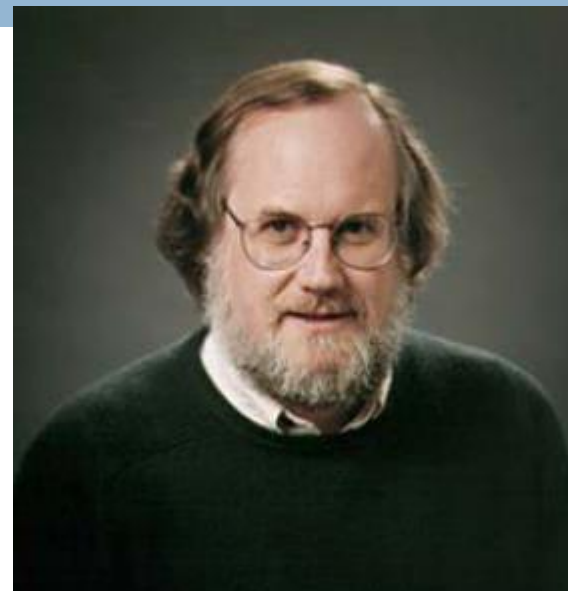


现代计算机网络

3.2 Multicasting-组播

2

- 20世纪80年代初，斯坦福大学的一个博士生Steve Deering正在为他的导师David Cheriton设计的分布式系统(V distributed system)解决独立计算机之间的通信问题。这些计算机最初在Ethernet的同一网段内，通过MAC层组播进行协同通信。
- 但随着项目进展，新加入的主机需跨越路由器进行连接。需要在网络层解决多个计算机之间的通信问题，以便构建一个跨越Ethernet网段的松散耦合系统。
- 在研究了OSPF和RIP协议之后，Steve认为，完全可以通过扩充OSPF协议，或者提供一个类似RIP的距离矢量协议来实现IP层的组播。该思想经过一系列研究和改进，最终发展为MOSPF和DVMRP协议
- Steve还提出主机成员资格协议，经过几个版本的修订最终确定为IGMP协议。Steve的这些研究工作奠定了IP组播技术的基础框架，并逐步发展为今天的组播协议集。



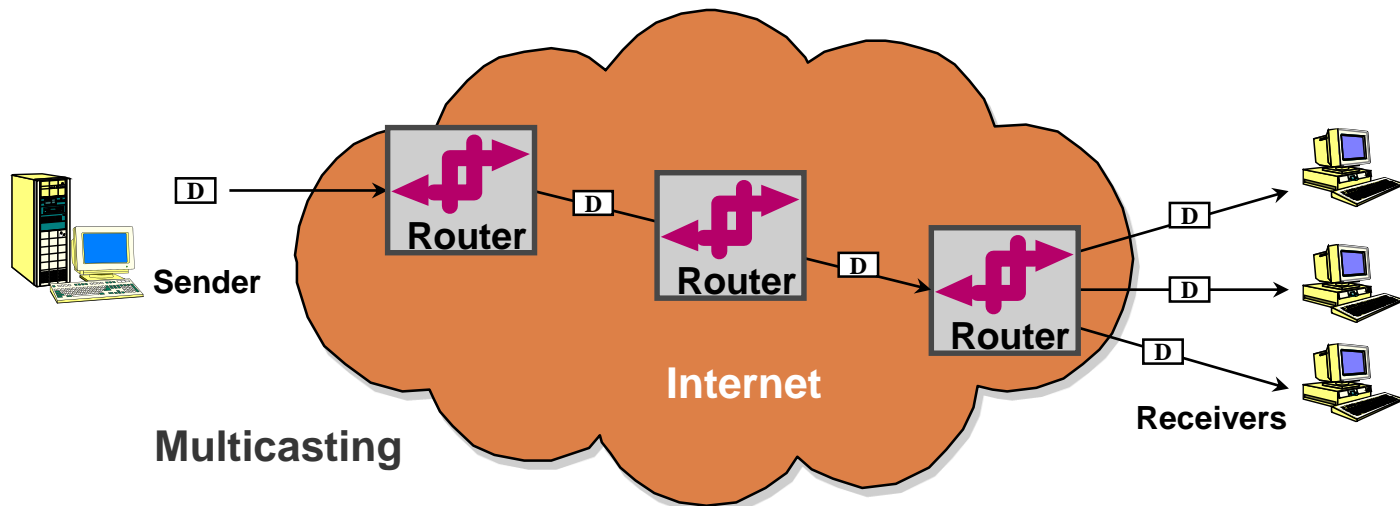
3.2 Multicasting-组播

3

- IP组播(IP multicast)又称为IP多播，特点：
 - ▣ IP组播是介于单播和广播之间的一种分组传送形式，它既不指定明确的接收者，也不是将数据分发给网络上的所有主机。
 - ▣ 发送者不关心接收者所处的位置，只要将数据发送到约定的目的地址，剩下的工作就交给网络去完成。
 - ▣ 这样做的好处在于，发送者只产生一份数据，而不用为每个接收者分别复制一份，而且发送者不必维护接收者的信息。
 - ▣ 网络中的路由器设备必须收集接收者的信息，并按照正确的路径实现组播报文的转发和复制。

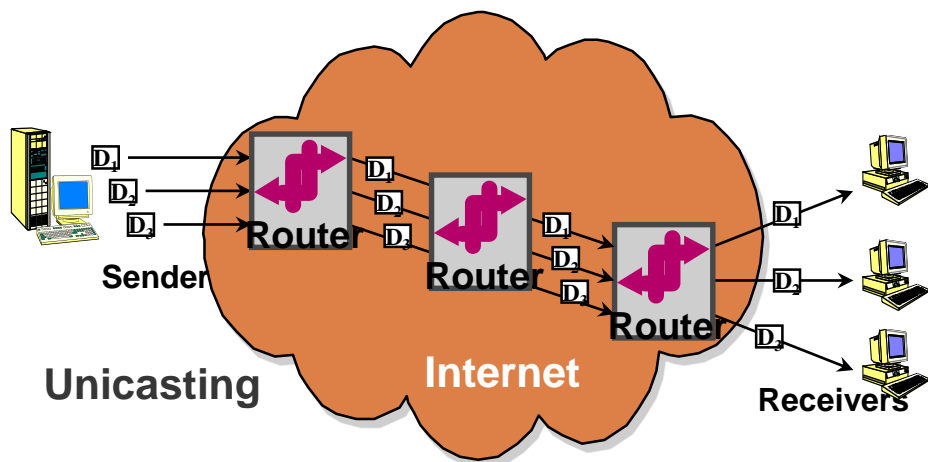
组播技术

4



Multicast

- 在 IP 网上一对多的传输
- 支持视频会议,
- e-learning, 培训等



3.2 Multicasting-组播

5

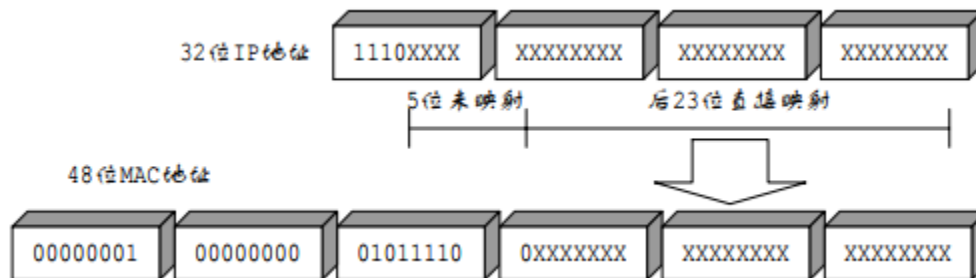
组播的三个组成部分

- 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）

1) Multicasting-IPv4组播地址

6

- 从数据格式来看，组播与单播使用共同的IP报文结构和上层封装，只是目的地址使用一段特殊的地址空间。这段地址在IPv4中被称为D类地址空间：224.0.0.0-239.255.255.255。
- 这段地址虽然看上去是连续地址，但不能聚合为网段地址，而是每个IP独立分配和使用。
- IANA 将 MAC 地址范围 01:00:5E:00:00:00 ~ 01:00:5E:7F:FF:FF 分配给组播使用 这就要求将28位的 IP 组播地址空间映射到 23 位的 MAC 地址空间中 具体的映射方法是将组播地址中的低 23 位放入 MAC 地址的低 23 位 如下图所示：



1) Multicasting-组播

7

特殊的组播地址

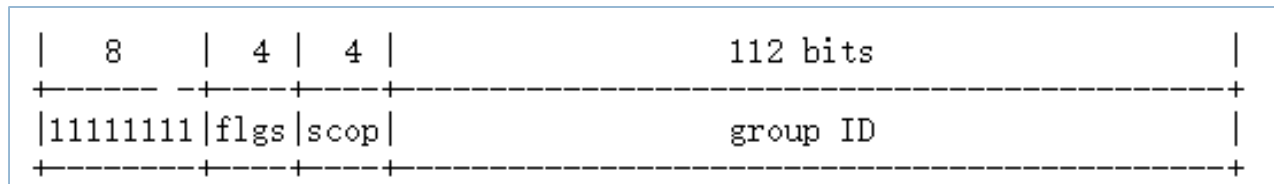
- 224.0.0.1 所有组播主机和路由器
- 224.0.0.2 所有组播路由器
- 224.0.0.4 DRMRP 路由器
- 224.0.0.5 所有 OSPF 的路由器
- 224.0.0.6 OSPF 指派路由器
- 224.0.0.9 RPIv2 路由器
- 224.0.0.10 EIGRP 路由器
- 224.0.0.13 PIM 路由器
- 224.0.0.22 IGMPv3
- 224.0.0.25 RGMP
- 224.0.1.1 NTP 网络时间协议

1) Multicasting-IPv6组播地址

8

IPv6的组播地址，RFC 4291规定，组播地址有特殊结构

Address type	Binary prefix	IPv6 notation	Section
Unspecified	00...0 (128 bits)	::/128	2.5.2
Loopback	00...1 (128 bits)	::1/128	2.5.3
Multicast	11111111	FF00::/8	2.7
Link-Local unicast	1111111010	FE80::/10	2.5.6
Global Unicast	(everything else)		



1) Multicasting-组播

9

IPv6的组播地址

- 组播地址高8bit为固定值FF
- 接着8个bit中4bit为flags位，4bit为组播组的泛洪范围。
- flags位为4bit：|0|R|P|T|。
 - ▣ flags位的第一个bit必须设置为0
 - ▣ T位如果为置0表示永久分配或者是well-known组播地址，如果置1表示临时分配动态的地址，不固定。
 - ▣ P位如果置1的话表示此组播地址是一个基于单播前缀的ipv6组播地址。默认为0，如果P位设置为1，那么T位必须为1。
 - ▣ R位如果置1的话表示此组播地址是一个内嵌RP地址的ipv6组播地址。默认为0。

1) Multicasting-组播

10

IPv6的组播地址

□ 4bit scope位来限制组播组的传播范围。

- 0 reserved
- 1 Interface-Local scope
- 2 Link-Local scope--链路本地范围
- 3 reserved
- 4 Admin-Local scope-管理本地范围
- 5 Site-Local scope--站点本地范围
- 6 (unassigned)
- 7 (unassigned)
- 8 Organization-Local scope-组织本地范围
- 9 (unassigned)
- A (unassigned)
- B (unassigned)
- C (unassigned)
- D (unassigned)
- E Global scope--全局范围的
- F reserved

2) 组成员机制

11

- 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，让路由器停止向其转发组播流量。

2) 组成员机制

12

- 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
- 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用来完成这样的工作。

2) 组成员机制

13

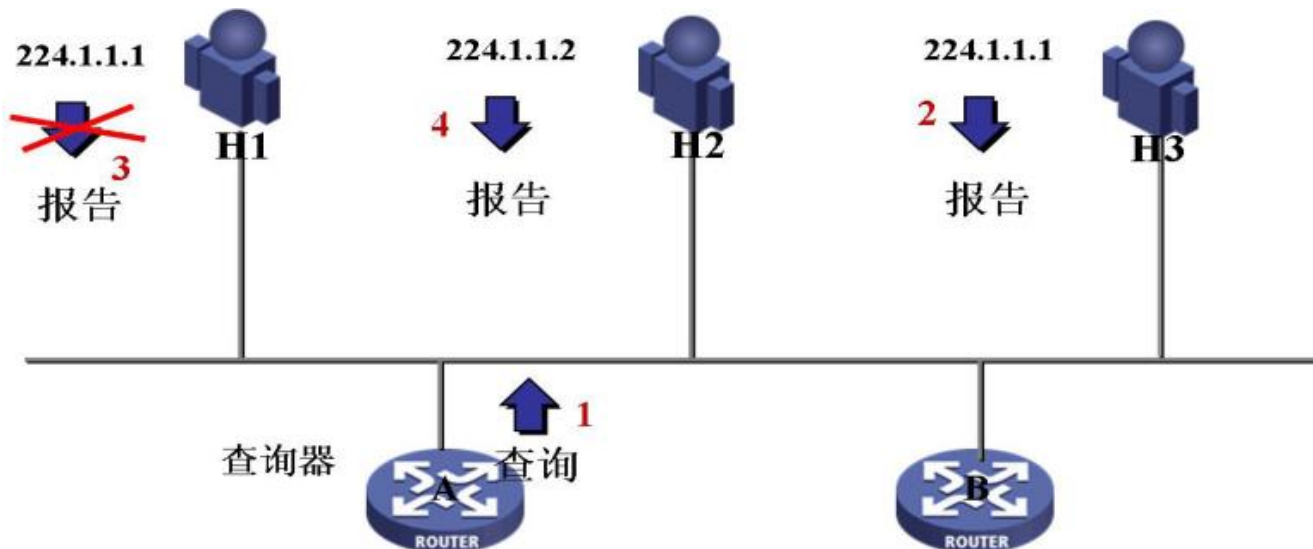
- 到目前为止，IGMP有三个版本：
 - IGMPv1（由RFC 1112定义）
 - IGMPv2（由RFC 2236定义）
 - IGMPv3（由RFC 3376定义）
- IGMP实现的功能是双向的：一方面，主机通过IGMP协议通知本地路由器希望加入并接收某个特定组播组的信息；另一方面，路由器通过IGMP协议周期性地查询局域网内某个已知组的成员是否处于活动状态（即该网段是否仍有属于某个组播组的成员），实现对所连网络组成员关系的收集与维护。
- 通过上述IGMP机制，在组播路由器里建立起一张表，其中包含路由器的各个接口以及在接口所对应的子网上都有哪些组的成员。当路由器接收到某个组的数据报文后，只向那些有该组成员的接口上转发数据报文。
- 至于数据报文在**路由器之间如何转发**则由组播路由协议决定，**IGMP协议并不负责**。

2) 组成员机制

14

加入/离开组播过程：

- 加入过程：当主机加入一个组播组时，不是等待路由器发出成员查询报文时再响应加入，而是立即主动地向一个该组播地址发出一个成员报告（类型为0x2）表示加入，这样在该组内没有其它组成员的情况下缩短了主机加入该组播组的延迟。
- 离开过程：在IGMPv1中没有定义特殊的离开机制。IGMPv1主机在任何时候“被动地”、“安静地”离开组播组而不用给路由器发任何通知。如果子网内还有同一组播组的其它成员，那么主机这样离开组播组不会出现什么问题；但如果该主机是子网内离开组播组的最后一个成员，那么路由器在一段时间仍会继续转发该组播组的数据到该网段（**IGMPv2，有了离开报告机制**）

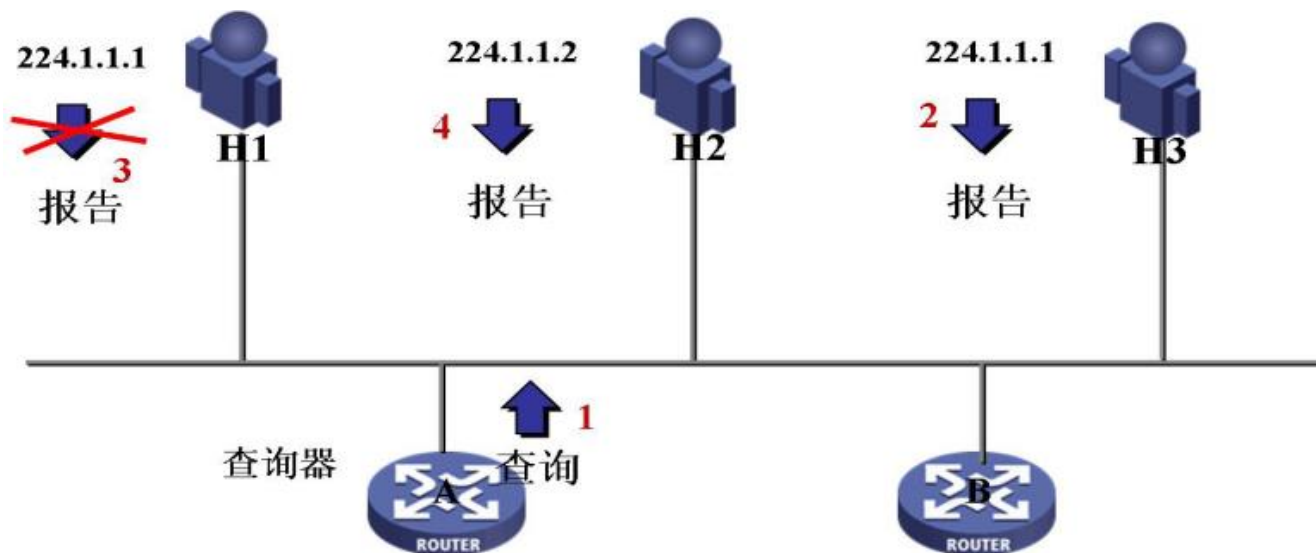


2) 组成员机制

15

维护组播成员过程：

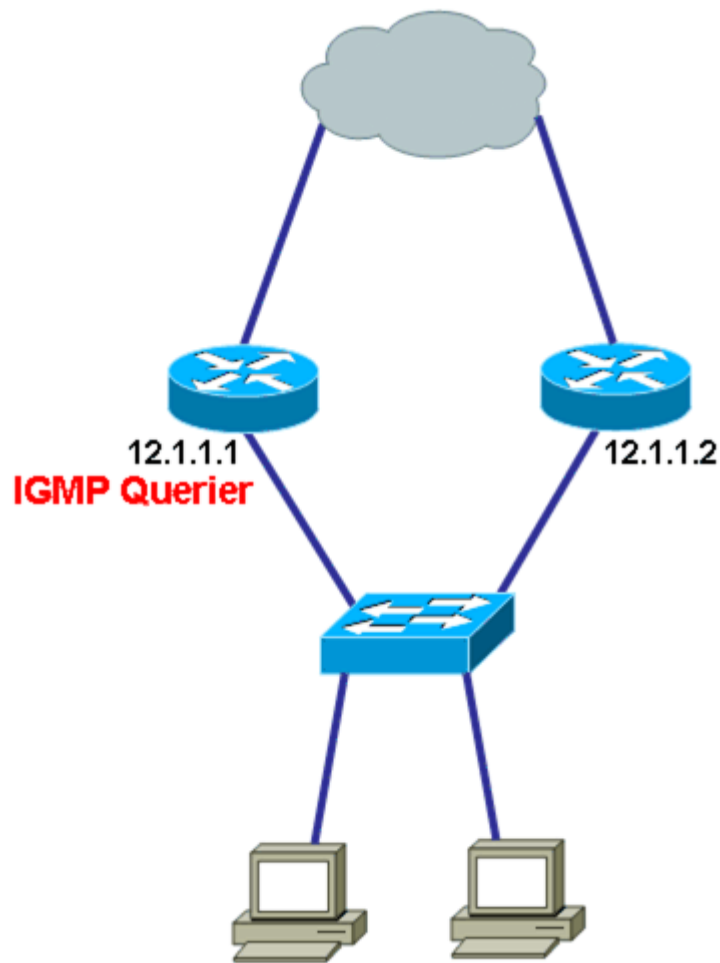
1. 查询器周期性（60s）的向子网组播组(224.0.0.1)的所有主机发送IGMPv1成员关系查询；
2. 所有主机收到IGMPv1的成员关系查询，并且某一主机(H3)首先通过向地址224.1.1.1发送成员关系报告做出响应，通知子网中的路由器有一个主机对接收组224.1.1.1感兴趣；
3. 由于主机H1正在关注组播组224.1.1.1,因此它收到主机H3发送的成员关系报告后就抑制它自己对组播组224.1.1.1的发送。有助于减少子网中信息流的数量；
4. 主机H2也收到成员关系查询，并且通过向地址224.1.1.2发送成员关系报告作出响应



2) 组成员机制

16

- 右图可以看出，如果网络中的两台路由器都发送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。
- IGMPv2为了改善性能，会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。

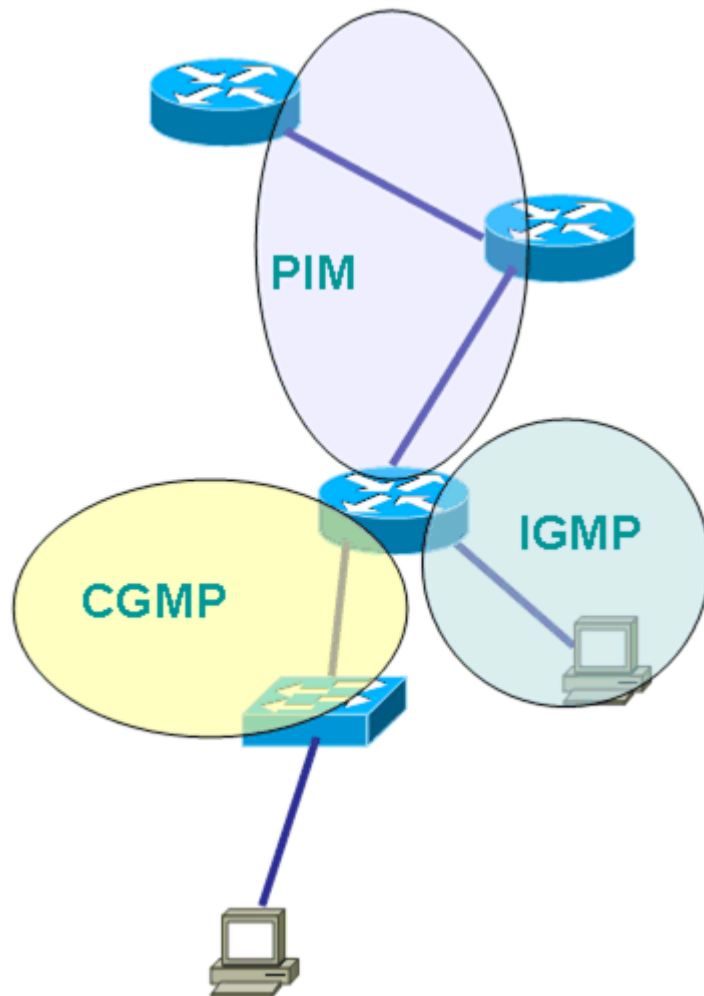


3) 组播路由协议

17

组播协议总体结构：

1. 路由器、主机之间：IGMP
2. 路由器、路由器之间：PIM
3. 路由器、交换机之间：CGMP
 - ▣ Cisco Group Management Protocol (CGMP) is a Cisco-developed protocol that allows Catalyst switches to leverage IGMP information on Cisco routers to make Layer 2 forwarding decisions.



3) 组播路由协议

18

PIM协议

- 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。
- 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）。

3) 组播路由协议

19

PIM基本概念：组播树

- 组播发送者到接收者之间的转发线路，被称为组播转发树。
- 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。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。
- 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，最终便形成了组播树。而记住组播的出口信息，这就是组播路由表的工作

3) 组播路由协议

20

PIM基本概念：防止组播反向路径转发

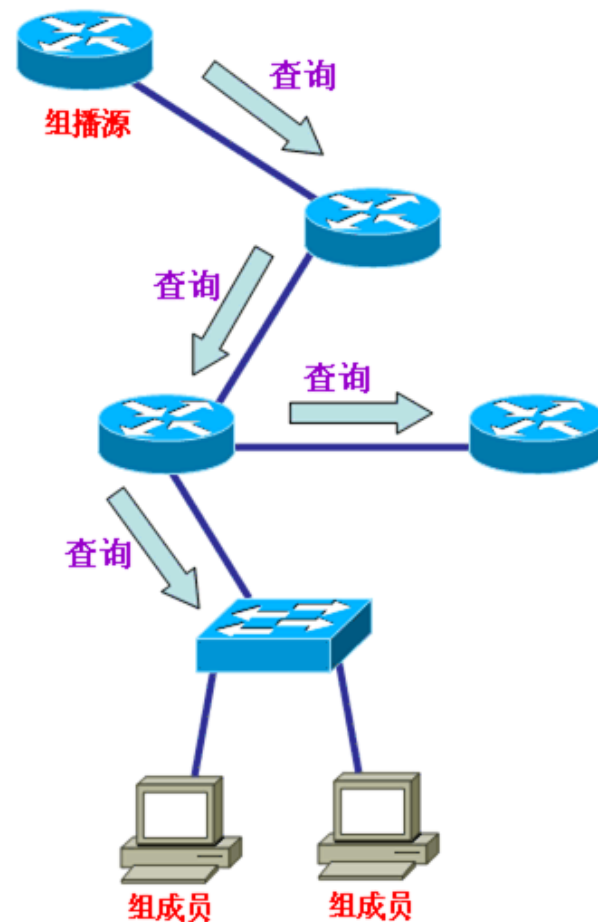
- 因为组播经常会有多个接收者分布在不同的网络，所以当路由器在转发组播时，也通常需要将同一份数据从多个接口发出去
- 由此可以看出，转发组播的路由器，必须拥有发现环路、避免环路的能力。
- 所有的组播路由器必须只将数据往接收者的方向转发，而不能往发送者的方向转发
- 向发送者转发称为RPF（Reverse Path Forwarding），组播路由器在收到组播数据后，都要对数据进行RPF检测，组播路由表记住接收者的接口，就能正常将数据发给组成员，而记住了发送者的接口后，就可以避免路由器将数据往回发而产生路由环路了。
- 组播路由器上朝发送者方向的接口被称为RPF接口，只有从RPF接口发来的数据才能被转发，从其它接口过来的数据被认为是无效的。

3) 组播路由协议

21

PIM基本概念：**PIM-Dense Mode（密集模式）**和**PIM-Sparse Mode（稀疏模式）**

- 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）
- **PIM-DM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邻居路由器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。
- 如果查询数据包到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。



3) 组播路由协议

22

PIM基本概念：PIM-DM（密集模式）

- 如果网络中有一个组，地址为：224.1.1.1，发送者为100.1.1.1，那么就应该记录路由为（100.1.1.1，224.1.1.1），还要记录的就是组播的出口，这些信息构成组播路由表。
- 路由器收到的224.1.1.1组播就会指定出口发出去。PIM-DM模式中这样记录组播路由的方式被称为(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)的路由记录方式，会因为组源地址的增加而增加记录条目。

3) 组播路由协议

23

PIM基本概念：PIM-DM（密集模式）

- 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。
- 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中记录个每个组播发送者的源地址，故又被称为源树。

3) 组播路由协议

PIM基本概念：PIM-SM（稀疏模式）

- 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。
- 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。

3) 组播路由协议

25

PIM基本概念：PIM-SM（稀疏模式）

- 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模式路由器记录为（*, 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为一个组播只记录一个路由条目，而不管这个组播有多少个发送者。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，使用密集模式（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。

3) 组播路由协议

26

PIM基本概念：PIM-SM（稀疏模式）

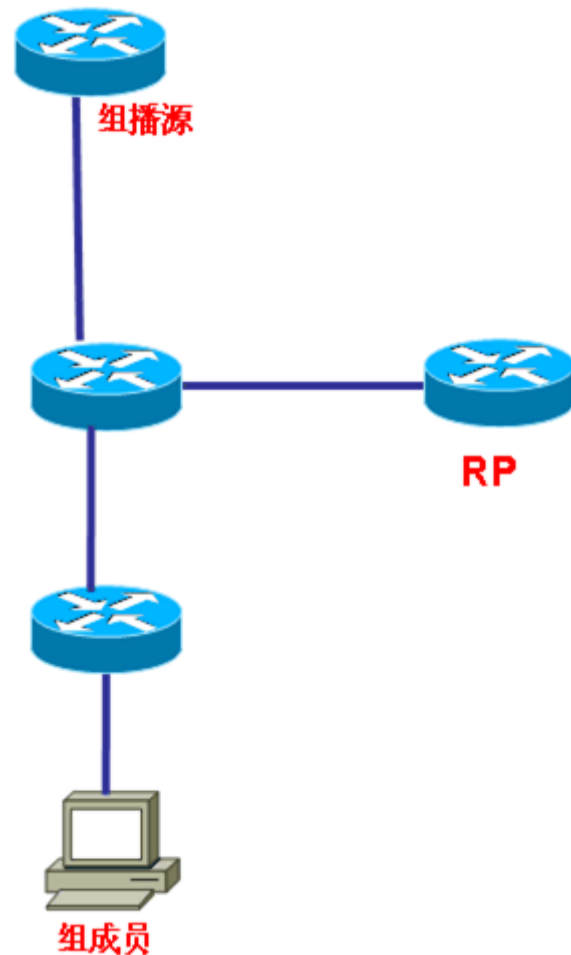
- PIM-SM模式路由器不仅需要记录(*, G)信息，也和PIM-DM模式一样需要记录在该组中，哪些接口是出口，从此接口将数据发给接收者。
- 但是与PIM-DM模式不同的是，PIM-SM模式只记录连接着接收者的接口，其它没有接收者，不需要接收组播的接口是不会被记录的。
- 比如路由器上有5个接口有PIM邻居，其中只有一个forwarding出口，再去掉1个RPF接口，还剩3个PIM接口是不需要接收组播的，在PIM-DM模式中，会记录下一个RPF口和一个forwarding状态的出口，以及三个不需要接收组播的pruning状态的接口
- 因此，PIM-SM模式的接口记录方式会比PIM-DM模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

3) 组播路由协议

27

PIM-SM RP (Rendezvous Point)

- 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。

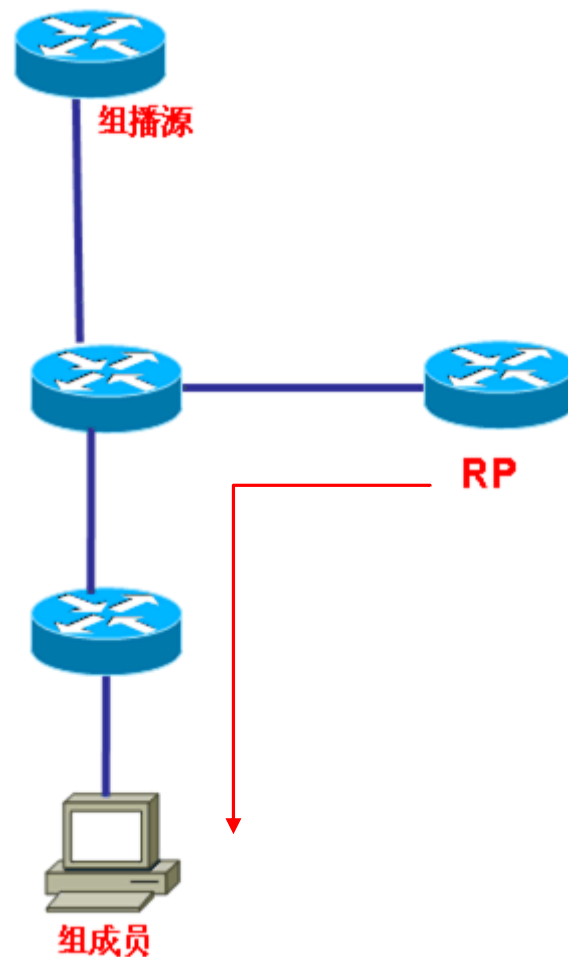


3) 组播路由协议

28

PIM-SM RP

- 路由器也就认为RP的地址，就是组播源的IP地址。
- 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
- PIM-SM中的组播树，即(*, G)形式的树被称为**共享树（RPT）**。



3) 组播路由协议

29

PIM DR（在PIM-DM和PIM-SM模式下都会选举）

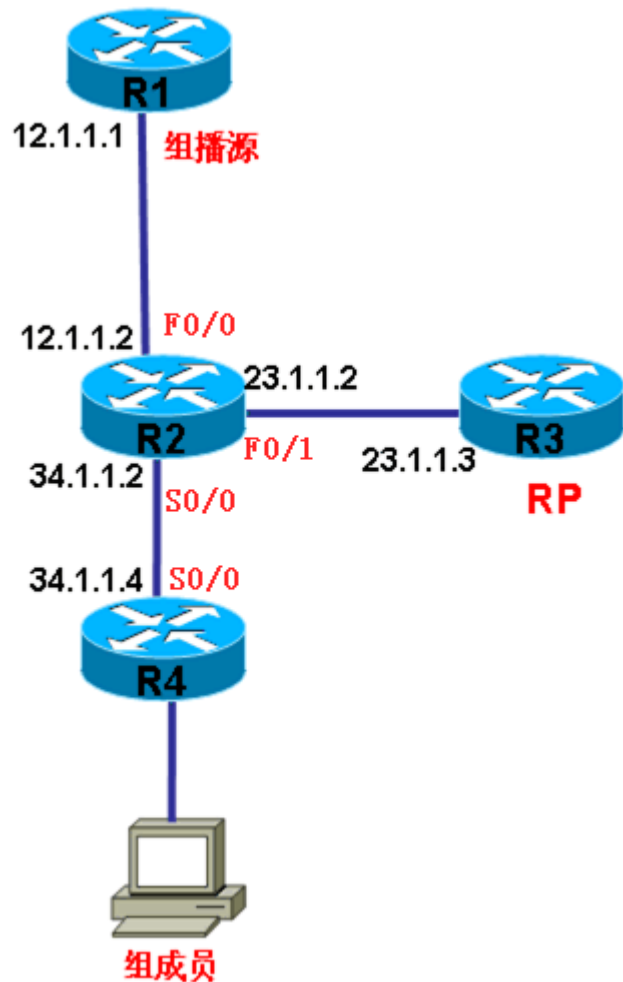
- 因为IGMP V1不选举IGMP查询器，如果有多台路由器连接同一个多路访问网段（以太网）时，就必须选举PIM Designated Router（DR）来避免重复查询
- 这个DR的选举，是由PIM协议进行的，选择IP地址高的为DR。PIM DR路由器在IGMP v1中充当IGMP查询器的功能，来向主机发送IGMP查询
- 选择DR的数据包为30秒一次，105秒保持时间，如果过了保持时间没有收到DR的数据包，则会重新选举IP地址最高的为DR。

3) 组播路由协议

30

PIM-SM DR

- 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。
- 例如图中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。

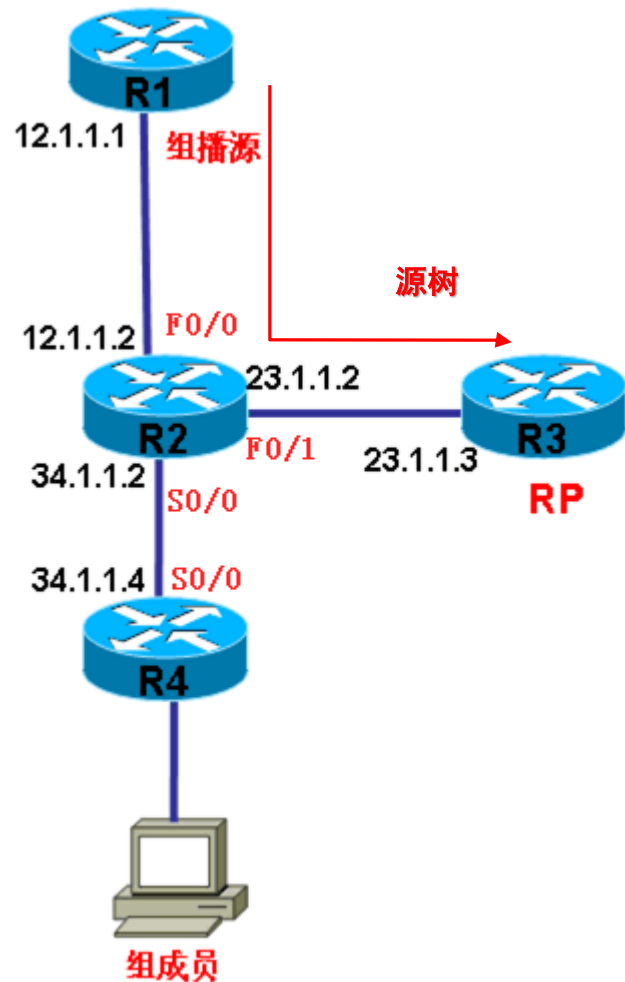


3) 组播路由协议

31

PIM-SM DR

- 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。

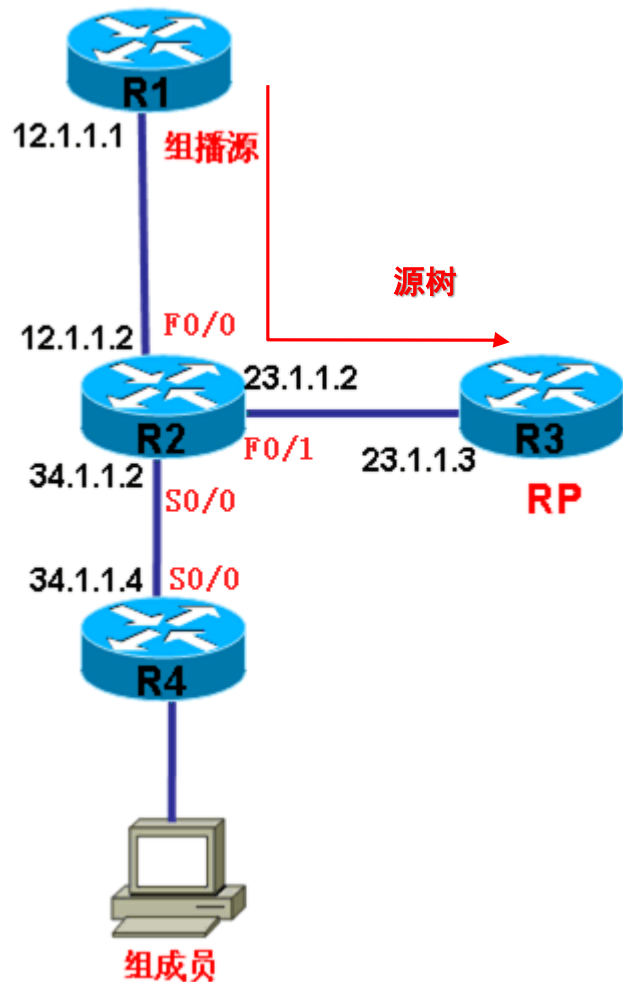


3) 组播路由协议

32

PIM-SM DR

- 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。
- 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。

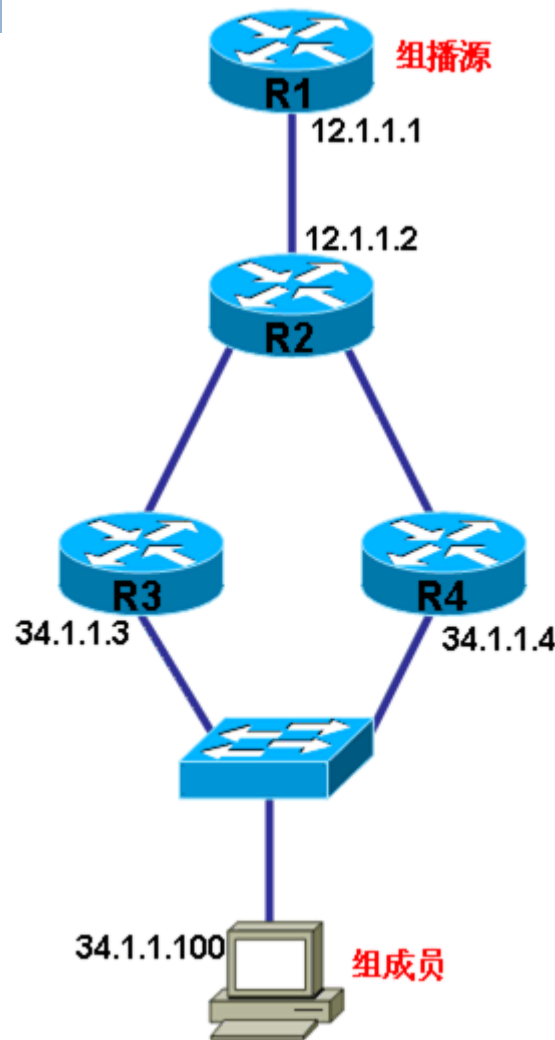


3) 组播路由协议

33

PIM前转器（在PIM-DM和PIM-SM模式下都会选举）

- 从图中可以看出，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。
- 假如R2选择将数据包以这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，数据包到达组成员就会出现明显的先后顺序，在视频直播会导致错误
- 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
- 两个下一跳中被R2选中的路由器，称为PIM前转器（Forwarder）。

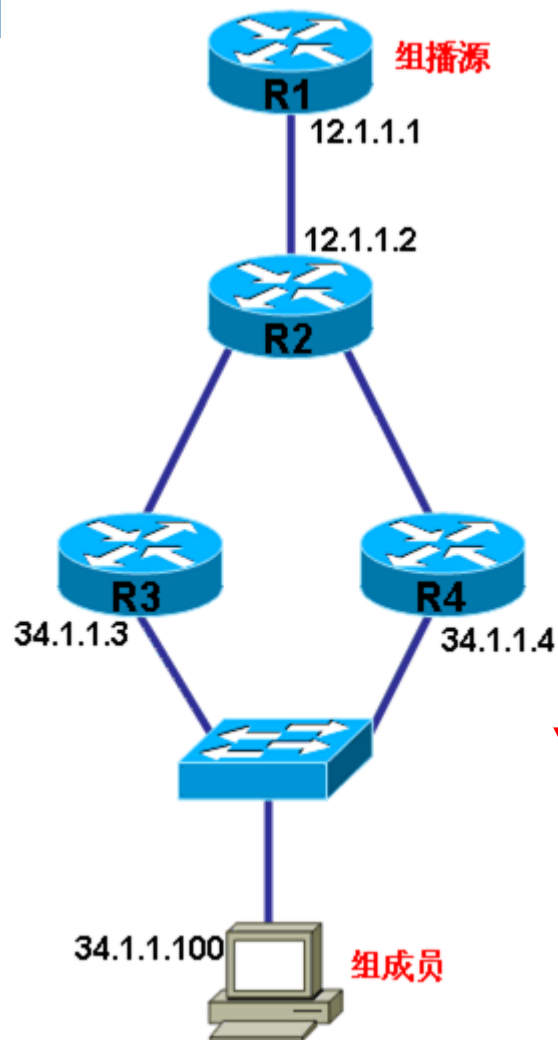


3) 组播路由协议

34

PIM-SM 前转器 (Forwarder)

- 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 (Administrative Distance, 不同路由协议该值不同) 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。



3) 组播路由协议

35

PIM-DM数据包

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

3) 组播路由协议 PIM-DM数据包

36

1. Hello

PIM路由器用来发现PIM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。

2. Join/prune

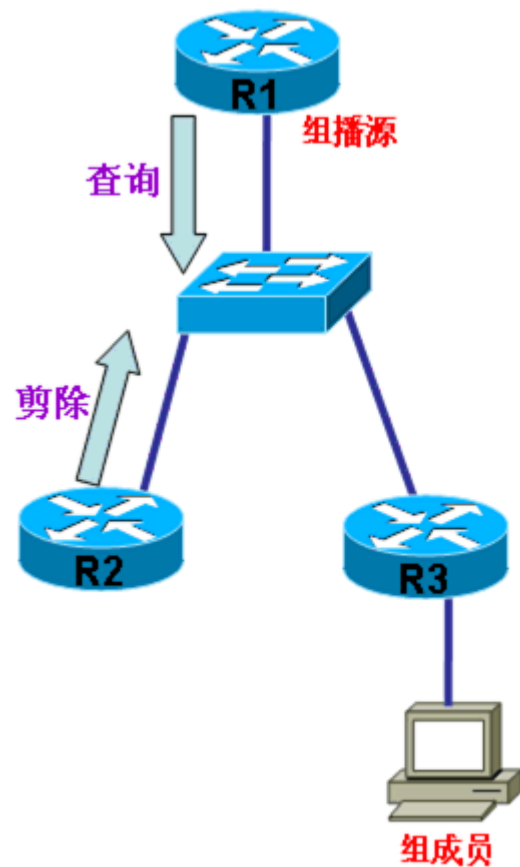
PIM-DM启动后，组播源开始向所有邻居发送查询消息，邻居会再向自己的邻居发送查询消息，如果有路由器连接着组成员，那么就会向上一跳邻居发送join数据包，用来通告自己需要接收组播，从而将自己接入组播树。其它没有连接组成员或不需要接收组播的路由器则向上一跳邻居发送prune数据包，用来通告自己不需要接收组播，状态被标为prune，而不是forward。

3) 组播路由协议 PIM-DM数据包

37

Prune 消息覆盖

- 图中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，在正常情况下，R1是应该将对应接口剪除
- 但是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
- 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。



3) 组播路由协议 PIM-DM数据包

38

3. Graft

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

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。

3) 组播路由协议 PIM-DM数据包

39

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（Administrative Distance）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。

3) 组播路由协议 PIM-SM数据包

40

PIM-SM数据包，PIM-SM采用了7种PIMv2消息：

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

3) 组播路由协议

41

1. Hello

和DM模式相同作用

2. Bootstrap

Bootstrap用于RP的选举。

3. Candidate-RP-Advertisement

Candidate-RP-Advertisement用于RP的选举。

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。

3) 组播路由协议

42

6. Register (组播源发给RP)

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 (RP发给组播源)

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。

3) 组播路由协议

43

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，有两种方法，分别是：

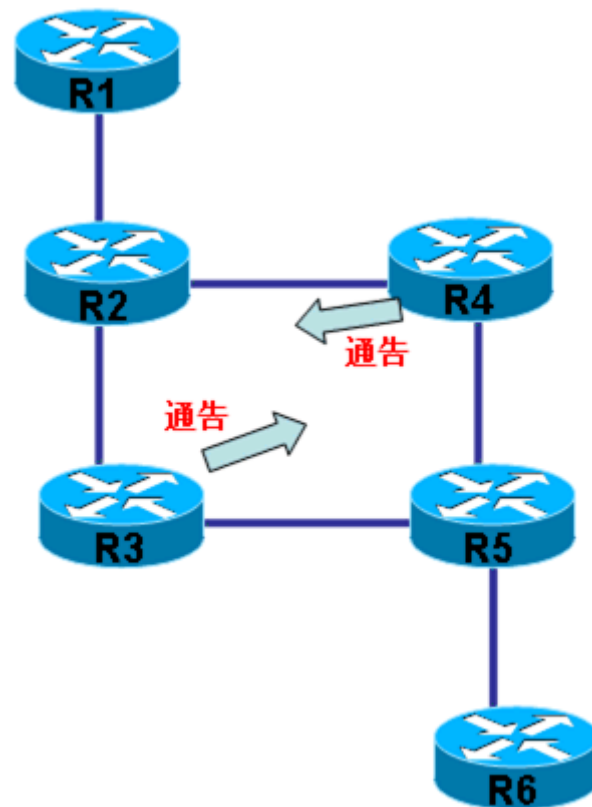
1. 手工静态配置
2. Bootstrap Router (BSR)来通告

3) 组播路由协议

44

2. Bootstrap Router (BSR)来通告

- 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

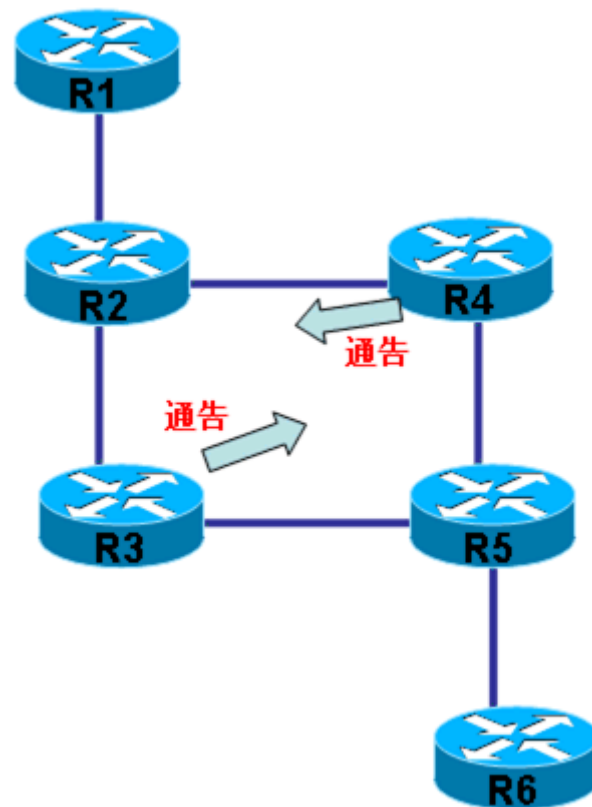


3) 组播路由协议

45

2. Bootstrap Router (BSR)

- 假设同时将R3和R4都配置为C-RP，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，而C-RP发出的竞选消息是Candidate-RP-Advertisement。
- BSR是通过路由器发送Bootstrap选出的。



4) 组播举例-DM

46

PIM Dense Mode例子：

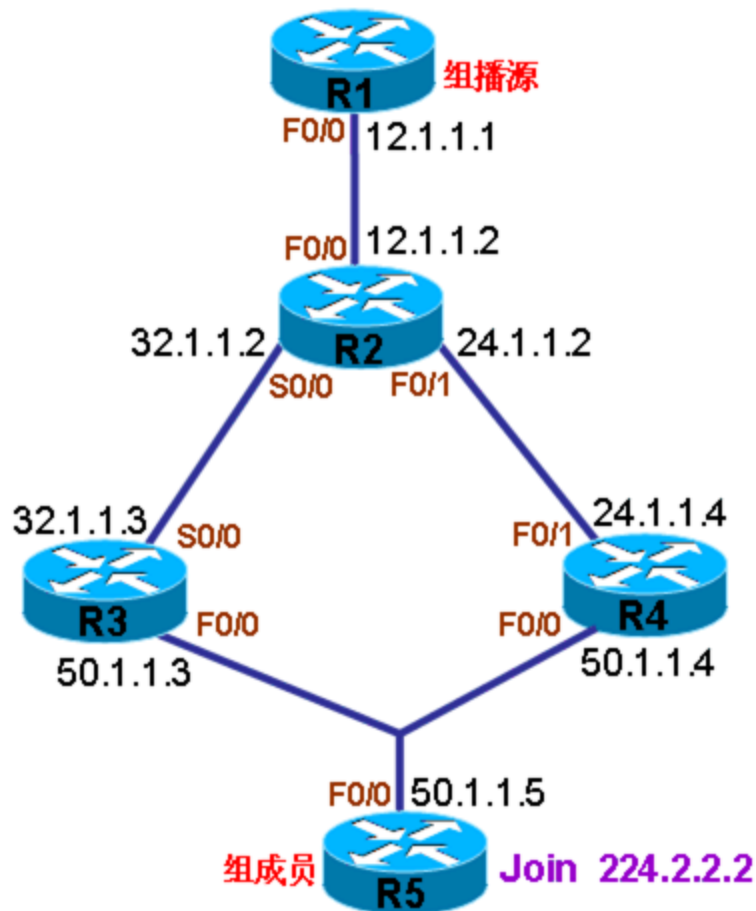
- 全网配置OSPF来完成单播通信
- 在每台路由器上开启组播路由功能：ip multicast-routing
- 在接口上开启PIM Dense-Mode：ip pim dense-mode

R1邻居R2，DR选择R2（IP大）

R5邻居R3，R4，DR为R5

同理：

R2邻居R3，R4，DR为R3，R4



4) 组播举例

47

PIM Dense Mode例子：

- 有主机在R5加入224.2.2.2
- 组播源为R1
- 从组播源ping 224.2.2.2
- 可以收到组成员回包，表示路由成功配置

在IGMPv2情况下，R3和R4、R5同网段，所以为了避免重复查询，R3成为查询器（IP最小）

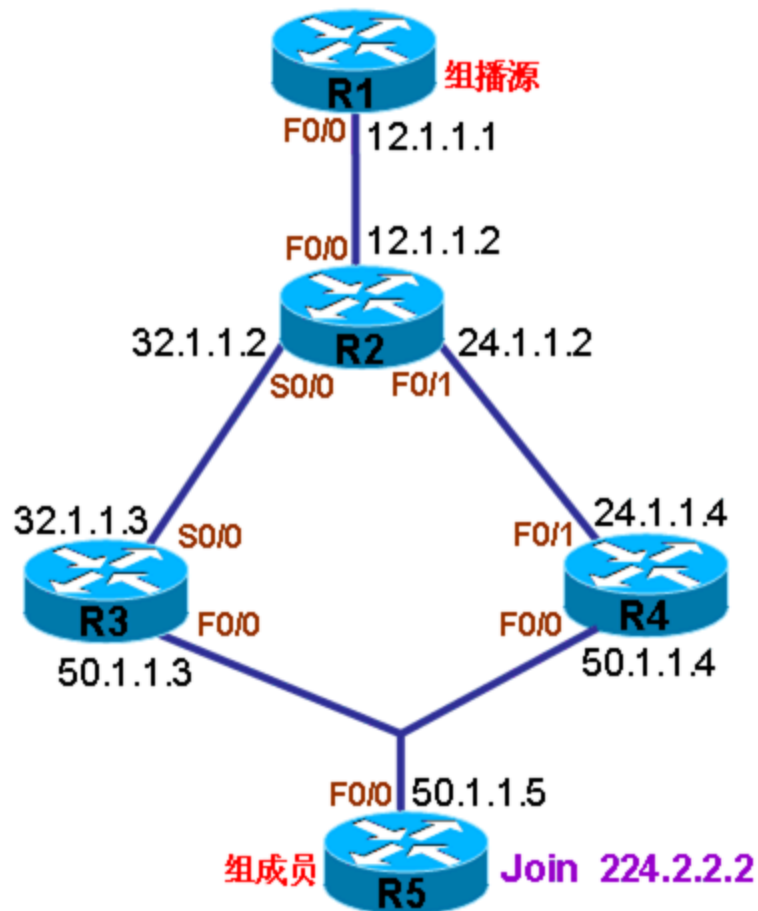
show ip mroute可以查看R2组路由表

(12.1.1.1, 224.2.2.2)这条路由中

Serial0/0, Prune/Dense

FastEthernet0/1, Forward/Dense

即，R3被prune了，R4被选为前转器



4) 组播举例

48

PIM Dense Mode例子:

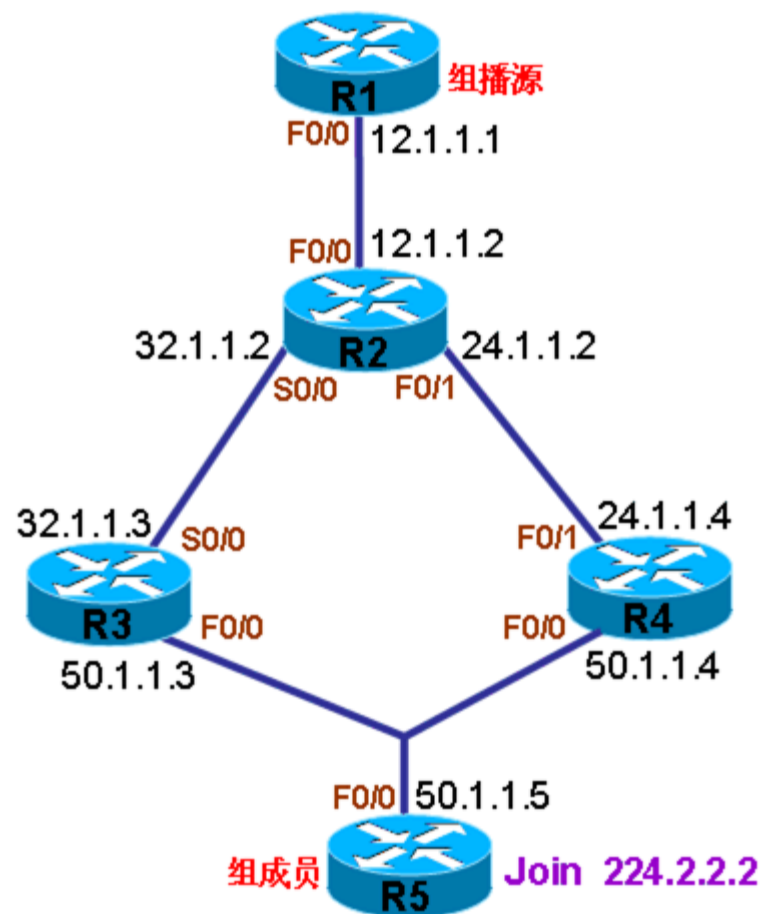
show ip mroute可以查看R4组播路由表:

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

Incoming interface: FastEthernet0/1, RPF nbr 24.1.1.2 (邻居来自组播源)

Outgoing interface list:

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



4) 组播举例-SM

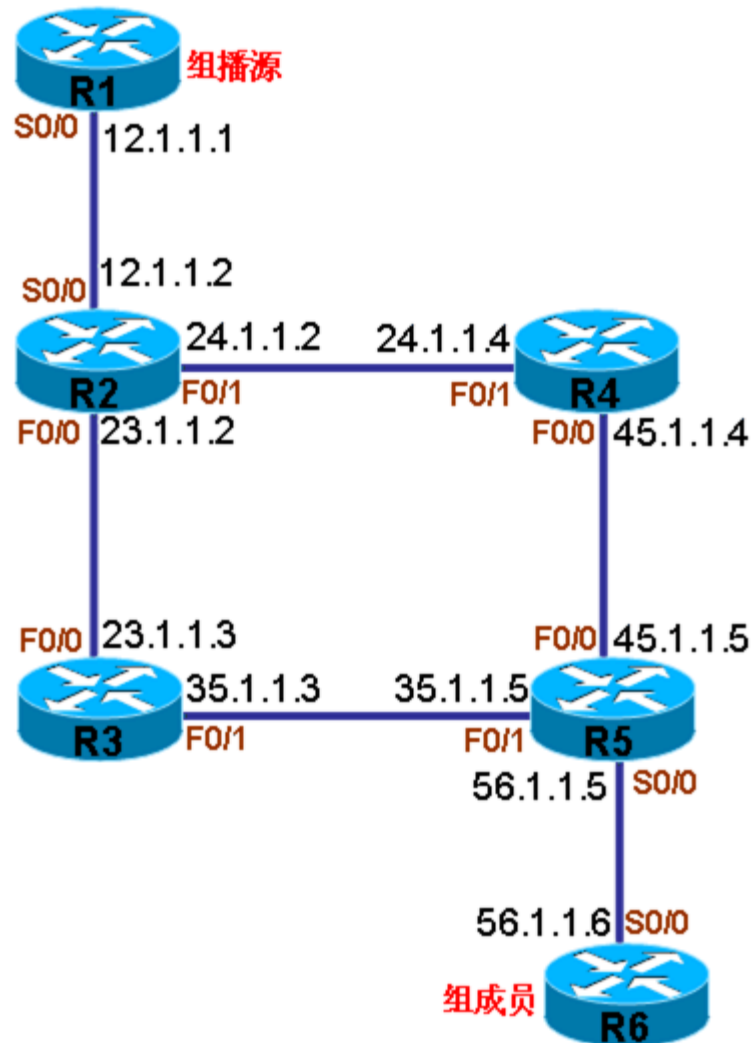
49

PIM Sparse Mode例子:

R6加入组播组224.6.6.6, 源为12.1.1.1

R4的loopback (虚接口) 地址4.4.4.4, 手工在每个路由器配置R4为RP

ip pim rp mapping可以在每个路由器上看到RP地址



4) 组播举例

50

PIM Sparse Mode例子:

同样，我们可以查看R4组播路由

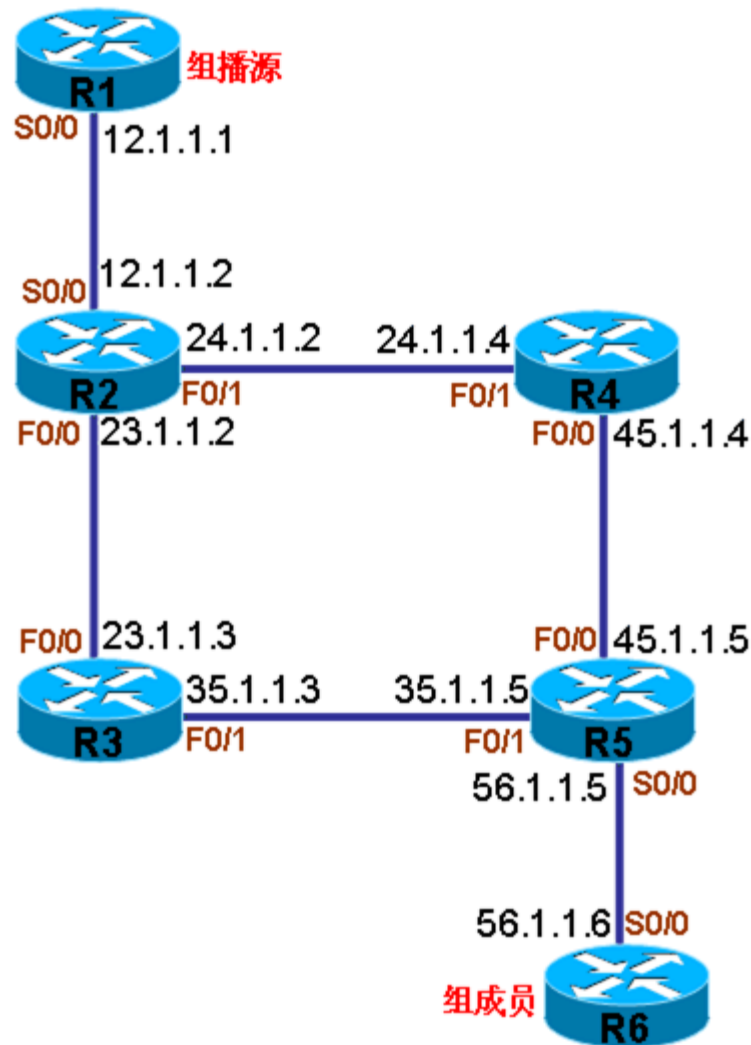
(*, 224.6.6.6), 00:02:16/00:03:07, RP 4.4.4.4,
flags: S

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

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

注意: R4直连RP, 所以Incoming interface为Null



4) 组播举例

51

PIM Sparse Mode例子:

同样, 我们可以查看R5组播路由

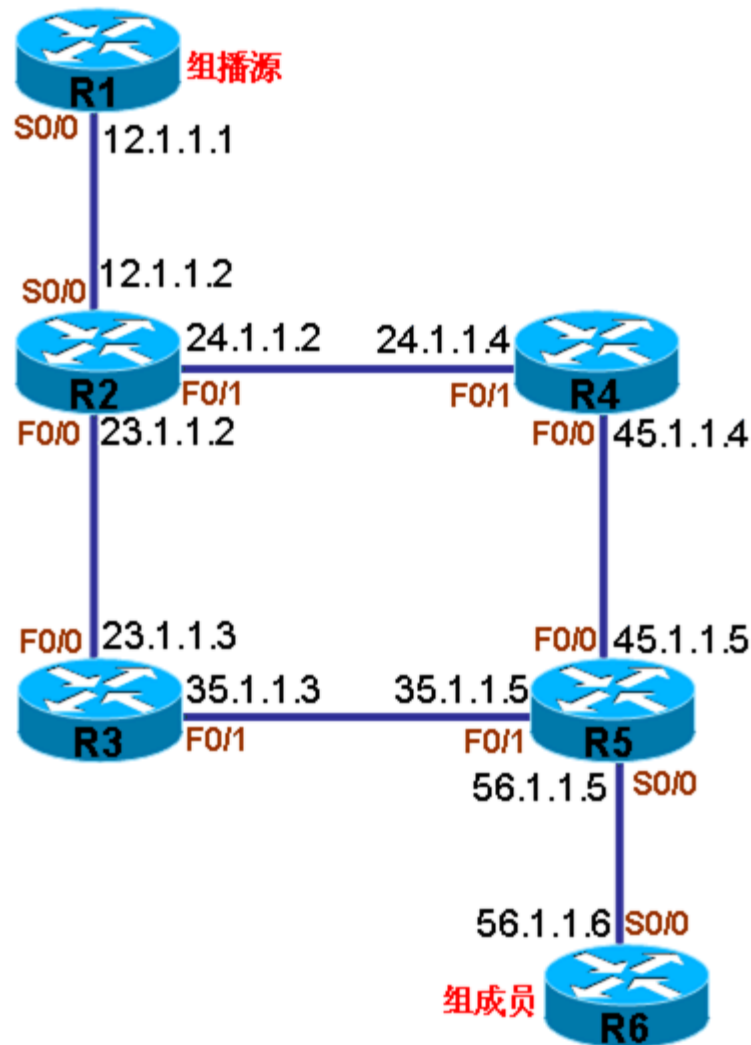
(*, 224.6.6.6), 00:05:36/00:02:45, RP 4.4.4.4,
flags: SJC (Sparse, Join SPT, Connected
)

Incoming interface: FastEthernet0/0, RPF
nbr 45.1.1.4

Outgoing interface list:

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

注意: 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。



4) 组播举例

52

PIM Sparse Mode例子:

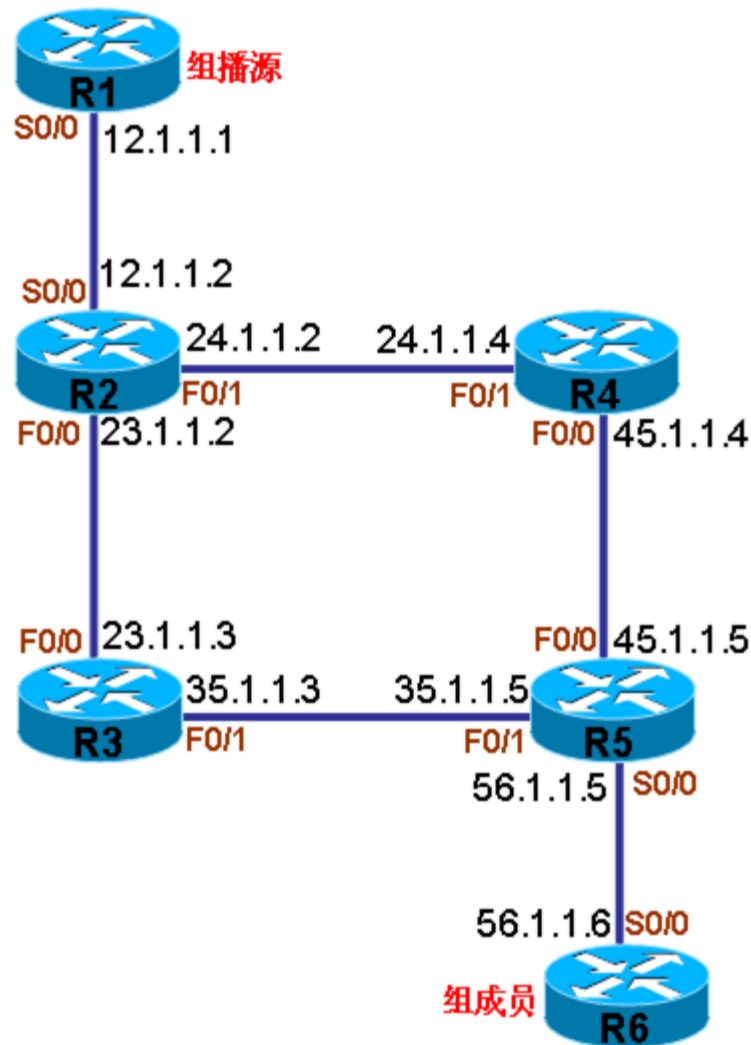
我们可以查看R2组播路由为 (S, G) 树

(12.1.1.1, 224.6.6.6), 00:02:05/00:03:24,
flags: FT (Register flag, SPT-bit set)

Incoming interface: Serial0/0, RPF nbr
0.0.0.0

Outgoing interface list:

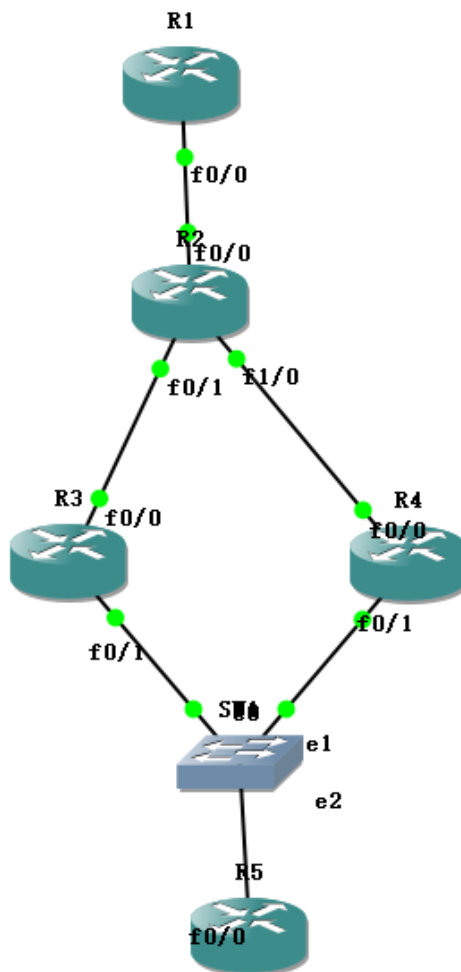
FastEthernet0/1, Forward/Sparse,
00:02:05/00:03:22



5) 如何在GNS3中配置组播

53

- 按照课件内容，建立下面路由器的拓扑结构



如何在GNS3中配置组播

54

- 利用ospf配置每个路由器的路由，例如R5路由器：

```
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 [110/20] via 50.1.1.3, 00:00:10, FastEthernet0/0
  24.0.0.0/24 is subnetted, 1 subnets
O       24.1.1.0 [110/20] via 50.1.1.4, 00:00:10, FastEthernet0/0
  12.0.0.0/24 is subnetted, 1 subnets
O       12.1.1.0 [110/30] via 50.1.1.4, 00:00:10, FastEthernet0/0
          [110/30] via 50.1.1.3, 00:00:10, FastEthernet0/0
```

如何在GNS3中配置组播

55

- 在每台路由器上开启组播路由功能

R1(config)#ip multicast-routing

- 在所有路由器的每个接口上开启 PIM:

R1(config)#int f0/0

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

- 查看PIM邻居

show ip pim neighbor

注意: mode的S表示State Refresh Capable

```
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           00:03:52/00:01:19 v2     1 / S
32.1.1.3           FastEthernet0/1           00:02:55/00:01:17 v2     1 / DR S
24.1.1.4           FastEthernet1/0           00:02:15/00:01:27 v2     1 / DR S
```

如何在GNS3中配置组播

56

- R5 加入组播（必须指定某个接口，在这个接口上加入组播组）

```
r5(config)#interface f0/0
```

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


如何在GNS3中配置组播

57

- 在R1路由器上ping 224.2.2.2组播，可以看到R5的回应

```
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, 56 ms
```

如何在GNS3中配置组播

58

- 再观察R2上的组播路由
 - 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。
 - 真正有效的是(12.1.1.1, 224.2.2.4)这条路由，可以清楚的看到去往R3的接口被Prune了

```
(* , 224.2.2.2), 00:00:06/stopped, RP 0.0.0.0, flags: D
Incoming interface: Null, RPF nbr 0.0.0.0
Outgoing interface list:
  FastEthernet1/0, Forward/Dense, 00:00:06/00:00:00
  FastEthernet0/1, Forward/Dense, 00:00:06/00:00:00
  FastEthernet0/0, Forward/Dense, 00:00:06/00:00:00

(12.1.1.1, 224.2.2.2), 00:00:06/00:02:53, flags: T
Incoming interface: FastEthernet0/0, RPF nbr 12.1.1.1
Outgoing interface list:
  FastEthernet0/1, Prune/Dense, 00:00:08/00:02:51
  FastEthernet1/0, Forward/Dense, 00:00:08/00:00:00
```

Multicasting-域间组播

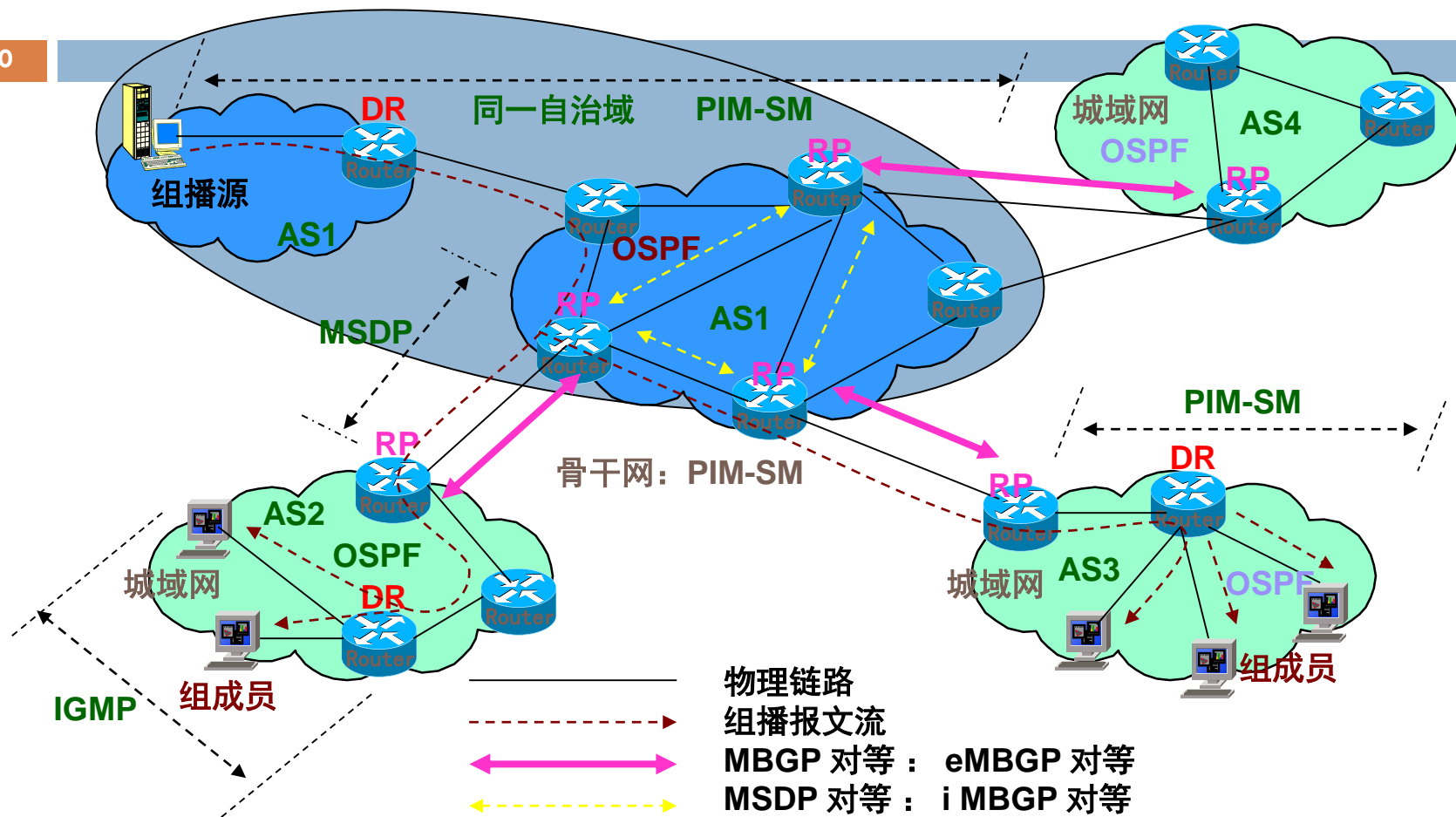
59

域间组播路由协议

- 域间组播目前仍然处于研究和试验阶段 目前比较成型的解决方案是下面三个协议的组合
 - ▣ MBGP 组播边界网关协议 用于在自治域之间交换组播路由信息
 - ▣ MSDP 组播信源发现协议 用于在 ISP 之间交换组播信源信息
 - ▣ PIM-SM 用作域内的组播路由协议

跨AS组播结构

60



IGMP: 完成路由器直连网段接收者发现
MSDP: 交换AS之间的源信息
PIM-SM: 完成AS内的源发现, 根据上述信息构建组播分发树, 达到域间组播转发

OSPF: 完成AS内单播拓扑发现
MBGP: 交换AS之间的拓扑信息

Multicasting-组播

61

◆ 相关缩略语

- IGMP:Internet Group Management Protocol,互联网组管协议
- MBGP:Multi-protocol Border Gateway Protocol, 多协议边界网关协议
- MSDP:Multicast Source Discovery Protocol,组播源发现协议
- **PIM-SM**:Protocol Independent Multicast-Sparse mode,协议独立组播—稀疏模式
- **PIM-DM**:Protocol Independent Multicast-Dense mode,协议独立组播—密集模式
- RP: Rendezvous Point,汇集点
- RPT: Rendezvous Point Tree , PIM-SM协议共享树
- SPT: Shortest Path Tree , 最短路径树
- DR: Designated Router, 指定路由器
- BSR:Boot Strap Router,自举路由器选举DR

组播的挑战、问题与发展

62

- 尽力传送（Best effort, UDP）
 - ▣ 会产生丢包
 - ▣ 不可能有很可靠的数据传输，应有针对性设计，可靠组播有待进一步研究
- 问题
 - ▣ 不能避免拥塞
 - 缺乏TCP滑动窗口，且“慢启动”导致拥塞，可尝试检测和避免机制
 - ▣ 复制：某些协议会导致偶尔生成重复的包
 - ▣ 无序发送：一些协议机制会导致无序发送

组播的挑战

63

- 没有更多的ISPs和OEM厂商开发有用的组播应用,缺乏组播工具和平台
- 防火墙阻隔
 - ▣ UDP能有效防止组成员ACKs的数量爆炸,但Firewall对其失去控制作用.解决办--应用网关?
- QoS:探讨用ATM,RSVP等?
- 基于Internet的组播实际上很少成功案例
- 应用层组播发展迅猛-P2P...IPTV...

组播的发展

64

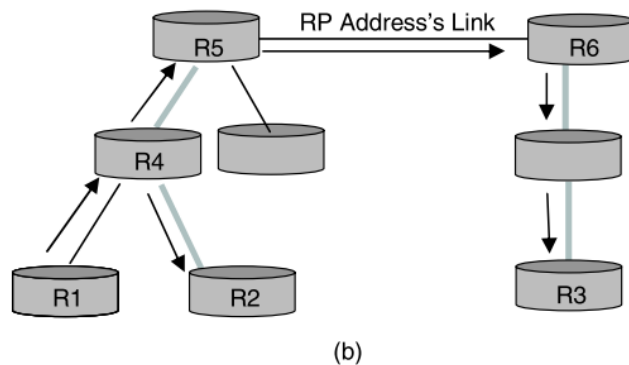
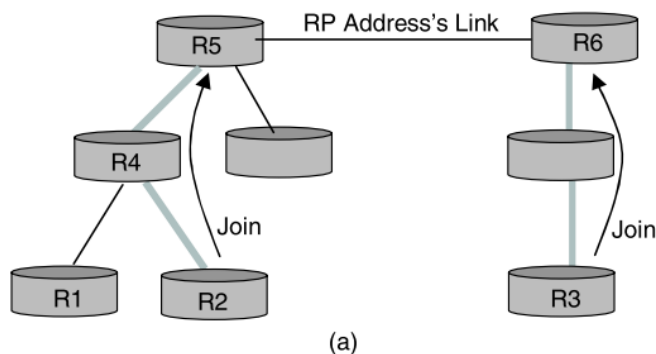
- 支持组播的主要高层应用协议:支持可靠数据传输
 - RTP:Real Time Transport Protocol
 - RTCP:Real Time Control Protocol
 - RTP:Real Time Streaming Protocol
 - RSVP:Resource Reservation Protocol
 - RMP: Reliable Multicast Protocol
 - RMF: Reliable Multicast Framework Protocol
 - RAMP: Reliable Adaptive Multicast Framework Protocol
 - Reliable Multicast Transport Protocol
 - Lucent 在其e-cast 用RMTP处理文件传输

组播的发展

65

教科书4.4，提到了PIM-SM扩展：

- Source-Specific Multicast (PIM-SSM) (one-many)
- Bidirectional Trees (BIDIR-PIM):



习题

66

◆ Ch2

25（序列号计算）； 40（以太网地址相同）；

◆ Ch3

14（交换机生成树算法）

◆ Ch4

4, 5（MTU）； 21, 22（路由）； 40, 46（子网）

◆ Ch5

12, 13（TCP序列号）； 39, 40（TCP协议）