本地组播流 — 任意源组播模型

目录

简介

背景信息

步骤1.当接收器处于活动状态时，它将发送IGMP报告消息

步骤2.当源处于活动状态时

步骤3.形成共享树

步骤4.(S，G)数据包到达FHR

步骤5.组播数据包的第一流，通过共享树到达接收方

步骤6. LHRR接收来自SPT的流量并向共享树发送修剪消息

简介

本文档介绍任意源组播(ASM)模型的数据包流。

背景信息

本文详细介绍了本地组播数据包流的详细数据包流及其输出分析，描述了控制平面和转发平面中的详细分析输出和数据包流。

ASM是接收方不了解发送方的模型。这意味着它可以接收来自任何来源的流量。接收方只知道发送方使用的组播组和Internet组管理协议(IGMP)，以便订阅接收发往该地址的所有流量。

本文档将介绍所有这些内容：

1.当接收器处于活动状态时会发生什么情况。

2.当源处于活动状态时会发生什么情况。

3.在交汇点(RP)收到注册信息后会发生什么情况。

4.(S，G)如何形成。直到第一跳路由器(FHR)。

5.第一个组播流采用哪条路径。

6.当两个流在最后一跳路由器(LHR)上接收时会发生什么情况。

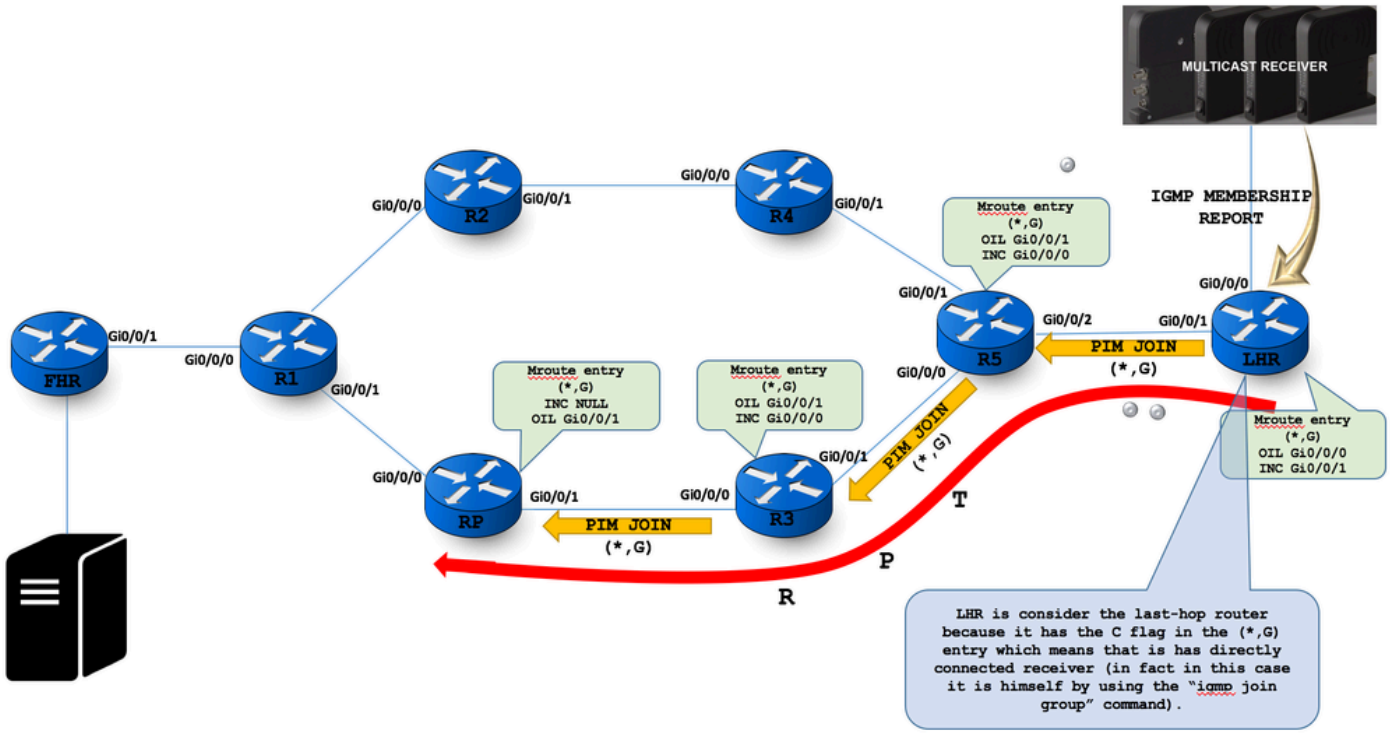
7.如何在共享树上形成最短路径树(SPT)。到底发生了什么以及发生切换的原因。

协议无关组播(PIM)用作源和接收器之间的组播路由协议，以创建组播树。在ASM中，使用(*,G)组播条目，其中*表示任意源，G是有兴趣接收流量的组播组地址接收器。

步骤1.当接收器处于活动状态时，它将发送IGMP报告消息

- 当收到接收方感兴趣的表述时，指定路由器(DR)随后会向该组播组的RP发送PIM加入消息。
- 此加入消息称为(*,G)加入，因为它将所有源的组G加入该组。
- (*,G)加入逐跳传输到组的RP，在它通过的每个路由器中，组G的组播树状态被实例化。

LHR被视为最后一跳路由器，因为它在(*,G)条目中带有C标志，这意味着它有直接连接的接收器(在本例中，使用igmp join group命令是它自己)。



Step 1 : On receiving the receiver's expression of interest, the DR then sends a PIM Join message towards the RP for that multicast group.
This Join message is known as a (*,G) Join because it joins group G for all sources to that group.
The (*,G) Join travels hop-by-hop towards the RP for the group, and in each router it passes through, multicast tree state for group G is instantiated.

```
LHR#sh ip igmp groups
IGMP Connected Group Membership
Group Address    Interface                Uptime    Expires   Last Reporter   Group Accounted
224.1.1.1        GigabitEthernet1/0       00:37:30  00:02:02  10.0.108.8
224.0.1.40       FastEthernet0/0          01:21:01  00:02:43  10.0.78.8


LHR#sh ip mroute
(*, 224.1.1.1), 00:00:29/00:02:30, RP 4.4.4.4, flags: SJCL
  Incoming interface: GigabitEthernet0/0/1, RPF nbr 10.0.78.7
  Outgoing interface list:
    GigabitEthernet0/0/0, Forward/Sparse

C flag in the (*,G) entry which means that is has directly connected receiver.


RP #sh ip mroute
(*, 224.1.1.1), 00:10:39/00:02:39, RP 4.4.4.4, flags: S
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/0, Forward/Sparse

S Flag Sparse mode created.


(*, 224.0.1.40), 01:56:40/00:02:58, RP 4.4.4.4, flags: SJPCL
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.78.7
  Outgoing interface list: Null(*, 224.0.1.40), 01:56:40/00:02:58, RP 4.4.4.4, flags: SJPCL
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.78.7
  Outgoing interface list: Null

There is a single (*,G) entry for the group 224.0.1.40 which is Auto-RP Discovery group address.
```

The value of "0.0.0.0" means self, and it appears in the output if the router is the RP itself

NOTE : To prevent a stale PIM-SM forwarding state from getting stuck in the routers, it is given a finite lifetime (3 minutes), after which it is deleted.
Routers refresh shared trees by periodically (once a minute) sending (*, G) Joins to the upstream neighbor in the direction of the RP.

Actually the PIM register message encapsulates the multicast packet sent by the source into a unicast packet.

```
▶ Frame 59: 68 bytes on wire (544 bits), 68 bytes captured (544 bits) on interface 0
▶ Ethernet II, Src: ca:08:fa:92:00:00 (ca:08:fa:92:00:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
▼ Internet Protocol Version 4, Src: 10.0.78.8, Dst: 224.0.0.13
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes
  ▶ Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
    Total Length: 54
    Identification: 0x0b27 (2855)
  ▶ Flags: 0x00
    Fragment offset: 0
    Time to live: 1
    Protocol: PIM (103)
  ▶ Header checksum: 0x7565 [validation disabled]
    Source: 10.0.78.8
    Destination: 224.0.0.13
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
▼ Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0011 = Type: Join/Prune (3)
    Reserved byte(s): 00
    Checksum: 0x87c7 [correct]
  ▼ PIM Options
      Upstream-neighbor: 10.0.78.7
      Reserved byte(s): 00
      Num Groups: 1
      Holdtime: 210
    ▼ Group 0: 224.10.10.10/32
      ▶ Num Joins: 1
        Num Prunes: 0
```

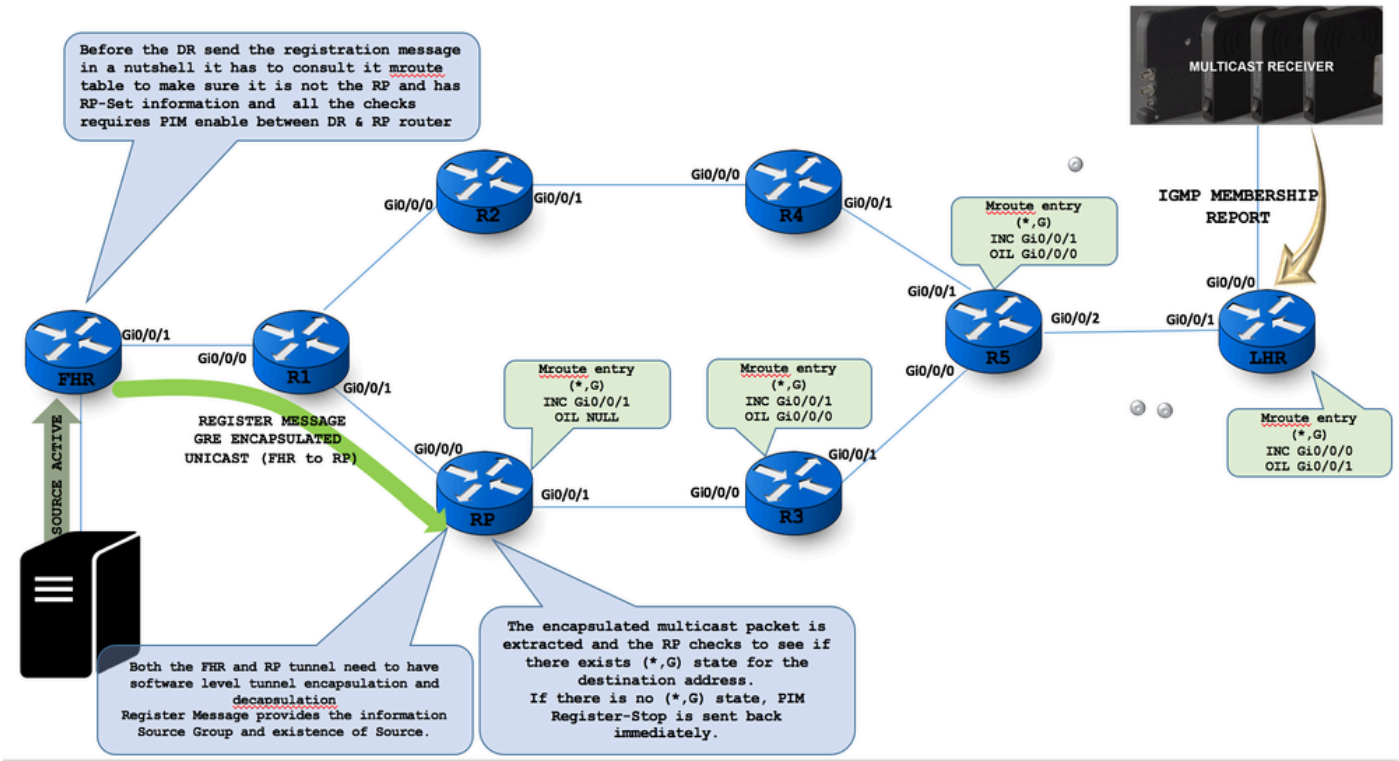
TTL is always 1. Which means it's a RP/RE destined packet.

PIM JOIN Message carries the active group address

步骤2.当源处于活动状态时

- 在DR发送注册消息之前，简而言之，它必须查阅mroute表以确保它不是RP并具有RP-Set信息，并且所有检查都要求在DR和RP路由器之间启用PIM。
- FHR和RP隧道都需要软件级隧道封装和解封。
- Register Message提供源组信息和源是否存在。
- 封装的组播数据包被提取后，RP将检查目标地址是否存在(*,G)状态。

- 如果没有(*,G)状态，PIM Register-Stop立即发送回。



Once Source is active :

```
FHR #
(1.1.1.1, 224.22.22.44), 00:03:15/00:00:02, flags: PFT
  Incoming interface: Loopback0, RPF nbr 0.0.0.0, Registering
  Outgoing interface list: Null
```

Register flag (F) is enabled for registration process in the FHR.

```
F flag: Source is directly connected and the register process must be used to notify the RP to this source.
P flag: Outgoing interface is null as no one has joined the SPT tree yet for this source
T flag: traffic is being received from the source.
```

PIM must enable between DR & RP router to send and receive the Register message.

```
▶ Frame 442: 142 bytes on wire (1136 bits), 142 bytes captured (1136 bits) on interface 0
▶ Ethernet II, Src: ca:01:c1:46:00:1c (ca:01:c1:46:00:1c), Dst: ca:02:c1:6a:00:00 (ca:02:c1:6a:00:00)
▶ Internet Protocol Version 4, Src: 10.0.12.1, Dst: 4.4.4.4
▼ Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0001 = Type: Register (1)
    Reserved byte(s): 00
    Checksum: 0xdeff [correct]
  ▼ PIM Options
    ▶ Flags: 0x00000000
      0100 .... = IP Version: IPv4 (4)
▶ Internet Protocol Version 4, Src: 1.1.1.1, Dst: 224.10.10.10
▶ Internet Control Message Protocol
```

If no active receiver present at RP, then RP sends REGISTER STOP DR will be silent for default 60 seconds may result in the so-called "join latency" where a newly Joined listener may have to way for almost a minute before it can discover a multicast source. This is why in many practical deployments with dynamic listeners you see PIM SSM being used in favor of complicated PIM SM mechanics.

1.1.1.1	224.22.22.44	PIMv2	142	Register
4.4.4.4	10.0.91.1	PIMv2	52	Register-stop

```
RP #
(1.1.1.1, 224.22.22.44), 00:00:43/00:02:16, flags: P
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.24.2
  Outgoing interface list: Null

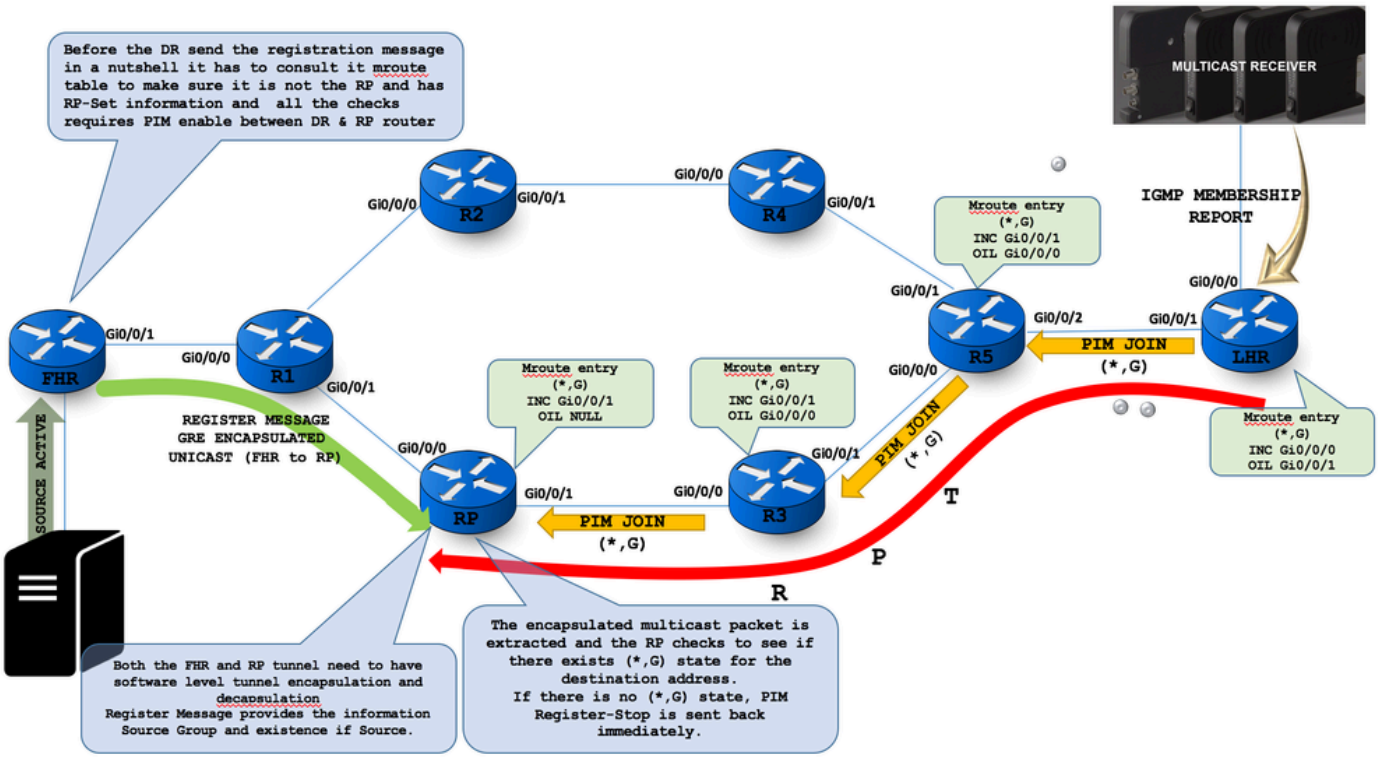
Prune Flag (P) is set as no active receiver (*,G) entry present in RP.
```

RP SENDS REGISTER STOP WHEN NO ACTIVE RECIEVER FOR THE GROUP AND DISCARD THE MULTICAST PACKET

```
▶ Frame 973: 52 bytes on wire (416 bits), 52 bytes captured (416 bits) on interface 0
▶ Ethernet II, Src: ca:02:c1:6a:00:00 (ca:02:c1:6a:00:00), Dst: ca:01:c1:46:00:1c (ca:01:c1:46:00:1c)
▶ Internet Protocol Version 4, Src: 4.4.4.4, Dst: 10.0.91.1
▼ Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0010 = Type: Register-stop (2)
    Reserved byte(s): 00
    Checksum: 0xe39a [correct]
  ▼ PIM Options
      Group: 224.22.22.44/32
      Source: 1.1.1.1
```

步骤3.形成共享树

- 在DR发送注册消息之前，简而言之，它必须查阅mroute表以确保它不是RP并具有RP-Set信息，并且所有检查都要求在DR和RP路由器之间启用PIM
- FHR和RP隧道都需要软件级隧道封装和解封
- Register Message提供源组和existence if Source的信息。
- 封装的组播数据包被提取后，RP将检查目标地址是否存在(*,G)状态。
- 如果没有(*,G)状态，PIM Register-Stop立即发送回。



The RP also sees that an active shared tree with a nonempty outgoing interface list exists and therefore sends the de-encapsulated packet down the shared tree.

```
RP #
(*, 224.1.1.1), 02:45:12/00:03:11, RP 4.4.4.4, flags: S
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/0, Forward/Sparse, 02:45:12/00:03:11

(10.0.12.1, 224.1.1.1), 00:02:42/00:00:21, flags: T
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.24.2
  Outgoing interface list: Null

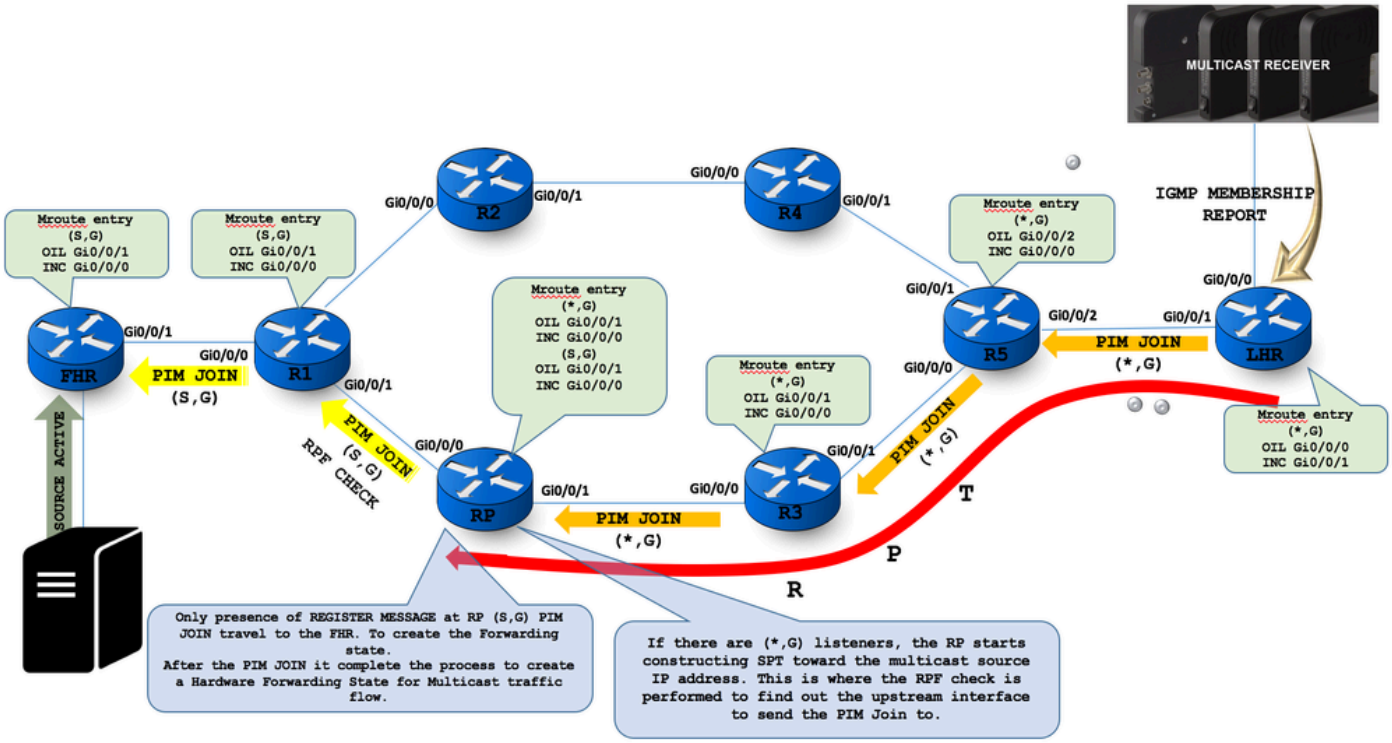
T Flag set for the shared tree.
```

Presence of (*,G) at RP means active receiver.

```
▶ Frame 29: 76 bytes on wire (608 bits), 76 bytes captured (608 bits) on interface 0
▶ Ethernet II, Src: ca:04:f1:9c:00:00 (ca:04:f1:9c:00:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
▶ Internet Protocol Version 4, Src: 10.0.24.4, Dst: 224.0.0.13
▼ Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0011 = Type: Join/Prune (3)
    Reserved byte(s): 00
    Checksum: 0xb4c2 [correct]
  ▼ PIM Options
      Upstream-neighbor: 10.0.24.2
      Reserved byte(s): 00
      Num Groups: 1
      Holdtime: 210
    ▼ Group 0: 224.1.1.1/32
      ▼ Num Joins: 2
          IP address: 1.1.1.1/32 (S)
          IP address: 10.0.12.1/32 (S)
        Num Prunes: 0
```

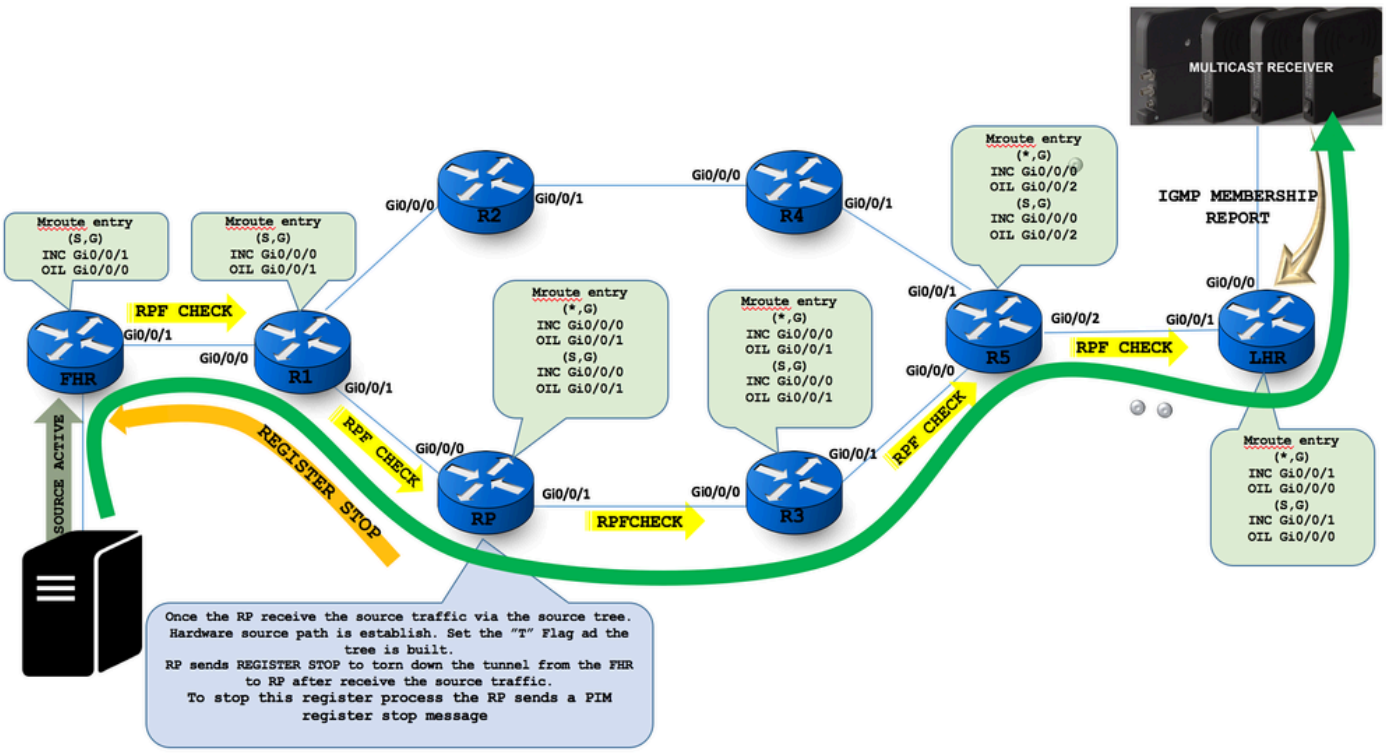
步骤4.(S，G)数据包到达FHR

- 只有在RP(S，G)PIM处存在注册消息才会传送到FHR。创建转发状态。
- 在PIM加入后，它完成为组播流量创建硬件转发状态的过程。
- 如果存在(*,G)侦听程序，RP将开始构建面向组播源IP地址的SPT。在此执行RPF检查以找出要向其发送PIM加入的上游接口。



步骤5.组播数据包的第一流，通过共享树到达接收方

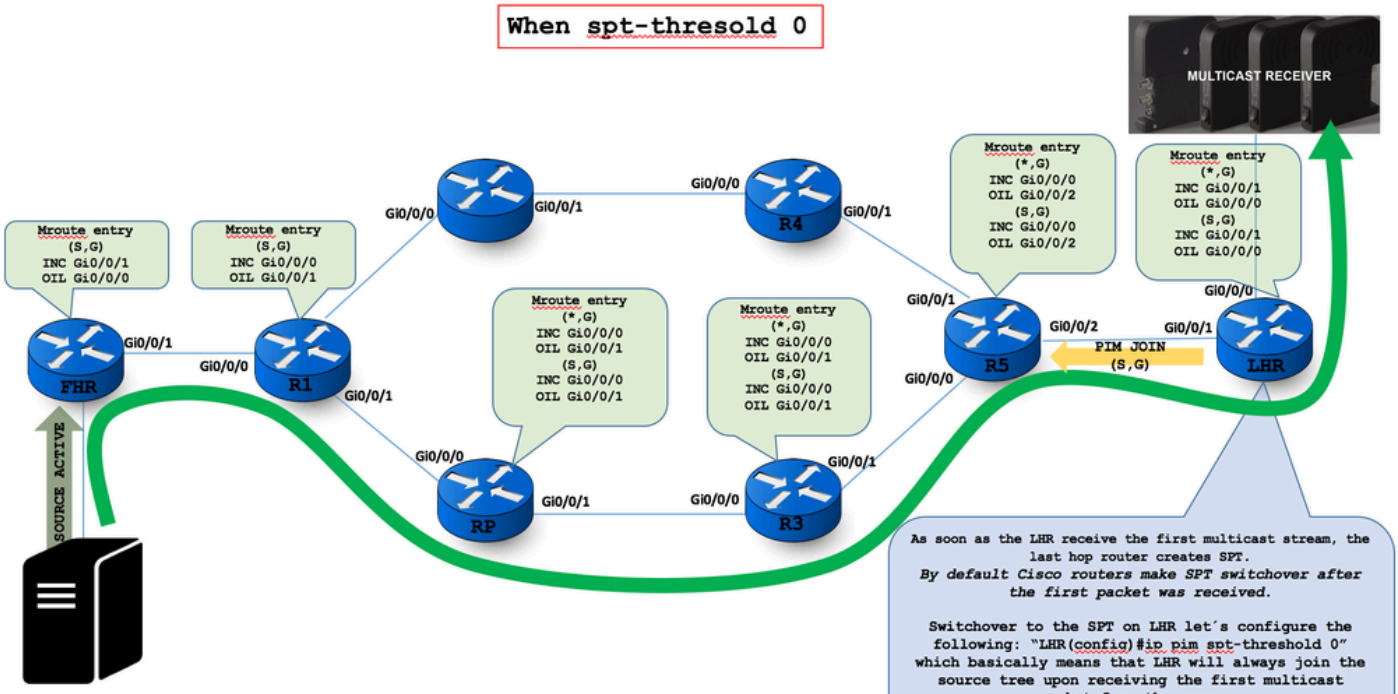
- 仅当在INC /RFP接口上收到组播流时，路由器才会转发组播流。
- 组播数据包源地址根据单播RT进行检查。
- 在发送加入的源方向确定接口和下一跳组播路由器。
- RP正在加入S的源特定树，数据包将继续封装到RP。当来自S的数据包也开始以本地方式到达RP时，RP将接收每个数据包的两个副本。
- 此时，RP开始丢弃这些数据包的封装副本，并向S的DR发送REGISTER STOP消息以防止DR不必要地封装数据包。



PIM-SM enables a last hop DR (that is, a DR with directly connected hosts that have joined a multicast group) to switch from the shared tree to the SPT for a specific source. This step is usually accomplished by specifying an SPT-Threshold in terms of bandwidth. If this threshold is exceeded, the last-hop DR joins the SPT. (Cisco routers have this threshold set to zero by default, which means that the SPT is joined as soon the first multicast packet from a source has been received via the shared tree.)

- 一旦RP通过源树接收源流量。建立硬件源路径。设置“T”标志，生成树。
- RP在收到源流量后发送REGISTER STOP以断开从FHR到RP的隧道。
- 要停止此注册过程，RP会发送PIM注册停止消息

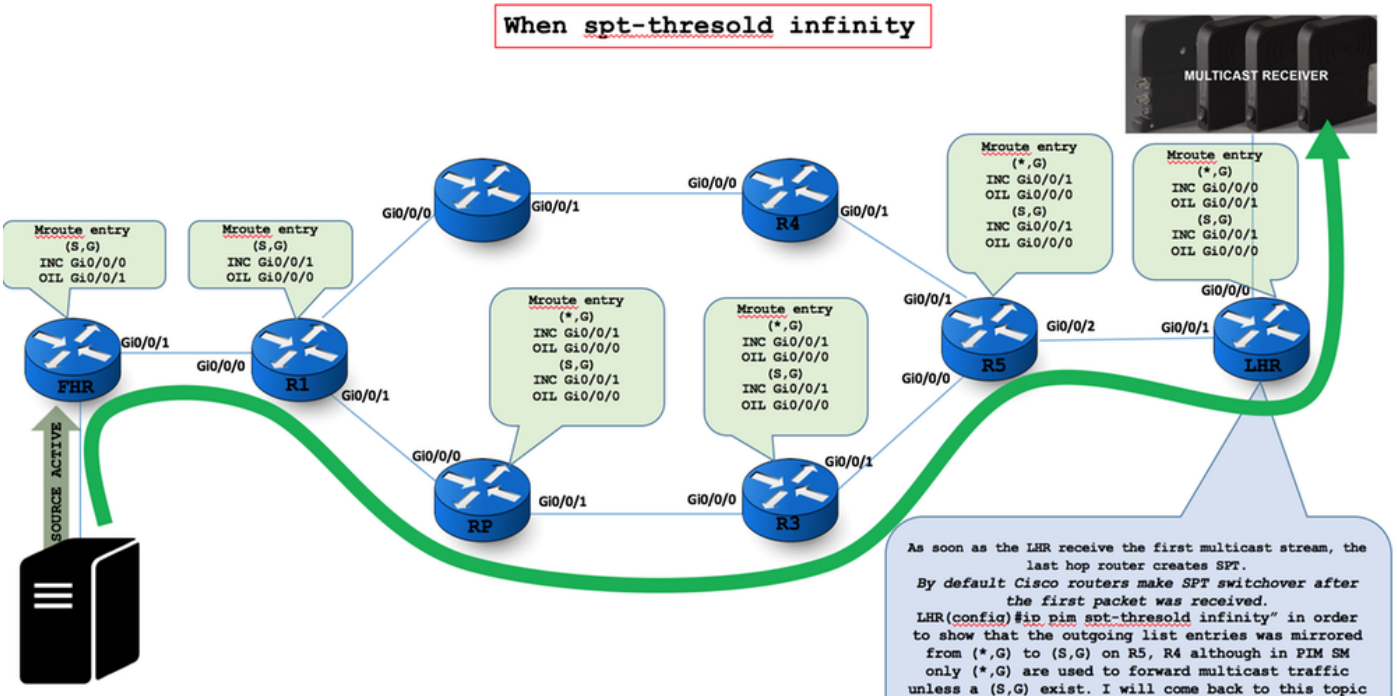
When spt-thresold 0



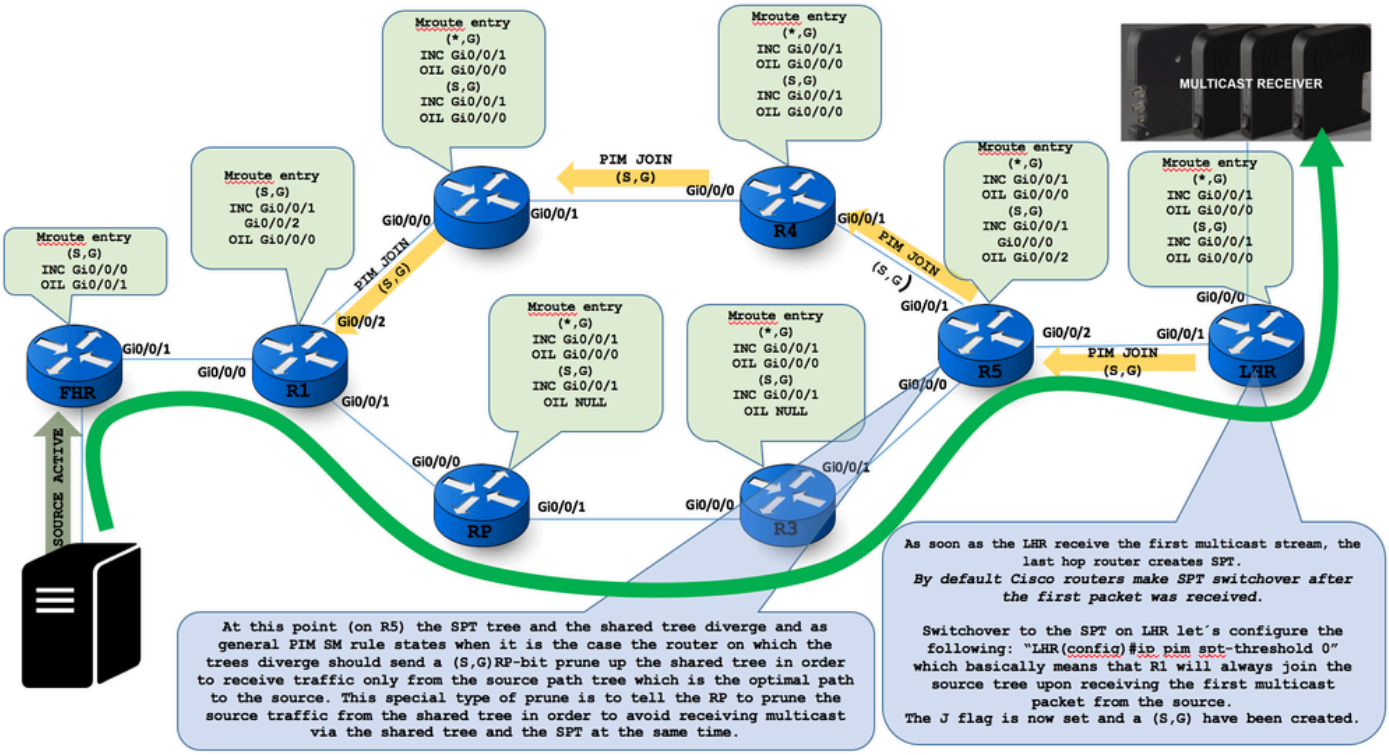
As soon as the LHR receive the first multicast stream, the last hop router creates SPT.
By default Cisco routers make SPT switchover after the first packet was received.

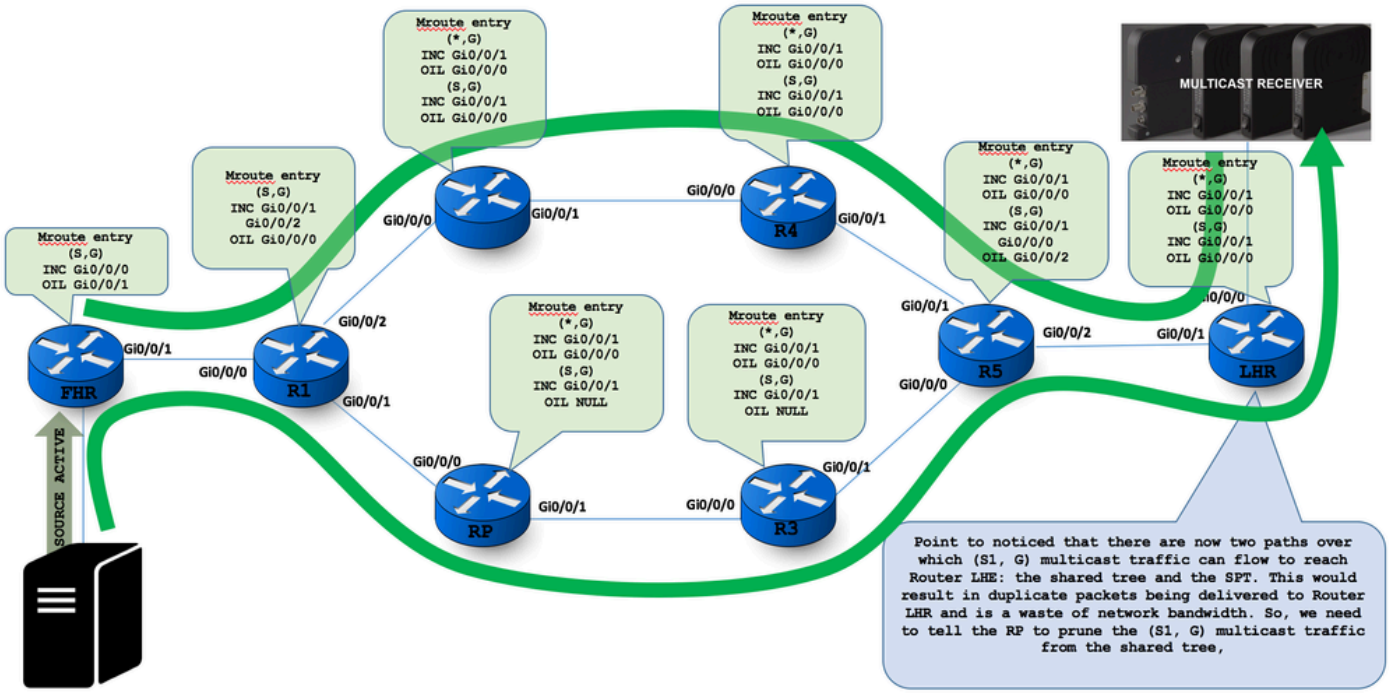
Switchover to the SPT on LHR let´s configure the following: "LHR(config)#ip pim spt-threshold 0" which basically means that LHR will always join the source tree upon receiving the first multicast packet from the source.
The J flag is now set and a (S,G) have been created.

When spt-thresold infinity



As soon as the LHR receive the first multicast stream, the last hop router creates SPT.
By default Cisco routers make SPT switchover after the first packet was received.
LHR(config)#ip pim spt-thresold infinity" in order to show that the outgoing list entries was mirrored from (*,G) to (S,G) on R5, R4 although in PIM SM only (*,G) are used to forward multicast traffic unless a (S,G) exist. I will come back to this topic later on.





步骤6. LHR从SPT接收流量并向共享树发送修剪消息

在收到两个组播流量流后，LHR开始接收来自SPT的流量并向共享树发送修剪消息。

J标志表示相应的(*,G)状态是通过枝叶路由器切换SPT。

LHR #

(10.0.12.1、239.1.1.1),00:00:38/00:02:21，标志：LJT

传入接口:FastEthernet0/0、RPF nbr 10.0.78.7

传出接口列表：

GigabitEthernet1/0、转发/稀疏，00:00:38/00:02:21

“F”标志通常用于在PIM DR路由器上创建的状态 — 它指示与向RP注册的流对应的转发状态。如果“F”标志仍然存在，则您的路由器很可能无法从RP接收PIM注册 — 停止消息，因此有些源尚未切换到SPT。

```
The J flag means the respective (*,G) state is to be switched the SPT by the leaf router.

LHR #
(10.0.12.1, 239.1.1.1), 00:00:38/00:02:21, flags: LJT
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.78.7
  Outgoing interface list:
    GigabitEthernet1/0, Forward/Sparse, 00:00:38/00:02:21

The "F" flag is typically found for the states created at the PIM DR router - it signalizes the forwarding states that correspond
to the flows being registered with the RP. If the "F" flag persists, then your router is most likely not receiving the PIM
Register-Stop messages back from the RP, and thus there are sources that has not switched to the SPT tree.

FHR #

(*, 239.1.1.1), 00:09:01/stopped, RP 4.4.4.4, flags: SPF
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list: Null

(1.1.1.1, 239.1.1.1), 00:03:02/00:00:15, flags: PFT
  Incoming interface: Loopback0, RPF nbr 0.0.0.0, Registering
  Outgoing interface list: Null

There is an (S,G) entry in this table, which has the flag "T" meaning it's a shortest-path and not a shared tree construct. The
incoming interface is set to Loopback0 and RPF neighbor to "0.0.0.0" which means the local router is the traffic source.
```

The receiver (or a router upstream of the receiver) will be receiving two copies of the data: one from the SPT and one from the RPT. When the first traffic starts to arrive from the SPT, the DR or upstream router starts to drop the packets for G from S that arrive via the RP tree. In addition, it sends an (S,G) Prune message towards the RP.
This is known as an (S,G,rpt) Prune. The Prune message travels hop-by-hop, instantiating state along the path towards the RP indicating that traffic from S for G should NOT be forwarded in this direction. The prune is propagated until it reaches the RP or a router that still needs the traffic from S for other receivers.

At this point (on R5) the SPT tree and the shared tree diverge and as general PIM SM rule states when it is the case the router on which the trees diverge should send a (S,G)RP-bit prune up the shared tree in order to receive traffic only from the source path tree which is the optimal path to the source. This special type of prune is to tell the RP to prune the source traffic from the shared tree in order to avoid receiving multicast via the shared tree and the SPT at the same time.

```
RP #
(10.0.12.1, 224.1.1.1), 00:00:10/00:02:53, flags: PTX
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.24.2
  Outgoing interface list: Null
```

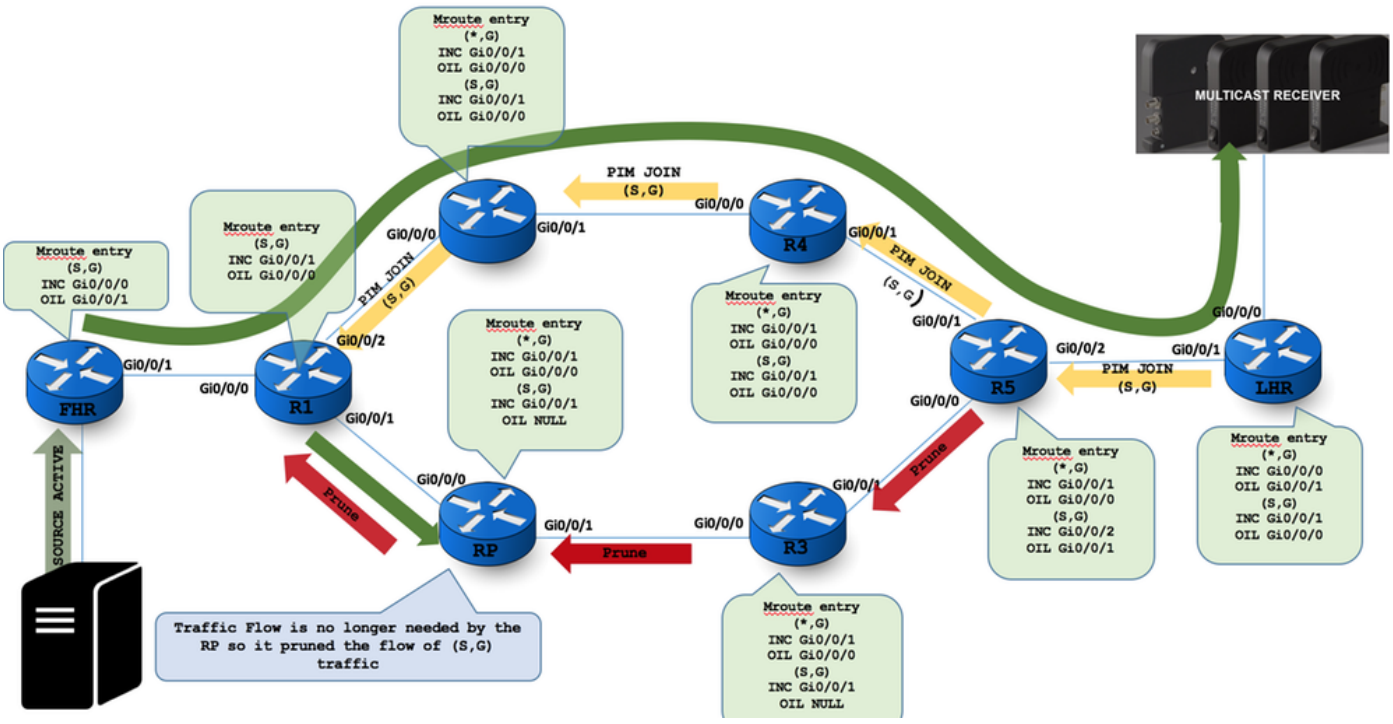
P Bit (Prune Flag) received from the diverge point.

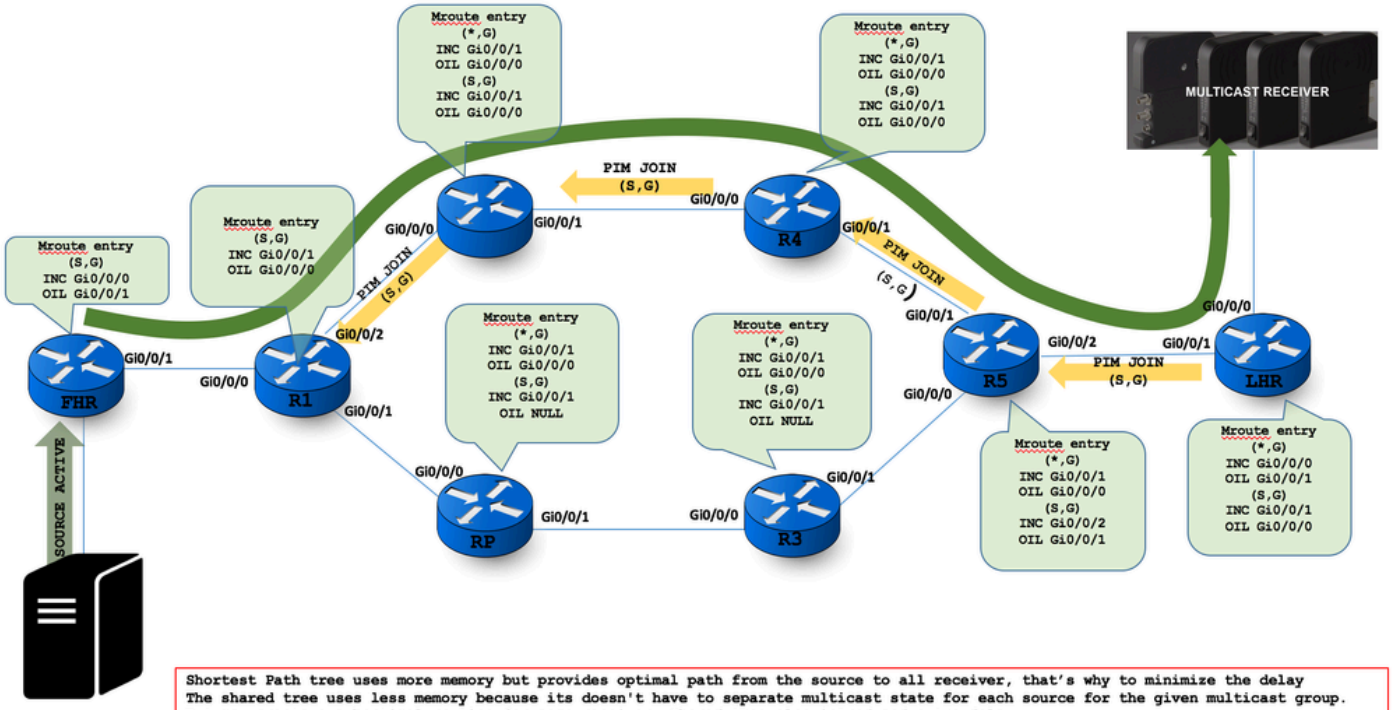
```
LHR #

(10.0.12.1, 224.1.1.1), 00:01:59/00:01:00, flags: LJT
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.78.7
  Outgoing interface list:
    GigabitEthernet1/0, Forward/Sparse, 00:01:59/00:02:57
```

J Flag Join the SPT// T Flag Tree formed

"PIM Join/Prune Messages" the RP flag (also referred to as the RP-bit) indicates that this message is applicable to the shared tree and should be forwarded up the shared tree toward the RP. Setting this flag/bit in an (S1, G) Prune and sending it up the shared tree tells the routers along the shared tree to prune Source S1 multicast traffic from the shared tree.





Shortest Path tree uses more memory but provides optimal path from the source to all receiver, that's why to minimize the delay
The shared tree uses less memory because its doesn't have to separate multicast state for each source for the given multicast group.
But may create a suboptimal routing for some receiver. Shared tree also introduced extra delay.

关于此翻译
思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。