻瓜式交换机虽然好用，但你无法在上面做太多的策略来控制流量，并且更遗憾的是在 CCIE 考试中，基本上不考你 PIM-DM。

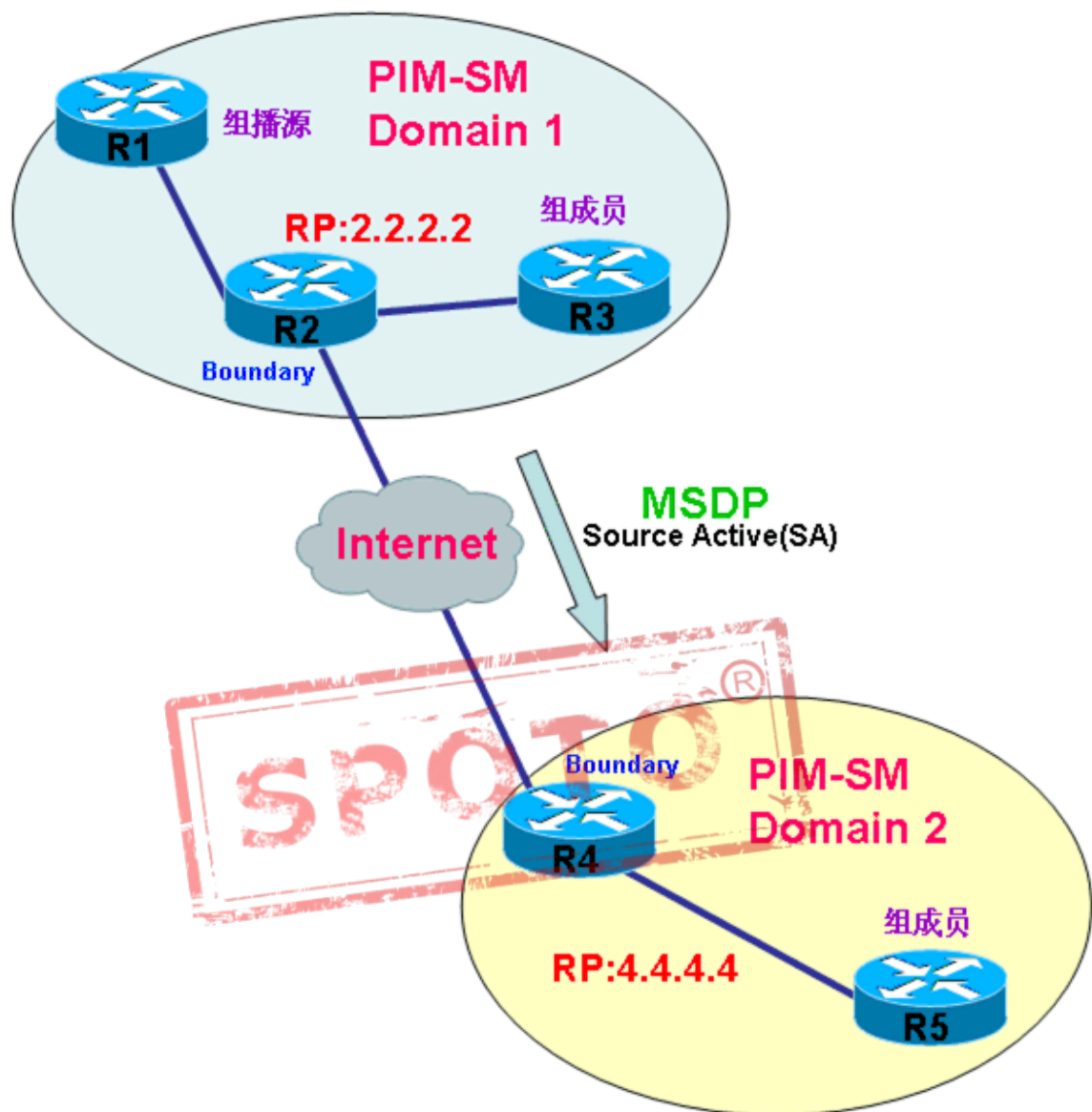
除 PIM-DM 之外，PIM-SM 是个重点，也是个难点，并且还是个考点，所以我们通常在解决组播问题的时候，可以说都是在解决 PIM-SM 的问题。在 PIM-SM 中，所有的流量都不是组播源直接发向组成员的，组播流量是由组播源发给 Rendezvous Point (RP)，然后 RP 再将流量转发给组成员，如果说没有了 RP，那么组播源也就不知道把流量该发给谁，而也就再也没有谁会给组成员发流量了，所以 RP 是否正常，直接关系到组播是否能通信。PIM 在建立组播树时，在 PIM-SM 中，这棵树是围绕 RP 来建立的，所以 RP 就是组播树的根，要完成组播树的建立，RP 必须要知道谁是组播源以及谁是组成员，这一点请牢记；那么 RP 是如何知道谁是组播源以及谁是组成员的呢？这一点也请牢牢地记住，过程是：组播源和组成员主动告诉 RP，因为 RP 是不会主动问它们的，事实上真正告诉 RP 的是离组播源最近的路由器以及离组成员最近的路由器，无论组播源和组成员本身是不是路由器都是这样，为什么要由路由器代为发起呢？因为通常组播员和组成员是一台 PC 或服务器，只是我们在做实验时，是用了路由器，而 PC 或服务器准确地讲是不会运行 PIM 协议的；组播源告诉 RP 自己是组播源的这个过程称为 register(注册)，而组成员告诉 RP 自己是组成员的这个过程称为 report(报告)，但也可以称为注册，Cisco 文档上也有时会把它们都称为是注册。

一个正常工作的组播网络中应该只有一个 RP，这样的只围绕一个 RP 建立组播树的组播网络被称为一个 PIM-SM 域。在某些情况下，比如一个大公司在全国各个城市都有分公司，而这个公司许多应用是靠组播来传递的，如视频会议，如果这个公司全国范围内都设计成一个 PIM-SM 域，那么就表示全国的分公司都需要使用一个共同的 RP，所以如果这个 RP 是放在北京分公司的，那么北京的这台 RP 路由器是不能关的，可能会面临着其它各个城市在开会时出现任何通信问题都往北京打电话询问 RP 路由器是否运行正常或已被关闭；对于各个分公司的 IT 人员来说也不希望如此，他们一定是希望 RP 路由器放在自己的分公司或者自己的分公司有自己独立可控的 RP 路由器，这样就不需要依靠于其它分公司的 RP 来通信，便于在出现通信故障时能够全面地排错。ISP 的某些业务也是通过组播来开展的，并且这些业务可能需要跨 ISP 来工作，可想而知，如果将所有 ISP 都规划成一个 PIM-SM 域，势必会造成某些 ISP 在网络控制上出现被动局面，对于 ISP 来说，它不可能希望自己的组播网络依赖竞争对手的 RP 路由器来维持通信，基于以上种种原因，需要在各环境下部署独立的 PIM-SM 域，即在各个分公司或各个 ISP 都建立各自独立的 PIM-SM 域，从而保证各个组播网络的独立性与自治性。但是请不要忘记，这些 PIM-SM 域之间还是要相互通信的。

MSDP 概述

在 PIM-SM 域之间要通信，是否会存在问题呢？答案是肯定的，究竟会存在什么样的问题呢？回想一下组播要正常通信的前提条件，那就是必须建立组播树，而组播树要正常建立，RP 就必须要知道组播源和组成员，如果是同一个 PIM-SM 域内，RP 想要知道所有的组播源和组成员是非常轻松的事，但 RP 却没有办法知道其它 PIM-SM 域中的组播源，那这个问题怎么解决呢？方法也很简单，那就是让一个 PIM-SM 域中的 RP 把自己知道的组播源信息告诉其它 PIM-SM 域中的 RP 就可以了，所以我们要想办法让不同 PIM-SM 之间的 RP 能够互相共享和交换组播源信息，这个信息被称为 Source Active (SA)，那么不同 PIM-SM 之间的 RP 是否会自动交换 SA 信息呢？当然不会，所以需要通过一个协议来强制这些 RP 之间交换 SA 信息，这个协议就是 Multicast Source Discovery Protocol (MSDP)，即组播源发现协议，如下图所示：





从上图中可以看出，PIM-SM Domain 1 的 RP 为 2.2.2.2，R1 是组播源，PIM-SM Domain 2 的 RP 为 4.4.4.4，R5 是组成员，而正常情况下，PIM-SM Domain 2 中的组成员 R5 想要收到 PIM-SM Domain 1 中组播源 R1 发出的数据包是不可能的，因为 PIM-SM Domain 2 中的 RP 虽然知道 R5 是组成员，却不知道谁是组播源，因此无法建立组播树。但当两个 PIM-SM 域的 RP 之间建立 MSDP 连接之后，PIM-SM Domain 1 中的 RP 就会将组播源信息通过 SA 数据包发给 PIM-SM Domain 2 中的 RP，当 PIM-SM Domain 2 中的 RP 学习到组播源和组成员信息之后，两个域之间就能正常建立组播组，最终实现 PIM-SM 域间组播通信。

说到这里，谨慎的你应该发现似乎漏掉了一个问题，上图中，因为 PIM-SM Domain 2 中的 RP 只知道组成员信息，而不知道组播源信息，所以当 PIM-SM Domain 1 中的

RP 通过 MSDP 的 SA 数据包将组播源信息发给它之后事情就完美了，但是不难看出，PIM-SM Domain 1 中的 RP 也只是知道谁是组播源，却不知道谁是组成员，为什么在只知道组播源而不知道组成员的情况下能够建立组播树的呢？为什么只考虑让 MSDP 共享组播源信息，却没有说共享组成员信息呢？难道只需要知道组播源，不知道谁是组成员也可以建立组播树吗？原因是这样的：试想一下如果你是个非常有钱的大富翁，现在你就是想让你的钱拿到街上去洒发给普通人，你能把你的钱顺利洒掉并让普通人捡到吗（当然这里不考虑街上是否有城管、ZF、以及怪兽）？其实答案咱们很清楚，是可以的，因为只要让普通人（接收者）知道你这个富翁（发送者）要洒钱，知道你是哪里洒钱即可，而根本就不需要你（发送者）去了解普通人（接收者）有哪些且现在在哪里，也就是说只需要他们知道怎么能够找到你就 OK，你不需要知道怎么找到他们，因为他们会自动来找你；这里的好好比就是组播源，而普通人就好比是组成员，只要组成员知道了谁是组播源，就会一拥而上去找组播源，不需要组播源费脑筋去找组成员。并且这也是因为 PIM-SM 模式是靠 PULL（拉）的方式来建组播树的，也就是谁要接收流量，是由需要接收组播的组成员自己去跟 RP 申请，而不是 RP 去追着问到底谁要接收组播，但如果换了是 PIM-DM 模式，它就是靠 PUSH（推）的方式来建组播树的，就是组播源去追着每个人问到底要不要收组播，这就是区别。

需要重点强调的是，MSDP 协议的功能和目的只有一个，就是把一个 PIM-SM 域内的组播源信息（SA）发送给其它 PIM-SM 域的 RP，从而让 PIM-SM 域间的组播通信正常，MSDP 只是在 PIM-SM 域之间传个消息而已，相当于给对方 RP 打个电话，告诉对方组播源是谁，然后挂了电话就再也没 MSDP 任何事了，至于后面组播是通还是不通，都不是 MSDP 的责任；在应用中，如果一个 PIM-SM 域的 RP 已经通过 MSDP 收到了组播源信息（SA），那么就表示 MSDP 的工作已经做到位了，MSDP 只负责 SA 的传递，而不负责组播流量是否正常，如果 PIM-SM 域之间的组播还是不通，这个时候请不要再怀疑是不是 MSDP 出了问题，问题绝对不在 MSDP 身上，所以这时你应该把精力放到其它地方，MSDP 只关系到 SA，而不关系到组播数据流。

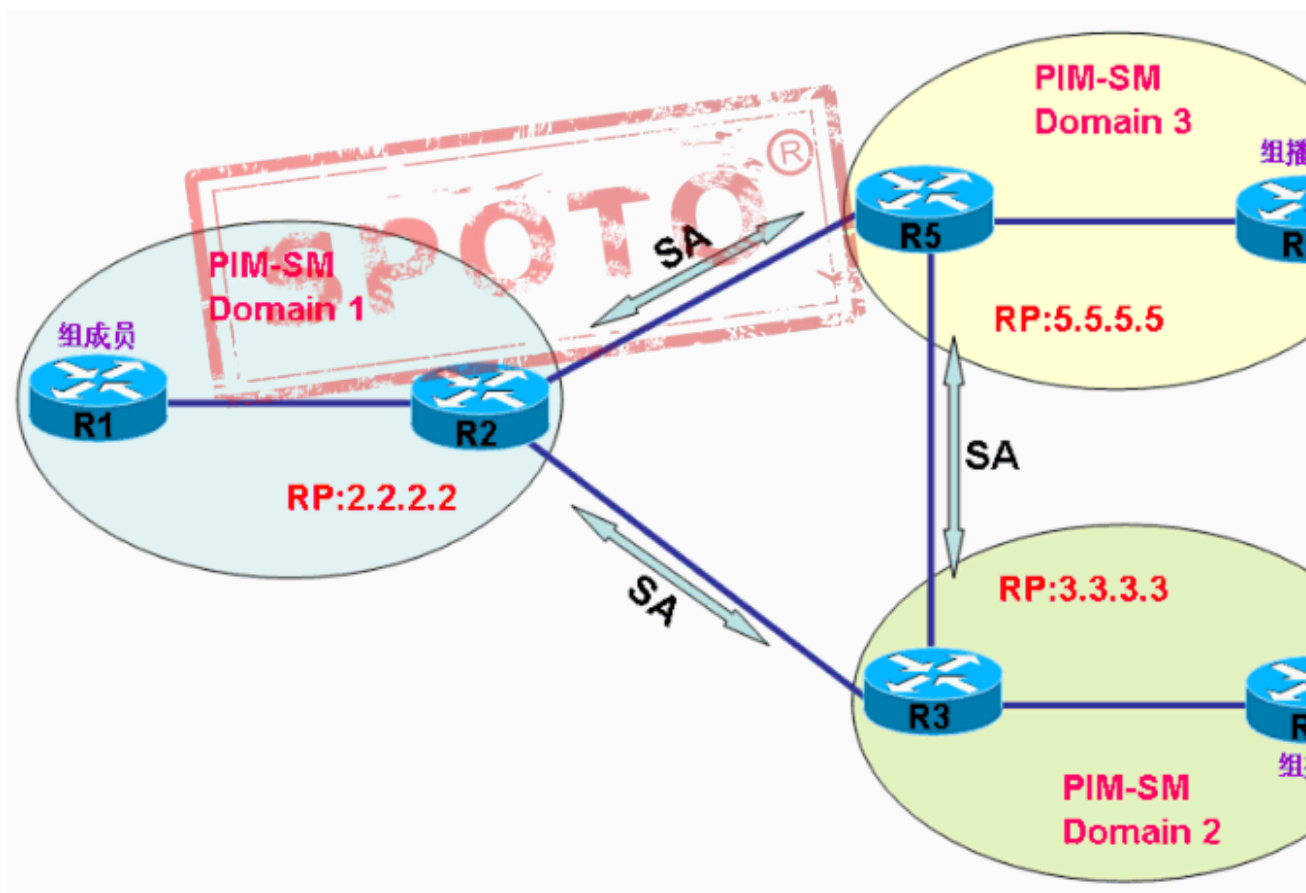
单个域内的 PIM-SM 要通信，是建立的（*, G）条目，如果 PIM-SM 域之间要通信，建立的是（S, G）条目，相当于是距离矢量的路径方式。

注：只有 PIM-SM 才有 MSDP，PIM-DM 是没有 MSDP 的，因为只有 PIM-SM 才有域的概念，才有域间组播通信的概念，PIM-DM 没有域的概念，不存在域间组播通信。

指定 MSDP peer 的命令为：ip msdp peer xxx （xxx 为对端 MSDP 路由器地址）。

MSDP RPF 检测

理论上，PIM-SM 域之间建立 MSDP 连接来交换组播源信息，是在 RP 与 RP 之间建立的，不建议在非 RP 上建立，MSDP 是通过 TCP 来建立的，端口号为 639，和 BGP 的邻居建立相似，所以要建立 MSDP，必须手工指定 MSDP peer 的地址，TCP 由地址小的一端主动发起连接，而地址高的一端则在 LISTEN 状态等待连接。待 MSDP peer 建立成功之后，便开始传递 SA 信息，MSDP 之间的 Keepalive 包每 60 秒一个，如果 75 秒没有收到 Keepalive 包，则会话被重置，而发送的 SA 信息也可以被当作 Keepalive，即如果 75 秒没有收到 Keepalive 但收到了 SA，也表示连接正常。当一个 RP 拥有多个 MSDP peer 时，在从一个 peer 那里收到 SA 之后，会转发给除发送者之外的其它所有 peer，如下图：



在上图中，PIM-SM Domain 1 中的 RP 路由器 R2 与 PIM-SM Domain 2 中的 RP 路由器 R3 以及 PIM-SM Domain 3 中的 RP 路由器 R5 之间建立全互联的 MSDP 连接，R2 从 R3 收到的 SA 会转发给 R5，从 R5 收到的 SA 也会转发给 R3，并且 R3 从 R2 收到的 SA 会转发给 R5，从 R5 收到的 SA 也会转发给 R2，R5 的动作和 R2 与 R3 一样，也会将自己收到的 SA 转发给其它 MSDP peer，这样的转发结果就是对于同一个组播

源信息会从多个 MSDP peer 收到多次重复的，比如对于 PIM-SM Domain 2 中的组播源信息，R5 会同时从 R3 和 R2 收到，那么对于哪一条是正确的，在收到的时候会做一个检测，然后将正常的 SA 缓存起来，如上图，R5 只应该接收 R3 发来的 SA，而 R3 也应该只接收 R5 发来的 SA，对于 PIM-SM Domain 2 的 SA，R2 只应该接收 R3 发来的，对于 PIM-SM Domain 3 的 SA，R2 只应该接收 R5 发来的；（至于 SA 是谁发来的，是不是从正常路径发来的，我本人认为这并不重要，全部接收好了，我觉得没什么，因为它只是一个消息而已，并且不管是谁发来的，消息的内容都是一模一样的，有什么好纠结的呢？是谁发来的只要消息是对的，并不影响组播流量的转发！），路由器从 MSDP peer 收到 SA 之后，需要做 Reverse Path Forwarding (RPF)检测，即反向路径检测，普通的 RPF 检测的方法是查看自己的路由表中去往发送者 IP 的数据包该从哪个接口出去，那么从哪个接口收到的数据包就被认为是合法有效的，这个合法的接口也被称为 RPF 接口，事实上 MSDP 对于 SA 数据包的 RPF 检测要比普通 RPF 检测复杂的多，但这是 MSDP 唯一的重点，也是唯一的难点，如果 RPF 检测失败，就表示 SA 信息被丢弃，那么组播树就无法建立，组播也就无法通信。

注：在当前的 IOS 版本中，将 MSDP 的 SA 信息缓存起来是强制执行的，不能人为手工开启或关闭，默认情况下，当 MSDP 邻居配置之后，命令“ip multicast cache-sa-state”将被自动添加到 running configuration 中，如果 IOS 版本早于 IOS Releases 12.1(7) 和 12.0(14)S1，默认是不开启 SA 信息缓存功能的，但可以通过手工输入命令“ip multicast cache-sa-state”来开启 SA 缓存功能。

对于 SA 的数据包内容，我们不用研究太深，也没必须研究太深，但有一样东西是必须掌握的，就是某个 PIM-SM 域把 SA 发来之后，SA 数据包里面明确写清楚了那个 PIM-SM 域内的 RP 地址是多少，这是 SA 数据包内我们唯一需要关心的内容，因为该 RP 地址对 SA 数据包的 RPF 检测至关重要，但是这个 SA 里面写的 RP 地址可以被人为更改，更改命令为 ip msdp originator-id 加接口，则使用相应接口的 IP 地址为 SA 内的 RP 地址。

对收到的 SA 数据包做 RPF 检测，和普通的 RPF 检测不一样，普通的 RPF 检测是根据所有的 IGP 路由表以及所有的 BGP 单播路由表来做检测的，即命令“show ip route”看到的路由表，而对 SA 数据包做的 RPF 检测只能根据 BGP 路由表来做检测，单播 BGP 和组播 BGP 都可以，即 Unicast Border Gateway Protocol (uBGP)和 Multicast Border Gateway Protocol (MBGP)，MBGP 是 Multiprotocol-Border Gateway Protocol (MP-BGP)应用的一种，因为 MP-BGP 不仅可以扩展到组播，还可以扩展到 IPv6 以及 MPLS。如果同时存在 MBGP 和单播 BGP，则 MBGP 优先，既然对 SA 数据包做的 RPF

检测必须依靠 BGP 路由表,那就意味着 PIM-SM 域之间必须运行 BGP,如果没有 BGP,那么 SA 的 RPF 检测将失败,最终导致 SA 数据包被丢弃。

MSDP RPF 检测详细规则

在以下 3 种情况下收到的 SA 是不需要做 RPF 检测的：（**相当重要**）

1. 发送方 MSDP peer 是 default MSDP peer 或是唯一的 1 个 MSDP peer(即只配置了 1 条 ip msdp peer 命令)的情况下；
2. 发送方 MSDP peer 是 MSDP Mesh Group 中的一员；
3. 发送方 MSDP peer 的 IP 地址与 SA 数据包中的 RP 地址相同。

所以在以上 3 种情况下，BGP 和 MBGP 都是不需要的。

通常情况下，MSDP peer 就是 BGP 或 MBGP 邻居，但也有不是的情况，因为 BGP 的邻居类型分 interior BGP (iBGP)和 exterior (eBGP)，所以 SA 的 RPF 具体检测过程也分如下 3 种：

1. MSDP peer 是 iBGP 邻居；
2. MSDP peer 是 eBGP 邻居；
3. MSDP peer 不是 BGP 邻居（但域之间也必须有 BGP）。

注：在以上 3 种检测情况下，都需要对比 SA 中 RP 的 IP 地址信息，所以务必保证将 SA 中 RP 的 IP 地址通告进 BGP 中；同样的，也建议使用建立 MSDP peer 的地址来建立 BGP 邻居，更能保证 RPF 的检测通过。

SA 的 RPF 具体检测细节如下：

1. 当 MSDP peer 是 iBGP 邻居

如果 BGP 路由表中显示去往 SA 数据包中 RP 地址的最佳路径就是走这个 iBGP 邻居，则检测通过，否则检测失败。（先查多播路由表(MRIB)，再查单播路由表(URIB)）

解释： BGP 路由表中去往 SA 数据包中 RP 地址的最佳路径的下一跳地址等于发送该 SA 数据包的 MSDP peer 的地址。

★所以建议建 BGP 和建 MSDP 的地址使用同一个地址。

2. 当 MSDP peer 是 eBGP 邻居；

如果 BGP 路由表中显示去往 SA 数据包中 RP 地址的最佳路径的下一跳 AS 号码就是这个 eBGP 邻居的 AS 号码，则检测通过，否则检测失败。（先查多播路由表(MRIB)，再查单播路由表(URIB)）

解释： 最佳路径的下一跳 AS 就是 AS_PATH 中的第 1 个 AS，即 AS_PATH 中最左边的 AS。

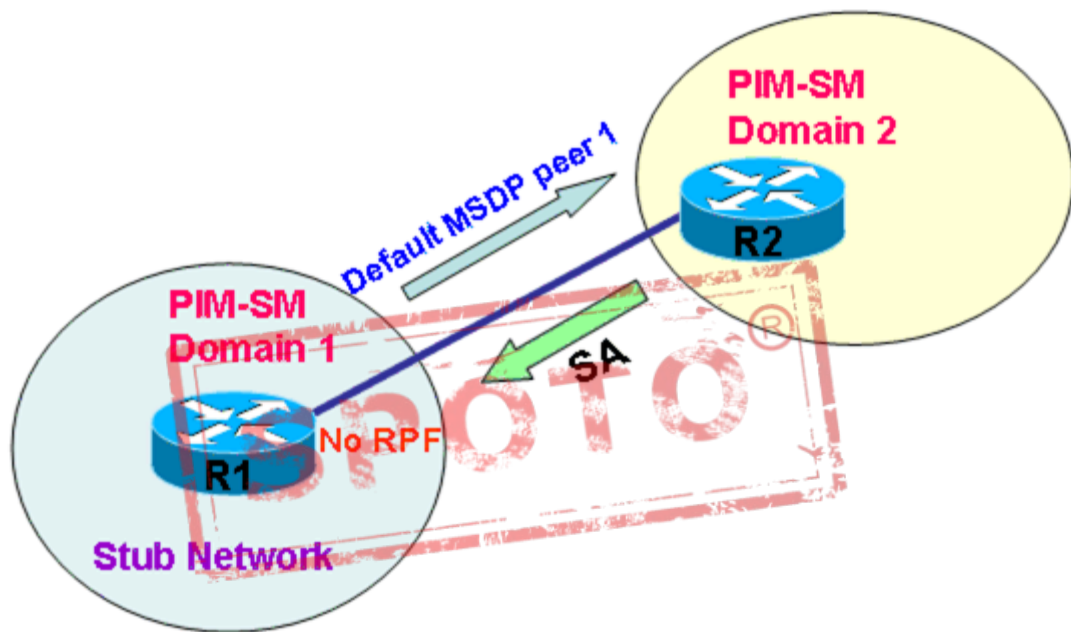
3. 当 MSDP peer 不是 BGP 邻居（但域之间也必须有 BGP）

如果 BGP 路由表中显示去往 SA 数据包中 RP 地址的最佳路径的下一跳 AS 号码和去往 MSDP peer 的最佳路径的下一跳 AS 号码相同，则检测通过，否则检测失败。（先查多播路由表(MRIB)，再查单播路由表(URIB)）

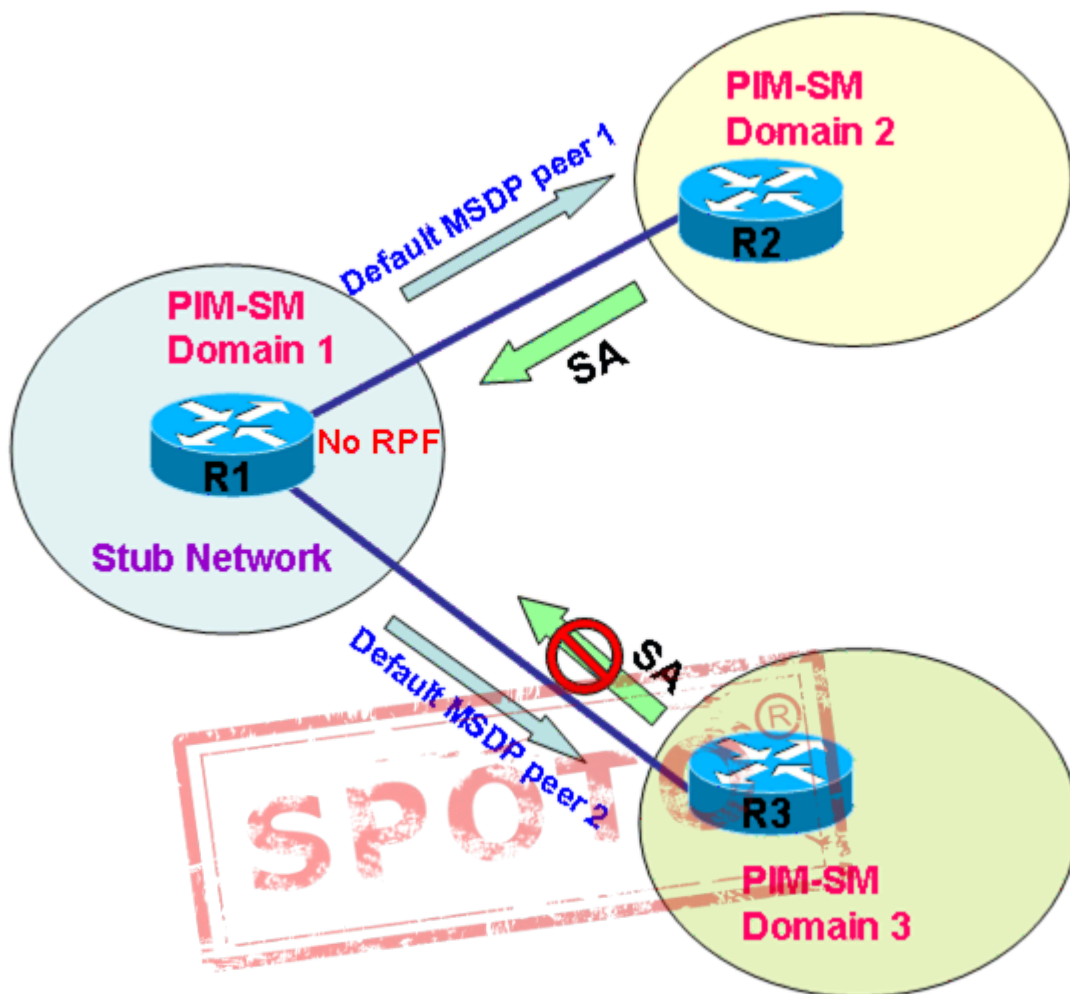
解释： 不需要再解释。

Default MSDP Peer

当 MSDP 路由器收到 SA 的时候，必须依靠 BGP 来做 RPF 检测，这就强制 PIM-SM 域中的路由器运行 BGP，否则 SA 数据包会因为 RPF 检测不通过而被丢弃。然而在很多时候，这是不现实的，例如这个网络是个末节网络，或者是用户与 ISP 互联的网络，通常这样的网络只有一个出口，只需要配一条默认路由指出去就足够了，根本用不着配 BGP，如下图：



如上图所示，PIM-SM Domain 1 是个末节网络，对外面只有一个出口，所以只需要对外写一条默认路由即可，不需要配置 BGP，但是很明显，对于 MSDP 收到的所有 SA 都是应该接收的，并且也没有必要对 SA 数据包做 RPF 检测，因为没有其它办法能够收到 SA 数据包了，所以从唯一的出口收到的 SA 统统都有效。在上述情况下，就可以通过在 PIM-SM Domain 1 的 MSDP 路由器 R1 上将它的 MSDP peer 指定为 Default MSDP Peer，也就是将 PIM-SM Domain 2 中的 R2 指定为 Default MSDP Peer，这样一来，R1 从 R2 收到的所有 SA 都不需要再做 RPF 检测就能够缓存起来并使用。在某些情况下，为了网络的冗余性，一个 PIM-SM 域也可能对连了多个 PIM-SM 域，但之间并没有运行 BGP，或者为了使 MSDP 有冗余功能，可以为一个 Default MSDP Peer 再配置一个备份 Default MSDP Peer，当主 Default MSDP Peer 失效之后，再从备份 Default MSDP Peer 接收 SA，如下图：



如上图所示，PIM-SM Domain 1 同时与 PIM-SM Domain 2 和 PIM-SM Domain 3 建立 MSDP 连接，但之间并未运行 BGP，在这种情况下，R1 可以同时将 R2 与 R3 都指定为 Default MSDP Peer，但配置命令会存在先后顺序，例如先指定了 R2，再指定了 R3，那么在正常情况下，只能从先指定的 R2 那里接收 SA，只有在 R2 失效后，才能从 R3 接收 SA；在指定了多个 Default MSDP Peer 的情况下，一开始只能从第一个 Default MSDP Peer 那里接收 SA，只有当第一个不可用了，才能从第二个 Default MSDP Peer 那里接收 SA，第二个不可用再从第三个接收，依此类推。

指定 Default MSDP Peer 的命令为：ip msdp default-peer xxx（xxx 为对端 MSDP 路由器地址）；

在配置此命令之前，必须已经通过命令 ip msdp peer xxx 指定过常规 MSDP peer。

但也可以在指定多个 Default MSDP Peer 的情况下从多个 Default MSDP Peer 那里同时接收 SA，在这里有个条件限制，那就是为每个 Default MSDP Peer 配置 Prefix List 来限制只从特定 Default MSDP Peer 接收特定的 SA。

为 Default MSDP Peer 指定 Prefix List 的命令为：ip msdp default-peer xxx yyy（xxx 为对端 MSDP 路由器地址，yyy 为 Prefix List 名）。

Default MSDP Peer 的优势：

与 Default MSDP Peer 之间不需要运行 BGP，从 Default MSDP Peer 收到的所有 SA 都不需要做 RPF 检测。

重点说明：

如果 MSDP 路由器上只使用了单条 ip msdp peer 命令指定了唯一的 1 个 MSDP peer，那么该 MSDP peer 等同于 Default MSDP Peer。



MSDP Mesh Group

MSDP Mesh Group 的中文意思是 MSDP 全互联组，从字面意思就可以看出，MSDP Peer 之间的连接应该是全互联的，何为全互联？全互联就是指每两个 MSDP Peer 之间都有一条 MSDP 连接，换句话说就是每一个 MSDP Peer 都和其它任何 MSDP Peer 拥有 MSDP 连接。在全互联的模式下，每一个 MSDP Peer 收到任何 SA 都不会转发给其它 MSDP Peer，为什么呢？例如有 A、B、C 共 3 个 MSDP 路由器，它们之间是全互联的，当 B 从 A 那里收到 SA 之后，是不会转发给 C 的，因为 A 和 B 有 MSDP 连接，A 和 C 也有 MSDP 连接，既然 B 能收到 A 的 SA，那说明 A 也将同样的 SA 发过给 C 了，既然 C 也能收到和 B 收到的一样的 SA，那么 B 也没必须多此一举将自己收到的 SA 发给别人，自己能收到，证明别人也能收到，因为这个网络是全互联的。

因为 MSDP Mesh Group 中的任何 MSDP Peer 在收到 SA 之后都不会转发给其它 MSDP Peer，这就表明对于同一份 SA 数据包来说，MSDP Peer 永远只能从单个 MSDP Peer 那里收到，并且发送 SA 的那个 MSDP Peer 就是初始 MSDP Peer，这也保证了 MSDP Peer 所收到的 SA 永远都是对的，MSDP Peer 永远不可能从一个错误的路径收到不可信任的 SA。既然如此，从 MSDP Mesh Group 中的 MSDP Peer 收到的所有 SA 都不再需要做 RPF 检测便被缓存起来使用。

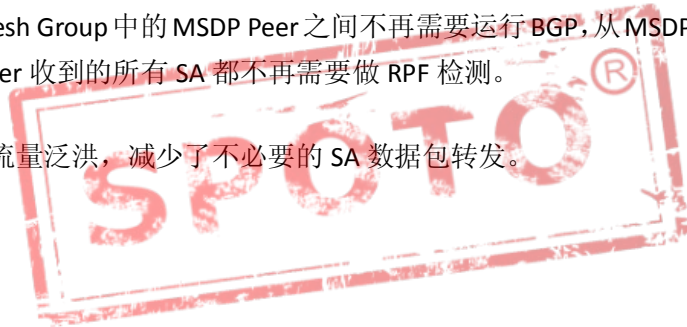
指定 MSDP Mesh Group 的命令为：ip msdp mesh-group test-mesh-group xxx（xxx 为对端 MSDP 路由器地址，test-mesh-group 为 Mesh Group 的组名，可以为每个 MSDP peer 配置不同的组名）。

在配置此命令之前，必须已经通过命令 ip msdp peer xxx 指定过常规 MSDP peer。

MSDP Mesh Group 的优势：

★与 MSDP Mesh Group 中的 MSDP Peer 之间不再需要运行 BGP，从 MSDP Mesh Group 中的 MSDP Peer 收到的所有 SA 都不再需要做 RPF 检测。

★杜绝了 SA 流量泛洪，减少了不必要的 SA 数据包转发。



MSDP SA Filter

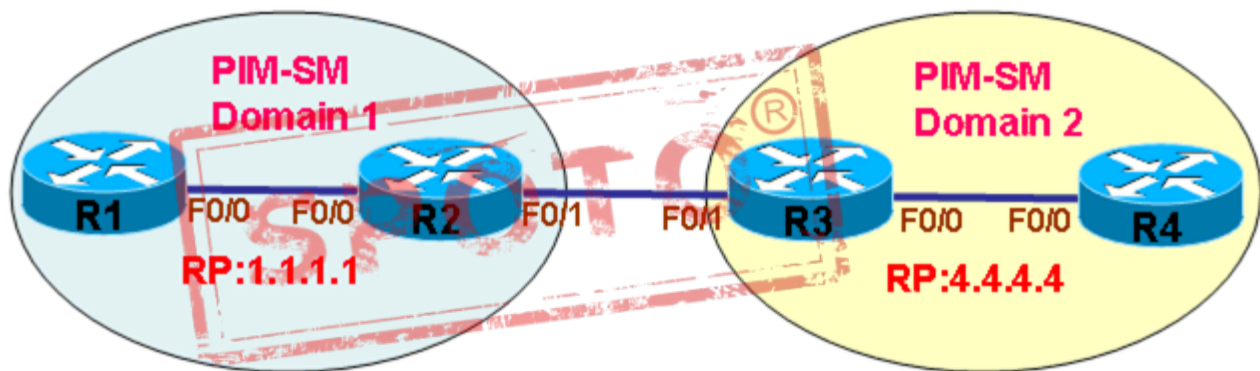
对于 SA，无论是发出去的还是收到的，无论是源自其它 MSDP Peer 还是自己产生的，都可以通过 Filter List 来做过滤，所过滤的条件可基于 ACL、route map、RP access list、RP route map，这里的 ACL 必须是扩展 ACL。

对接收的 SA 做过滤的命令为：ip msdp sa-filter in xxx yyy（xxx 为对端 MSDP 路由器地址，yyy 为 ACL、route map、RP access list、RP route map，如果不跟 yyy，则表示所有 SA）；

对发出去的 SA 做过滤的命令为：ip msdp sa-filter out xxx yyy（xxx 为对端 MSDP 路由器地址，yyy 为 ACL、route map、RP access list、RP route map，如果不跟 yyy，则表示所有 SA）。

PIM-SM 域边界

我们所定义的 PIM-SM 域,其实就是所有组播路由器都拥有同一个 RP 的网络范围,如果在网络里一部分组播路由器使用某一个 RP,而另一部分组播路由器使用另一个 RP,那么它们就属于两个不同的 PIM-SM 域。为组播网络指定 RP 的方法有 3 种,分别为:手工静态指定, Auto-RP(Cisco 私有), Bootstrap Router (BSR), 如果使用手工静态指定的方法,要明确划分 PIM-SM 域是非常简单的,因为想让某台路由器使用某个 RP 只要使用命令强制指定即可,并且任何时候都不会变动,如下图:



如上图所示,网络中有 4 台路由器,要让 R1 和 R2 在 PIM-SM Domain 1 中,使用 R1 (1.1.1.1) 作为 PIM-SM Domain 1 的 RP,让 R3 和 R4 在 PIM-SM Domain 2 中,使用 R4 (4.4.4.4) 作为 PIM-SM Domain 2 的 RP,在使用手工静态指定的情况下,只需要使用命令强制指定 R1 和 R2 的 RP 为 1.1.1.1,强制指定 R3 和 R4 的 RP 为 4.4.4.4 即可。

而在使用 Auto-RP 时,我们知道它是个动态分发 RP 的协议,组播网络中的 RP 是谁,需要靠协议自动去计算,然后自动通告出去,所以结果可能并非我们所预期的那样,如果整个组播网络中都使用 Auto-RP 来分发 RP,那么结果可想而知,这里只可能出现一个 PIM-SM 域,因为全网通过 Auto-RP 所获得的 RP 信息都是一致的,如果在使用 Auto-RP 的情况下也要划分出像上图所示的 2 个 PIM-SM 域,就必须让 Auto-RP 的协议流量只在单个域范围内通告,在上图中就是 PIM-SM Domain 1 内的 Auto-RP 通告信息只能在 R1 和 R2 之间传递,无法穿过 R2 的 F0/1 接口到达 PIM-SM

CCIE LAB认证经验分享千人群：539730342

Domain 2, 而 PIM-SM Domain 2 内的 Auto-RP 通告信息也只能在 R3 和 R4 之间传递, 无法穿过 R3 的 F0/1 接口到达 PIM-SM Domain 1, 这样两个 PIM-SM 域在使用 Auto-RP 的情况下也能保证互不影响, 相互独立。要让 Auto-RP 的通告信息无法穿过 R2 的 F0/1 或者无法穿过 R3 的 F0/1, 可以使用过滤组播流量的方法来实现, 我们都知道 Auto-RP 的流量使用的组播地址为 224.0.1.39 和 224.0.1.40, 所以我们只要在 R2 或 R3 的 F0/1 接口上将这两个组的流量过滤掉即可, 但其它组的流量必须放行, 要不然正常的组播流量也被过滤掉了, 那么组播就不通了。过滤组播流量的方法为在接口下使用命令 “ip multicast boundary xxx”, 其中 xxx 表示匹配组播地址的 ACL 名, 该命令在接口上应用之后, 将同时过滤掉接口进来和出去的所有流量, 按上面所提到的需求, 那就只需要在 R2 或 R3 上做如下配置即可:

```
access-list 1 deny 224.0.1.39
```

```
access-list 1 deny 224.0.1.40
```

```
access-list 1 permit any
```

```
interface f0/1
```

```
ip multicast boundary 1
```



该配置定义了将组播地址为 224.0.1.39 和 224.0.1.40 的流量过滤掉, 而放行其它所有组播流量, 并将该过滤应用到接口 F0/1 上, 这样一来, Auto-RP 的流量就无法通过 R2 或 R3 的 F0/1 接口了, 从而实现了 PIM-SM Domain 1 与 PIM-SM Domain 2 的相互独立, 互不干扰。

对于在使用 BSR 的情况下划分 PIM-SM 域, 方法同使用 Auto-RP 的情况类似, 区别就是使用的命令不同, 过滤 BSR 信息的方法为在接口下使用命令 “ip pim bsr-border”, 后面不需要跟任何参数, 包括 ACL, 这条命令将阻止接口上收到或发出的所有 BSR 数据, 按上面所提到的需求, 那就只需要在 R2 或 R3 上做如下配置即可:

```
interface f0/1
```

```
ip pim bsr-border
```

该配置定义了 BSR 的流量在 R2 或 R3 的 F0/1 接口上被过滤掉, 但放行其它所有组播流量, 这样一来, BSR 的流量就无法通过, 从而实现了 PIM-SM Domain 1 与 PIM-SM Domain 2 的相互独立, 互不干扰。

重点说明：

在划分多个 PIM-SM 域之后，如果 PIM-SM 域之间还需要组播通信，这两个 PIM-SM 域之间必须还有 PIM-SM 连接通路，即还有 PIM-SM 邻居，否则两个域之间连最基本的 PIM-SM 通路都没了，组播流量就无从转发；虽然 PIM-SM 域之间有 MSDP 连接，但不要忘记，MSDP 的作用是传递 SA 的，它与普通的组播流量无关。如果两个 PIM-SM 域之间被 ISP 隔开，则可通过创建 Generic Routing Encapsulation (GRE) Tunnel 的方式来实现 PIM-SM 互联，所以也必须在 GRE Tunnel 上配开启 PIM-SM。

组播流 RPF 检测详细规则

这里所提到的组播流 RPF 检测是指对普通组播流量的 RPF 检测，即对用户组播流量的 RPF 检测，它与 SA 数据包的 RPF 检测毫无关系，两个是相互独立，毫无关联的。

之前我们已经重点强调过，MSDP 协议的功能和目的只有一个，就是把一个 PIM-SM 域内的组播源信息（SA）发送给其它 PIM-SM 域的 RP，从而让 PIM-SM 域间的组播通信正常，MSDP 只是在 PIM-SM 域之间传个消息而已，只要 PIM-SM 域之间的 SA 已经确认收到了，那么就表示 MSDP 的工作已经做到位了，至于后面组播是通还是不通，都不是 MSDP 的责任；所以这时你应该把精力放到其它地方，在 PIM-SM 域之间的 SA 数据包已经正常的情况下，最有可能影响到组播流量不通的情况就是普通组播流的 RPF 检测，所以下面我们来详细介绍普通组播流的 RPF 检测如何工作。

前面所提到的对 SA 数据包的 RPF 检测只能基于 BGP 路由表，而普通组播流的 RPF 检测可以基于任何路由表，其中包括了 BGP 和 MBGP 路由表，可用的路由表如下：

1. 单播路由表，即命令“show ip route”看到的所有路由表，包括 BGP。
2. MBGP 路由表，也就是组播 BGP 专用的路由表。
3. Distance Vector Multicast Routing Protocol (DVMRP) 路由表（目前无须掌握）。
4. 静态组播路由表（就像静态指定单播路由表一样指定组播路由表）。

在对普通组播流做 RPF 检测时，是根据以上所有路由表 AD 值来决定先后顺利的，当收到组播流量以后，先查看发送该数据的源 IP 地址是什么，然后在以上路由表中查看去往源 IP 的数据包该从哪个接口出去，那么从哪个接口收到的数据包就被认为是合法有效的，这个合法的接口也被称为 RPF 接口，只有从 RPF 接口收到的流量才能转发，从其它所有接口收到的组播流量都将被丢弃。

注：如果是共享树（Share-Tree），则 RP 的地址被视为源 IP 地址的。

当以上路由表中路由条目的 AD 值完全一样时，则按以下顺序来做 RPF 检测：

如果都一样，那就最长匹配

1. 静态组播路由表

2. DVMRP 路由表

3. MBGP 路由表

4. 单播路由表



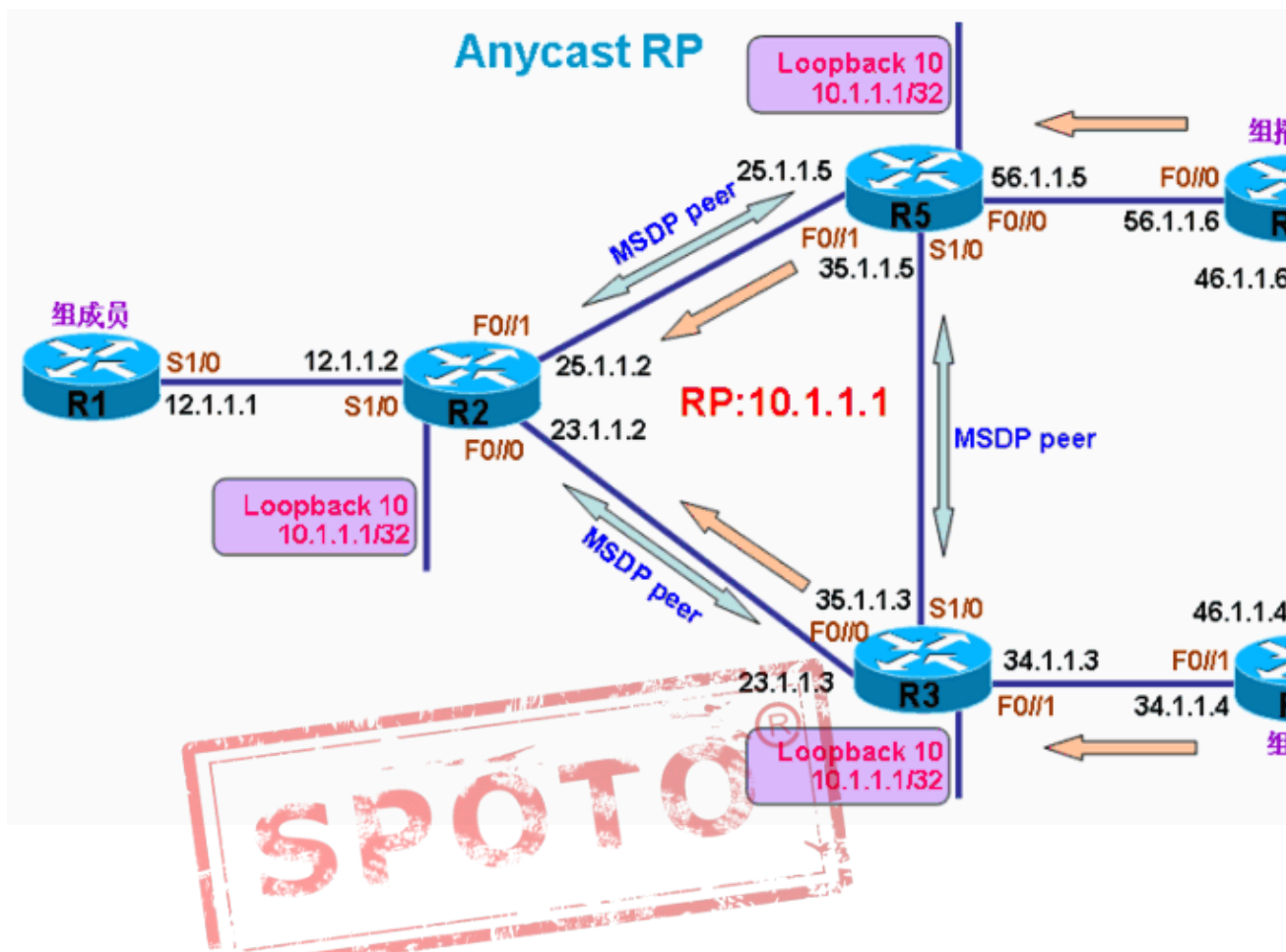
如果在同一个路由表中发现多条去往源 IP 的路由，则选择最长掩码匹配，因为在这里同一个路由表中不可能出现多条掩码相同的路由，所以一定能够选出结果，如果在负载均衡时路由出口有多个，则每个接口都被认为有效。

注：如果接口没有开启 PIM，将不参与 RPF 检测，如果需要看到更多情况下的 RPF 检测效果，请尽可能多的开启组播接口。

Anycast RP

在 PIM-SM 模式下，组播想要通信，就必须拥有正常运行的 RP，指定 RP 的方法有 3 种：手工静态指定，Auto-RP，BSR，其中通过手工静态指定的 RP 在任何时候都不会变动，只要该 RP 路由器出现故障，组播流量将中断，不能拥有备份 RP；如果使用 Auto-RP 或 BSR，则可以设置多个 RP，但同一时间只能有一个 RP 是工作的，只要工作的 RP 出现故障，就会选举备份 RP 接替之前的工作，使用 Auto-RP 或 BSR 能够实现多个 RP 的冗余备份功能，在出现故障时能够快速切换，但是多个 RP 中只能有一个 RP 可以工作，可想而知这个 RP 在大流量组播的情况下就会出现瓶颈，虽然配置了多个 RP，但这些 RP 并不能分担组播流量。

Auto-RP 和 BSR 最大的优势是能够通过多个 RP 实现备份功能，但是却不能实现多个 RP 负载均衡，为了能够在网络中配置多个 RP 既能实现备份功能，又能让所有 RP 同时工作实现流量负载功能，所以在 MSDP 的应用下扩展出了 Anycast RP。Anycast RP 采用了一种很新颖的思路来实现多个 RP 的负载与冗余，它通过在组播网络中将 2 个或者 2 个以上的 RP 配置成相同的 RP 地址来实现冗余和备份功能，每个 RP 路由器上都创建一个 loopback 接口来充当 RP，并且该接口地址的掩码必须是 /32 位的，然后将该地址通过动态路由协议发布到网络中，因为所有 RP 的这个 loopback 接口地址全部是一样的，所以能够实现冗余功能，当其中某一台 RP 出现故障后，流量可以很平滑地被转移到另一台 RP 上，这都是因为所有 RP 的地址是同一个地址，如下图：



在上图中，全网运行 OSPF 路由协议，其中 R2、R3、R5 共 3 台路由器上都创建了接口 Loopback 10，并且配置的 IP 地址都是 10.1.1.1/32，然后将该地址发布到 OSPF 中，3 台路由器都以接口 Loopback 10 的地址发布为 RP，最终所有路由器都认为 RP 地址是 10.1.1.1，这样一来，3 个 RP 都能为网络中提供组播流量转发，无论其中哪台 RP 出现故障，只要网络中还剩一台可用 RP，那么组播流量还能正常通信，这就是 Anycast RP 实现的 RP 冗余功能。并且因为所有 RP 的 Loopback 10 接口地址都发布进了 OSPF，所以所有路由器选择去往 10.1.1.1 都是通过 OSPF 来选择最近的路径，例如 R4 一定是从 F0/1 出去选择 R3 的 10.1.1.1，而 R6 一定是从 F0/0 出去选择 R5 的 10.1.1.1，而 R1 一定是从 S1/0 出去选择 R2 的 10.1.1.1，这样一来，Anycast RP 不仅实现了 RP 冗余功能，同时还实现了 RP 的流量负载均衡功能，因为这时所有的流量是被分摊到多个 RP 同时传输的，如果当某个 RP 出现故障，那么流量将被转移至下一台最近的路径，例如当 R5 出现故障后，R6 将从 S1/0 出去选择 R3 的 10.1.1.1，从而维持了组播通信。

从上述环境中我们还可以看出，组播源 R4 和 R6 以及组成员 R1 都注册到了不同的 RP，这么一来，就没有一个 RP 拥有完整的组播源和组成员信息，这样就不可能建立正常的组播树，所以必须寻求一种办法让所有的 RP 都拥有完整一致的组播源

和组成员信息，这个目标可以通过在 RP 之间创建 MSDP 来实现，在上图中，当在 R2、R3 和 R5 之间建立 MSDP 之后，通过相互交换 SA 信息，就能够让所有的 RP 都学习到完整的组播源和组成员信息，从而保证组播正常通信。这里的 MSDP 可以是全互联，也可以不全互联，只要保护 SA 信息能够让每个 RP 都收到即可。

之前提到过，MSDP 路由器发出的 SA 数据包里面明确写清楚了 RP 地址是多少，但在布署 Anycast RP 的环境下，所有 MSDP 发出的 SA 数据包里面的 RP 地址全是相同的，这种情况可能会导致 SA 无法传递，因为在遇到对 SA 数据包做 RPF 检测时，肯定是会失败的，所以需要手工配置命令将其改掉，更改命令为 `ip msdp originator-id` 加接口，则使用相应接口的 IP 地址为 SA 数据包内的 RP 地址。

因为不同的 PIM-SM 域之间的 RP 地址是不同的，而 Anycast RP 要求所有的 RP 地址相同，所以 Anycast RP 只能在单个 PIM-SM 域范围内设计和使用，在 PIM-SM 域之间不需要实施 Anycast RP

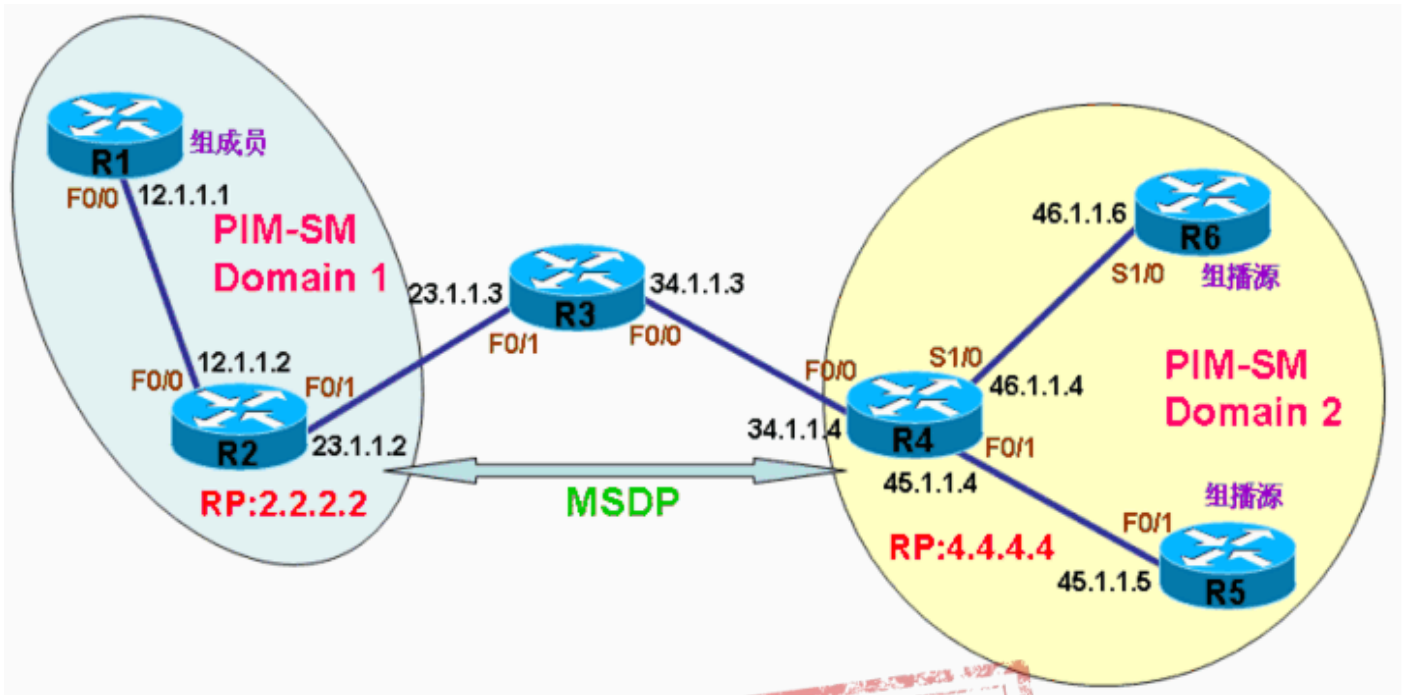
注：无论在配置 PIM-SM 单域还是 PIM-SM 多域，都应该事先保证全网单播互通。



配置一对一 PIM-SM 域的 MSDP 实验

实验说明：

以下图为环境配置一对一 PIM-SM 域的 MSDP 实验：



上图中两个 PIM-SM 域通过 Internet 相连，其中 R1 和 R2 在 PIM-SM Domain 1 中，R4、R5 和 R6 在 PIM-SM Domain 2 中，R3 模拟 Internet 路由器，所以 R3 不需要参与组播，也不允许参与组播；PIM-SM Domain 1 和 PIM-SM Domain 2 之间需要创建 Tunnel 来穿越 Internet 实现组播通信，其中 R5 和 R6 为组 224.1.1.1 的组播源，R1 为组成员。

除了图上所标识出的接口地址外，R2 和 R4 还配有 Loopback 接口，分别为：

R2 (Loopback 0 :2.2.2.2/24)

R4 (Loopback 0 :4.4.4.4/24)

PIM-SM Domain 1 的 RP 为 2.2.2.2，即 R2，PIM-SM Domain 2 的 RP 为 4.4.4.4，即 R4，

所有路由器上都运行 OSPF，并将所有接口都发布进 OSPF，以保证全网单播互通。

1. 配置初始网络环境

(1) 配置 R1:

```
r1(config)#int f0/0
```

```
r1(config-if)#ip add 12.1.1.1 255.255.255.0
```

```
r1(config-if)#no shutdown
```

```
r1(config-if)#exit
```

```
r1(config)#router ospf 1
```

```
r1(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r1(config-router)#exit
```

说明：为 R1 的 F0/0 配置接口地址，并将所有接口发布进 OSPF。

(2) 配置 R2:

```
r2(config)#int loopback 0
```

```
r2(config-if)#ip add 2.2.2.2 255.255.255.0
```

```
r2(config-if)#ip ospf network point-to-point
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int f0/0
```

```
r2(config-if)#ip add 12.1.1.2 255.255.255.0
```



```
r2(config-if)#no shutdown
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/1
```

```
r2(config-if)#ip add 23.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#router ospf 1
```

```
r2(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r2(config-router)#exit
```

说明：为 R2 的 Loopback 0，F0/0，F0/1 配置接口地址，并将所有接口发布进 OSPF。

(3) 配置 R3:

```
r3(config)#int f0/1
```

```
r3(config-if)#ip add 23.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#exit
```

```
r3(config)#
```

```
r3(config)#int f0/0
```



```
r3(config-if)#ip add 34.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#exit
```

```
r3(config)#router ospf 1
```

```
r3(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r3(config-router)#exit
```

说明：为 R3 的 F0/1，F0/0 配置接口地址，并将所有接口发布进 OSPF。

(4) 配置 R4:

```
r4(config)#int loopback 0
```

```
r4(config-if)#ip add 4.4.4.4 255.255.255.0
```

```
r4(config-if)#ip ospf network point-to-point
```

```
r4(config-if)#exit
```

```
r4(config)#
```

```
r4(config)#int f0/0
```

```
r4(config-if)#ip add 34.1.1.4 255.255.255.0
```

```
r4(config-if)#no shutdown
```

```
r4(config-if)#exit
```

```
r4(config)#int f0/1
```

```
r4(config-if)#ip add 45.1.1.4 255.255.255.0
```

```
r4(config-if)#no shutdown
```

```
r4(config-if)#exit
```

```
r4(config)#
```

```
r4(config)#int s1/0
```

```
r4(config-if)#encapsulation frame-relay
```

```
r4(config-if)#no frame-relay inverse-arp
```

```
r4(config-if)#no arp frame-relay
```

```
r4(config-if)#ip add 46.1.1.4 255.255.255.0
```

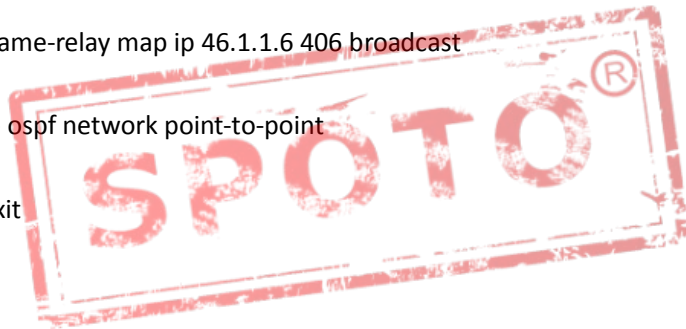
```
r4(config-if)#no shutdown
```

```
r4(config-if)#frame-relay map ip 46.1.1.6 406 broadcast
```

```
r4(config-if)#ip ospf network point-to-point
```

```
r4(config-if)#exit
```

```
r4(config)#
```



说明：为 R4 的 Loopback 0，F0/0，F0/1，S1/0 配置接口地址，并将所有接口发布进 OSPF。

（5）配置 R5：

```
r5(config)#int f0/1
```

```
r5(config-if)#ip add 45.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#exit
```

```
r5(config)#router ospf 1
```

```
r5(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r5(config-router)#exit
```

说明：为 R5 的 F0/1 配置接口地址，并将所有接口发布进 OSPF。

(6) 配置 R6:

```
r6(config)#int s1/0
```

```
r6(config-if)#encapsulation frame-relay
```

```
r6(config-if)#no frame-relay inverse-arp
```

```
r6(config-if)#no arp frame-relay
```

```
r6(config-if)#ip add 46.1.1.6 255.255.255.0
```

```
r6(config-if)#no shutdown
```

```
r6(config-if)#frame-relay map ip 46.1.1.4 604 broadcast
```

```
r6(config-if)#ip ospf network point-to-point
```

```
r6(config-if)#exit
```

```
r6(config)#router ospf 1
```

```
r6(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r6(config-router)#exit
```

说明：为 R6 的 S1/0 配置接口地址，并将所有接口发布进 OSPF。

(7) 查看 R1 的路由学习情况:

```
r1#sh ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

CCIE LAB认证经验分享千人群：539730342

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/3] via 12.1.1.2, 00:02:19, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/2] via 12.1.1.2, 00:02:19, FastEthernet0/0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/4] via 12.1.1.2, 00:02:19, FastEthernet0/0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/2] via 12.1.1.2, 00:02:19, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/67] via 12.1.1.2, 00:02:19, FastEthernet0/0

CCIE LAB认证经验分享千人群：539730342

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/4] via 12.1.1.2, 00:02:19, FastEthernet0/0

r1#

说明：R1 已经学习到全网的每一条路由。

(8) 查看 R2 的路由学习情况：

r2#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:02:50, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/3] via 23.1.1.3, 00:02:50, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

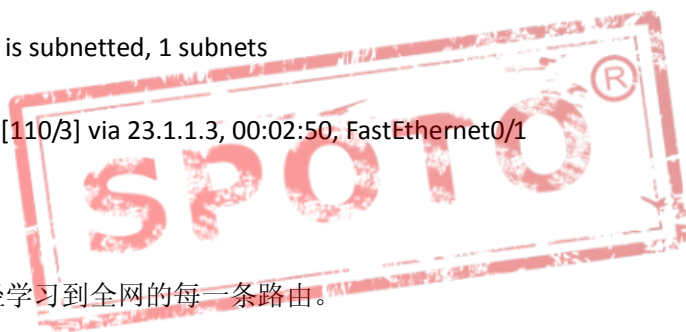
O 46.1.1.0 [110/66] via 23.1.1.3, 00:02:50, FastEthernet0/1

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/3] via 23.1.1.3, 00:02:50, FastEthernet0/1

r2#

说明：R2 已经学习到全网的每一条路由。



(9) 查看 R3 的路由学习情况：

R3:

r3#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

CCIE LAB认证经验分享千人群：539730342

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

C 34.1.1.0 is directly connected, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/2] via 23.1.1.2, 00:03:14, FastEthernet0/1

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/2] via 34.1.1.4, 00:03:14, FastEthernet0/0

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/2] via 23.1.1.2, 00:03:14, FastEthernet0/1

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/65] via 34.1.1.4, 00:03:14, FastEthernet0/0

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/2] via 34.1.1.4, 00:03:15, FastEthernet0/0

r3#

说明：R3 已经学习到全网的每一条路由。

(10) 查看 R4 的路由学习情况：

R4:

r4#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

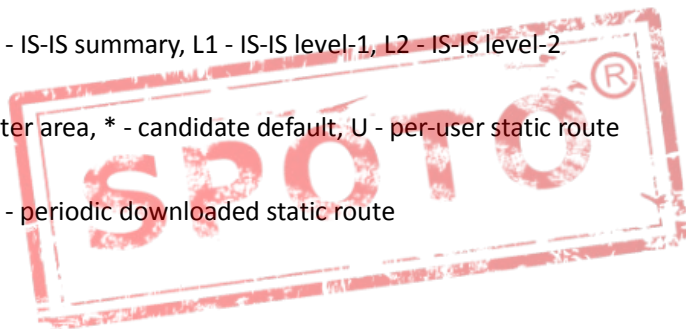
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route



Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

C 34.1.1.0 is directly connected, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/3] via 34.1.1.3, 00:03:22, FastEthernet0/0

4.0.0.0/24 is subnetted, 1 subnets

C 4.4.4.0 is directly connected, Loopback0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/2] via 34.1.1.3, 00:03:22, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/3] via 34.1.1.3, 00:03:22, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

C 46.1.1.0 is directly connected, Serial1/0

45.0.0.0/24 is subnetted, 1 subnets

C 45.1.1.0 is directly connected, FastEthernet0/1

r4#

说明：R4 已经学习到全网的每一条路由。

(11) 查看 R5 的路由学习情况：

R5:

r5#sh ip route



Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 45.1.1.4, 00:03:48, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/4] via 45.1.1.4, 00:03:48, FastEthernet0/1

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/2] via 45.1.1.4, 00:03:48, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/3] via 45.1.1.4, 00:03:48, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/4] via 45.1.1.4, 00:03:48, FastEthernet0/1

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/65] via 45.1.1.4, 00:03:48, FastEthernet0/1

45.0.0.0/24 is subnetted, 1 subnets

C 45.1.1.0 is directly connected, FastEthernet0/1

r5#

说明：R5 已经学习到全网的每一条路由。

(12) 查看 R6 的路由学习情况：

R6:

r6#sh ip route

CCIE LAB认证经验分享千人群：539730342

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/65] via 46.1.1.4, 00:04:11, Serial1/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/67] via 46.1.1.4, 00:04:11, Serial1/0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/65] via 46.1.1.4, 00:04:11, Serial1/0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/66] via 46.1.1.4, 00:04:11, Serial1/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/67] via 46.1.1.4, 00:04:11, Serial1/0

46.0.0.0/24 is subnetted, 1 subnets



C 46.1.1.0 is directly connected, Serial1/0

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/65] via 46.1.1.4, 00:04:11, Serial1/0

r6#

说明：R6 已经学习到全网的每一条路由。

(13) 在 R1 上测试到其它网段的连通性：

r1#ping

r1#ping 2.2.2.2

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 2.2.2.2, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 28/47/80 ms

r1#

r1#ping 4.4.4.4

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 4.4.4.4, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 108/191/240 ms

r1#

r1#ping 45.1.1.5

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 45.1.1.5, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 124/216/288 ms

r1#

r1#ping 46.1.1.6

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 46.1.1.6, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 140/212/268 ms

r1#

说明：R1 与 2.2.2.0/24、4.4.4.0/24、45.1.1.0/24、46.1.1.0/24 通信正常，说明已经全网单播互通。

2. 配置 PIM-SM

(1) 在 R1 上配置 PIM-SM:

```
r1(config)#ip multicast-routing
```

```
r1(config)#
```

```
r1(config)#int f0/0
```

```
r1(config-if)#ip pim sparse-mode
```

```
r1(config-if)#ip igmp join-group 224.1.1.1
```

```
r1(config-if)#exit
```

```
r1(config)#
```

说明：在 R1 上开启组播功能，并且在接口 F0/0 下开启 PIM-SM 模式，接口 F0/0 加入组 224.1.1.1，成为组成员。

(2) 在 R2 上配置 PIM-SM:

```
r2(config)#ip multicast-routing
```

```
r2(config)#int loopback 0
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/0
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/1
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#
```

说明：在 R2 上开启组播功能，并且在接口 Loopback 0，F0/0，F0/1 下开启 PIM-SM 模式；虽然 R3 不参与组播，但连接 R3 的接口 F0/1 还是开了 PIM，是因为如果接口没有开启 PIM，将不参与 RPF 检测，所以为了尽量看到 RPF 检测的效果，尽可能多的开了组播接口。

(3) 在 R4 上配置 PIM-SM:

```
r4(config)#ip multicast-routing
```

```
r4(config)#int loopback 0
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```

```
r4(config)#int f0/1
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```

```
r4(config)#int s1/0
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```



```
r4(config)#
```

说明：在 R3 上开启组播功能，并且在接口 Loopback 0，F0/1，S1/0 下开启 PIM-SM 模式。

（4）在 R5 上配置 PIM-SM：

```
r5(config)#ip multicast-routing
```

```
r5(config)#
```

```
r5(config)#int f0/1
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

说明：在 R5 上开启组播功能，并且在接口 F0/1 下开启 PIM-SM 模式。

（5）在 R6 上配置 PIM-SM：

```
r6(config)#ip multicast-routing
```

```
r6(config)#
```

```
r6(config)#int s1/0
```

```
r6(config-if)#ip pim sparse-mode
```

```
r6(config-if)#exit
```

说明：在 R6 上开启组播功能，并且在接口 S1/0 下开启 PIM-SM 模式。

（6）在 R2 上查看 PIM 邻居情况：

```
r2#sh ip pim neighbor
```


CCIE LAB认证经验分享千人群：539730342

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
12.1.1.1	FastEthernet0/0	00:02:42/00:01:30	v2	1 / S

r2#

说明：在 PIM-SM Domain 1 中，只有 R2 和 R1 是 PIM 邻居。

(7) 在 R4 上查看 PIM 邻居情况：

r4#sh ip pim neighbor

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
45.1.1.5	FastEthernet0/1	00:01:21/00:01:22	v2	1 / DR S
46.1.1.6	Serial1/0	00:00:53/00:01:20	v2	1 / DR S

r4#

说明：在 PIM-SM Domain 2 中，只有 R4 和 R5 与 R6 是 PIM 邻居。

(8) 将 R2 配置为 PIM-SM Domain 1 的 RP:

```
r2(config)#access-list 24 permit 224.1.1.1
```

```
r2(config)#ip pim send-rp-announce loopback 0 scope 16 group-list 24
```

```
r2(config)#ip pim send-rp-discovery loopback 0 scope 16
```

说明: R2 以 Loopback 0 作为组 224.1.1.1 的 RP 和映射代理。

(9) 将 R4 配置为 PIM-SM Domain 1 的 RP:

```
r4(config)#access-list 24 permit 224.1.1.1
```

```
r4(config)#ip pim send-rp-announce loopback 0 scope 16 group-list 24
```

```
r4(config)#ip pim send-rp-discovery loopback 0 scope 16
```

说明: R4 以 Loopback 0 作为组 224.1.1.1 的 RP 和映射代理。

(10) 在 R1 上查看 RP 学习情况:

```
r1#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 2.2.2.2 (?), v2v1

Info source: 2.2.2.2 (?), elected via Auto-RP

Uptime: 00:00:47, expires: 00:02:09

```
r1#
```

CCIE LAB认证经验分享千人群：539730342

说明：和预期一样，PIM-SM Domain 1 的 R1 学习到的 RP 正是 2.2.2.2，因为目前 PIM-SM Domain 1 与 PIM-SM Domain 2 之间没有 PIM 通路，所以两边的 RP 信息互不传递，互不影响。

(11) 在 R2 上查看 RP 学习情况：

```
r2#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback0)

Group(s) 224.1.1.1/32

RP 2.2.2.2 (?), v2v1

Info source: 2.2.2.2 (?), elected via Auto-RP

Uptime: 00:01:03, expires: 00:02:58

```
r2#
```

说明：和预期一样，PIM-SM Domain 1 的 R2 学习到的 RP 也是 2.2.2.2。

(12) 在 R4 上查看 RP 学习情况：

```
r4#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback0)

Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:00:40, expires: 00:02:16

r4#

说明：和预期一样，PIM-SM Domain 2 的 R4 学习到的 RP 正是 4.4.4.4，因为 与 PIM-SM Domain 1 之间没有 PIM 通路，所以两边保持着 PIM-SM 域独立。

(13) 在 R5 上查看 RP 学习情况：

r5#sh ip pim rp mapping

PIM Group-to-RP Mappings



Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:00:58, expires: 00:02:00

r5#

说明：和预期一样，PIM-SM Domain 2 的 R5 学习到的 RP 也是 4.4.4.4。

(14) 在 R6 上查看 RP 学习情况：

r6#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:01:13, expires: 00:02:42

r6#

说明：和预期一样，PIM-SM Domain 2 的 R6 学习到的 RP 也是 4.4.4.4。

(15) 测试 R1 到组 224.1.1.1 的通信情况：

r1#ping 224.1.1.1

Type escape sequence to abort.

Sending 1, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 4 ms

r1#

说明：在单个域内，组播通信没问题。

(16) 测试 R2 到组 224.1.1.1 的通信情况：

CCIE LAB认证经验分享千人群：539730342

r2#ping 224.1.1.1

Type escape sequence to abort.

Sending 1, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 168 ms

Reply to request 0 from 12.1.1.1, 180 ms

r2#

(17) 测试 PIM-SM Domain 2 到组 224.1.1.1 的通信情况:

r4#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r4#

r5#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r5#

R6:

r6#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r6#

说明：因为 PIM-SM Domain 2 与 PIM-SM Domain 1 之间连最基本的 PIM 通路都没有，所以组播不通是正常的，要通了才怪了。



3. 配置 GRE Tunnel 隧道

(1) 在 R2 上配置连接 R4 的 GRE Tunnel 隧道:

r2(config)#int tunnel 24

r2(config-if)#ip address 24.1.1.2 255.255.255.0

r2(config-if)#tunnel source loopback 0

r2(config-if)#tunnel destination 4.4.4.4

r2(config-if)#exit

CCIE LAB认证经验分享千人群：539730342

说明：配置连接 PIM-SM Domain 1 和 PIM-SM Domain 2 的 GRE Tunnel 隧道，使用双方的 Loopback 0 地址作为 GRE Tunnel 隧道两端的端点。

(2) 在 R4 上配置连接 R2 的 GRE Tunnel 隧道：

```
r4(config)#int tunnel 24
```

```
r4(config-if)#ip add 24.1.1.4 255.255.255.0
```

```
r4(config-if)#tunnel source loopback 0
```

```
r4(config-if)#tunnel destination 2.2.2.2
```

```
r4(config-if)#exit
```

说明：配置连接 PIM-SM Domain 1 和 PIM-SM Domain 2 的 GRE Tunnel 隧道

(3) 测试 GRE Tunnel 隧道的连通性：

```
r2#ping 24.1.1.4
```

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 24.1.1.4, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 124/223/360 ms

```
r2#
```

```
r4#ping 24.1.1.2
```


CCIE LAB认证经验分享千人群：539730342

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 24.1.1.2, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 72/145/276 ms

r4#

说明：目前连接 PIM-SM Domain 1 和 PIM-SM Domain 2 的 GRE Tunnel 隧道通信正常。

(4) 再次测试 PIM-SM Domain 2 到组 224.1.1.1 的通信情况：

r4#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r4#

r5#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:



.....

r5#

说明：即使配置了连接 PIM-SM Domain 1 和 PIM-SM Domain 2 的 GRE Tunnel 隧道，但域间组播还是无法通信，因为两个 PIM-SM 域之间连最基本的 PIM 通路都没有，如果没有 PIM 通路，这种做了 GRE Tunnel 隧道的环境跟没做是一样的。

(5) 在 R4 上查看关于组 224.1.1.1 的 PIM-SM 组播树情况：

r4#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:06:46/stopped, RP 4.4.4.4, flags: SP

CCIE LAB认证经验分享千人群：539730342

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(4.4.4.4, 224.1.1.1), 00:01:01/00:02:24, flags: PT

Incoming interface: Loopback0, RPF nbr 0.0.0.0

Outgoing interface list: Null

(45.1.1.4, 224.1.1.1), 00:01:01/00:02:16, flags: P

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list: Null

(45.1.1.5, 224.1.1.1), 00:00:50/00:02:34, flags: PT

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list: Null

(46.1.1.4, 224.1.1.1), 00:01:00/00:02:16, flags: P

Incoming interface: Serial1/0, RPF nbr 0.0.0.0

Outgoing interface list: Null

r4#

说明：由以上组播树可以看出，R4 根本就没有能够去往 PIM-SM Domain 1 的组播出口。

(6) 将 GRE Tunnel 隧道加入 PIM，形成 PIM-SM Domain 1 和 PIM-SM Domain 2 的 PIM 通道：

```
r2(config)#int tunnel 24
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r4(config)#int tunnel 24
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```

说明：R2 和 R4 都已经在 GRE Tunnel 隧道上开启 PIM 功能，但不要忘记，GRE Tunnel 隧道接口也已经被通告进了 OSPF 进程，因为在配置 OSPF 时，是将所有接口通告进去的。

(7) 查看 R4 的 PIM 邻居情况：

```
r4#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
45.1.1.5	FastEthernet0/1	00:14:49/00:01:42	v2	1 / DR S

46.1.1.6 Serial1/0 00:14:22/00:01:39 v2 1 / DR S

24.1.1.2 Tunnel24 00:00:25/00:01:19 v2 1 / S

r4#

说明：R4 已经通过 GRE Tunnel 隧道和 R2 形成 PIM 邻居关系，说明 PIM-SM Domain 1 已经通过 PIM 和 PIM-SM Domain 2 正式连通。

（8）查看所有组播路由器的 RP 情况：

r1#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 2.2.2.2 (?), elected via Auto-RP

Uptime: 00:09:09, expires: 00:02:58

r1#

r2#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback0)

Group(s) 224.1.1.1/32

CCIE LAB认证经验分享千人群：539730342

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 00:12:08, expires: 00:02:00

RP 2.2.2.2 (?), v2v1

Info source: 2.2.2.2 (?), via Auto-RP

Uptime: 01:21:36, expires: 00:02:55

r2#

r4#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback0)



Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 01:21:02, expires: 00:02:45

r4#

r5#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 01:21:14, expires: 00:02:19

r5#

r5#

r6#sh ip pim rp mapping

PIM Group-to-RP Mappings



Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 01:21:29, expires: 00:02:06

r6#

说明: 因为 PIM-SM Domain 1 和 PIM-SM Domain 2 已经在 GRE Tunnel 隧道上通过 PIM 连通，所以所有路由器就等于在同一个 PIM-SM 大域里面，由于 RP 4.4.4.4 的地址高于 RP 2.2.2.2，导致所有组播路由器将得到 RP 是 4.4.4.4 的消息，因此要规划 PIM-SM 域间组播，就必须在

PIM-SM Domain 1 和 PIM-SM Domain 2 之间划分域边界来隔离两个组播网络。

(9) 在 R2 上划分 PIM-SM 域边界:

```
r2(config)#access-list 1 deny 224.0.1.39
```

```
r2(config)#access-list 1 deny 224.0.1.40
```

```
r2(config)#access-list 1 permit any
```

```
r2(config)#int tunnel 24
```

```
r2(config-if)#ip multicast boundary 1
```

```
r2(config-if)#exit
```

```
r2(config)#
```

说明: 因为 GRE Tunnel 隧道是导致 PIM-SM Domain 1 和 PIM-SM Domain 2 成为同一个 PIM-SM 域的根本原因, 所以在 GRE Tunnel 隧道上将分发 RP 信息的 Auto-RP 流量过滤掉, 从而实现 PIM-SM 域的分割, 因为过滤是对进和出流量同时生效, 所以只需要在 R2 一边做即可, 不需要在 R4 上重复做。

(10) 再次查看所有组播路由器的 RP 情况:

```
r1#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 2.2.2.2 (?), v2v1

Info source: 2.2.2.2 (?), elected via Auto-RP

CCIE LAB认证经验分享千人群：539730342

Uptime: 00:00:44, expires: 00:02:15

r1#

r2#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback0)

Group(s) 224.1.1.1/32

RP 2.2.2.2 (?), v2v1

Info source: 2.2.2.2 (?), elected via Auto-RP

Uptime: 01:26:30, expires: 00:02:00

r2#

r4#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback0)

Group(s) 224.1.1.1/32



RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 01:25:52, expires: 00:02:58

r4#

r5#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 01:26:05, expires: 00:02:26

r5#

r5#

r6#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32



RP 4.4.4.4 (?), v2v1

Info source: 4.4.4.4 (?), elected via Auto-RP

Uptime: 01:26:16, expires: 00:02:19

r6#

说明：划分 PIM-SM 域边界后，PIM-SM Domain 1 中的 R1 和 R2 学习到 2.2.2.2 是 RP，PIM-SM Domain 2 中的 R4，R5 和 R6 学习到 4.4.4.4 是 RP，和预期的一样，从而实现了 PIM-SM Domain 1 和 PIM-SM Domain 2 的 PIM-SM 域独立。

4. 配置 MSDP

(1) 在 PIM-SM Domain 1 与 PIM-SM Domain 2 之间配置 MSDP 连接：

```
r2(config)#ip msdp peer 4.4.4.4 connect-source loopback 0
```

```
r4(config)#ip msdp peer 2.2.2.2 connect-source loopback 0
```

说明：在 PIM-SM Domain 1 的 RP 路由器 R2 和 PIM-SM Domain 2 的路由器 R4 上使用 Loopback 0 建立 MSDP 连接。

(2) 查看 MSDP 的连接情况：

```
r2#sh ip msdp summary
```

MSDP Peer Status Summary

CCIE LAB认证经验分享千人群：539730342

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

		Downtime	Count	Count	
--	--	----------	-------	-------	--

4.4.4.4	?	Down	00:00:48	0	0	?
---------	---	------	----------	---	---	---

r2#

r4#sh ip msdp summary

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

		Downtime	Count	Count	
--	--	----------	-------	-------	--

2.2.2.2	?	Listen	00:00:27	0	0	?
---------	---	--------	----------	---	---	---

r4#

说明：因为 R4 的 IP 地址大，所以停留在 Listen 状态对方发起连接。

(3) 再次查看 MSDP 的连接情况：

r2#sh ip msdp summary

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

		Downtime	Count	Count	
--	--	----------	-------	-------	--

4.4.4.4	?	Up	00:01:07	0	0	?
---------	---	----	----------	---	---	---

r2#

CCIE LAB认证经验分享千人群：539730342

```
r4#sh ip msdp summary
```

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

		Downtime	Count	Count	
--	--	----------	-------	-------	--

2.2.2.2	?	Up	00:01:16	0	?
---------	---	----	----------	---	---

```
r4#
```

说明：MSDP 已经正常连接，说明已经可以传递 SA 信息。

(4) 在 R2 上查看 SA 信息：

```
r2#sh ip msdp sa-cache
```

MSDP Source-Active Cache - 0 entries

```
r2#
```

说明：因为目前没有路由器发起组播流量，也就没有组播源，那么也就不会产生 SA 信息。

(5) 在组播源路由器 R5 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

在 MSDP 路由器 R2 上开启 Debug:

```
r2#debug ip msdp peer
```

MSDP Peer debugging is on

```
r2#
```

```
r2#debug ip msdp detail
```

CCIE LAB认证经验分享千人群：539730342

MSDP Detail debugging is on

r2#

在组播源路由器 R5 上发起组播流量：

r5#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r5#

说明：组播不通，后面将根据实际情况一步一步解决。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 00:49:41.563: MSDP(0): Received 20-byte TCP segment from 4.4.4.4

*Mar 1 00:49:41.567: MSDP(0): Append 20 bytes to 0-byte msg 6 from 4.4.4.4, qs 1

*Mar 1 00:49:41.567: MSDP(0): 4.4.4.4: Received 20-byte msg 6 from peer

*Mar 1 00:49:41.567: MSDP(0): 4.4.4.4: SA TLV, len: 20, ec: 1, RP: 4.4.4.4

*Mar 1 00:49:41.571: MSDP(0): 4.4.4.4: Peer RPF check passed for single peer

*Mar 1 00:49:41.571: MSDP(0): WAVL Insert SA Source 45.1.1.5 Group 224.1.1.1 RP 4.4.4.4 Successful

r2#

说明：因为目前 R2 只有单个 MSDP peer，所以对于收到的 SA 也就没必要进行 RPF

检测，最后 R2 顺利将起始 PIM-SM 域的 MSDP 路由器 4.4.4.4 发来的 SA 缓存起来，从 Debug 信息中可以看出，SA 信息中的组播源是 45.1.1.5，正是我们发起组播流量的 R5，而组地址 224.1.1.1 也是正确的，对方的 RP 是 4.4.4.4 也正确。

(6) 在 MSDP 路由器 R2 上查看 SA 信息的接收情况：

```
r2#sh ip msdp sa-cache
```

```
MSDP Source-Active Cache - 1 entries
```

```
(45.1.1.5, 224.1.1.1), RP 4.4.4.4, AS 1, 00:01:04/00:04:55, Peer 4.4.4.4
```

```
Learned from peer 4.4.4.4, RPF peer 4.4.4.4,
```

```
SAs received: 1, Encapsulated data received: 0
```

```
r2#
```

说明：MSDP 路由器 R2 已经正常接收到 PIM-SM Domain 1 的 SA，但组播还是不通，这个原因在理论部分我们已经重点强调过，MSDP 的责任只是负责 SA 的传递，只要 SA 已经收到，那么 MSDP 的工作就已经完成，如果组播仍然不通，就不再是 MSDP 的问题，一定是其它问题，而通常影响组播不通的最大问题可能就是组播流的 RPF 检测，所以下面我们着手检查组播流 RPF 的情况，从而解决组播的通信问题。

因为普通组播流量的 RPF 检测是先查看发送该数据的源 IP 地址是什么，然后在路由表中查看去往源 IP 的数据包该从哪个接口出去，那么从哪个接口收到的数据包就被认为是合法有效的，而查看的路由表共包括：单播路由表，MBGP 路由表，DVMRP 路由表以及静态组播路由表，但目前我们只有单播路由表，其它都没有配，所以我们首先从单播路由表下手，并且也只能从单播路由表下手。

(7) 在 R2 上查看去往组播源 45.1.1.5 的单播路由表情况：

```
r2#sh ip route
```

```
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
```

```
D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
```

CCIE LAB认证经验分享千人群：539730342

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:06:01, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/3] via 23.1.1.3, 00:06:01, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/1

24.0.0.0/24 is subnetted, 1 subnets

C 24.1.1.0 is directly connected, Tunnel24

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

CCIE LAB认证经验分享千人群：539730342

O 46.1.1.0 [110/66] via 23.1.1.3, 00:06:01, FastEthernet0/1

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/3] via 23.1.1.3, 00:06:01, FastEthernet0/1

r2#

说明：从 R2 的单播路由表可知，R2 认为去往组播源 45.1.1.5 的流量应该从接口 F0/1 出去，那么源地址为 45.1.1.5 的流量也必须从接口 F0/1 进来，如果从其它接口进来就被认为是不合法的，统统都会被丢弃，但是我们应该知道，从 PIM-SM Domain 1 发来的组播流量一定是从 GRE Tunnel 隧道进来的，因为是 GRE Tunnel 隧道通过 PIM 将 PIM-SM Domain 1 与 PIM-SM Domain 2 连接起来的。

(8) 查看 GRE Tunnel 隧道在 OSPF 下的情况：

r2#sh ip ospf interface tunnel 24

Tunnel24 is up, line protocol is up

Internet Address 24.1.1.2/24, Area 0

Process ID 1, Router ID 2.2.2.2, Network Type POINT_TO_POINT, Cost: 11111

Transmit Delay is 1 sec, State POINT_TO_POINT

Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5

oob-resync timeout 40

Hello due in 00:00:02

Supports Link-local Signaling (LLS)

Cisco NSF helper support enabled

IETF NSF helper support enabled

Index 4/4, flood queue length 0

CCIE LAB认证经验分享千人群：539730342

Next 0x0(0)/0x0(0)

Last flood scan length is 1, maximum is 2

Last flood scan time is 0 msec, maximum is 4 msec

Neighbor Count is 1, Adjacent neighbor count is 1

Adjacent with neighbor 4.4.4.4

Suppress hello for 0 neighbor(s)

r2#

说明：虽然 GRE Tunnel 隧道接口已经被通告进 OSPF 进程，但 R2 去往组播源 45.1.1.5 的路径还是通过物理接口 F0/1 出去的，仍然没有走 GRE Tunnel 隧道，从上可以看出，这是因为 GRE Tunnel 隧道的 OSPF Cost 为 11111，远大于走 F0/1 的路径。

(9) 查看 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

r2#sh ip rpf 45.1.1.5

RPF information for ? (45.1.1.5)

RPF interface: FastEthernet0/1

RPF neighbor: ? (23.1.1.3)

RPF route/mask: 45.1.1.0/24

RPF type: unicast (ospf 1)

RPF recursion count: 0

Doing distance-preferred lookups across tables

r2#

说明：从以上可以看出，R2 的系统对 45.1.1.5 的 RPF 检测结果为从物理接口 F0/1 出

去，和我们之前所看到的一样，并且这个结果是通过 unicast (ospf 1)的信息获得的，但事实上这个结果与真实的组播路径不吻合，所以组播流量遭到了丢弃，如果要让组播正常通信，就要改变系统的 RPF 检测结果，让其认为从 GRE Tunnel 隧道进来的组播流量才是正确的。

5. 通过单播路由表解决 PIM-SM 域间组播通信问题

说明：因为 R2 的系统对 45.1.1.5 的 RPF 检测结果是通过 unicast (ospf 1)的信息获得的，OSPF 路由的 AD 值为 110，这时我们完全可以通过手工配置单播静态路由来取代 OSPF 学习到的路由，手工配置单播静态路由将 45.1.1.5 指向 GRE Tunnel 隧道，让从 GRE Tunnel 隧道过来的组播流量能够通过 RPF 检测。

(1) 在 R2 上配置单播静态路由：

```
r2(config)#ip route 45.1.1.5 255.255.255.255 tunnel 24
```

说明：手工配置单播静态路由将组播源地址 45.1.1.5 指向 GRE Tunnel 隧道，让从 GRE Tunnel 隧道过来的组播流量能够通过 RPF 检测。

(2) 在 R2 上再次查看去往组播源 45.1.1.5 的单播路由表情况：

```
r2#sh ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

CCIE LAB认证经验分享千人群：539730342

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:13:37, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/3] via 23.1.1.3, 00:13:37, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/1

24.0.0.0/24 is subnetted, 1 subnets

C 24.1.1.0 is directly connected, Tunnel24

12.0.0.0/24 is subnetted, 1 subnets

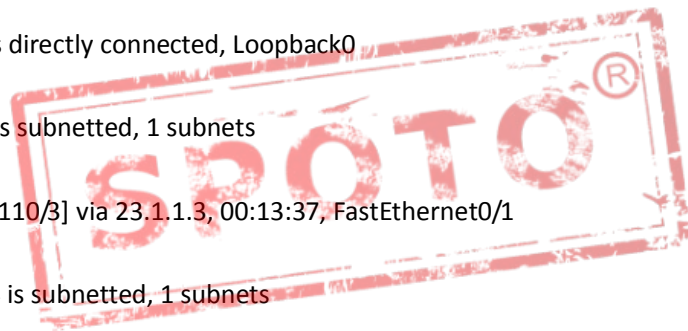
C 12.1.1.0 is directly connected, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/66] via 23.1.1.3, 00:13:37, FastEthernet0/1

45.0.0.0/8 is variably subnetted, 2 subnets, 2 masks

S 45.1.1.5/32 is directly connected, Tunnel24



CCIE LAB认证经验分享千人群：539730342

```
O    45.1.1.0/24 [110/3] via 23.1.1.3, 00:13:37, FastEthernet0/1
```

```
r2#
```

说明：手工配置的将组播源地址 45.1.1.5 指向 GRE Tunnel 隧道的单播静态路由已经生效，虽然 OSPF 路由保持不变，但静态路由的子网掩码为/32 位，优于任何路由。

（3）再次查看 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

```
r2#sh ip rpf 45.1.1.5
```

```
RPF information for ? (45.1.1.5)
```

```
RPF interface: Tunnel24
```

```
RPF neighbor: ? (24.1.1.4)
```

```
RPF route/mask: 45.1.1.5/32
```

```
RPF type: unicast (static)
```

```
RPF recursion count: 0
```

```
Doing distance-preferred lookups across tables
```

```
r2#
```

说明：通过手工配置将组播源地址 45.1.1.5 指向 GRE Tunnel 隧道的单播静态路由之后，目前系统的 RPF 检测结果已经成功认为从 GRE Tunnel 隧道接口进来的组播流量是合法的了，这样一来，PIM-SM Domain 1 与 PIM-SM Domain 2 之间的组播就可以通信了。

（4）再次测试从 PIM-SM Domain 2 的组播源 45.1.1.5 向 PIM-SM Domain 1 发送组播：

```
r5#ping 224.1.1.1 repeat 10
```

CCIE LAB认证经验分享千人群：539730342

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 328 ms

Reply to request 0 from 12.1.1.1, 516 ms

Reply to request 1 from 12.1.1.1, 316 ms

Reply to request 2 from 12.1.1.1, 236 ms

Reply to request 3 from 12.1.1.1, 236 ms

Reply to request 4 from 12.1.1.1, 268 ms

Reply to request 5 from 12.1.1.1, 196 ms

Reply to request 6 from 12.1.1.1, 236 ms

Reply to request 7 from 12.1.1.1, 212 ms

Reply to request 8 from 12.1.1.1, 268 ms

Reply to request 9 from 12.1.1.1, 204 ms

r5#

说明：已经通过手工配置单播静态路由成功解决 PIM-SM Domain 1 与 PIM-SM Domain 2 的域间组播通信。

查看当前环境中所有路由器的配置信息请点击 “[running-config](#)”

6. 通过静态组播路由表解决 PIM-SM 域间组播通信问题

(1) 将之前手工配置的单播静态路由删除，以恢复组播故障问题：

```
r2(config)#no ip route 45.1.1.5 255.255.255.255 tunnel 24
```

```
r2(config)#exit
```

```
r2#sh ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:15:10, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/3] via 23.1.1.3, 00:15:10, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/1

24.0.0.0/24 is subnetted, 1 subnets

C 24.1.1.0 is directly connected, Tunnel24

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/66] via 23.1.1.3, 00:15:10, FastEthernet0/1

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/3] via 23.1.1.3, 00:15:11, FastEthernet0/1

r2#

说明：之前手工配置的将组播源地址 45.1.1.5 指向 GRE Tunnel 隧道的单播静态路由已被删除，目前又重新选择了错误的 OSPF 路由。

(2) 查看当前 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

r2#sh ip rpf 45.1.1.5

RPF information for ? (45.1.1.5)

RPF interface: FastEthernet0/1

RPF neighbor: ? (23.1.1.3)

RPF route/mask: 45.1.1.0/24

RPF type: unicast (ospf 1)

RPF recursion count: 0

CCIE LAB认证经验分享千人群：539730342

Doing distance-preferred lookups across tables

r2#

说明：R2 的系统对 45.1.1.5 的 RPF 检测结果又回到了依靠 OSPF 做检测的状态，此状态为错误状态。

(3) 测试从 PIM-SM Domain 2 的组播源 45.1.1.5 向 PIM-SM Domain 1 发送组播：

r5#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r5#

说明：删除了手工配置的单播静态路由之后，组播问题依旧。

(4) 在 R2 上配置组播静态路由：

r2(config)#ip mroute 45.1.1.5 255.255.255.255 tunnel 24

说明：手工配置组播静态路由将组播源地址 45.1.1.5 指向 GRE Tunnel 隧道，让从 GRE Tunnel 隧道过来的组播流量能够通过 RPF 检测。

(5) 在 R2 上查看组播静态路由表：

r2#sh ip mroute static

Mroute: 45.1.1.5/32, interface: Tunnel24

CCIE LAB认证经验分享千人群：539730342

Protocol: none, distance: 0, route-map: none

r2#

说明：手工配置的将组播源地址 45.1.1.5 指向 GRE Tunnel 隧道的组播静态路由已生效。

（6）再次查看 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

r2#sh ip rpf 45.1.1.5

RPF information for ? (45.1.1.5)

RPF interface: Tunnel24

RPF neighbor: ? (24.1.1.4)

RPF route/mask: 45.1.1.5/32

RPF type: unicast (static)

RPF recursion count: 0

Doing distance-preferred lookups across tables

r2#

说明：和预期的一样，系统的 RPF 检测结果已经成功认为从 GRE Tunnel 隧道接口进来的组播流量是合法的。

（7）再次测试从 PIM-SM Domain 2 的组播源 45.1.1.5 向 PIM-SM Domain 1 发送组播：

r5#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 384 ms

Reply to request 1 from 12.1.1.1, 236 ms

Reply to request 2 from 12.1.1.1, 264 ms

Reply to request 3 from 12.1.1.1, 236 ms

Reply to request 4 from 12.1.1.1, 224 ms

Reply to request 5 from 12.1.1.1, 248 ms

Reply to request 6 from 12.1.1.1, 280 ms

Reply to request 7 from 12.1.1.1, 220 ms

Reply to request 8 from 12.1.1.1, 232 ms.

r5#

说明：通过手工配置组播静态路由也能成功解决 PIM-SM Domain 1 与 PIM-SM Domain 2 的域间组播通信。

查看当前环境中所有路由器的配置信息请点击 “[running-config](#)”

7. 通过 MBGP 解决 PIM-SM 域间组播通信问题

说明：除了单播路由表和静态组播路由表之外，还可以通过 MBGP 来解决组播流的 RPF 检测问题，当然不要忘记还有 DVMRP，但这不是我们讨论的范围，如果这 4 种

CCIE LAB认证经验分享千人群：539730342

路由表同时出现，则靠路由条目的 AD 值来决定选谁，但如果 AD 值全部一样，请注意他们的先后顺序是：

1. 静态组播路由表
2. DVMRP 路由表
3. MBGP 路由表
4. 单播路由表

(1) 在 R2 上删除组播静态路由，恢复组播故障：

```
r2(config)#no ip mroute 45.1.1.5 255.255.255.255 tunnel 24
```

```
r2(config)#exit
```

```
r2#
```

```
r2#
```

```
r2#sh ip mroute static
```



```
r2#
```

```
r2#sh ip rpf 45.1.1.5
```

```
RPF information for ? (45.1.1.5)
```

```
RPF interface: FastEthernet0/1
```

```
RPF neighbor: ? (23.1.1.3)
```

```
RPF route/mask: 45.1.1.0/24
```

```
RPF type: unicast (ospf 1)
```

```
RPF recursion count: 0
```

Doing distance-preferred lookups across tables

r2#

r5#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r5#

说明：从上面看到，删除组播静态路由后，系统对组播源 45.1.1.5 的 RPF 检测又回到了选择 OSPF 的错误状态，并且最后组播不通。

(2) 在 R2 和 R4 之间配置 MBGP:

R2:

r2(config)#router bgp 2

r2(config-router)#bgp router-id 2.2.2.2

r2(config-router)#neighbor 4.4.4.4 remote-as 4

r2(config-router)#neighbor 4.4.4.4 update-source loopback 0

r2(config-router)#neighbor 4.4.4.4 ebgp-multihop

r2(config-router)#address-family ipv4 multicast

r2(config-router-af)#neighbor 4.4.4.4 activate

```
r2(config-router-af)#exit
```

```
r2(config-router)#exit
```

```
r2(config)#
```

R4:

```
r4(config)#router bgp 4
```

```
r4(config-router)#bgp router-id 4.4.4.4
```

```
r4(config-router)#neighbor 2.2.2.2 remote-as 2
```

```
r4(config-router)#neighbor 2.2.2.2 update-source loopback 0
```

```
r4(config-router)#neighbor 2.2.2.2 ebgp-multihop
```

```
r4(config-router)#address-family ipv4 multicast
```

```
r4(config-router-af)#neighbor 2.2.2.2 activate
```

```
r4(config-router-af)#exit
```

```
r4(config-router)#exit
```

```
r4(config)#
```

说明：R2 与 R4 通过双方的 Loopback 0 来建 MBGP 邻居，所以请注意，通过 Loopback 0 来建 MBGP 邻居，那么 Loopback 0 的地址就会影响到 MBGP 的路由走向，因为 BGP 的路由是根据到达邻居的地址然后去 IGP 路由表做递归路由查询的。

(3) 查看 BGP 与 MBGP 邻居状态:

```
r2#sh ip bgp summary
```

CCIE LAB认证经验分享千人群：539730342

BGP router identifier 2.2.2.2, local AS number 2

BGP table version is 1, main routing table version 1

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
----------	---	----	---------	---------	--------	-----	------	---------	--------------

4.4.4.4	4	4	4	4	1	0	0	00:01:11	0
---------	---	---	---	---	---	---	---	----------	---

r2#

r2#show ip bgp ipv4 multicast summary

BGP router identifier 2.2.2.2, local AS number 2

BGP table version is 1, main routing table version 1

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
----------	---	----	---------	---------	--------	-----	------	---------	--------------

4.4.4.4	4	4	5	5	1	0	0	00:02:09	0
---------	---	---	---	---	---	---	---	----------	---

r2#

说明：BGP 和 MBGP 的邻居均已建立。

(4) 查看 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

r2#sh ip rpf 45.1.1.5

RPF information for ? (45.1.1.5)

RPF interface: FastEthernet0/1

RPF neighbor: ? (23.1.1.3)

RPF route/mask: 45.1.1.0/24

RPF type: unicast (ospf 1)

RPF recursion count: 0

Doing distance-preferred lookups across tables

r2#

说明：虽然已经配置了 MBGP，但是 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果还是选择了使用 OSPF 路由，导致最终结果错误。

(5) 查看 MBGP 路由表：

r2#show ip bgp ipv4 multicast

r2#

说明：因为之前并没有在 MBGP 里面通告关于组播源地址 45.1.1.5 的路由信息，所以虽然配置了 MBGP，但毫无作用，所以需要关于组播源地址 45.1.1.5 的路由发布进 MBGP。

(6) 在 MBGP 中发布路由：

r4(config)#router bgp 4

r4(config-router)#address-family ipv4 multicast

r4(config-router-af)#network 45.1.1.0 mask 255.255.255.0

r4(config-router-af)#exit


```
r4(config-router)#exit
```

说明：在 R4 上将组播源地址 45.1.1.5 的路由发布进 MBGP，使 45.1.1.5 能够依靠 MBGP 路由顺利通过 RPF 检测。

(7) 再次查看 MBGP 路由表：

```
r2#show ip bgp ipv4 multicast
```

BGP table version is 2, local router ID is 2.2.2.2

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,

r RIB-failure, S Stale

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 45.1.1.0/24	4.4.4.4	0		0 4	i

```
r2#
```

说明：已经从 MBGP 正常学习到关于组播源地址 45.1.1.5 的路由，该路由可用作对 45.1.1.5 的 RPF 检测。

(8) 再次查看 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

```
r2#sh ip rpf 45.1.1.5
```

RPF information for ? (45.1.1.5)

RPF interface: FastEthernet0/1

RPF neighbor: ? (23.1.1.3)

CCIE LAB认证经验分享千人群：539730342

RPF route/mask: 45.1.1.0/24

RPF type: mbgp

RPF recursion count: 0

Doing distance-preferred lookups across tables

r2#

说明：从上面可以看出，R2 的系统已经使用 MBGP 的路由对组播源地址 45.1.1.5 进行 RPF 检测，但是检测的结果却是错误的，这里我们需要注意，从之前的 MBGP 路由表可以得知，MBGP 认为去往 45.1.1.5 的数据包应该发到 4.4.4.4，即发给它的 MBGP 邻居 R4，然后 4.4.4.4 并不是与之直连的网络，所以 MBGP 就得需要到单播路由表中做递归查询去往 4.4.4.4 究竟该如何去。

(9) 查询 R2 去往 4.4.4.4 的单播路由表：

r2#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

CCIE LAB认证经验分享千人群：539730342

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:10:15, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

4.0.0.0/24 is subnetted, 1 subnets

O 4.4.4.0 [110/3] via 23.1.1.3, 00:10:15, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/1

24.0.0.0/24 is subnetted, 1 subnets

C 24.1.1.0 is directly connected, Tunnel24

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/66] via 23.1.1.3, 00:10:16, FastEthernet0/1

45.0.0.0/24 is subnetted, 1 subnets

O 45.1.1.0 [110/3] via 23.1.1.3, 00:10:16, FastEthernet0/1

r2#

说明：从单播路由表中可以看出，去 4.4.4.4 的数据包应该从物理接口 F0/1 出去，所以 MBGP 路由表的结果就是去 45.1.1.5 的数据包也应该顺着 4.4.4.4 的路径从物理接口 F0/1 出去，这就与我们预期的从 GRE Tunnel 隧道出去不相符，所以我们只有想办法让 R2 从 GRE Tunnel 隧道接口去 4.4.4.4，才能改变让 MBGP 去往 45.1.1.5 从 GRE Tunnel 隧道出去，这样对组播源 45.1.1.5 的 RPF 检测才能变成 GRE Tunnel 隧道。

CCIE LAB认证经验分享千人群：539730342

(10) 在 R2 上手工配置从 GRE Tunnel 隧道去 4.4.4.4 的静态路由：

```
r2(config)#ip route 4.4.4.4 255.255.255.255 24.1.1.4
```

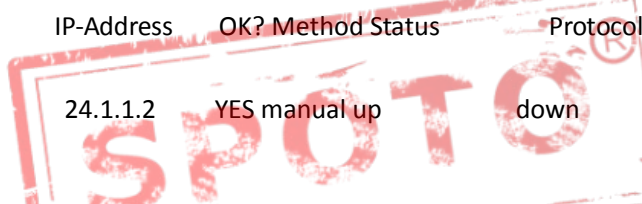
说明：配置静态路由从 GRE Tunnel 隧道去 4.4.4.4，在这里，请你注意，请你一定要注意，如果你不注意这个，那么通过这种方式配置的 MBGP 将不生效，将前功尽弃，需要注意的就是这条静态路由一定要指定下一跳地址，而不能指定出接口，所以你千万不要配成 “ip route 4.4.4.4 255.255.255.255 tunnel 24”，如果你这么配了，你可以试一下，你的组播看似正常，但却不通。

(11) 在 R2 上查看 GRE Tunnel 隧道情况：

```
r2#sh ip int brief tunnel 24
```

Interface	IP-Address	OK?	Method	Status	Protocol
Tunnel24	24.1.1.2	YES	manual up	up	down

r2#



说明：无意间，两个 PIM-SM 域之间的关键点 GRE Tunnel 隧道的状态居然变成了 “down”，这是因为我们配置的 GRE Tunnel 隧道是靠 R2 与 R4 双方的 Loopback 0 来通信的，而刚才我们配置的静态路由居然要让去往 R4 的 Loopback 0（4.4.4.4）靠 GRE Tunnel 隧道来通信，这里就形成了死循环，因为必须双方的 Loopback 0 通了，GRE Tunnel 隧道才能通，但现在又变成了必须 GRE Tunnel 隧道通了，Loopback 0 才能过，这就前后矛盾了，所以我们既然强制 Loopback 0 靠 GRE Tunnel 隧道来通信，那么我们就必须让 GRE Tunnel 隧道靠其它接口来通信而不能靠 Loopback 0 口，那我们就可以更改 GRE Tunnel 隧道靠 R2 与 R4 的物理口来通信。

(12) 更改 GRE Tunnel 隧道的通信方式：

R2:

```
r2(config)#int tunnel 24
```

```
r2(config-if)#tunnel source f0/1
```

```
r2(config-if)#tunnel destination 34.1.1.4
```

```
r2(config-if)#exit
```

```
r2(config)#
```

R4:

```
r4(config)#int tunnel 24
```

```
r4(config-if)#tunnel source f0/0
```

```
r4(config-if)#tunnel destination 23.1.1.2
```

```
r4(config-if)#exit
```

说明：更改 GRE Tunnel 隧道靠 R2 与 R4 的物理口来通信。

(13) 再次查看 GRE Tunnel 隧道情况：

```
r2#sh ip int brief tunnel 24
```

Interface	IP-Address	OK?	Method	Status	Protocol
Tunnel24	24.1.1.2	YES	manual	up	up

```
r2#
```

```
r2#ping 24.1.1.4
```

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 24.1.1.4, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 88/164/216 ms

r2#

说明：更改 GRE Tunnel 隧道靠 R2 与 R4 的物理口来通信之后，目前 GRE Tunnel 隧道工作正常。

（14）再次查看 R2 的系统对组播源 45.1.1.5 的 RPF 检测结果：

r2#sh ip rpf 45.1.1.5

RPF information for ? (45.1.1.5)

RPF interface: Tunnel24

RPF neighbor: ? (24.1.1.4)

RPF route/mask: 45.1.1.0/24

RPF type: mbgp

RPF recursion count: 1

Doing distance-preferred lookups across tables

r2#

说明：现在 R2 的系统已经靠 MBGP 的路由正确对组播源 45.1.1.5 做出了 RPF 检测，如果您之前将从 GRE Tunnel 隧道去 4.4.4.4 的静态路由 “ip route 4.4.4.4 255.255.255.255 24.1.1.4” 误写成指向出接口的 “ip route 4.4.4.4 255.255.255.255 tunnel 24”，那么 R2 的系统对组播源 45.1.1.5 做出的 RPF 检测结果将如下：

r2#sh ip rpf 45.1.1.5

RPF information for ? (45.1.1.5)

RPF interface: Tunnel24



CCIE LAB认证经验分享千人群：539730342

RPF neighbor: ? (4.4.4.4)

RPF route/mask: 45.1.1.0/24

RPF type: mbgp

RPF recursion count: 0

Doing distance-preferred lookups across tables

r2#

而这个结果将使域间组播无法通信，您应该能发现两者的 RPF neighbor 有所区别，正确的应该是真正的直连地址，错误的却是远程非直连地址。

(15) 再次测试从 PIM-SM Domain 2 的组播源 45.1.1.5 向 PIM-SM Domain 1 发送组播：

r5#ping 224.1.1.1 repeat 10



Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 232 ms

Reply to request 1 from 12.1.1.1, 188 ms

Reply to request 2 from 12.1.1.1, 236 ms

Reply to request 2 from 12.1.1.1, 316 ms

Reply to request 3 from 12.1.1.1, 200 ms

Reply to request 4 from 12.1.1.1, 204 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 5 from 12.1.1.1, 356 ms.

Reply to request 7 from 12.1.1.1, 236 ms

Reply to request 8 from 12.1.1.1, 216 ms

Reply to request 9 from 12.1.1.1, 204 ms

r5#

说明：通过配置 MBGP 也成功解决了 PIM-SM Domain 1 与 PIM-SM Domain 2 的域间组播通信，请认真观察，认真体会三种方法解决组播通信的关键所在。

附：在这里需要说明：在配置 MBGP 解决 PIM-SM 域间组播通信时，我们遇到的问题是由于 MBGP 的路由依靠单播路由表对 MBGP 路由的下一跳做递归查询时，错将组播源地址走向了物理接口而并非 GRE Tunnel 隧道接口，所以我们将被 MBGP 路由用来作递归查询的单播路由静态指向了 GRE Tunnel 隧道口，从而使得 MBGP 在单播路由表做递归查询时能被正确引导到 GRE Tunnel 隧道接口，在这里还有另外一个做法就是直接使用 GRE Tunnel 隧道接口地址来建立 MBGP 邻居，这样 MBGP 的路由不需要做递归查询就知道直接走 GRE Tunnel 隧道，从而就不会有以上那么多步骤，但还请不要忘记，并不建议使用 GRE Tunnel 隧道建立 BGP 或 MBGP 邻居，通常是保持建立 MSDP 连接的地址和建立 BGP 或 MBGP 的地址为相同地址，因为这样做可以保证在出现 3 个或 3 个以上的 PIM-SM 域时对 SA 数据包所做的 RPF 检测能够顺利通过，虽然我们这里只有两个 PIM-SM 域。

如果使用 GRE Tunnel 隧道建邻居，则 R2 和 R4 的配置如下：

```
r2(config)#router bgp 2
```

```
r2(config-router)#no neighbor 4.4.4.4
```

```
r2(config-router)#neighbor 24.1.1.4 remote-as 4
```

```
r2(config-router)#address-family ipv4 multicast
```

```
r2(config-router-af)#neighbor 24.1.1.4 activate
```



```
r2(config-router-af)#exit
```

```
r4(config)#router bgp 4
```

```
r4(config-router)#no neighbor 2.2.2.2
```

```
r4(config-router)#neighbor 24.1.1.2 remote-as 2
```

```
r4(config-router)#address-family ipv4 multicast
```

```
r4(config-router-af)#neighbor 24.1.1.2 activate
```

```
r4(config-router-af)#network 45.1.1.0 mask 255.255.255.0
```

```
r4(config-router-af)#exit
```



(16) 测试从 PIM-SM Domain 2 的组播源 46.1.1.6 (即 R6) 向 PIM-SM Domain 1 发送组播:

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

```
r6#
```

说明：种种原因，PIM-SM Domain 2 的组播源 46.1.1.6 (即 R6) 与 PIM-SM Domain 1 的组播还无法通信。

(17) 在 R2 上查看 SA 信息：

```
r2#sh ip msdp sa-cache
```

```
MSDP Source-Active Cache - 1 entries
```

```
(45.1.1.5, 224.1.1.1), RP 4.4.4.4, BGP/AS 0, 00:02:53/00:05:43, Peer 4.4.4.4
```

```
r2#
```

说明：R2 上连关于组播源 46.1.1.6 的 SA 都没有，问题很严重，其实我们不难看出，R5 和 R6 的架构完全相同，但唯一的区别就是 R6 是使用的帧中继链路与 R4 相连的，而很不巧的是，R6 的 S1/0 地址大于 R4 的 S1/0 地址，所以在 PIM 选举 DR 时，结果 R6 成了 DR，如下：

```
r4#sh ip pim neighbor
```

```
PIM Neighbor Table
```

```
Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,
```

```
S - State Refresh Capable
```

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
24.1.1.2	Tunnel24	00:06:04/00:01:34	v2	1 / S
45.1.1.5	FastEthernet0/1	00:07:33/00:01:33	v2	1 / DR S
46.1.1.6	Serial1/0	00:05:59/00:01:41	v2	1 / DR S

```
r4#
```

(18) 提高 R4 的 S1/0 的 DR 优先级：

```
r4(config)#int s1/0
```

CCIE LAB认证经验分享千人群：539730342

```
r4(config-if)#ip pim dr-priority 4
```

```
r4(config-if)#exit
```

```
r4#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
----------	-----------	----------------	-----	----

Address			Prio/Mode	
---------	--	--	-----------	--

24.1.1.2	Tunnel24	00:06:25/00:01:44	v2 1 / S	
----------	----------	-------------------	----------	--

45.1.1.5	FastEthernet0/1	00:07:54/00:01:43	v2 1 / DR S	
----------	-----------------	-------------------	-------------	--

46.1.1.6	Serial1/0	00:06:19/00:01:21	v2 1 / S	
----------	-----------	-------------------	----------	--

```
r4#
```

说明：提高 R4 的 S1/0 接口的 DR 优先级后，R4 成为了 DR，这里不再详细讨论到底是因为什么，不清楚的请复习组播前部分的理论知识。

(19) R6 再次发起组播数据并在 R2 上查看 SA 信息：

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r6#

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 2 entries

(45.1.1.5, 224.1.1.1), RP 4.4.4.4, BGP/AS 0, 00:04:21/00:05:14, Peer 4.4.4.4

(46.1.1.6, 224.1.1.1), RP 4.4.4.4, BGP/AS 0, 00:00:45/00:05:14, Peer 4.4.4.4

r2#

说明：组播仍然不通，但关于组播源 46.1.1.6 的 SA 已经正常传递，那么剩下的问题就不可能是 MSDP 的问题。

(20) 查看 R2 的系统关于组播源 46.1.1.6 的 RPF 检测结果：

r2#sh ip rpf 46.1.1.6

RPF information for ? (46.1.1.6)

RPF interface: FastEthernet0/1

RPF neighbor: ? (23.1.1.3)

RPF route/mask: 46.1.1.0/24

RPF type: unicast (ospf 1)

RPF recursion count: 0

Doing distance-preferred lookups across tables

r2#

说明：由于 R2 的系统还是基于 OSPF 路由对组播源 46.1.1.6 做的 RPF 检测，所以路

径错误。

(21)在 R2 上配置组播静态路由解决 PIM-SM Domain 2 的组播源 46.1.1.6 与 PIM-SM Domain 1 的组播通信：

```
r2(config)#ip mroute 46.1.1.6 255.255.255.255 tunnel 24
```

说明：解决 PIM-SM Domain 2 的组播源 46.1.1.6 与 PIM-SM Domain 1 组播通信的方法有

单播路由、MBGP、组播静态路由，而我们这里采用了组播静态路由。

(22) 再次查看 R2 的系统关于组播源 46.1.1.6 的 RPF 检测结果：

```
r2#sh ip rpf 46.1.1.6
```

RPF information for ? (46.1.1.6)

RPF interface: Tunnel24

RPF neighbor: ? (24.1.1.4)

RPF route/mask: 46.1.1.6/32

RPF type: static

RPF recursion count: 0

Doing distance-preferred lookups across tables

```
r2#
```

说明：配置了组播静态路由后，R2 对组播源 46.1.1.6 做的 RPF 检测成功指向 GRE Tunnel 隧道。

(23) 再次测试从 PIM-SM Domain 2 的组播源 46.1.1.6 向 PIM-SM Domain 1 发送组播：

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 420 ms

Reply to request 1 from 12.1.1.1, 312 ms

Reply to request 2 from 12.1.1.1, 312 ms

Reply to request 3 from 12.1.1.1, 204 ms

Reply to request 4 from 12.1.1.1, 200 ms

Reply to request 5 from 12.1.1.1, 280 ms

Reply to request 6 from 12.1.1.1, 216 ms

Reply to request 7 from 12.1.1.1, 248 ms

Reply to request 8 from 12.1.1.1, 248 ms

Reply to request 9 from 12.1.1.1, 220 ms

```
r6#
```

说明：所有问题解决后，PIM-SM Domain 1 与 PIM-SM Domain 2 之间的组播畅通无阻。

(24) 查看组播树：

R2:

CCIE LAB认证经验分享千人群：539730342

```
r2#sh ip mroute 224.1.1.1
```

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 00:22:53/00:02:32, RP 2.2.2.2, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:22:53/00:02:40

(45.1.1.5, 224.1.1.1), 00:03:29/00:02:56, flags: MT

Incoming interface: Tunnel24, RPF nbr 24.1.1.4, Mbgp

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:03:29/00:02:40

(46.1.1.6, 224.1.1.1), 00:16:55/00:02:56, flags: MT

Incoming interface: Tunnel24, RPF nbr 24.1.1.4, Mroute

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:16:55/00:02:40

r2#

R4:

r4#sh ip mroute 224.1.1.1



IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

CCIE LAB认证经验分享千人群：539730342

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:21:18/stopped, RP 4.4.4.4, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(45.1.1.5, 224.1.1.1), 00:04:02/00:03:23, flags: TA

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list:

Tunnel24, Forward/Sparse, 00:02:00/00:03:18



(46.1.1.6, 224.1.1.1), 00:21:06/00:03:23, flags: TA

Incoming interface: Serial1/0, RPF nbr 0.0.0.0

Outgoing interface list:

Tunnel24, Forward/Sparse, 00:16:38/00:02:51

r4#

说明：从组播树中可以看见，PIM-SM 域之间的通信使用的是 (S, G)，而非 PIM-SM 模式传统的 (*, G)，如果您之前将从 GRE Tunnel 隧道去 4.4.4.4 的静态路由“ip route 4.4.4.4 255.255.255.255 24.1.1.4 ”误写成指向出接口的“ip route 4.4.4.4 255.255.255.255 tunnel 24”，那么组播树将如下：

CCIE LAB认证经验分享千人群：539730342

r4#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:22:34/stopped, RP 4.4.4.4, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(45.1.1.5, 224.1.1.1), 00:05:18/00:02:56, flags: PTA

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list: Null

CCIE LAB认证经验分享千人群：539730342

(46.1.1.6, 224.1.1.1), 00:22:22/00:03:26, flags: TA

Incoming interface: Serial1/0, RPF nbr 0.0.0.0

Outgoing interface list:

Tunnel24, Forward/Sparse, 00:17:55/00:02:33

r4#

R2:

r2#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode



CCIE LAB认证经验分享千人群：539730342

(* , 224.1.1.1), 00:25:03/00:00:22, RP 2.2.2.2, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:25:03/00:02:27

(45.1.1.5, 224.1.1.1), 00:05:39/00:02:26, flags: M

Incoming interface: Tunnel24, RPF nbr 4.4.4.4, Mbgp

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:05:39/00:02:27

(46.1.1.6, 224.1.1.1), 00:19:05/00:02:56, flags: MT

Incoming interface: Tunnel24, RPF nbr 24.1.1.4, Mroute

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:19:05/00:02:27

r2#

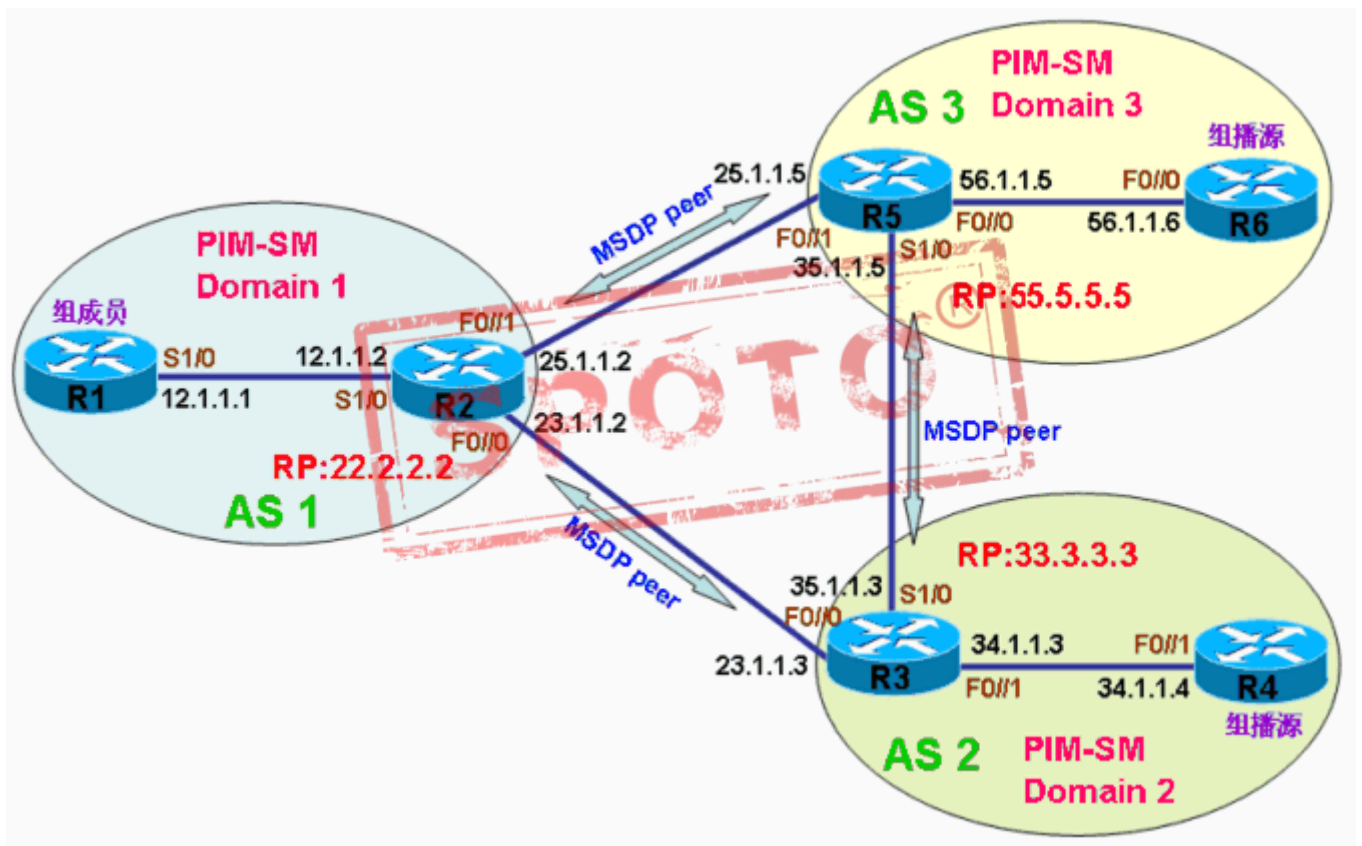
说明：R2 上正常的 flags 为 MT，而写错静态路由的 flags 为 M；R4 上正常的 flags 为 TA，而写错静态路由的 flags 为 PTA。

查看当前环境中所有路由器的配置信息请点击 “[running-config](#)”

配置 3 个 PIM-SM 域全互联的 MSDP 实验

实验说明：

以下图为环境配置 3 个 PIM-SM 域全互联的 MSDP 实验：



上图中 3 个 PIM-SM 域互连，其中 R1 和 R2 在 PIM-SM Domain 1 中，R3 和 R4 在 PIM-SM Domain 2 中，R5 和 R6 在 PIM-SM Domain 3 中；R4 和 R6 为组 224.1.1.1 的组播源，R1 为组成员。

除了图上所标识出的接口地址外，R2、R3 和 R5 还配有 2 个 Loopback 接口，分别为：

R2 (Loopback 0 :2.2.2.2/24; Loopback 1 :22.2.2.2/24)

R3 (Loopback 0 :3.3.3.3/24; Loopback 1 :33.3.3.3/24)

R5 (Loopback 0 :5.5.5.5/24; Loopback 1 :55.5.5.5/24)

PIM-SM Domain 1 的 RP 为 22.2.2.2，即 R2，PIM-SM Domain 2 的 RP 为 33.3.3.3，即 R3，PIM-SM Domain 3 的 RP 为 55.5.5.5，即 R5；

3 个 PIM-SM 域之间建立全互联的 MSDP 连接，连接情况如下：

R2 (2.2.2.2) —— R3 (3.3.3.3)

R2 (2.2.2.2) —— R5 (55.5.5.5)

R3 (3.3.3.3) —— R5 (55.5.5.5)

所有路由器上都运行 OSPF，并将所有接口都发布进 OSPF，以保证全网单播互通。



1. 配置初始网络环境

(1) 配置 R1:

```
r1(config)#int s1/0
```

```
r1(config-if)#encapsulation frame-relay
```

```
r1(config-if)#no frame-relay inverse-arp
```

```
r1(config-if)#no arp frame-relay
```

```
r1(config-if)#ip add 12.1.1.1 255.255.255.0
```

```
r1(config-if)#no shutdown
```

```
r1(config-if)#frame-relay map ip 12.1.1.2 102 broadcast
```

```
r1(config-if)#ip ospf network point-to-point
```

```
r1(config-if)#exit
```

```
r1(config)#router ospf 1
```

```
r1(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r1(config-router)#exit
```

说明：为 R1 的 S1/0 配置接口地址，并将所有接口发布进 OSPF。

（2）配置 R2：

```
r2(config)#int loopback 0
```

```
r2(config-if)#ip add 2.2.2.2 255.255.255.0
```

```
r2(config-if)#ip ospf network point-to-point
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int loopback 22
```

```
r2(config-if)#ip add 22.2.2.2 255.255.255.0
```

```
r2(config-if)#ip ospf network point-to-point
```

```
r2(config-if)#exit
```

```
r2(config)#
```

CCIE LAB认证经验分享千人群：539730342

```
r2(config)#int s1/0
```

```
r2(config-if)#encapsulation frame-relay
```

```
r2(config-if)#no frame-relay inverse-arp
```

```
r2(config-if)#no arp frame-relay
```

```
r2(config-if)#ip add 12.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#frame-relay map ip 12.1.1.1 201 broadcast
```

```
r2(config-if)#ip ospf network point-to-point
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int f0/0
```

```
r2(config-if)#ip add 23.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/1
```

```
r2(config-if)#ip add 25.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#exit
```

```
r2(config)#router ospf 1
```

```
r2(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r2(config-router)#exit
```



说明：为 R2 的 Loopback 0， Loopback 22， S1/0， F0/0， F0/1 配置接口地址，并将所有接口发布进 OSPF。

(3) 配置 R3:

```
r3(config)#int loopback 0
```

```
r3(config-if)#ip add 3.3.3.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#ip ospf network point-to-point
```

```
r3(config-if)#exit
```

```
r3(config)#
```

```
r3(config)#int loopback 33
```

```
r3(config-if)#ip add 33.3.3.3 255.255.255.0
```

```
r3(config-if)#ip ospf network point-to-point
```

```
r3(config-if)#exit
```

```
r3(config)#int f0/0
```

```
r3(config-if)#ip add 23.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#exit
```

```
r3(config)#
```

```
r3(config)#int f0/1
```

```
r3(config-if)#ip add 34.1.1.3 255.255.255.0
```



```
r3(config-if)#no shutdown
```

```
r3(config-if)#exit
```

```
r3(config)#int s1/0
```

```
r3(config-if)#encapsulation frame-relay
```

```
r3(config-if)#no frame-relay inverse-arp
```

```
r3(config-if)#no arp frame-relay
```

```
r3(config-if)#ip add 35.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#frame-relay map ip 35.1.1.5 305 broadcast
```

```
r3(config-if)#ip ospf network point-to-point
```

```
r3(config-if)#exit
```

```
r3(config)#router ospf 1
```

```
r3(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r3(config-router)#exit
```

```
r3(config)#
```

说明：为 R3 的 Loopback 0，Loopback 33，S1/0，F0/1，F0/0 配置接口地址，并将所有接口发布进 OSPF。

(4) 配置 R4:

```
r4(config)#int f0/1
```

```
r4(config-if)#ip add 34.1.1.4 255.255.255.0
```

```
r4(config-if)#no shutdown
```

```
r4(config-if)#exit
```

```
r4(config)#router ospf 1
```

```
r4(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r4(config-router)#exit
```

```
r4(config)#
```

说明：为 R4 的 F0/1 配置接口地址，并将所有接口发布进 OSPF。

(5) 配置 R5:

```
r5(config)#int loopback 0
```

```
r5(config-if)#ip address 5.5.5.5 255.255.255.0
```

```
r5(config-if)#ip ospf network point-to-point
```

```
r5(config-if)#exit
```

```
r5(config)#
```

```
r5(config)#int loopback 55
```

```
r5(config-if)#ip add 55.5.5.5 255.255.255.0
```

```
r5(config-if)#ip ospf network point-to-point
```

```
r5(config-if)#exit
```

```
r5(config)#int f0/0
```

```
r5(config-if)#ip address 56.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#exit
```

```
r5(config)#int f0/1
```

```
r5(config-if)#ip add 25.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#exit
```

```
r5(config)#
```

```
r5(config)#int s1/0
```

```
r5(config-if)#encapsulation frame-relay
```

```
r5(config-if)#no frame-relay inverse-arp
```

```
r5(config-if)#no arp frame-relay
```

```
r5(config-if)#ip add 35.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#frame-relay map ip 35.1.1.3 503 broadcast
```

```
r5(config-if)#ip ospf network point-to-point
```

```
r5(config-if)#exit
```

说明：为 R5 的 Loopback 0，Loopback 55，S1/0，F0/1，F0/0 配置接口地址，并将所有接口发布进 OSPF。

（6）配置 R6：

```
r6(config)#int f0/0
```

```
r6(config-if)#ip address 56.1.1.6 255.255.255.0
```

```
r6(config-if)#no shutdown
```

```
r6(config-if)#exit
```

```
r6(config)#
```

说明：为 R6 的 F0/0 配置接口地址，并将所有接口发布进 OSPF。

(7) 查看 R1 的路由学习情况：

```
r1#sh ip route
```

```
*Mar  1 00:28:10.043: %SYS-5-CONFIG_I: Configured from console by console
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/66] via 12.1.1.2, 00:01:23, Serial1/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/129] via 12.1.1.2, 00:01:23, Serial1/0

CCIE LAB认证经验分享千人群：539730342

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/65] via 12.1.1.2, 00:01:23, Serial1/0

33.0.0.0/24 is subnetted, 1 subnets

O 33.3.3.0 [110/66] via 12.1.1.2, 00:01:23, Serial1/0

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/66] via 12.1.1.2, 00:01:23, Serial1/0

55.0.0.0/24 is subnetted, 1 subnets

O 55.5.5.0 [110/66] via 12.1.1.2, 00:01:23, Serial1/0

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/66] via 12.1.1.2, 00:01:24, Serial1/0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/65] via 12.1.1.2, 00:01:24, Serial1/0

22.0.0.0/24 is subnetted, 1 subnets

O 22.2.2.0 [110/65] via 12.1.1.2, 00:01:24, Serial1/0

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/65] via 12.1.1.2, 00:01:24, Serial1/0

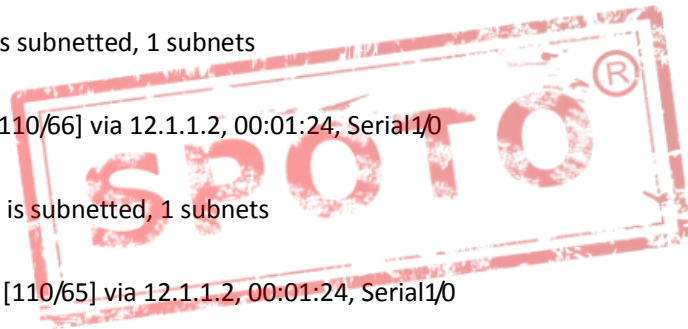
56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/66] via 12.1.1.2, 00:01:24, Serial1/0

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, Serial1/0

r1#



说明： R1 已经学习到全网的每一条路由。

(8) 查看 R2 的路由学习情况：

r2#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

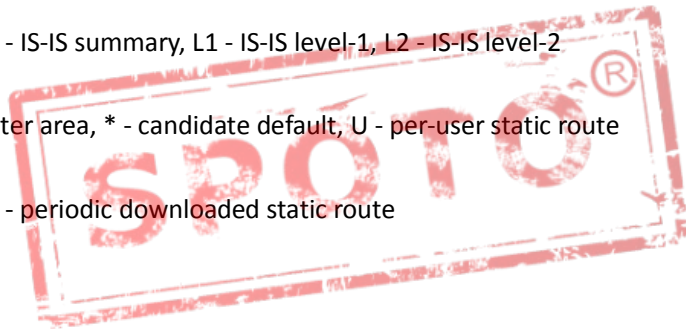
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route



Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:01:41, FastEthernet0/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 25.1.1.5, 00:01:41, FastEthernet0/1

[110/65] via 23.1.1.3, 00:01:41, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

CCIE LAB认证经验分享千人群：539730342

33.0.0.0/24 is subnetted, 1 subnets

O 33.3.3.0 [110/2] via 23.1.1.3, 00:01:41, FastEthernet0/0

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/2] via 23.1.1.3, 00:01:41, FastEthernet0/0

55.0.0.0/24 is subnetted, 1 subnets

O 55.5.5.0 [110/2] via 25.1.1.5, 00:01:41, FastEthernet0/1

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/2] via 25.1.1.5, 00:01:42, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/0

22.0.0.0/24 is subnetted, 1 subnets

C 22.2.2.0 is directly connected, Loopback22

25.0.0.0/24 is subnetted, 1 subnets

C 25.1.1.0 is directly connected, FastEthernet0/1

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/2] via 25.1.1.5, 00:01:42, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, Serial1/0

r2#

说明：R2 已经学习到全网的每一条路由。

(9) 查看 R3 的路由学习情况：

```
r3#sh ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

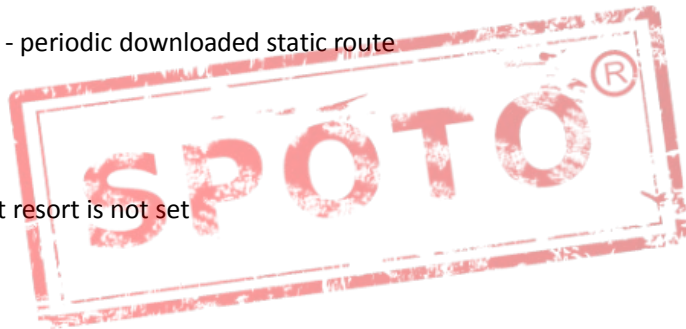
E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set



34.0.0.0/24 is subnetted, 1 subnets

C 34.1.1.0 is directly connected, FastEthernet0/1

35.0.0.0/24 is subnetted, 1 subnets

C 35.1.1.0 is directly connected, Serial1/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/2] via 23.1.1.2, 00:01:56, FastEthernet0/0

33.0.0.0/24 is subnetted, 1 subnets

C 33.3.3.0 is directly connected, Loopback33

3.0.0.0/24 is subnetted, 1 subnets

CCIE LAB认证经验分享千人群：539730342

C 3.3.3.0 is directly connected, Loopback0

55.0.0.0/24 is subnetted, 1 subnets

O 55.5.5.0 [110/3] via 23.1.1.2, 00:01:56, FastEthernet0/0

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/3] via 23.1.1.2, 00:01:57, FastEthernet0/0

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/0

22.0.0.0/24 is subnetted, 1 subnets

O 22.2.2.0 [110/2] via 23.1.1.2, 00:01:57, FastEthernet0/0

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/2] via 23.1.1.2, 00:01:57, FastEthernet0/0

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/3] via 23.1.1.2, 00:01:57, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/65] via 23.1.1.2, 00:01:57, FastEthernet0/0

r3#

说明：R3 已经学习到全网的每一条路由。

(10) 查看 R4 的路由学习情况：

r4#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

CCIE LAB认证经验分享千人群：539730342

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

C 34.1.1.0 is directly connected, FastEthernet0/1

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 34.1.1.3, 00:02:09, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/3] via 34.1.1.3, 00:02:09, FastEthernet0/1

33.0.0.0/24 is subnetted, 1 subnets

O 33.3.3.0 [110/2] via 34.1.1.3, 00:02:09, FastEthernet0/1

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/2] via 34.1.1.3, 00:02:09, FastEthernet0/1

55.0.0.0/24 is subnetted, 1 subnets

O 55.5.5.0 [110/4] via 34.1.1.3, 00:02:09, FastEthernet0/1



CCIE LAB认证经验分享千人群：539730342

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/4] via 34.1.1.3, 00:02:10, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/2] via 34.1.1.3, 00:02:10, FastEthernet0/1

22.0.0.0/24 is subnetted, 1 subnets

O 22.2.2.0 [110/3] via 34.1.1.3, 00:02:10, FastEthernet0/1

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/3] via 34.1.1.3, 00:02:10, FastEthernet0/1

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/4] via 34.1.1.3, 00:02:10, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/66] via 34.1.1.3, 00:02:10, FastEthernet0/1

r4#

说明：R4 已经学习到全网的每一条路由。

(11) 查看 R5 的路由学习情况：

r5#sh ip route

*Mar 1 00:29:03.083: %SYS-5-CONFIG_I: Configured from console by console

r5#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

CCIE LAB认证经验分享千人群：539730342

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/3] via 25.1.1.2, 00:02:25, FastEthernet0/1

35.0.0.0/24 is subnetted, 1 subnets

C 35.1.1.0 is directly connected, Serial1/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/2] via 25.1.1.2, 00:02:25, FastEthernet0/1

33.0.0.0/24 is subnetted, 1 subnets

O 33.3.3.0 [110/3] via 25.1.1.2, 00:02:25, FastEthernet0/1

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/3] via 25.1.1.2, 00:02:25, FastEthernet0/1

55.0.0.0/24 is subnetted, 1 subnets

C 55.5.5.0 is directly connected, Loopback55

CCIE LAB认证经验分享千人群：539730342

5.0.0.0/24 is subnetted, 1 subnets

C 5.5.5.0 is directly connected, Loopback0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/2] via 25.1.1.2, 00:02:25, FastEthernet0/1

22.0.0.0/24 is subnetted, 1 subnets

O 22.2.2.0 [110/2] via 25.1.1.2, 00:02:25, FastEthernet0/1

25.0.0.0/24 is subnetted, 1 subnets

C 25.1.1.0 is directly connected, FastEthernet0/1

56.0.0.0/24 is subnetted, 1 subnets

C 56.1.1.0 is directly connected, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/65] via 25.1.1.2, 00:02:25, FastEthernet0/1

r5#

说明：R5 已经学习到全网的每一条路由。

(12) 查看 R6 的路由学习情况：

r6#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

CCIE LAB认证经验分享千人群：539730342

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/4] via 56.1.1.5, 00:02:38, FastEthernet0/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 56.1.1.5, 00:02:38, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/3] via 56.1.1.5, 00:02:38, FastEthernet0/0

33.0.0.0/24 is subnetted, 1 subnets

O 33.3.3.0 [110/4] via 56.1.1.5, 00:02:38, FastEthernet0/0

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/4] via 56.1.1.5, 00:02:38, FastEthernet0/0

55.0.0.0/24 is subnetted, 1 subnets

O 55.5.5.0 [110/2] via 56.1.1.5, 00:02:38, FastEthernet0/0

5.0.0.0/24 is subnetted, 1 subnets

CCIE LAB认证经验分享千人群：539730342

O 5.5.5.0 [110/2] via 56.1.1.5, 00:02:38, FastEthernet0/0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/3] via 56.1.1.5, 00:02:38, FastEthernet0/0

22.0.0.0/24 is subnetted, 1 subnets

O 22.2.2.0 [110/3] via 56.1.1.5, 00:02:38, FastEthernet0/0

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/2] via 56.1.1.5, 00:02:38, FastEthernet0/0

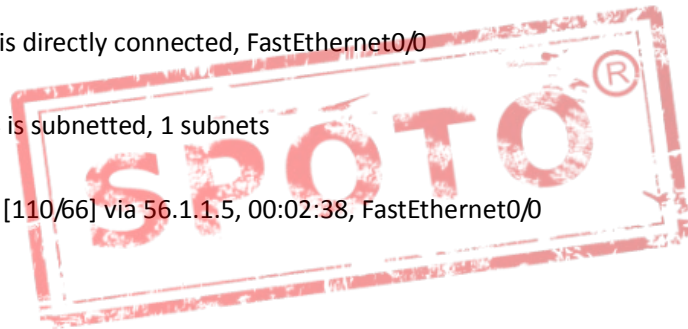
56.0.0.0/24 is subnetted, 1 subnets

C 56.1.1.0 is directly connected, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/66] via 56.1.1.5, 00:02:38, FastEthernet0/0

r6#



说明：R6 已经学习到全网的每一条路由。

(13) 在 R1 上测试到其它网段的连通性：

r1#ping 2.2.2.2

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 2.2.2.2, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 8/98/200 ms

CCIE LAB认证经验分享千人群：539730342

r1#

r1#ping 22.2.2.2

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 22.2.2.2, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 64/136/196 ms

r1#

r1#

r1#ping 3.3.3.3

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 3.3.3.3, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 76/124/172 ms

r1#

r1#ping 33.3.3.3

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 33.3.3.3, timeout is 2 seconds:

!!!!



CCIE LAB认证经验分享千人群：539730342

Success rate is 100 percent (5/5), round-trip min/avg/max = 60/136/240 ms

r1#

r1#

r1#ping 5.5.5.5

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 5.5.5.5, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 64/135/252 ms

r1#

r1#ping 55.5.5.5



Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 55.5.5.5, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 64/80/104 ms

r1#

r1#ping 34.1.1.4

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 34.1.1.4, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 76/191/276 ms

r1#

r1#

r1#ping 56.1.1.6

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 56.1.1.6, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 88/134/220 ms

r1#

说明：R1 与 2.2.2.0/24、22.2.2.0/24、3.3.3.0/24、33.3.3.0/24、5.5.5.0/24、55.5.5.0/24、34.1.1.0/24、56.1.1.0/24 通信正常，说明已经全网单播互通。

2. 配置 PIM-SM

(1) 在 R1 上配置 PIM-SM:

```
r1(config)#ip multicast-routing
```

```
r1(config)#int s1/0
```

```
r1(config-if)#ip pim sparse-mode
```

```
r1(config-if)#ip igmp join-group 224.1.1.1
```

```
r1(config-if)#exit
```

说明：在 R1 上开启组播功能，并且在接口 S1/0 下开启 PIM-SM 模式，接口 S1/0 加入组 224.1.1.1，成为组成员。

（2）在 R2 上配置 PIM-SM：

```
r2(config)#ip multicast-routing
```

```
r2(config)#int s1/0
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/0
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int f0/1
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#int loopback 22
```

```
r2(config-if)#ip pim sparse-mode
```


```
r2(config-if)#exit
```

```
r2(config)#
```

说明：在 R2 上开启组播功能，并且在接口 Loopback 22, F0/0, F0/1, S1/0 下开启 PIM-SM 模式，因为 Loopback 0 只用来建 MSDP 邻居，并不运行组播，所以没有必要运行 PIM。

（3）在 R3 上配置 PIM-SM:

```
r3(config)#int f0/0  
  
r3(config-if)#ip pim sparse-mode  
  
r3(config-if)#exit  
  
r3(config)#  
  
r3(config)#int f0/1  
  
r3(config-if)#ip pim sparse-mode  
  
r3(config-if)#exit  
  
r3(config)#int s1/0  
  
r3(config-if)#ip pim sparse-mode  
  
r3(config-if)#exit  
  
r3(config-if)#int loopback 33  
  
r3(config-if)#ip pim sparse-mode  
  
r3(config-if)#exit  
  
r3(config)#
```



说明: 在 R3 上开启组播功能, 并且在接口 Loopback 33, F0/0, F0/1, S1/0 下开启 PIM-SM 模式, 因为 Loopback 0 只用来建 MSDP 邻居, 并不运行组播, 所以没有必要运行 PIM。

（4）在 R4 上配置 PIM-SM:

```
r4(config)#ip multicast-routing
```

```
r4(config)#int f0/1
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```

说明：在 R4 上开启组播功能，并且在接口 F0/1 下开启 PIM-SM 模式。

（5）在 R5 上配置 PIM-SM：

```
r5(config)#ip multicast-routing
```

```
r5(config)#int f0/1
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

```
r5(config)#int f0/0
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

```
r5(config)#
```

```
r5(config)#int s1/0
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

```
r5(config)#int loopback 55
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```



说明：在 R5 上开启组播功能，并且在接口 Loopback 55, F0/0, F0/1, S1/0 下开启 PIM-SM 模式。

(6) 在 R6 上配置 PIM-SM:

```
r6(config)#ip multicast-routing
```

```
r6(config)#int f0/0
```

```
r6(config-if)#ip pim sparse-mode
```

```
r6(config-if)#exit
```

说明：在 R6 上开启组播功能，并且在接口 F0/0 下开启 PIM-SM 模式

(7) 在 R2 上查看 PIM 邻居情况:

```
r2#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
12.1.1.1	Serial1/0	00:09:22/00:01:15	v2	1 / S
23.1.1.3	FastEthernet0/0	00:04:22/00:01:19	v2	1 / DR S
25.1.1.5	FastEthernet0/1	00:01:59/00:01:43	v2	1 / DR S

```
r2#
```

CCIE LAB认证经验分享千人群：539730342

说明： R2 与 R1， R3 以及 R5 建立 PIM 邻居关系，一切正常。

(8) 在 R3 上查看 PIM 邻居情况：

```
r3#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address			Prio/Mode	
23.1.1.2	FastEthernet0/0	00:04:41/00:01:29	v2	1 / S
34.1.1.4	FastEthernet0/1	00:02:42/00:01:29	v2	1 / DR S
35.1.1.5	Serial1/0	00:02:07/00:01:35	v2	1 / DR S

r3#

说明： R3 与 R2， R4 以及 R5 建立 PIM 邻居关系，一切正常。

(9) 在 R5 上查看 PIM 邻居情况：

```
r5#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
----------	-----------	----------------	-----	----

CCIE LAB认证经验分享千人群：539730342

Address		Prio/Mode
25.1.1.2	FastEthernet0/1	00:02:02/00:01:39 v2 1 / S
56.1.1.6	FastEthernet0/0	00:01:08/00:01:35 v2 1 / DR S
35.1.1.3	Serial1/0	00:02:20/00:01:22 v2 1 / S

r5#

说明：R5 与 R2, R3 以及 R6 建立 PIM 邻居关系，一切正常。

(10) 将 R2 配置为 PIM-SM Domain 1 的 RP:

r2(config)#access-list 1 deny 224.0.1.39

r2(config)#access-list 1 deny 224.0.1.40

r2(config)#access-list 1 permit any

r2(config)#

r2(config)#access-list 24 permit 224.1.1.1

r2(config)#

r2(config)#int f0/0

r2(config-if)#ip multicast boundary 1

r2(config-if)#exit

r2(config)#

r2(config)#int f0/1

r2(config-if)#ip multicast boundary 1

r2(config-if)#exit



CCIE LAB认证经验分享千人群：539730342

```
r2(config)#ip pim send-rp-announce loopback 22 scope 16 group-list 24
```

```
r2(config)#ip pim send-rp-discovery loopback 22 scope 16
```

说明：R2 以 Loopback 22 作为组 224.1.1.1 的 RP 和映射代理；因为网络划分为 PIM-SM Domain 1、PIM-SM Domain 2 以及 PIM-SM Domain 3 共 3 个 PIM-SM 域，所以必须配置 PIM-SM 域边界，这里将 R2 连接 PIM-SM Domain 2 与 PIM-SM Domain 3 的接口 F0/0 和 F0/1 都配置成了 PIM-SM 域边界，后面还需在 PIM-SM Domain 2 与 PIM-SM Domain 3 之间配置 PIM-SM 域边界，在 R3 或 R5 上配置都可以。

(11) 将 R3 配置为 PIM-SM Domain 2 的 RP：

```
r3(config)#access-list 1 deny 224.0.1.39
```

```
r3(config)#access-list 1 deny 224.0.1.40
```

```
r3(config)#access-list 1 permit any
```

```
r3(config)#
```

```
r3(config)#access-list 24 permit 224.1.1.1
```

```
r3(config)#
```

```
r3(config)#int s1/0
```

```
r3(config-if)#ip multicast boundary 1
```

```
r3(config-if)#exit
```

```
r3(config)#ip pim send-rp-announce loopback 33 scope 16 group-list 24
```

```
r3(config)#ip pim send-rp-discovery loopback 33 scope 16
```

说明：R3 以 Loopback 33 作为组 224.1.1.1 的 RP 和映射代理；并将连接 PIM-SM Domain 3 的接口 S1/0 都配置成了 PIM-SM 域边界，到这里为止，PIM-SM Domain 1、PIM-SM Domain 2 以及 PIM-SM Domain 3 已经被成功分割开，相互独立。

(12) 将 R5 配置为 PIM-SM Domain 3 的 RP:

```
r5(config)#access-list 24 permit 224.1.1.1
```

```
r5(config)#
```

```
r5(config)#ip pim send-rp-announce loopback 55 scope 16 group-list 24
```

```
r5(config)#ip pim send-rp-discovery loopback 55 scope 16
```

说明: R5 以 Loopback 55 作为组 224.1.1.1 的 RP 和映射代理。

(13) 在 R1 上查看 RP 学习情况:

```
r1#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 22.2.2.2 (?), v2v1

Info source: 22.2.2.2 (?), elected via Auto-RP

Uptime: 00:02:37, expires: 00:02:21

```
r1#
```

说明: PIM-SM Domain 1 中的 R1 学习到的 RP 正是 22.2.2.2, 和预期一样, 也证明 PIM-SM 域已被成功隔离。

(14) 在 R2 上查看 RP 学习情况:

```
r2#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback22)

Group(s) 224.1.1.1/32

RP 22.2.2.2 (?), v2v1

Info source: 22.2.2.2 (?), elected via Auto-RP

Uptime: 00:02:50, expires: 00:02:07

r2#

说明：PIM-SM Domain 1 中的 R2 学习到的 RP 正是 22.2.2.2，和预期一样，也证明 PIM-SM 域已被成功隔离。



(15) 在 R3 上查看 RP 学习情况：

```
r3#sh ip pim rp mapping
```

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback33)

Group(s) 224.1.1.1/32

RP 33.3.3.3 (?), v2v1

Info source: 33.3.3.3 (?), elected via Auto-RP

Uptime: 00:01:34, expires: 00:02:23

r3#

说明：PIM-SM Domain 2 中的 R3 学习到的 RP 正是 33.3.3.3，和预期一样，也证明 PIM-SM 域已被成功隔离。

(16) 在 R4 上查看 RP 学习情况：

r4#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 33.3.3.3 (?), v2v1

Info source: 33.3.3.3 (?), elected via Auto-RP

Uptime: 00:01:47, expires: 00:02:12

r4#

说明：PIM-SM Domain 2 中的 R4 学习到的 RP 正是 33.3.3.3，和预期一样，也证明 PIM-SM 域已被成功隔离。

(17) 在 R5 上查看 RP 学习情况：

r5#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback55)

Group(s) 224.1.1.1/32

RP 55.5.5.5 (?), v2v1

Info source: 55.5.5.5 (?), elected via Auto-RP

Uptime: 00:01:11, expires: 00:02:47

r5#

说明：PIM-SM Domain 3 中的 R5 学习到的 RP 正是 55.5.5.5，和预期一样，也证明 PIM-SM 域已被成功隔离。

(18) 在 R6 上查看 RP 学习情况：

说明：PIM-SM Domain 3 中的 R6 学习到的 RP 正是 55.5.5.5，和预期一样，也证明 PIM-SM 域已被成功隔离。

(19) 测试组播源 R4 和 R6 到组 224.1.1.1 的通信情况：

r4#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r4#

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

```
r6#
```

说明：因为 PIM-SM 域之间已经被隔离，各自的 RP 都没有完整的组播源和组成员信息，所以在没有 MSDP 共享组播源信息的情况下，域间组播是不可能过的。

3. 配置 MSDP



(1) 在 2 个 PIM-SM 域之间建立 MSDP 连接：

```
r2(config)#ip msdp peer 3.3.3.3 connect-source loopback 0
```

```
r2(config)#ip msdp peer 55.5.5.5 connect-source loopback 0
```

```
r3(config)#ip msdp peer 2.2.2.2 connect-source loopback 0
```

```
r3(config)#ip msdp peer 55.5.5.5 connect-source loopback 0
```

```
r5(config)#ip msdp peer 2.2.2.2 connect-source loopback 55
```

```
r5(config)#ip msdp peer 3.3.3.3 connect-source loopback 55
```

CCIE LAB认证经验分享千人群：539730342

说明：R2, R3 以及 R5 之间建立全互联的 MSDP 邻居关系，其中 R2 和 R3 都是使用接口 Loopback 0 的地址作为 MSDP 源地址，而 R5 使用接口 Loopback 55 的地址作为 MSDP 源地址，连接情况如下：

R2 (2.2.2.2) —— R3 (3.3.3.3)

R2 (2.2.2.2) —— R5 (55.5.5.5)

R3 (3.3.3.3) —— R5 (55.5.5.5)

正因为 R5 的 RP 地址也是 55.5.5.5，而 MSDP 的源地址也是 55.5.5.5，如果收到的 SA 数据包的发送方 MSDP peer 的 IP 地址与 SA 数据包中的 RP 地址相同时，是不对 SA 数据包作 RPF 检测的，所以在任何 MSDP peer 收到 R5 的 SA 时都不做 RPF 检测。

(2) 查看 MSDP 的连接情况：

r2#sh ip msdp summary

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

			Downtime	Count	Count
--	--	--	----------	-------	-------

3.3.3.3	?	Up	00:01:17	0	?
---------	---	----	----------	---	---

55.5.5.5	?	Up	00:00:17	0	?
----------	---	----	----------	---	---

r2#

r3#sh ip msdp summary

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

Downtime Count Count

2.2.2.2	?	Up	00:01:41	0	0	?
---------	---	----	----------	---	---	---

55.5.5.5	?	Up	00:01:03	0	0	?
----------	---	----	----------	---	---	---

r3#

R5:

r5#sh ip msdp summary

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

Downtime Count Count

2.2.2.2	?	Up	00:00:53	0	0	?
---------	---	----	----------	---	---	---

3.3.3.3	?	Up	00:01:16	0	0	?
---------	---	----	----------	---	---	---

r5#

说明：R2，R3 以及 R5 之间的全互联 MSDP 连接已经全部正常建立。

(3) 在 R2 上查看 SA 信息：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 0 entries

r2#

CCIE LAB认证经验分享千人群：539730342

说明：因为目前没有路由器发起组播流量，也就没有组播源，那么也就不会产生 SA 信息。

(4) 在组播源路由器 R4 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

在 MSDP 路由器 R2 上开启 Debug:

```
r2#debug ip msdp peer
```

```
MSDP Peer debugging is on
```

```
r2#
```

```
r2#debug ip msdp detail
```

```
MSDP Detail debugging is on
```

```
r2#
```



在组播源路由器 R4 上发起组播流量：

```
r4#ping 224.1.1.1 repeat 10
```

```
Type escape sequence to abort.
```

```
Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:
```

```
.....
```

```
r4#
```

说明：组播不通，后面将根据实际情况一步一步解决。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 01:02:35.471: MSDP(0): Received 20-byte TCP segment from 3.3.3.3

*Mar 1 01:02:35.475: MSDP(0): Append 20 bytes to 0-byte msg 16 from 3.3.3.3, qs 1

*Mar 1 01:02:35.475: MSDP(0): 3.3.3.3: Received 20-byte msg 16 from peer

*Mar 1 01:02:35.475: MSDP(0): 3.3.3.3: SA TLV, len: 20, ec: 1, RP: 33.3.3.3

*Mar 1 01:02:35.479: MSDP(0): 3.3.3.3: Peer RPF check failed for 33.3.3.3, used ?
route's peer 0.0.0.0

r2#

说明：因为目前 R2 有 1 个以上的 MSDP peer（R3 和 R5 共 2 个），所以对于收到的 SA 必须进行 RPF 检测，而对 SA 的 RPF 检测必须依靠 BGP 或 MBGP 路由表，而我们并没有配置 BGP 和 MBGP，所以 Debug 信息中显示对于 MSDP peer 3.3.3.3 发来的 SA 做 RPF 检测失败了，在这里还请记住，如果收到的 SA 数据包的发送方 MSDP peer 的 IP 地址与 SA 数据包中的 RP 地址相同时，是不对 SA 数据包作 RPF 检测的，但因为 R3 的 MSDP 地址是 3.3.3.3，而它发出的 SA 里面的 RP 地址是 33.3.3.3，不满足不做 RPF 检测的要求，所以必须对 R3 发来的 SA 做 RPF 检测。

（5）在 R2 上再次查看 SA 信息：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 0 entries

r2#

说明：由于对 R3 发来的 SA 数据包 RPF 检测失败，所以 R2 并没有缓存任何可用的 SA 信息。

（6）尝试在组播源路由器 R6 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观

CCIE LAB认证经验分享千人群：539730342

察 MSDP 信息：

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 368 ms

Reply to request 1 from 12.1.1.1, 296 ms

Reply to request 2 from 12.1.1.1, 208 ms

Reply to request 3 from 12.1.1.1, 264 ms

Reply to request 4 from 12.1.1.1, 248 ms

Reply to request 5 from 12.1.1.1, 264 ms

Reply to request 6 from 12.1.1.1, 248 ms

Reply to request 7 from 12.1.1.1, 224 ms

Reply to request 8 from 12.1.1.1, 248 ms

Reply to request 9 from 12.1.1.1, 264 ms

```
r6#
```

说明：PIM-SM Domain 3 与 PIM-SM Domain 1 的组播居然通了，观察 R2 上的 Debug 信息，看看发生了什么：

```
r2#
```

```
*Mar 1 01:04:39.351: MSDP(0): Received 120-byte TCP segment from 55.5.5.5
```

```
*Mar 1 01:04:39.355: MSDP(0): Append 120 bytes to 0-byte msg 21 from 55.5.5.5, qs 1
```

CCIE LAB认证经验分享千人群：539730342

```
*Mar 1 01:04:39.355: MSDP(0): 55.5.5.5: Received 120-byte msg 21 from peer

*Mar 1 01:04:39.355: MSDP(0): 55.5.5.5: SA TLV, len: 120, ec: 1, RP: 55.5.5.5, with data

*Mar 1 01:04:39.359: MSDP(0): 55.5.5.5: Peer RPF check passed for 55.5.5.5, peer is RP

*Mar 1 01:04:39.359: MSDP(0): WAVL Insert SA Source 56.1.1.6 Group 224.1.1.1 RP
55.5.5.5 Successful
```

r2#

说明：因为 R5 的 MSDP peer 地址 55.5.5.5 与 SA 数据包中的 RP 地址 55.5.5.5 相同，满足不对 SA 数据包作 RPF 检测的条件，所以 R5 发来的 SA 数据包可以直接使用，从 R2 的 Debug 信息中也看出确如此，这就是最终 PIM-SM Domain 3 与 PIM-SM Domain 1 组播通信成功的原因。

(7) 在 R2 上再次查看 SA 信息：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 1 entries

(56.1.1.6, 224.1.1.1), RP 55.5.5.5, AS ?,00:03:54/00:04:42, Peer 55.5.5.5

Learned from peer 55.5.5.5, RPF peer 55.5.5.5,

SAs received: 4, Encapsulated data received: 1

r2#

说明：R5 发来的 SA 信息已被成功缓存并使用。

4. 配置 MBGP 帮助 MSDP 的 SA 通过 RPF 检测

说明：因为对 SA 的 RPF 检测必须依靠 BGP 或 MBGP 路由表来完成，所以我们在这里选择采取配置 MBGP 的方式，来帮助 R2 对 PIM-SM Domain 3 的 R3 发来的 SA 通过 RPF 检测，从而实现两个 PIM-SM 域之间的组播通信。

(1) 在 R2 和 R3 之间配置 MBGP：

R2：

```
r2(config)#router bgp 1
```

```
r2(config-router)#neighbor 3.3.3.3 remote-as 2
```

```
r2(config-router)#neighbor 3.3.3.3 update-source loopback 0
```

```
r2(config-router)#neighbor 3.3.3.3 ebgp-multihop
```

```
r2(config-router)#address-family ipv4 multicast
```

```
r2(config-router-af)#neighbor 3.3.3.3 activate
```

```
r2(config-router-af)#exit
```

```
r2(config-router)#exit
```

```
r2(config)#
```

R3：

```
r3(config)#router bgp 2
```

```
r3(config-router)#neighbor 2.2.2.2 remote-as 1
```

```
r3(config-router)#neighbor 2.2.2.2 update-source loopback 0
```

```
r3(config-router)#neighbor 2.2.2.2 ebgp-multihop
```

```
r3(config-router)#address-family ipv4 multicast
```

```
r3(config-router-af)#neighbor 2.2.2.2 activate
```

```
r3(config-router-af)#network 33.3.3.0 mask 255.255.255.0
```

```
r3(config-router-af)#exit
```

```
r3(config-router)#exit
```

```
r3(config)#
```

说明：为了更好的帮助 SA 通过 RPF 检测，所以使用了与建立 MSDP peer 相同的地址来建立 MBGP；因为 SA 的 RPF 检测是基于 SA 中 RP 的 IP 地址的，所以这里将 RP 的地址段 33.3.3.0/24 发布进了 MBGP，但并没有将组播源的地址段 34.1.1.0/24 发布进 MBGP，因为这条路由在 IGP 里面已经有了，对普通组播流的 RPF 检测可以基于静态组播路由表、DVMRP 路由表、MBGP 路由表、单播路由表等多种形式。

(2) 查看 BGP 与 MBGP 邻居状态：

```
r2#show ip bgp summary
```

```
BGP router identifier 22.2.2.2, local AS number 1
```

```
BGP table version is 1, main routing table version 1
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
3.3.3.3	4	2	2	2	0	0	0	00:00:26	0

```
r2#
```

```
r2#sh ip bgp ipv4 multicast summary
```

BGP router identifier 22.2.2.2, local AS number 1

BGP table version is 2, main routing table version 2

1 network entries using 129 bytes of memory

1 path entries using 48 bytes of memory

2/1 BGP path/bestpath attribute entries using 248 bytes of memory

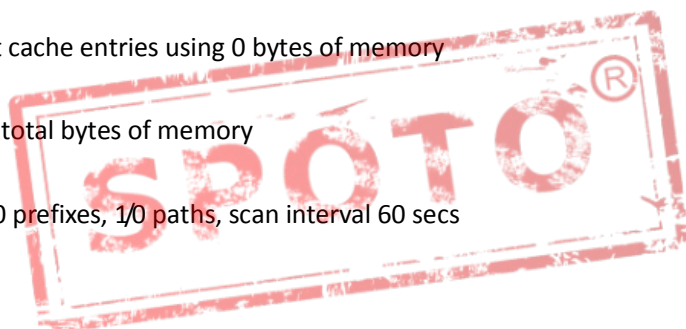
1 BGP AS-PATH entries using 24 bytes of memory

0 BGP route-map cache entries using 0 bytes of memory

0 BGP filter-list cache entries using 0 bytes of memory

BGP using 449 total bytes of memory

BGP activity 1/0 prefixes, 1/0 paths, scan interval 60 secs



Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
----------	---	----	---------	---------	--------	-----	------	---------	--------------

3.3.3.3	4	2	4	3	2	0	0	00:00:32	1
---------	---	---	---	---	---	---	---	----------	---

```
r2#
```

说明：R2 与 R3 之间的 BGP 和 MBGP 邻居均已建立，并且邻居类型为 eBGP，请注意对方 R3 的 AS 号码为 2，这个 AS 号码关系到它发来的 SA 能不能通过 RPF 检测，因为当 MSDP peer 是 eBGP 邻居的情况下，是看去往 SA 数据包中 RP 地址的最佳路径的下一跳 AS 号码是不是这个 eBGP 邻居的 AS 号码，如果是，则检测通过，否则检测失败，这里只要去往 RP 的地址 33.3.3.3 的 BGP 最佳路径的下一跳 AS 号码和 R3 的 AS 号码 2 一样，那么就能 SA 通过 RPF 检测，否则失败。

(3) 查看 MBGP 路由表：

CCIE LAB认证经验分享千人群：539730342

```
r2#sh ip bgp ipv4 multicast
```

```
BGP table version is 8, local router ID is 22.2.2.2
```

```
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
```

```
             r RIB-failure, S Stale
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 33.3.3.0/24	3.3.3.3	0		0 2	i

```
r2#
```

说明：MBGP 路由表中已经正常收到关于 R3 发来的 SA 数据包中的 RP 地址段，并且显示最佳路径的下一跳 AS 号码正是 eBGP 邻居 R3 的 AS 号码 2 相同，所以 RPF 检测一定通过。

（4）再次在组播源路由器 R4 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

```
r4#ping 224.1.1.1 repeat 10
```

```
Type escape sequence to abort.
```

```
Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:
```

```
Reply to request 0 from 12.1.1.1, 396 ms
```

```
Reply to request 0 from 12.1.1.1, 444 ms
```

CCIE LAB认证经验分享千人群：539730342

Reply to request 1 from 12.1.1.1, 240 ms

Reply to request 2 from 12.1.1.1, 288 ms

Reply to request 3 from 12.1.1.1, 220 ms

Reply to request 4 from 12.1.1.1, 216 ms

Reply to request 5 from 12.1.1.1, 288 ms

Reply to request 6 from 12.1.1.1, 304 ms

Reply to request 7 from 12.1.1.1, 240 ms

Reply to request 8 from 12.1.1.1, 196 ms

Reply to request 9 from 12.1.1.1, 188 ms

r4#

说明：和预期一样，通过 MBGP 解决 SA 的 RPF 检测问题之后，PIM-SM Domain 2 与 PIM-SM Domain 1 的组播通信已经变的正常。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 01:19:43.883: MSDP(0): Received 120-byte TCP segment from 3.3.3.3

*Mar 1 01:19:43.883: MSDP(0): Append 120 bytes to 0-byte msg 56 from 3.3.3.3, qs 1

*Mar 1 01:19:43.887: MSDP(0): 3.3.3.3: Received 120-byte msg 56 from peer

*Mar 1 01:19:43.887: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 33.3.3.3, with data

*Mar 1 01:19:43.887: MSDP(0): 3.3.3.3: Peer RPF check passed for 33.3.3.3, used EMBGP peer

*Mar 1 01:19:43.891: MSDP(0): Forward decapsulated SA data for (34.1.1.4, 224.1.1.1)

on Serial1/0

r2#

说明：和预期一样，R2 基于 MBGP 对 R3 发来的 SA 做 RPF 检测成功。

(5) 在 R2 上再次查看 SA 信息：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 2 entries

(34.1.1.4, 224.1.1.1), RP 33.3.3.3, MBGP/AS 2, 00:00:16/00:05:43, Peer 3.3.3.3

(56.1.1.6, 224.1.1.1), RP 55.5.5.5, BGP/AS 0, 00:00:27/00:05:32, Peer 55.5.5.5

r2#

说明：R3 和 R5 发来的 SA 信息已被 R2 成功缓存并使用，所以现在 3 个 PIM-SM 域之间的组播通信均正常。

(6) 查看组播树情况：

R3:

r3#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

CCIE LAB认证经验分享千人群：539730342

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:08:20/stopped, RP 33.3.3.3, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(34.1.1.4, 224.1.1.1), 00:00:13/00:03:24, flags: TA

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:00:13/00:03:16

r3#

R2:

r2#sh ip mroute 224.1.1.1

IP Multicast Routing Table



CCIE LAB认证经验分享千人群：539730342

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:49:23/00:02:28, RP 22.2.2.2, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:49:23/00:02:54

(34.1.1.4, 224.1.1.1), 00:00:31/00:02:37, flags: MT

Incoming interface: FastEthernet0/0, RPF nbr 23.1.1.3, Mbgp

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:00:31/00:02:54

(56.1.1.6, 224.1.1.1), 00:02:43/00:01:07, flags: T

Incoming interface: FastEthernet0/1, RPF nbr 25.1.1.5

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:02:43/00:02:53

r2#

说明：从组播树中可以看见，PIM-SM 域之间的通信使用的是 (S, G)，而非 PIM-SM 模式传统的 (*, G)，各组流量走向进出口完全和拓扑中一致。

查看当前环境中所有路由器的配置信息请点击“[running-config](#)”

5. 测试 default MSDP peer 帮助 MSDP 的 SA 通过 RPF 检测

说明：目前 3 个 PIM-SM 域之间的组播通信完全正常，所以想要测试 default MSDP peer 帮助 MSDP 的 SA 通过 RPF 检测，首先必须制造 SA 无法通过 RPF 检测的环境，才能够利用 default MSDP peer 帮助 MSDP 的 SA 通过 RPF 检测；根据上图的实验环境，我们采用过滤掉 PIM-SM Domain 3 的 R5 直接向 PIM-SM Domain 1 的 R2 发送 SA，但还是正常发给 PIM-SM Domain 2 的 R3，这样一来，PIM-SM Domain 3 和 PIM-SM Domain 2 的 SA 都必须通过 R3 发给 R2，这就势必会造成 PIM-SM Domain 3 的 SA 因为 RPF 检测不通过而被 R2 丢弃，但 PIM-SM Domain 2 的 SA 和之前一样保持正常。

(1) 将 PIM-SM Domain 3 的 R5 向 PIM-SM Domain 1 的 R2 发送的 SA 全部过滤掉：

```
r5(config)#ip msdp sa-filter out 2.2.2.2
```

说明：如果命令后面不跟任何匹配工具，则表示所有 SA。

(2) 清除 R2 上的 SA 缓存信息：

```
r2#clear ip msdp sa-cache
```

```
r2#
```

```
r2#sh ip msdp sa-cache
```

```
MSDP Source-Active Cache - 0 entries
```

```
r2#
```

说明： R2 上的 SA 缓存信息已被全部清空。

(3) 在组播源路由器 R6 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

```
r6#ping 224.1.1.1 repeat 10
```

```
Type escape sequence to abort.
```

```
Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:
```

```
.....
```

```
r6#
```

说明： 和预期一样，PIM-SM Domain 3 与 PIM-SM Domain 1 的组播不通，这一定是 SA 数据包无法通过 RPF 检测引起的。

观察 R2 上的 Debug 信息：

```
r2#
```

```
*Mar  1 01:35:56.891: MSDP(0): Received 120-byte TCP segment from 3.3.3.3
```

CCIE LAB认证经验分享千人群：539730342

*Mar 1 01:35:56.891: MSDP(0): Append 120 bytes to 0-byte msg 102 from 3.3.3.3, qs 1

*Mar 1 01:35:56.891: MSDP(0): 3.3.3.3: Received 120-byte msg 102 from peer

*Mar 1 01:35:56.891: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 55.5.5.5, with data

*Mar 1 01:35:56.895: MSDP(0): 3.3.3.3: Peer RPF check failed for 55.5.5.5, used IBGP route's peer 0.0.0.0

r2#

说明：正因为 PIM-SM Domain 3 的 R5 直接向 PIM-SM Domain 1 的 R2 发送的 SA 被过滤掉了，所以这时 PIM-SM Domain 3 的 SA 是先发给 PIM-SM Domain 2 的 R3，然后由 R3 再发给 R2 的，这样一来，在对收到的 SA 做 RPF 检测时，由于 SA 数据包中 RP 地址 55.5.5.5 根本就无法在 BGP 或 MBGP 路由表中找到，也就无法做 RPF 检测，所以最终因检测失败而被丢弃了。

（4）再次在组播源路由器 R4 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

r4#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 296 ms

Reply to request 1 from 12.1.1.1, 188 ms

Reply to request 2 from 12.1.1.1, 344 ms

Reply to request 3 from 12.1.1.1, 320 ms

Reply to request 4 from 12.1.1.1, 392 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 5 from 12.1.1.1, 336 ms

Reply to request 6 from 12.1.1.1, 168 ms

Reply to request 7 from 12.1.1.1, 172 ms

Reply to request 8 from 12.1.1.1, 324 ms

Reply to request 9 from 12.1.1.1, 308 ms

r4#

说明：因为 PIM-SM Domain 2 与 PIM-SM Domain 1 之间并没有任何变化，所以组播通信仍然保持正常。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 01:37:11.759: MSDP(0): Received 120-byte TCP segment from 3.3.3.3

*Mar 1 01:37:11.763: MSDP(0): Append 120 bytes to 0-byte msg 105 from 3.3.3.3, qs 1

*Mar 1 01:37:11.763: MSDP(0): 3.3.3.3: Received 120-byte msg 105 from peer

*Mar 1 01:37:11.767: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 33.3.3.3, with data

*Mar 1 01:37:11.767: MSDP(0): 3.3.3.3: Peer RPF check passed for 33.3.3.3, used EMBGP peer

*Mar 1 01:37:11.767: MSDP(0): WAVL Insert SA Source 34.1.1.4 Group 224.1.1.1 RP 33.3.3.3 Successful

r2#

说明：PIM-SM Domain 2 中的 SA 还是被 R2 正常接收并缓存。

(5) 在 R2 上再次查看 SA 信息：

CCIE LAB认证经验分享千人群：539730342

```
r2#sh ip msdp sa-cache
```

MSDP Source-Active Cache - 1 entries

(34.1.1.4, 224.1.1.1), RP 3.3.3.3, MBGP/AS 2, 00:00:49/00:05:44, Peer 3.3.3.3

Learned from peer 3.3.3.3, RPF peer 3.3.3.3,

SAs received: 2, Encapsulated data received: 1

```
r2#
```

说明：从结果中看出，PIM-SM Domain 2 中的 SA 被 R2 正常接收并缓存，而 PIM-SM Domain 3 还在故障中。

(6) 通过 default MSDP peer 帮助 MSDP 的 SA 通过 RPF 检测

```
r2(config)#ip msdp default-peer 3.3.3.3
```

说明：因为从 Default MSDP Peer 收到的所有 SA 都不需要做 RPF 检测，所以无论是 PIM-SM Domain 2 中的 SA，还是 PIM-SM Domain 3 先发给 PIM-SM Domain 2 的 R3，然后再由 R3 发过来的 SA，R2 都不再做 RPF 检测，结果就是所有 SA 都被有效缓存。

(7) 再次在组播源路由器 R6 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 456 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 1 from 12.1.1.1, 408 ms

Reply to request 2 from 12.1.1.1, 220 ms

Reply to request 3 from 12.1.1.1, 200 ms

Reply to request 4 from 12.1.1.1, 192 ms

Reply to request 5 from 12.1.1.1, 280 ms

Reply to request 6 from 12.1.1.1, 264 ms

Reply to request 7 from 12.1.1.1, 248 ms

Reply to request 8 from 12.1.1.1, 312 ms

Reply to request 9 from 12.1.1.1, 276 ms

r6#

说明：default MSDP peer 已经成功解决 PIM-SM 域间的 SA 检测与组播通信问题。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 01:44:04.047: MSDP(0): Received 120-byte TCP segment from 3.3.3.3

*Mar 1 01:44:04.047: MSDP(0): Append 120 bytes to 0-byte msg 118 from 3.3.3.3, qs 1

*Mar 1 01:44:04.047: MSDP(0): 3.3.3.3: Received 120-byte msg 118 from peer

*Mar 1 01:44:04.051: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 55.5.5.5, with data

*Mar 1 01:44:04.051: MSDP(0): 3.3.3.3: Peer RPF check passed for 55.5.5.5, used default peer

*Mar 1 01:44:04.051: MSDP(0): WAVL Insert SA Source 56.1.1.6 Group 224.1.1.1 RP 55.5.5.5 Successful

r2#

说明：由于配置了 default MSDP peer 的原因，原本 R5 发给 R3 再发给 R2 的 SA 不能通过 RPF 检测，现在已经被 R2 正常接收并缓存了。

(8) 在 R2 上再次查看 SA 信息：

```
r2#sh ip msdp sa-cache
```

MSDP Source-Active Cache - 2 entries

(34.1.1.4, 224.1.1.1), RP 33.3.3.3, MBGP/AS 2, 00:07:52/00:03:56, Peer 3.3.3.3

Learned from peer 3.3.3.3, RPF peer 3.3.3.3,

SAs received: 3, Encapsulated data received: 1

(56.1.1.6, 224.1.1.1), RP 55.5.5.5, BGP/AS 0, 00:01:00/00:05:43, Peer 3.3.3.3

Learned from peer 3.3.3.3, RPF peer 3.3.3.3,

SAs received: 2, Encapsulated data received: 1

```
r2#
```

说明：无论 PIM-SM Domain 2 还是 PIM-SM Domain 3 的 SA，在 R3 发给 R2 后都被正常接收并缓存了。

查看当前环境中所有路由器的配置信息请点击“[running-config](#)”

6. 测试 MSDP Mesh Group 帮助 MSDP 的 SA 通过 RPF 检测

说明：目前 3 个 PIM-SM 域之间的组播通信完全正常，所以想要测试 MSDP Mesh Group 帮助 MSDP 的 SA 通过 RPF 检测，首先必须制造 SA 无法通过 RPF 检测的环境，才能

够利用 MSDP Mesh Group 帮助 MSDP 的 SA 通过 RPF 检测；根据上图的实验环境，我们采用在 PIM-SM Domain 1 的 R2 与 PIM-SM Domain 3 的 R5 之间增加 BGP 连接，然后对于 PIM-SM Domain 2 中的 RP 地址 33.3.3.3 的网段信息，不再通过 R3 发给 R2，而通过 R5 发给 R2，从而使 SA 中 RP 地址（33.3.3.3）因为最佳路径的下一跳 AS 号码（现在为 3）与发送方 MSDP peer（R3）的 AS 号码（为 2）不匹配而导致 RPF 检测失败，从而被丢弃，然后我们在 R2 上将 R3 指定为 MSDP Mesh Group 中的成员，让 R3 发来的 SA 因为发送 MSDP peer 是 MSDP Mesh Group 中的成员而跳过 RPF 检测，最终解决 PIM-SM Domain 2 与 PIM-SM Domain 1 的域间组播通信。

（1）删除 default MSDP peer 配置：

```
r2(config)#no ip msdp default-peer 3.3.3.3
```

说明：因为 R2 对 R3 配置的 default MSDP peer 会让所有从 R3 发来的 SA 都跳过 RPF 检测而被缓存，所以必须先删除。

（2）删除 R5 对 R2 的 SA 过滤配置：

```
r5(config)#no ip msdp sa-filter out 2.2.2.2
```

说明：因为 R5 已经过滤掉发给 R2 的所有 SA，所以先将配置删除。

（3）配置 BGP 制造 RPF 检测故障（注：采用 MBGP 代替 BGP 也可以，两者一样）：

R3 不再通告关于 RP 地址 33.3.3.3 的网段信息：

```
r3(config)#router bgp 2
```

```
r3(config-router)#address-family ipv4 multicast
```

```
r3(config-router-af)#no network 33.3.3.0 mask 255.255.255.0
```

R2 增加与 R5 之间的 BGP：

```
r2(config)#router bgp 1
```

```
r2(config-router)#neighbor 5.5.5.5 remote-as 3
```

```
r2(config-router)#neighbor 5.5.5.5 update-source loopback 0
```

```
r2(config-router)#neighbor 5.5.5.5 ebgp-multihop
```

```
r2(config-router)#exit
```

R3 取消在 MBGP 里通告关于 RP 地址 33.3.3.3 的网段：

```
r3(config)#router bgp 2
```

```
r3(config-router)#address-family ipv4 multicast
```

```
r3(config-router-af)#no network 33.3.3.0 mask 255.255.255.0
```

R5 替 R3 通告关于 RP 地址 33.3.3.3 的网段信息：

```
r5(config)#router bgp 3
```

```
r5(config-router)#neighbor 2.2.2.2 remote-as 1
```

```
r5(config-router)#neighbor 2.2.2.2 update-source loopback 0
```

```
r5(config-router)#neighbor 2.2.2.2 ebgp-multihop
```

```
r5(config-router)#network 33.1.1.0 mask 255.255.255.0
```

```
r5(config-router)#exit
```

说明：将 R5 配置为 BGP AS3，并将关于 RP 地址 33.3.3.3 的网段通告进 BGP*（因为 R5 已经通过 OSPF 从 R3 学习到该路由，所以可以使用 network 命令来通告），这就像是背地里对 R3 干的一个恶作剧。

(4) 在 R2 上查看路由情况:

```
r2#sh ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

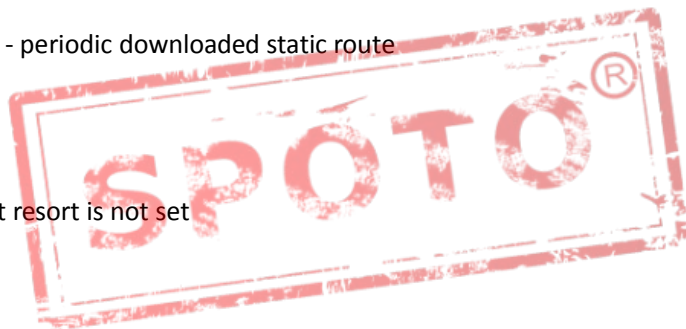
E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set



34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:03:45, FastEthernet0/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 25.1.1.5, 00:21:26, FastEthernet0/1

[110/65] via 23.1.1.3, 00:21:26, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

33.0.0.0/24 is subnetted, 1 subnets

B 33.3.3.0 [20/3] via 5.5.5.5, 00:01:54

CCIE LAB认证经验分享千人群：539730342

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/2] via 23.1.1.3, 00:21:26, FastEthernet0/0

55.0.0.0/24 is subnetted, 1 subnets

O 55.5.5.0 [110/2] via 25.1.1.5, 00:21:26, FastEthernet0/1

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/2] via 25.1.1.5, 00:21:27, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/0

22.0.0.0/24 is subnetted, 1 subnets

C 22.2.2.0 is directly connected, Loopback22

25.0.0.0/24 is subnetted, 1 subnets

C 25.1.1.0 is directly connected, FastEthernet0/1

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/2] via 25.1.1.5, 00:21:27, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, Serial1/0

r2#

说明：R2 上已经通过 BGP 从 R5 学习到关于 RP 地址 33.3.3.3 的网段。

(5) 在 R2 上查看 MBGP 的路由情况：

r2#sh ip bgp ipv4 multicast

r2#

说明：因为 R3 已经取消在 MBGP 里通告关于 RP 地址 33.3.3.3 的网段，所以 R2 上的 MBGP 路由为空。

(6)清除掉 R2 上的 SA 缓存信息，然后在组播源路由器 R4 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

R2 上清除 SA：

r2#clear ip msdp sa-cache

r2#

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 0 entries

r2#

R4 上发组播：

r4#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r4#

CCIE LAB认证经验分享千人群：539730342

说明：和预期一样，PIM-SM Domain 2 与 PIM-SM Domain 1 之间的组播不通。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 02:09:02.451: MSDP(0): Received 120-byte TCP segment from 3.3.3.3

*Mar 1 02:09:02.455: MSDP(0): Append 120 bytes to 0-byte msg 173 from 3.3.3.3, qs 1

*Mar 1 02:09:02.455: MSDP(0): 3.3.3.3: Received 120-byte msg 173 from peer

*Mar 1 02:09:02.455: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 33.3.3.3, with data

*Mar 1 02:09:02.459: MSDP(0): 3.3.3.3: Peer RPF check failed for 33.3.3.3, EBGP route/peer in AS 3/2

r2#

说明：R2 的 Debug 信息也表明 R3 发给 R2 的 SA 信息在基于 EBGP 做 RPF 检测时失败了，因而被丢弃。



(7) 确认 R3 发给 R2 的 SA 数据包 RPF 检测失败的详细过程：

在 R2 上查看 MSDP peer 信息：

r2#show ip msdp summary

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

			Downtime	Count	Count
--	--	--	----------	-------	-------

3.3.3.3	2	Up	00:30:53	1	?
---------	---	----	----------	---	---

55.5.5.5	?	Up	01:16:54	0	?
----------	---	----	----------	---	---

r2#

说明：R3 作为 R2 的 MSDP peer，它的 AS 号码是 2，当 R2 对 R3 发来的 SA 做 RPF 检测时，SA 数据包中的 RP 地址在 BGP 路由表中最佳路径的下一跳 AS 号码必须也和 R3 的号码一样为 2，如果是其它号码就不能通过 RPF 检测。

在 R2 上查看 BGP 路由表中关于 RP 地址 33.3.3.3 的最佳路径情况：

r2#sh ip bgp

BGP table version is 6, local router ID is 22.2.2.2

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,

r RIB-failure, S Stale

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 33.3.3.0/24	5.5.5.5	3		0	3 i

r2#

说明：正是因为关于 RP 地址 33.3.3.3 的最佳路径的下一跳 AS 号码是 R5 的 AS 号码 3，与发送 MSDP peer R3 的 AS 号码 2 不一致，所以 R3 发来的 SA 不能通过 RPF 检测而被丢弃了。

(8) 在 R2 上查看 SA 信息：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 0 entries

r2#

CCIE LAB认证经验分享千人群：539730342

说明：由于 R3 发来的 SA 不能通过 RPF 检测，所以目前 R2 的 SA 缓存为空。

(9) 通过 MSDP Mesh Group 帮助 MSDP 的 SA 通过 RPF 检测：

```
r2(config)#ip msdp mesh-group CCIE 3.3.3.3
```

说明：因为从 MSDP Mesh Group 收到的所有 SA 都不需要做 RPF 检测，所以后面 R3 发过来的 SA，R2 都不再做 RPF 检测就能被有效缓存。

(10) 再次在组播源路由器 R4 上发起组播，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

```
r4#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 360 ms

Reply to request 1 from 12.1.1.1, 344 ms

Reply to request 2 from 12.1.1.1, 264 ms

Reply to request 3 from 12.1.1.1, 304 ms

Reply to request 4 from 12.1.1.1, 300 ms

Reply to request 5 from 12.1.1.1, 308 ms

Reply to request 6 from 12.1.1.1, 388 ms

Reply to request 7 from 12.1.1.1, 320 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 8 from 12.1.1.1, 264 ms

Reply to request 9 from 12.1.1.1, 452 ms

r4#

说明： MSDP Mesh Group 已经成功解决 PIM-SM 域间的 SA 检测与组播通信问题.

观察 R2 上的 Debug 信息：

r2#

*Mar 1 02:13:54.195: MSDP(0): Received 120-byte TCP segment from 3.3.3.3

*Mar 1 02:13:54.195: MSDP(0): Append 120 bytes to 0-byte msg 184 from 3.3.3.3, qs 1

*Mar 1 02:13:54.195: MSDP(0): 3.3.3.3: Received 120-byte msg 184 from peer

*Mar 1 02:13:54.199: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 33.3.3.3, with data

*Mar 1 02:13:54.199: MSDP(0): 3.3.3.3: Peer RPF check bypassed, peer 3.3.3.3 in mesh-group CCIE

r2#

说明： 由于配置了 MSDP Mesh Group 的原因，原本 R3 发给 R2 的 SA 不能通过 RPF 检测，现在已经被 R2 正常接收并缓存了。

(11) 在 R2 上再次查看 SA 信息：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 1 entries

(34.1.1.4, 224.1.1.1), RP 33.3.3.3, BGP/AS 3, 00:03:56/00:05:50, Peer 3.3.3.3

Learned from peer 3.3.3.3, RPF peer 0.0.0.0,

SAs received: 4, Encapsulated data received: 1

r2#

说明：由于配置了 MSDP Mesh Group 的原因，原本 R3 发给 R2 的 SA 不能通过 RPF 检测，现在已经被 R2 正常接收并缓存了。

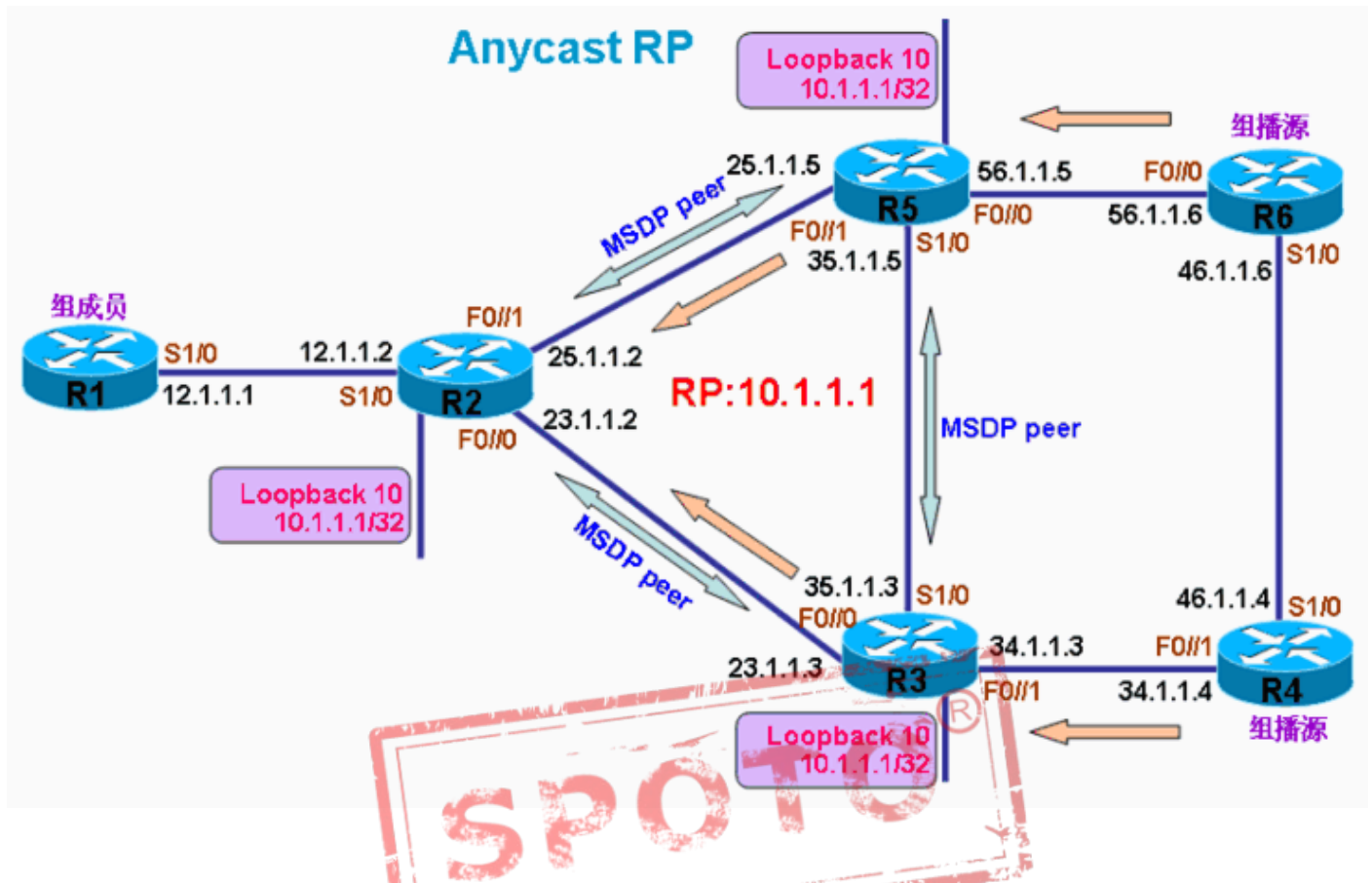
查看当前环境中所有路由器的配置信息请点击 “[running-config](#)”

配置 Anycast RP 实验

实验说明：



以下图为环境配置 Anycast RP 实验：



上图中所有路由器都在同一个 PIM-SM 域，其中 R4 和 R6 为组 224.1.1.1 的组播源，R1 为组成员；R2，R3 以及 R5 都配有 Loopback 10，并且地址都为 10.1.1.1/32；

除了图上所标识出的接口地址外，R2、R3 和 R5 均配有 Loopback 0 接口，分别为：

R2（Loopback 0 :2.2.2.2/24）

R3（Loopback 0 :3.3.3.3/24）

R5（Loopback 0 :5.5.5.5/24）

所有路由器上都运行 OSPF，并将所有接口都发布进 OSPF，以保证全网单播互通。

R2、R3 和 R5 都以接口 Loopback 10 的地址 10.1.1.1 配置为 RP，从而形成 Anycast RP，然后还需要在 RP 之间创建 MSDP 连接，连接情况如下：

R2（2.2.2.2）—— R3（3.3.3.3）

R2 (2.2.2.2) —— R5 (55.5.5.5)

因为从一个 MSDP peer 收到的 SA 会转发给其它所有 MSDP peer, 所以 R3 和 R5 都能互相收到对方的 SA, 也就没必要在 R3 和 R5 之间创建 MSDP 连接, 但也可以创建。

1. 配置初始网络环境

(1) 配置 R1:

```
r1(config)#int s1/0
```

```
r1(config-if)#encapsulation frame-relay
```

```
r1(config-if)#no frame-relay inverse-arp
```

```
r1(config-if)#no arp frame-relay
```

```
r1(config-if)#ip add 12.1.1.1 255.255.255.0
```

```
r1(config-if)#no shutdown
```

```
r1(config-if)#frame-relay map ip 12.1.1.2 102 broadcast
```

```
r1(config-if)#ip ospf network point-to-point
```

```
r1(config-if)#exit
```

```
r1(config)#router ospf 1
```

```
r1(config-router)#router-id 1.1.1.1
```

```
r1(config-router)#network 0.0.0.0 0.0.0.0 area 0
```



```
r1(config-router)#exit
```

说明：为 R1 的 S1/0 配置接口地址，并将所有接口发布进 OSPF。

(2) 配置 R2:

```
r2(config)#int loopback 0
```

```
r2(config-if)#ip add 2.2.2.2 255.255.255.0
```

```
r2(config-if)#ip ospf network point-to-point
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int loopback 10
```

```
r2(config-if)#ip address 10.1.1.1 255.255.255.255
```

```
r2(config-if)#exit
```

```
r2(config)#int s1/0
```

```
r2(config-if)#encapsulation frame-relay
```

```
r2(config-if)#no frame-relay inverse-arp
```

```
r2(config-if)#no arp frame-relay
```

```
r2(config-if)#ip add 12.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#frame-relay map ip 12.1.1.1 201 broadcast
```



```
r2(config-if)#ip ospf network point-to-point
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int f0/0
```

```
r2(config-if)#ip add 23.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/1
```

```
r2(config-if)#ip add 25.1.1.2 255.255.255.0
```

```
r2(config-if)#no shutdown
```

```
r2(config-if)#exit
```

```
r2(config)#router ospf 1
```

```
r2(config-router)#router-id 2.2.2.2
```

```
r2(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r2(config-router)#exit
```

说明：为 R2 的 Loopback 0，Loopback 10，S1/0，F0/0，F0/1 配置接口地址，并将所有接口发布进 OSPF，OSPF 的 router-id 改为 Loopback 0 的地址，因为默认为 Loopback 10 的地址，但 Loopback 10 的地址在网络中是重复的。

(3) 配置 R3:

```
r3(config)#int loopback 0
```

```
r3(config-if)#ip add 3.3.3.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#ip ospf network point-to-point
```

```
r3(config-if)#exit
```

```
r3(config)#
```

```
r3(config)#int loopback 10
```

```
r3(config-if)#ip address 10.1.1.1 255.255.255.255
```

```
r3(config-if)#exit
```

```
r3(config)#int f0/0
```

```
r3(config-if)#ip add 23.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#exit
```

```
r3(config)#
```

```
r3(config)#int f0/1
```

```
r3(config-if)#ip add 34.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#exit
```

```
r3(config)#int s1/0
```

```
r3(config-if)#encapsulation frame-relay
```

```
r3(config-if)#no frame-relay inverse-arp
```

```
r3(config-if)#no arp frame-relay
```



```
r3(config-if)#ip add 35.1.1.3 255.255.255.0
```

```
r3(config-if)#no shutdown
```

```
r3(config-if)#frame-relay map ip 35.1.1.5 305 broadcast
```

```
r3(config-if)#ip ospf network point-to-point
```

```
r3(config-if)#exit
```

```
r3(config)#router ospf 1
```

```
r3(config-router)#router-id 3.3.3.3
```

```
r3(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r3(config-router)#exit
```

```
r3(config)#
```

说明：为 R3 的 Loopback 0， Loopback 10， S1/0， F0/1， F0/0 配置接口地址，并将所有接口发布进 OSPF，OSPF 的 router-id 改为 Loopback 0 的地址，因为默认为 Loopback 10 的地址，但 Loopback 10 的地址在网络中是重复的。

(4) 配置 R4:

```
r4(config)#int f0/1
```

```
r4(config-if)#ip add 34.1.1.4 255.255.255.0
```

```
r4(config-if)#no shutdown
```

```
r4(config-if)#exit
```

```
r4(config)#int s1/0
```

```
r4(config-if)#encapsulation frame-relay
```

```
r4(config-if)#no frame-relay inverse-arp

r4(config-if)#no arp frame-relay

r4(config-if)#ip address 46.1.1.4 255.255.255.0

r4(config-if)#no shutdown

r4(config-if)#frame-relay map ip 46.1.1.6 406 broadcast

r4(config-if)#ip ospf network point-to-point

r4(config-if)#exit

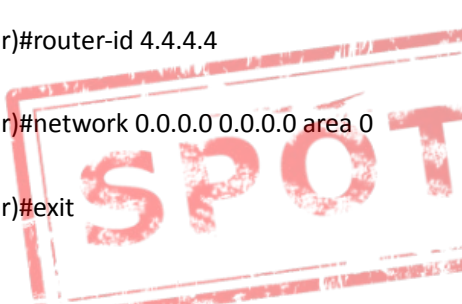
r4(config)#router ospf 1

r4(config-router)#router-id 4.4.4.4

r4(config-router)#network 0.0.0.0 0.0.0.0 area 0

r4(config-router)#exit

r4(config)#
```



说明：为 R4 的 F0/1，S1/0 配置接口地址，并将所有接口发布进 OSPF。

(5) 配置 R5:

```
r5(config)#int loopback 0

r5(config-if)#ip address 5.5.5.5 255.255.255.0

r5(config-if)#ip ospf network point-to-point

r5(config-if)#exit

r5(config)#

r5(config)#int loopback 10
```

CCIE LAB认证经验分享千人群：539730342

```
r5(config-if)#ip address 10.1.1.1 255.255.255.255
```

```
r5(config-if)#exit
```

```
r5(config)#int f0/0
```

```
r5(config-if)#ip address 56.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#exit
```

```
r5(config)#int f0/1
```

```
r5(config-if)#ip add 25.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#exit
```

```
r5(config)#
```

```
r5(config)#int s1/0
```

```
r5(config-if)#encapsulation frame-relay
```

```
r5(config-if)#no frame-relay inverse-arp
```

```
r5(config-if)#no arp frame-relay
```

```
r5(config-if)#ip add 35.1.1.5 255.255.255.0
```

```
r5(config-if)#no shutdown
```

```
r5(config-if)#ip ospf network point-to-point
```

```
r5(config-if)#frame-relay map ip 35.1.1.3 503 broadcast
```

```
r5(config-if)#exit
```



说明：为 R5 的 Loopback 0， Loopback 10， S1/0， F0/1， F0/0 配置接口地址，并将所有接口发布进 OSPF，OSPF 的 router-id 改为 Loopback 0 的地址，因为默认为 Loopback 10 的地址，但 Loopback 10 的地址在网络中是重复的。

(6) 配置 R6:

```
r6(config)#int f0/0
```

```
r6(config-if)#ip address 56.1.1.6 255.255.255.0
```

```
r6(config-if)#no shutdown
```

```
r6(config-if)#exit
```

```
r6(config)#
```

```
r6(config)#int s1/0
```

```
r6(config-if)#encapsulation frame-relay
```

```
r6(config-if)#no frame-relay inverse-arp
```

```
r6(config-if)#no arp frame-relay
```

```
r6(config-if)#ip add 46.1.1.6 255.255.255.0
```

```
r6(config-if)#no shutdown
```

```
r6(config-if)#frame-relay map ip 46.1.1.4 604 broadcast
```

```
r6(config-if)#ip ospf network point-to-point
```

```
r6(config-if)#exit
```

```
r6(config)#
```

```
r6(config)#router ospf 1
```



CCIE LAB认证经验分享千人群：539730342

```
r6(config-router)#router-id 6.6.6.6
```

```
r6(config-router)#network 0.0.0.0 0.0.0.0 area 0
```

```
r6(config-router)#exit
```

说明：为 R6 的 F0/0, S1/0 配置接口地址，并将所有接口发布进 OSPF。

(7) 查看 R1 的路由学习情况：

```
r1#sh ip route
```

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/66] via 12.1.1.2, 00:00:38, Serial1/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/129] via 12.1.1.2, 00:00:38, Serial1/0

CCIE LAB认证经验分享千人群：539730342

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/65] via 12.1.1.2, 00:00:38, Serial1/0

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/66] via 12.1.1.2, 00:00:38, Serial1/0

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/66] via 12.1.1.2, 00:00:38, Serial1/0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/65] via 12.1.1.2, 00:00:38, Serial1/0

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/65] via 12.1.1.2, 00:00:39, Serial1/0

10.0.0.0/32 is subnetted, 1 subnets

O 10.1.1.1 [110/65] via 12.1.1.2, 00:00:39, Serial1/0

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/66] via 12.1.1.2, 00:00:39, Serial1/0

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, Serial1/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/130] via 12.1.1.2, 00:00:39, Serial1/0

r1#

说明：R1 已经学习到全网的每一条路由。

(8) 查看 R2 的路由学习情况：

r2#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

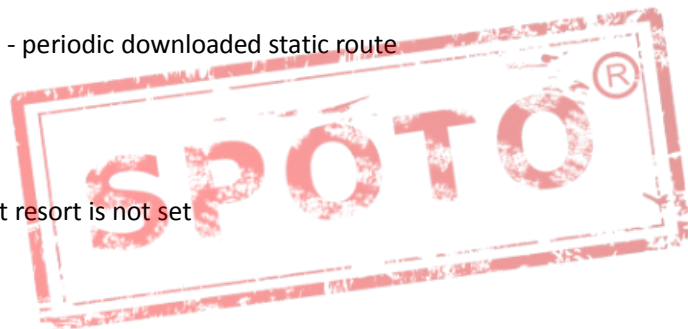
E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set



34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/2] via 23.1.1.3, 00:00:55, FastEthernet0/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 25.1.1.5, 00:00:55, FastEthernet0/1

[110/65] via 23.1.1.3, 00:00:55, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

C 2.2.2.0 is directly connected, Loopback0

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/2] via 23.1.1.3, 00:00:55, FastEthernet0/0

CCIE LAB认证经验分享千人群：539730342

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/2] via 25.1.1.5, 00:00:55, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/0

25.0.0.0/24 is subnetted, 1 subnets

C 25.1.1.0 is directly connected, FastEthernet0/1

10.0.0.0/32 is subnetted, 1 subnets

C 10.1.1.1 is directly connected, Loopback10

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/2] via 25.1.1.5, 00:00:56, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

C 12.1.1.0 is directly connected, Serial1/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/66] via 25.1.1.5, 00:00:56, FastEthernet0/1

[110/66] via 23.1.1.3, 00:00:56, FastEthernet0/0

r2#

说明：R2 已经学习到全网的每一条路由。

(9) 查看 R3 的路由学习情况：

r3#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

CCIE LAB认证经验分享千人群：539730342

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

C 34.1.1.0 is directly connected, FastEthernet0/1

35.0.0.0/24 is subnetted, 1 subnets

C 35.1.1.0 is directly connected, Serial1/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/2] via 23.1.1.2, 00:01:12, FastEthernet0/0

3.0.0.0/24 is subnetted, 1 subnets

C 3.3.3.0 is directly connected, Loopback0

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/3] via 23.1.1.2, 00:01:12, FastEthernet0/0

23.0.0.0/24 is subnetted, 1 subnets

C 23.1.1.0 is directly connected, FastEthernet0/0

CCIE LAB认证经验分享千人群：539730342

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/2] via 23.1.1.2, 00:01:12, FastEthernet0/0

10.0.0.0/32 is subnetted, 1 subnets

C 10.1.1.1 is directly connected, Loopback10

56.0.0.0/24 is subnetted, 1 subnets

O 56.1.1.0 [110/3] via 23.1.1.2, 00:01:12, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/65] via 23.1.1.2, 00:01:12, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/65] via 34.1.1.4, 00:01:12, FastEthernet0/1

r3#

说明：R3 已经学习到全网的每一条路由。



(10) 查看 R4 的路由学习情况：

r4#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

CCIE LAB认证经验分享千人群：539730342

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

C 34.1.1.0 is directly connected, FastEthernet0/1

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 34.1.1.3, 00:01:27, FastEthernet0/1

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/3] via 34.1.1.3, 00:01:27, FastEthernet0/1

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/2] via 34.1.1.3, 00:01:27, FastEthernet0/1

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/4] via 34.1.1.3, 00:01:27, FastEthernet0/1

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/2] via 34.1.1.3, 00:01:27, FastEthernet0/1

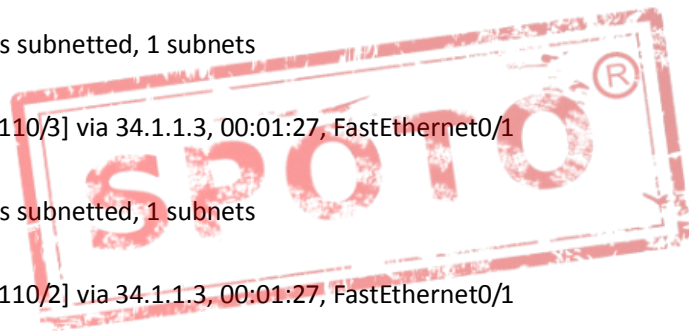
25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/3] via 34.1.1.3, 00:01:28, FastEthernet0/1

10.0.0.0/32 is subnetted, 1 subnets

O 10.1.1.1 [110/2] via 34.1.1.3, 00:01:28, FastEthernet0/1

56.0.0.0/24 is subnetted, 1 subnets



O 56.1.1.0 [110/4] via 34.1.1.3, 00:01:28, FastEthernet0/1

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/66] via 34.1.1.3, 00:01:28, FastEthernet0/1

46.0.0.0/24 is subnetted, 1 subnets

C 46.1.1.0 is directly connected, Serial1/0

r4#

说明：R4 已经学习到全网的每一条路由。

(11) 查看 R5 的路由学习情况：

r5#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

CCIE LAB认证经验分享千人群：539730342

O 34.1.1.0 [110/3] via 25.1.1.2, 00:00:23, FastEthernet0/1

35.0.0.0/24 is subnetted, 1 subnets

C 35.1.1.0 is directly connected, Serial1/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/2] via 25.1.1.2, 00:00:23, FastEthernet0/1

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/3] via 25.1.1.2, 00:00:23, FastEthernet0/1

5.0.0.0/24 is subnetted, 1 subnets

C 5.5.5.0 is directly connected, Loopback0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/2] via 25.1.1.2, 00:00:23, FastEthernet0/1

25.0.0.0/24 is subnetted, 1 subnets

C 25.1.1.0 is directly connected, FastEthernet0/1

10.0.0.0/32 is subnetted, 1 subnets

C 10.1.1.1 is directly connected, Loopback10

56.0.0.0/24 is subnetted, 1 subnets

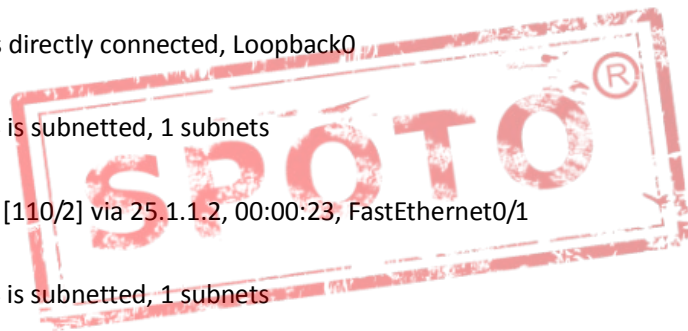
C 56.1.1.0 is directly connected, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/65] via 25.1.1.2, 00:00:23, FastEthernet0/1

46.0.0.0/24 is subnetted, 1 subnets

O 46.1.1.0 [110/65] via 56.1.1.6, 00:00:23, FastEthernet0/0



r5#

说明： R5 已经学习到全网的每一条路由。

(12) 查看 R6 的路由学习情况：

r6#sh ip route

Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP

D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area

N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

E1 - OSPF external type 1, E2 - OSPF external type 2

i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2

ia - IS-IS inter area, * - candidate default, U - per-user static route

o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

34.0.0.0/24 is subnetted, 1 subnets

O 34.1.1.0 [110/4] via 56.1.1.5, 00:00:39, FastEthernet0/0

35.0.0.0/24 is subnetted, 1 subnets

O 35.1.1.0 [110/65] via 56.1.1.5, 00:00:39, FastEthernet0/0

2.0.0.0/24 is subnetted, 1 subnets

O 2.2.2.0 [110/3] via 56.1.1.5, 00:00:39, FastEthernet0/0

CCIE LAB认证经验分享千人群：539730342

3.0.0.0/24 is subnetted, 1 subnets

O 3.3.3.0 [110/4] via 56.1.1.5, 00:00:39, FastEthernet0/0

5.0.0.0/24 is subnetted, 1 subnets

O 5.5.5.0 [110/2] via 56.1.1.5, 00:00:39, FastEthernet0/0

23.0.0.0/24 is subnetted, 1 subnets

O 23.1.1.0 [110/3] via 56.1.1.5, 00:00:39, FastEthernet0/0

25.0.0.0/24 is subnetted, 1 subnets

O 25.1.1.0 [110/2] via 56.1.1.5, 00:00:39, FastEthernet0/0

10.0.0.0/32 is subnetted, 1 subnets

O 10.1.1.1 [110/2] via 56.1.1.5, 00:00:39, FastEthernet0/0

56.0.0.0/24 is subnetted, 1 subnets

C 56.1.1.0 is directly connected, FastEthernet0/0

12.0.0.0/24 is subnetted, 1 subnets

O 12.1.1.0 [110/66] via 56.1.1.5, 00:00:39, FastEthernet0/0

46.0.0.0/24 is subnetted, 1 subnets

C 46.1.1.0 is directly connected, Serial1/0

r6#

说明：R6 已经学习到全网的每一条路由。

(13) 在 R1 上测试到其它网段的连通性：

r1#ping 2.2.2.2

CCIE LAB认证经验分享千人群：539730342

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 2.2.2.2, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 48/98/132 ms

r1#

r1#ping 3.3.3.3

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 3.3.3.3, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 60/120/232 ms

r1#

r1#ping 5.5.5.5

Type escape sequence to abort.

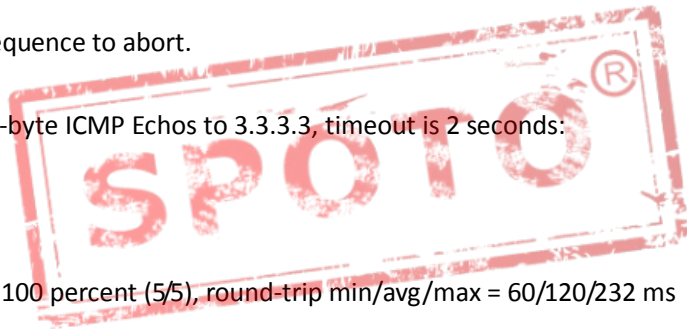
Sending 5, 100-byte ICMP Echos to 5.5.5.5, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 84/166/244 ms

r1#

r1#ping 10.1.1.1



CCIE LAB认证经验分享千人群：539730342

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 10.1.1.1, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 28/108/200 ms

r1#

r1#

r1#ping 34.1.1.4

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 34.1.1.4, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 64/170/220 ms

r1#

r1#

r1#

r1#ping 56.1.1.6

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 56.1.1.6, timeout is 2 seconds:

!!!!



CCIE LAB认证经验分享千人群：539730342

Success rate is 100 percent (5/5), round-trip min/avg/max = 160/195/280 ms

r1#

r1#ping 46.1.1.6

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 46.1.1.6, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 84/156/268 ms

r1#

说明：R1 与 2.2.2.0/24、3.3.3.0/24、5.5.5.0/24、10.1.1.1/32、34.1.1.0/24、56.1.1.0/24、46.1.1.0/24 通信正常，说明已经全网单播互通。



(14) 在 R4 上测试到 RP 地址 10.1.1.1 的连通性和路径走向：

r4#ping 10.1.1.1

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 10.1.1.1, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 24/45/76 ms

r4#

CCIE LAB认证经验分享千人群：539730342

r4#

r4#traceroute 10.1.1.1

Type escape sequence to abort.

Tracing the route to 10.1.1.1

1 34.1.1.3 96 msec * 112 msec

r4#

说明：R4 到 10.1.1.1/32 通信正常，并且是选择的 R3，路径和预期相同。

(15) 在 R6 上测试到 RP 地址 10.1.1.1 的连通性和路径走向：

r6#ping 10.1.1.1

Type escape sequence to abort.

Sending 5, 100-byte ICMP Echos to 10.1.1.1, timeout is 2 seconds:

!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 60/103/160 ms

r6#

r6#traceroute 10.1.1.1

Type escape sequence to abort.

Tracing the route to 10.1.1.1

```
1 56.1.1.5 136 msec * 116 msec
```

r6#

说明：R6 到 10.1.1.1/32 通信正常，并且是选择的 R5，路径和预期相同。

2. 配置 PIM-SM

(1) 在 R1 上配置 PIM-SM:

```
r1(config)#ip multicast-routing
```

```
r1(config)#int s1/0
```

```
r1(config-if)#ip pim sparse-mode
```

```
r1(config-if)#ip igmp join-group 224.1.1.1
```

```
r1(config-if)#exit
```

说明：在 R1 上开启组播功能，并且在接口 S1/0 下开启 PIM-SM 模式，接口 S1/0 加入组 224.1.1.1，成为组成员。

(2) 在 R2 上配置 PIM-SM:

```
r2(config)#ip multicast-routing
```

```
r2(config)#int s1/0
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#int f0/0
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#
```

```
r2(config)#int f0/1
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#int loopback 10
```

```
r2(config-if)#ip pim sparse-mode
```

```
r2(config-if)#exit
```

```
r2(config)#
```



说明：在 R2 上开启组播功能，并且在接口 Loopback 10, F0/0, F0/1, S1/0 下开启 PIM-SM 模式，因为 Loopback 0 只用来建 MSDP 邻居，并不运行组播，所以没有必要运行 PIM。

(3) 在 R3 上配置 PIM-SM:

```
r3(config)#int f0/0
```

```
r3(config-if)#ip pim sparse-mode
```

```
r3(config-if)#exit
```

```
r3(config)#
```

```
r3(config)#int f0/1
```

```
r3(config-if)#ip pim sparse-mode
```



```
r3(config-if)#exit
```

```
r3(config)#int s1/0
```

```
r3(config-if)#ip pim sparse-mode
```

```
r3(config-if)#exit
```

```
r3(config-if)#int loopback 10
```

```
r3(config-if)#ip pim sparse-mode
```

```
r3(config-if)#exit
```

```
r3(config)#
```

说明：在 R3 上开启组播功能，并且在接口 Loopback 10, F0/0, F0/1, S1/0 下开启 PIM-SM 模式，因为 Loopback 0 只用来建 MSDP 邻居，并不运行组播，所以没有必要运行 PIM。

(4) 在 R4 上配置 PIM-SM:

```
r4(config)#ip multicast-routing
```

```
r4(config)#int f0/1
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```

```
r4(config)#int s1/0
```

```
r4(config-if)#ip pim sparse-mode
```

```
r4(config-if)#exit
```

说明：在 R4 上开启组播功能，并且在接口 F0/1, S1/0 下开启 PIM-SM 模式。

(5) 在 R5 上配置 PIM-SM:

```
r5(config)#ip multicast-routing
```

```
r5(config)#int f0/1
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

```
r5(config)#int f0/0
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

```
r5(config)#
```

```
r5(config)#int s1/0
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

```
r5(config)#int loopback 10
```

```
r5(config-if)#ip pim sparse-mode
```

```
r5(config-if)#exit
```

说明：在 R5 上开启组播功能，并且在接口 Loopback 10, F0/0, F0/1, S1/0 下开启 PIM-SM 模式，因为 Loopback 0 只用来建 MSDP 邻居，并不运行组播，所以没有必要运行 PIM。

(6) 在 R6 上配置 PIM-SM:

```
r6(config)#ip multicast-routing
```

```
r6(config)#int f0/0
```

```
r6(config-if)#ip pim sparse-mode
```

```
r6(config-if)#exit
```

CCIE LAB认证经验分享千人群：539730342

```
r6(config)#int s1/0
```

```
r6(config-if)#ip pim sparse-mode
```

```
r6(config-if)#exit
```

说明：在 R6 上开启组播功能，并且在接口 F0/1，S1/0 下开启 PIM-SM 模式。

(7) 在 R2 上查看 PIM 邻居情况：

```
r2#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address		Prio/Mode		
23.1.1.3	FastEthernet0/0	00:03:05/00:01:38	v2	1 / DR S
25.1.1.5	FastEthernet0/1	00:01:22/00:01:21	v2	1 / DR S
12.1.1.1	Serial1/0	00:03:36/00:01:37	v2	1 / S

```
r2#
```

说明：R2 与 R1，R3 以及 R5 建立 PIM 邻居关系，一切正常。

(8) 在 R3 上查看 PIM 邻居情况：

```
r3#sh ip pim neighbor
```

PIM Neighbor Table

CCIE LAB认证经验分享千人群：539730342

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address				Prio/Mode
23.1.1.2	FastEthernet0/0	00:02:48/00:01:25	v2	1 / S
34.1.1.4	FastEthernet0/1	00:01:54/00:01:19	v2	1 / DR S
35.1.1.5	Serial1/0	00:01:32/00:01:40	v2	1 / DR S

r3#

说明：R3 与 R2，R4 以及 R5 建立 PIM 邻居关系，一切正常。

(9) 在 R5 上查看 PIM 邻居情况：

r5#sh ip pim neighbor

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address				Prio/Mode
56.1.1.6	FastEthernet0/0	00:01:28/00:01:15	v2	1 / DR S
25.1.1.2	FastEthernet0/1	00:02:00/00:01:42	v2	1 / S
35.1.1.3	Serial1/0	00:01:57/00:01:16	v2	1 / S

r5#

CCIE LAB认证经验分享千人群：539730342

说明： R5 与 R2, R3 以及 R6 建立 PIM 邻居关系，一切正常。

(10) 在 R4 上查看 PIM 邻居情况：

```
r4#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
Address			Prio/Mode	
34.1.1.3	FastEthernet0/1	00:01:37/00:01:36	v2	1 / S
46.1.1.6	Serial1/0	00:01:14/00:01:30	v2	1 / DR S

r4#

说明： R4 与 R3 和 R6 建立 PIM 邻居关系，一切正常。

(11) 分别将 R2, R3 以及 R5 都以 Loopback 10 配置为 RP，从而形成 Anycast RP:

```
r2(config)#access-list 24 permit 224.1.1.1
```

```
r2(config)#ip pim send-rp-announce loopback 10 scope 16 group-list 24
```

```
r2(config)#ip pim send-rp-discovery loopback 10 scope 16
```

```
r3(config)#access-list 24 permit 224.1.1.1
```

```
r3(config)#ip pim send-rp-announce loopback 10 scope 16 group-list 24
```

```
r3(config)#ip pim send-rp-discovery loopback 10 scope 16
```

```
r5(config)#access-list 24 permit 224.1.1.1
```

```
r5(config)#ip pim send-rp-announce loopback 10 scope 16 group-list 24
```

```
r5(config)#ip pim send-rp-discovery loopback 10 scope 16
```

说明：R2，R3 以及 R5 都以 Loopback 10 配置为 RP，形成 Anycast RP。

(12) 查看全网路由器的 RP 学习情况：

```
r1#sh ip pim rp mapping
```

```
PIM Group-to-RP Mappings
```

```
Group(s) 224.1.1.1/32
```

```
RP 10.1.1.1 (?), v2v1
```

```
Info source: 10.1.1.1 (?), elected via Auto-RP
```

```
Uptime: 00:01:22, expires: 00:02:35
```

```
r1#
```

```
r2#sh ip pim rp mapping
```

```
PIM Group-to-RP Mappings
```

```
This system is an RP (Auto-RP)
```



This system is an RP-mapping agent (Loopback10)

Group(s) 224.1.1.1/32

RP 10.1.1.1 (?), v2v1

Info source: 10.1.1.1 (?), elected via Auto-RP

Uptime: 00:01:44, expires: 00:02:13

r2#

r3#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback10)

Group(s) 224.1.1.1/32

RP 10.1.1.1 (?), v2v1

Info source: 10.1.1.1 (?), elected via Auto-RP

Uptime: 00:01:53, expires: 00:02:04

r3#

r3#



r4#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 10.1.1.1 (?), v2v1

Info source: 10.1.1.1 (?), elected via Auto-RP

Uptime: 00:03:41, expires: 00:02:13

r4#



r5#sh ip pim rp mapping

PIM Group-to-RP Mappings

This system is an RP (Auto-RP)

This system is an RP-mapping agent (Loopback10)

Group(s) 224.1.1.1/32

RP 10.1.1.1 (?), v2v1

Info source: 10.1.1.1 (?), elected via Auto-RP

Uptime: 00:03:51, expires: 00:02:09

r5#

r6#sh ip pim rp mapping

PIM Group-to-RP Mappings

Group(s) 224.1.1.1/32

RP 10.1.1.1 (?), v2v1

Info source: 10.1.1.1 (?), elected via Auto-RP

Uptime: 00:04:41, expires: 00:02:12

r6#

说明：全网的路由器学习到的 RP 地址都为 10.1.1.1，然而 R2，R3 以及 R5 就是以 10.1.1.1 地址为 RP，所以每台路由器都走拓扑上离自己最近的 RP，如 R1 选择 R2，R4 选择 R3，而 R6 选择 R5，这就是 Anycast RP。

（13）测试组播源 R4 和 R6 到组 224.1.1.1 的通信情况：

r4#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

CCIE LAB认证经验分享千人群：539730342

r4#

r6#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

r6#

说明：因为 R1 选择 R2 作 RP，R4 选择 R3 作 RP，而 R6 选择 R5 作 RP，所以每一台 RP 上都没有完整的组播源和组成员信息，最终全网的组播是无法通信的，只要通过 MSDP 让这些 RP 之间共享自己已知的组播源或组成员信息，这样才能保证每一台 RP 都有完整的组播源和组成员信息，才能保证全网组播互通。

(14) 查看 R3 和 R2 的组播树情况：

r3#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

CCIE LAB认证经验分享千人群：539730342

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 00:01:17/stopped, RP 10.1.1.1, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(34.1.1.4, 224.1.1.1), 00:01:17/00:02:01, flags: PT

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list: Null

r3#

r2#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:13:45/00:02:39, RP 10.1.1.1, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:13:45/00:02:39

r2#

说明：正是因为 R3 没有组成员信息，R2 没有组播源信息，所以全网的组播树无法形成，导致全网组播无法通信。

3. 配置 MSDP

(1) 在 R2 和 R3 之间建立 MSDP 连接：

CCIE LAB认证经验分享千人群：539730342

```
r2(config)#ip msdp peer 3.3.3.3 connect-source loopback 0
```

```
r2(config)#ip msdp originator-id loopback 0
```

```
r3(config)#ip msdp peer 2.2.2.2 connect-source loopback 0
```

```
r3(config)#ip msdp originator-id loopback 0
```

说明：在 R2 和 R3 之间以 loopback 0 的地址为源建立 MSDP 连接，并且将自己发出去的 SA 数据包中的 RP 地址改成 loopback 0 的地址，因为默认都是 loopback 10 的地址，这样会造成全网都一样，并且会导致 RPF 检测失败。

(2) 查看 MSDP 的连接情况：

```
r2#sh ip msdp summary
```

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

			Downtime	Count	Count
--	--	--	----------	-------	-------

3.3.3.3	?	Up	00:00:27	0	?
---------	---	----	----------	---	---

```
r2#
```

说明：R2 与 R3 之间的 MSDP 连接已经正常建立。

(3) 再次测试组播源 R4 到组 224.1.1.1 的通信情况，并在 MSDP 路由器 R2 上通过 Debug 观察 MSDP 信息：

在 MSDP 路由器 R2 上开启 Debug:

CCIE LAB认证经验分享千人群：539730342

```
r2#debug ip msdp peer
```

MSDP Peer debugging is on

```
r2#
```

```
r2#debug ip msdp detail
```

MSDP Detail debugging is on

```
r2#
```

在组播源路由器 R4 上发起组播流量：

```
r4#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 276 ms

Reply to request 1 from 12.1.1.1, 208 ms

Reply to request 2 from 12.1.1.1, 220 ms

Reply to request 3 from 12.1.1.1, 204 ms

Reply to request 4 from 12.1.1.1, 240 ms

Reply to request 5 from 12.1.1.1, 236 ms

Reply to request 6 from 12.1.1.1, 200 ms

Reply to request 7 from 12.1.1.1, 264 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 8 from 12.1.1.1, 236 ms

Reply to request 9 from 12.1.1.1, 296 ms

r4#

说明：和预期一样，在 RP 之间建立 MSDP 共享 SA 信息之后，组播通信正常。

观察 R2 上的 Debug 信息：

r2#

*Mar 1 00:50:00.291: MSDP(0): Received 120-byte TCP segment from 3.3.3.3

*Mar 1 00:50:00.291: MSDP(0): Append 120 bytes to 0-byte msg 6 from 3.3.3.3, qs 1

*Mar 1 00:50:00.291: MSDP(0): 3.3.3.3: Received 120-byte msg 6 from peer

*Mar 1 00:50:00.295: MSDP(0): 3.3.3.3: SA TLV, len: 120, ec: 1, RP: 3.3.3.3, with data

*Mar 1 00:50:00.295: MSDP(0): 3.3.3.3: Peer RPF check passed for single peer

*Mar 1 00:50:00.299: MSDP(0): WAVL Insert SA Source 34.1.1.4 Group 224.1.1.1 RP 3.3.3.3 Successful

*Mar 1 00:50:00.299: MSDP(0): Forward decapsulated SA data for (34.1.1.4, 224.1.1.1) on Serial1/0

r2#

说明：因为目前 R2 只有单个 MSDP peer，所以对于 R3 发来的 SA 也就没必要进行 RPF 检测，最后 R2 顺利将关于组播源 34.1.1.4（R4）的 SA 缓存起来。

（4）在 R2 上查看 SA 信息的接收情况：

r2#sh ip msdp sa-cache

MSDP Source-Active Cache - 1 entries

(34.1.1.4, 224.1.1.1), RP 3.3.3.3, AS 7, 00:01:52/00:05:41, Peer 3.3.3.3

Learned from peer 3.3.3.3, RPF peer 3.3.3.3,

SAs received: 3, Encapsulated data received: 1

r2#

说明：R3 发来的 SA 信息已被成功缓存并使用。

(5) 查看 R4 的组播树情况：

r4#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 00:00:18/stopped, RP 10.1.1.1, flags: SPF

CCIE LAB认证经验分享千人群：539730342

Incoming interface: FastEthernet0/1, RPF nbr 34.1.1.3

Outgoing interface list: Null

(34.1.1.4, 224.1.1.1), 00:00:14/00:03:22, flags: T

Incoming interface: Serial1/0, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:00:14/00:03:15

r4#

说明：从 R4 的组播树可以看出，流量是从接口 F0/1 发出去的，即是走 R3 的，与预期一样。



(6) 查看 R3 的组播树情况：

r3#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

CCIE LAB认证经验分享千人群：539730342

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:00:23/stopped, RP 10.1.1.1, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(34.1.1.4, 224.1.1.1), 00:00:23/00:03:24, flags: TA

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:00:19/00:03:10

r3#

说明：从 R3 的组播树可以看出，R4 的组播流量路径正常，与预期一样。

(7) 查看 R2 的组播树情况：

r2#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

CCIE LAB认证经验分享千人群：539730342

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:00:51/stopped, RP 10.1.1.1, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:00:51/00:02:08

(34.1.1.4, 224.1.1.1), 00:00:51/00:02:25, flags: MT

Incoming interface: FastEthernet0/0, RPF nbr 23.1.1.3

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:00:51/00:02:08

r2#

说明：从 R2 的组播树可以看出，R4 的组播流量路径正常，与预期一样。

(8) 测试组播源 R6 到组 224.1.1.1 的通信情况：

```
r6#ping 224.1.1.1 repeat 10
```

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

.....

```
r6#
```

说明：因为 R4 是选择 R3 作为 RP 的，而 R6 是选择 R5 作为 RP 的，R2 与 R3 之间的 MSDP 正常，但 R2 与 R5 却没有 MSDP，所以 R6 的组播不通是正常的，要解决该问题，应该先在 R2 与 R5 之间创建 MSDP 连接。

(9) 在 R2 和 R5 之间建立 **MSDP** 连接：

```
r2(config)#ip msdp peer 5.5.5.5 connect-source loopback 0
```

```
r5(config)#ip msdp peer 2.2.2.2 connect-source loopback 0
```

```
r5(config)#ip msdp originator-id loopback 0
```

说明：在 R2 和 R5 之间以 loopback 0 的地址为源建立 MSDP 连接，R5 也将自己发出去的 SA 数据包中的 RP 地址改成 loopback 0 的地址。

(10) 在 R2 上再次查看 **MSDP** 的连接情况：

```
r2#sh ip msdp summary
```

CCIE LAB认证经验分享千人群：539730342

MSDP Peer Status Summary

Peer Address	AS	State	Uptime/	Reset SA	Peer Name
--------------	----	-------	---------	----------	-----------

			Downtime	Count	Count
--	--	--	----------	-------	-------

3.3.3.3	?	Up	00:08:05	0	1 ?
---------	---	----	----------	---	-----

5.5.5.5	?	Up	00:00:05	0	1 ?
---------	---	----	----------	---	-----

r2#

说明：R2 与 R5 之间的 MSDP 也已经正常建立。

(11) 再次测试组播源 R6 到组 224.1.1.1 的通信情况，并观察 R2 上的 Debug 信息：

r6#ping 224.1.1.1 repeat 10

Type escape sequence to abort.

Sending 10, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 212 ms

Reply to request 1 from 12.1.1.1, 256 ms

Reply to request 2 from 12.1.1.1, 212 ms

Reply to request 3 from 12.1.1.1, 304 ms

Reply to request 4 from 12.1.1.1, 212 ms

Reply to request 5 from 12.1.1.1, 152 ms

Reply to request 6 from 12.1.1.1, 192 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 7 from 12.1.1.1, 184 ms

Reply to request 8 from 12.1.1.1, 160 ms

Reply to request 9 from 12.1.1.1, 212 ms

r6#

说明：在 R2 与 R5 之间建立 MSDP 连接之后，便能发送组播源信息，所以此时 R6 的组播通信正常。

观察 R2 上的 Debug 信息：

r2#

r2#Received 20-byte msg 11 from peer

*Mar 1 00:53:52.855: MSDP(0): 5.5.5.5: SA TLV, len: 20, ec: 1, RP: 5.5.5.5

*Mar 1 00:53:52.859: MSDP(0): 5.5.5.5: Peer RPF check passed for 5.5.5.5, peer is RP

*Mar 1 00:53:52.859: MSDP(0): WAVL Insert SA Source 56.1.1.6 Group 224.1.1.1 RP 5.5.5.5 Successful

*Mar 1 00:53:52.863: MSDP(0): 3.3.3.3: Add SA entry (56.1.1.6, 224.1.1.1) RP: 5.5.5.5

*Mar 1 00:53:52.867: MSDP(0): 3.3.3.3: Send 20-byte SA message to 3.3.3.3, 1 entries

r2#

说明：虽然 R2 上有两个 MSDP 连接，但是因为 R3 和 R5 建立 MSDP 的地址和 SA 数据包中的 RP 地址都是 Loopback 0，所以能够正常通过 RPF 检测，这就是为什么之前要将 SA 数据包中 RP 地址更改为 Loopback 0 的地址的原因。

(12) 在 R2 上再次查看 SA 信息：

CCIE LAB认证经验分享千人群：539730342

```
r2#sh ip msdp sa-cache
```

MSDP Source-Active Cache - 2 entries

(34.1.1.4, 224.1.1.1), RP 3.3.3.3, AS ?,00:07:38/00:00:51, Peer 3.3.3.3

Learned from peer 3.3.3.3, RPF peer 3.3.3.3,

SAs received: 4, Encapsulated data received: 1

(56.1.1.6, 224.1.1.1), RP 5.5.5.5, AS ?,00:03:45/00:03:14, Peer 5.5.5.5

Learned from peer 5.5.5.5, RPF peer 5.5.5.5,

SAs received: 2, Encapsulated data received: 0

```
r2#
```

说明：R3 和 R5 发来的 SA 信息已被 R2 成功缓存并使用，所以现在全网的组播通信均正常。



(13) 查看 R6 的组播树情况：

```
r6#sh ip mroute 224.1.1.1
```

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

CCIE LAB认证经验分享千人群：539730342

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:00:05/stopped, RP 10.1.1.1, flags: SPF

Incoming interface: FastEthernet0/0, RPF nbr 56.1.1.5

Outgoing interface list: Null

(56.1.1.6, 224.1.1.1), 00:00:05/00:02:59, flags: T

Incoming interface: FastEthernet0/0, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:00:14/00:02:16

r6#

说明：从 R6 的组播树可以看出，流量是从接口 F0/0 发出去的，即是走 R5 的，与预期一样。

（14）查看 R5 的组播树情况：

r5#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

CCIE LAB认证经验分享千人群：539730342

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

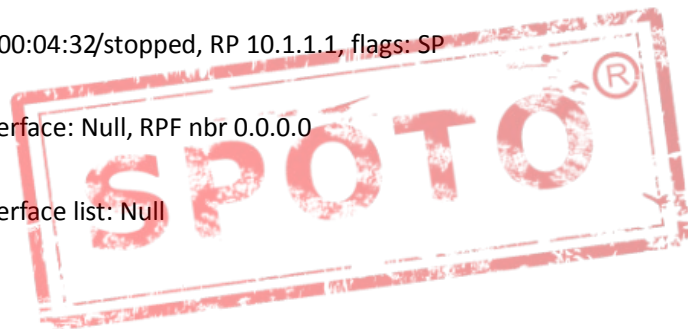
Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:04:32/stopped, RP 10.1.1.1, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null



(46.1.1.4, 224.1.1.1), 00:01:33/00:01:56, flags:

Incoming interface: FastEthernet0/0, RPF nbr 56.1.1.6

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:01:33/00:02:55

(56.1.1.6, 224.1.1.1), 00:04:32/00:02:50, flags: TA

Incoming interface: FastEthernet0/0, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:04:32/00:03:07

r5#

说明：从 R5 的组播树可以看出，R6 的组播流量路径正常，与预期一样，从目前的情况来看，Anycast RP 确实为组播网络带来了负载均衡的效果。

4. 测试 Anycast RP 冗余性

说明：因为 R2, R3 以及 R5 的 RP 地址都配置为 10.1.1.1，并且该地址在 OSPF 里面全网可达，如果当 R4 使用的 RP 路由器 R3 出现故障，那么 R4 应该能够根据 OSPF 路由的变化从而切换到 R5，实际冗余性，对于 R6 也一样，当 R6 使用的 RP 路由器 R5 出现故障，那么 R6 应该能够根据 OSPF 路由的变化从而切换到 R3，下面我们来测试让 R6 使用的 RP 路由器 R5 出现故障，看看 R6 的组播通信情况。

(1) 调整 R4 与 R6 之间的 PIM 关系：

```
r4(config)#int s1/0
```

```
r4(config-if)#ip pim dr-priority 4
```

```
r4(config-if)#exit
```

```
r6#sh ip pim neighbor
```

PIM Neighbor Table

Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,

S - State Refresh Capable

Neighbor	Interface	Uptime/Expires	Ver	DR
----------	-----------	----------------	-----	----

CCIE LAB认证经验分享千人群：539730342

Address		Prio/Mode
56.1.1.5	FastEthernet0/0	00:01:38/00:01:35 v2 1 / S
46.1.1.4	Serial1/0	00:07:17/00:01:21 v2 4 / DR S

r6#

说明：提高 R4 的 S1/0 接口的 DR 优先级后，R4 成为了 DR。

(2) 关闭 RP 路由器 R5，并观察 R4 的组播流量情况：

r4#ping 224.1.1.1 repeat 1000

Type escape sequence to abort.

Sending 1000, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 144 ms

Reply to request 0 from 12.1.1.1, 444 ms

Reply to request 1 from 12.1.1.1, 312 ms

Reply to request 2 from 12.1.1.1, 328 ms

Reply to request 3 from 12.1.1.1, 308 ms

Reply to request 4 from 12.1.1.1, 272 ms

Reply to request 5 from 12.1.1.1, 208 ms

Reply to request 6 from 12.1.1.1, 132 ms

CCIE LAB认证经验分享千人群：539730342

Reply to request 7 from 12.1.1.1, 208 ms

Reply to request 8 from 12.1.1.1, 216 ms

Reply to request 9 from 12.1.1.1, 200 ms

Reply to request 10 from 12.1.1.1, 216 ms

Reply to request 11 from 12.1.1.1, 216 ms

Reply to request 12 from 12.1.1.1, 200 ms

Reply to request 13 from 12.1.1.1, 172 ms

Reply to request 14 from 12.1.1.1, 188 ms

Reply to request 15 from 12.1.1.1, 188 ms

Reply to request 16 from 12.1.1.1, 188 ms

Reply to request 17 from 12.1.1.1, 156 ms

Reply to request 18 from 12.1.1.1, 156 ms

Reply to request 19 from 12.1.1.1, 208 ms

Reply to request 20 from 12.1.1.1, 280 ms

Reply to request 21 from 12.1.1.1, 200 ms

Reply to request 22 from 12.1.1.1, 216 ms

Reply to request 23 from 12.1.1.1, 280 ms

Reply to request 24 from 12.1.1.1, 216 ms

Reply to request 25 from 12.1.1.1, 216 ms

Reply to request 26 from 12.1.1.1, 216 ms

Reply to request 27 from 12.1.1.1, 296 ms



CCIE LAB认证经验分享千人群：539730342

Reply to request 28 from 12.1.1.1, 216 ms

Reply to request 29 from 12.1.1.1, 172 ms

Reply to request 30 from 12.1.1.1, 172 ms

Reply to request 31 from 12.1.1.1, 192 ms

Reply to request 32 from 12.1.1.1, 196 ms

r4#

说明：在关闭 RP 路由器 R5 期间，R4 的组播通信不受影响，因为 R4 是靠 R3 来通信的，与 R5 无关。

(3) 观察关闭 RP 路由器 R5 后，R6 的组播流量情况：

r6#ping 224.1.1.1 repeat 1000



Type escape sequence to abort.

Sending 1000, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:

Reply to request 0 from 12.1.1.1, 440 ms

Reply to request 0 from 12.1.1.1, 692 ms

Reply to request 1 from 12.1.1.1, 200 ms

Reply to request 2 from 12.1.1.1, 216 ms

Reply to request 3 from 12.1.1.1, 192 ms.....

*Mar 1 00:10:42.131: %OSPF-5-ADJCHG: Process 1, Nbr 5.5.5.5 on FastEthernet0/0
from FULL to DOWN, Neighbor Down: Dead timer expired...

Reply to request 23 from 12.1.1.1, 460 ms

Reply to request 24 from 12.1.1.1, 264 ms

Reply to request 25 from 12.1.1.1, 404 ms

Reply to request 26 from 12.1.1.1, 360 ms

Reply to request 27 from 12.1.1.1, 324 ms

Reply to request 28 from 12.1.1.1, 372 ms

Reply to request 29 from 12.1.1.1, 388 ms

Reply to request 30 from 12.1.1.1, 328 ms

r6#

说明：因为 R6 是优先选择 RP 路由器 R5 来通信的，所以当 R5 出现故障后，R6 的组播被中断，但最终又重新选择另一台可用的 RP 路由器 R3，从而又恢复了组播通信，体现了 Anycast RP 的冗余性。

（4）查看 R4 的组播树情况：

r4#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

CCIE LAB认证经验分享千人群：539730342

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:01:43/stopped, RP 10.1.1.1, flags: SPF

Incoming interface: FastEthernet0/1, RPF nbr 34.1.1.3

Outgoing interface list: Null

(34.1.1.4, 224.1.1.1), 00:01:43/00:02:21, flags: PFT

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list: Null

(46.1.1.6, 224.1.1.1), 00:01:35/00:02:59, flags: FT

Incoming interface: Serial1/0, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/1, Forward/Sparse, 00:00:50/00:02:54

(56.1.1.6, 224.1.1.1), 00:00:50/00:02:39, flags:

Incoming interface: Serial1/0, RPF nbr 46.1.1.6

Outgoing interface list:

CCIE LAB认证经验分享千人群：539730342

FastEthernet0/1, Forward/Sparse, 00:00:50/00:02:39

r4#

说明：在 R6 优先选择 RP 路由器 R5 出现故障后，重新选择了从备用路径 R4 走。

(5) 查看 R3 的组播树情况：

r3#sh ip mroute 224.1.1.1

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:02:55/stopped, RP 10.1.1.1, flags: SP

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list: Null

(34.1.1.4, 224.1.1.1), 00:02:55/00:02:21, flags: TA

Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:02:55/00:02:32

(46.1.1.6, 224.1.1.1), 00:02:08/00:02:31, flags: TA

Incoming interface: FastEthernet0/1, RPF nbr 34.1.1.4

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:01:24/00:03:20



(56.1.1.6, 224.1.1.1), 00:01:24/00:02:05, flags:

Incoming interface: FastEthernet0/1, RPF nbr 34.1.1.4

Outgoing interface list:

FastEthernet0/0, Forward/Sparse, 00:01:24/00:03:19

r3#

说明：R3 上的组播树也体现了 R6 的流量是从 R4 再到 R3 被转发给 R2 的。

(6) 查看 R2 的组播树情况：

CCIE LAB认证经验分享千人群：539730342

```
r2#sh ip mroute 224.1.1.1
```

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,

L - Local, P - Pruned, R - RP-bit set, F - Register flag,

T - SPT-bit set, J - Join SPT, M - MSDP created entry,

X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,

U - URD, I - Received Source Specific Host Report,

Z - Multicast Tunnel, z - MDT-data group sender,

Y - Joined MDT-data group, y - Sending to MDT-data group

Outgoing interface flags: H - Hardware switched, A - Assert winner

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(* , 224.1.1.1), 00:10:20/00:00:28, RP 10.1.1.1, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:10:20/00:02:48

(34.1.1.4, 224.1.1.1), 00:08:15/00:01:29, flags: MT

Incoming interface: FastEthernet0/0, RPF nbr 23.1.1.3

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:08:15/00:02:48

(46.1.1.6, 224.1.1.1), 00:02:31/00:01:29, flags: MT

Incoming interface: FastEthernet0/0, RPF nbr 23.1.1.3

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:02:31/00:02:48

(56.1.1.6, 224.1.1.1), 00:02:31/00:00:39, flags: M

Incoming interface: FastEthernet0/0, RPF nbr 23.1.1.3

Outgoing interface list:

Serial1/0, Forward/Sparse, 00:02:31/00:02:48



r2#

说明：R2 上的组播树表明 R4 和 R6 的流量都是从 R3 过来的，和我们预期的路径相同。

查看当前环境中所有路由器的配置信息请点击“[running-config](#)”

IPv6 Multicast

IPv6 Multicast 在此主题中不再详细讨论,具体内容请阅读 IPv6 部分的 IPv6 Multicast, [点此进入](#).

附：

在测试组播时，除了使用ICMP之外，Cisco IOS中还内置了专门测试组播的工具-MRM(Multicast Routing Monitor),由于时间关系，本站不再对MRM以及其它组播技术进行详细介绍。对于Cisco MRM的应用及配置，请自行参阅思科文档，[点此进入](#)

