Native Multicast Flow - Any-Source Multicast Model

Contents

Introduction

This document describes the packet flow of the Any-Source Multicast (ASM) Model.

Background Information

This document provides the detail packet flow of the Native Multicast packet flow and analysis of its output. This describes about the detail analysis output and packet flow in control plane and forwarding plane.

The ASM is the model in which the receiver does not have the knowledge of sender. It means it can receive traffic from any source. Receiver is only aware of the multicast group that sender uses and Internet Group Management Protocol (IGMP) in order to subscribe to receive all the traffic destined for this address.

All of this is covered in this document:

1. What happens when Receiver is active.

2. What happens when Source is active.

3. What happens when Register is received at Rendezvous Point (RP).

4. How (S,G) formed. Till First Hop Router (FHR).

5. Which path it takes for the first Multicast Stream.

6. What happens when two streams receive at the Last Hop Router (LHR).

7. How Shortest Path Tree (SPT) is formed over Shared Tree. Exactly what happens and the reason why the switchover takes place.

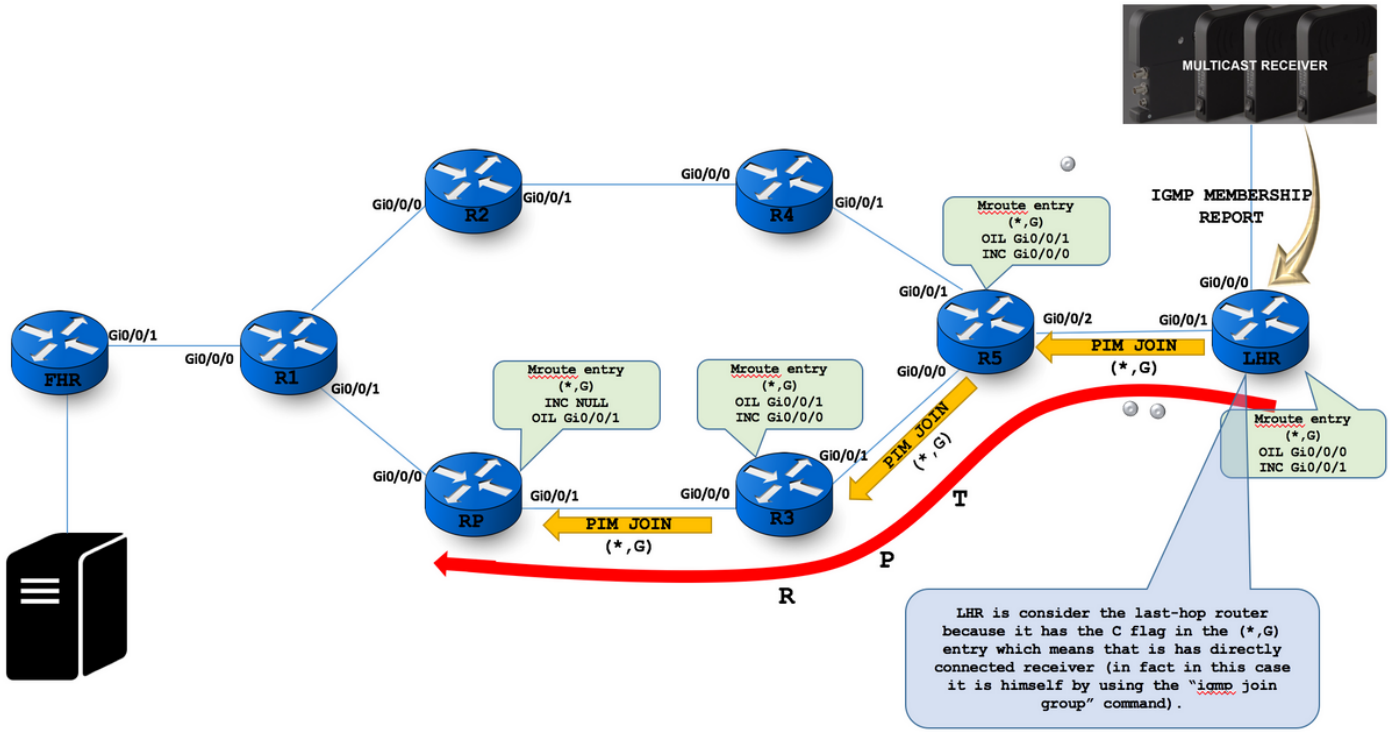
Protocol Independent Multicast (PIM) is used as a multicast routing protocol between source and

receiver to create the multicast tree. In ASM, (*,G) multicast entry is used where * represents any-source and G is the multicast group address receiver that is interested in order to receive the traffic.

Step 1. When Receiver is Active, it Sends IGMP Report Message

- When the receiver's expression of interest is received, the Designated Router (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 is consider the last-hop router because it has the C flag in the (*,G) entry which means that it has directly connected receiver (in fact in this case it is himself with the use of the **igmp join group** command).



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: GigabitEthernet2/0/1, 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.

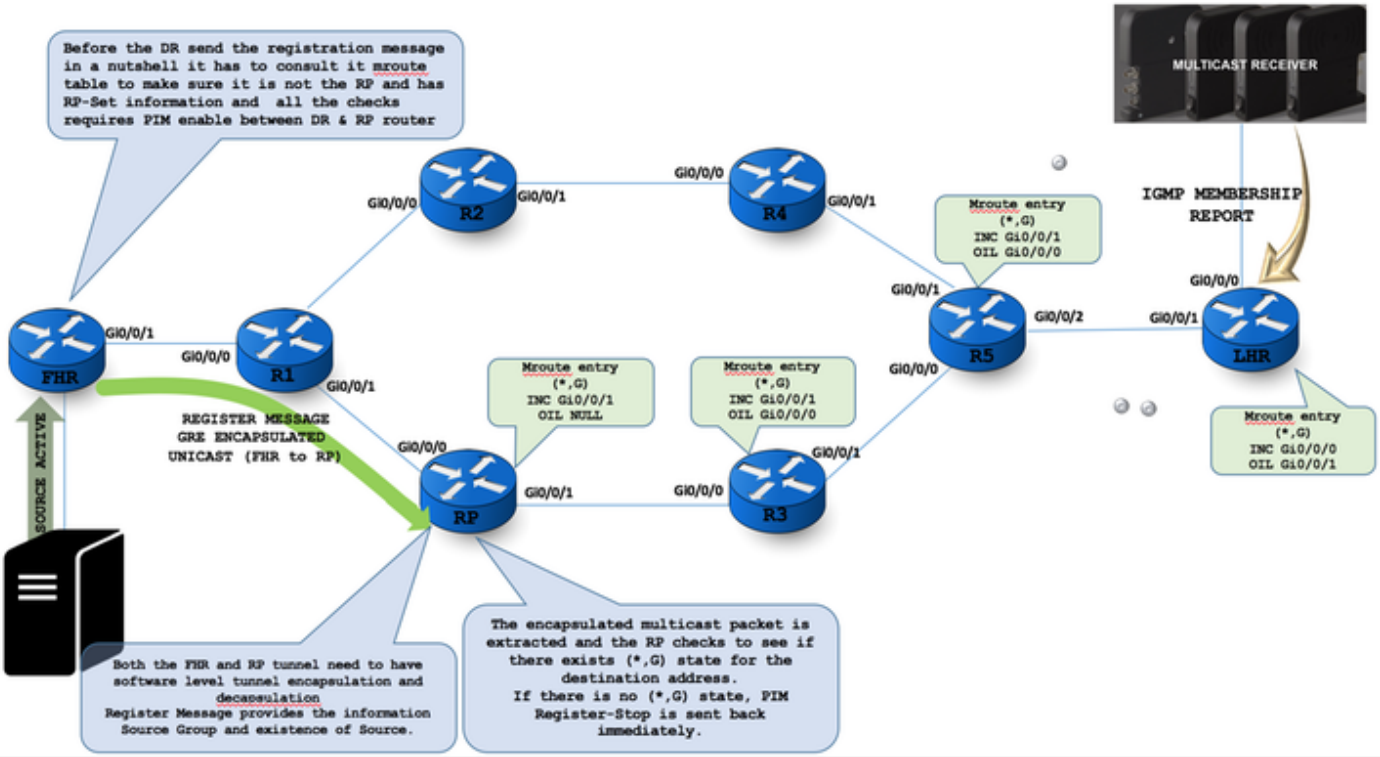
```
▶ Frame 59: 68 bytes on wire (544 bits), 68 bytes captured (544 bits) on interface 0
▶ Ethernet II, Src: ca:08:fa:92:00:00 (ca:08:fa:92:00:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
▼ Internet Protocol Version 4, Src: 10.0.78.8, Dst: 224.0.0.13
    0100 .... = Version: 4
    .... 0101 = Header Length: 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

Step 2. When Source is Active

- Before the DR sends the registration message, in a nutshell, it has to consult the mroute table to ensure that it is not the RP and has RP-Set information and all the checks require PIM to be enabled between DR and RP router.
- Both the FHR and RP tunnel need to have software level tunnel encapsulation and decapsulation.
- Register Message provides the information Source Group and existence of Source.
- The encapsulated multicast packet is extracted and the RP checks to see if there exists (*,G) state for the destination address.
- If there is no (*,G) state, PIM Register-Stop is sent back immediately.



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

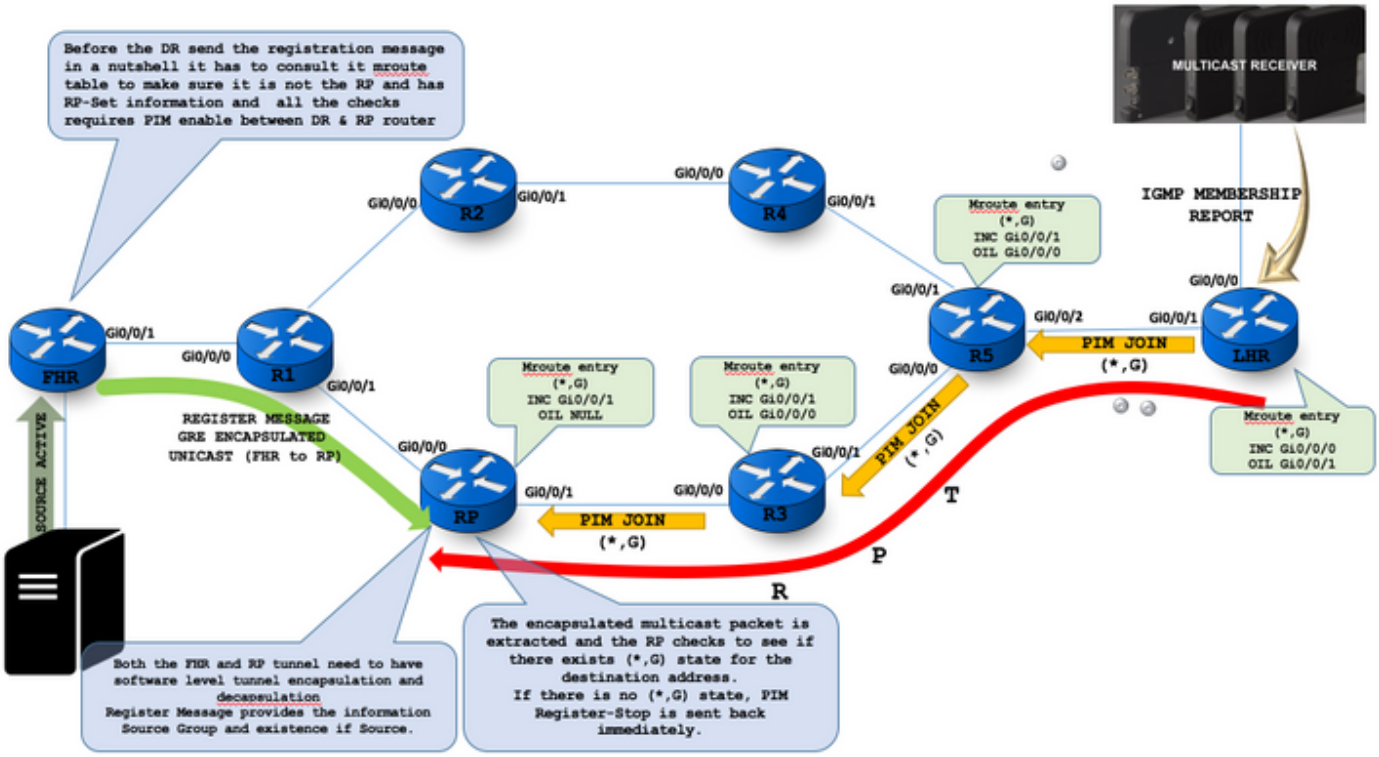
```
RP #
(1.1.1.1, 224.22.22.44), 00:00:43/00:02:16, flags: P
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.24.2
  Outgoing interface list: Null

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
```

Step 3. Form Shared Tree

- Before the DR sends the registration message, in a nutshell, it has to consult the mroute table to ensure that it is not the RP and has RP-Set information and all the checks require PIM to be enabled between DR and RP router
- Both the FHR and RP tunnel need to have software level tunnel encapsulation and decapsulation
- Register Message provides the information Source Group and existence if Source.
- The encapsulated multicast packet is extracted and the RP checks to see if there exists (*,G) state for the destination address.
- If there is no (*,G) state, PIM Register-Stop is sent back immediately.



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.
```

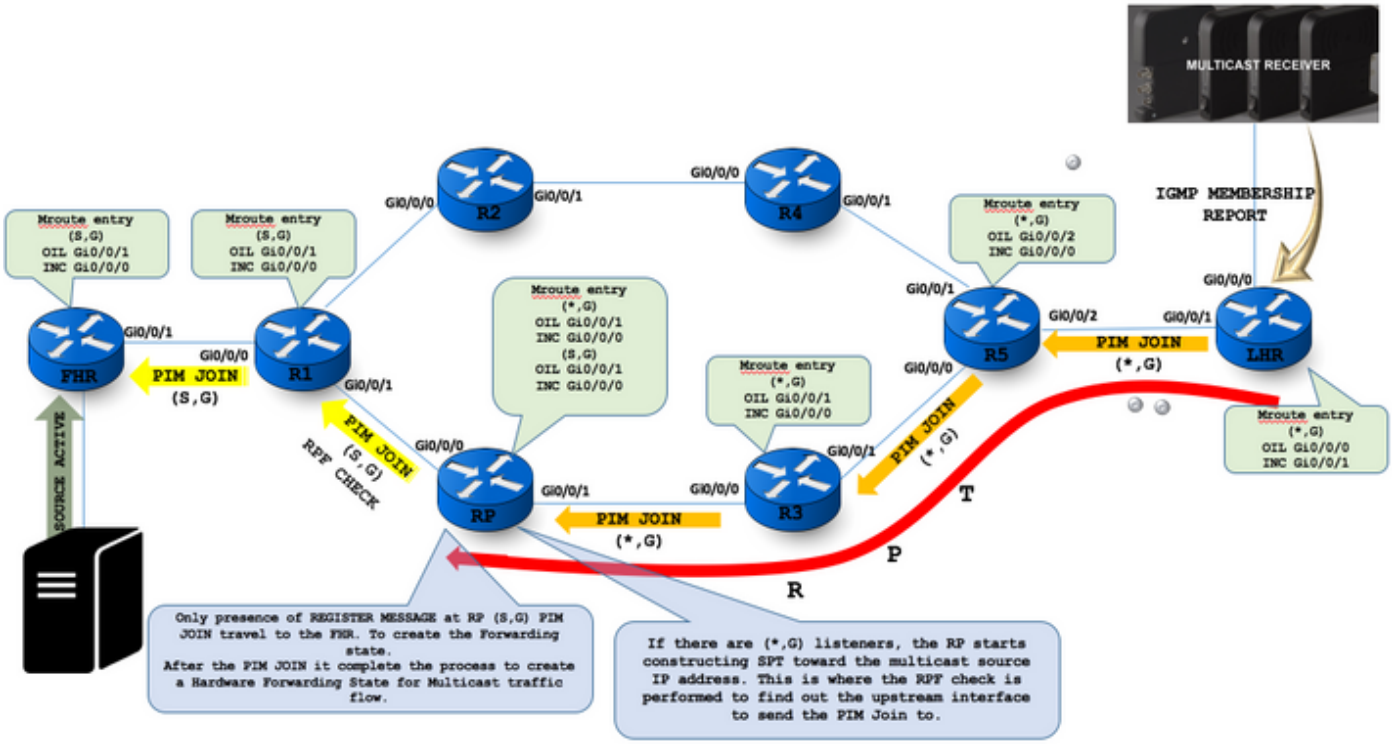
Presence of (*,G) at RP means active receiver.

```
▶ 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
```

Step 4. (S,G) Packet Reaching towards the FHR

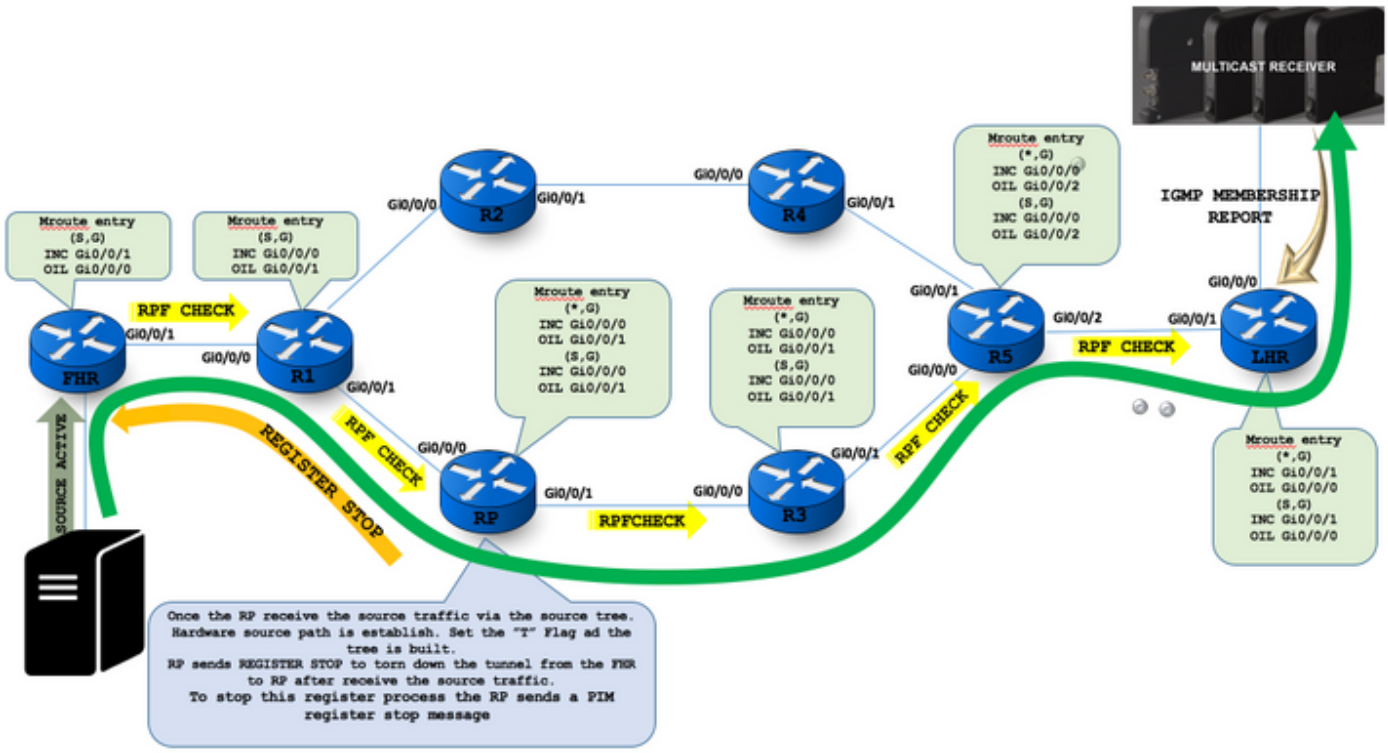
- Only presence of REGISTER MESSAGE at RP (S,G) PIM JOIN travel to the FHR. To create the Forwarding state.
- After the PIM JOIN it complete the process to create a Hardware Forwarding State for Multicast traffic flow.

- If there are (*,G) listeners, the RP starts constructing SPT toward the multicast source IP address. This is where the RPF check is performed to find out the upstream interface to send the PIM Join to.



Step 5. First Stream of Multicast Packet, reach to Receiver via Shared Tree

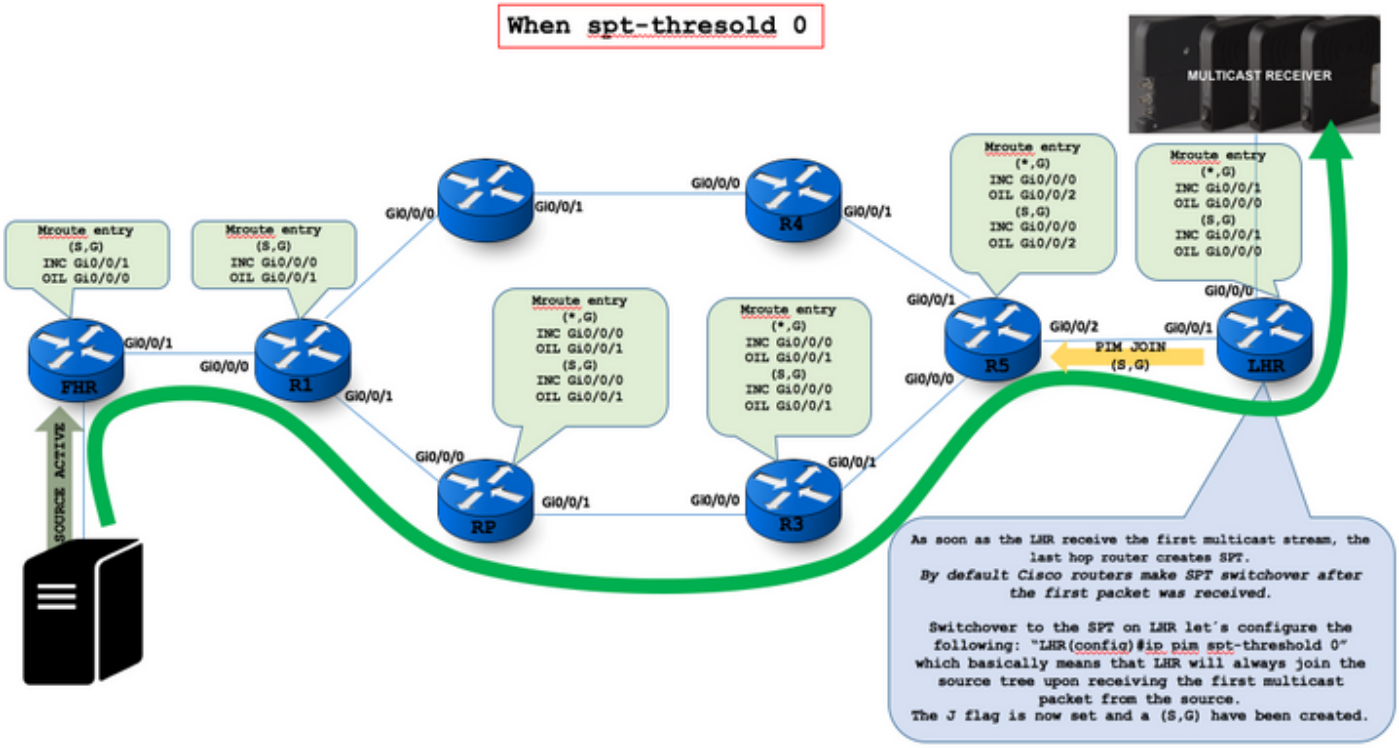
- A router forwards multicast stream only if received on the INC /RFP interface.
- Multicast Packets source address is checked against unicast RT.
- Determine the interface and next hop multicast router in the direction of the source where the join sent.
- RP is in the process of joining the source-specific tree for S, the data packets will continue being encapsulated to the RP. When packets from S also start to arrive natively at the RP, the RP will be receiving two copies of each of these packets.
- At this point, the RP starts to discard the encapsulated copy of these packets, and it sends a REGISTER STOP message back to S's DR to prevent the DR from unnecessarily encapsulating the packets.

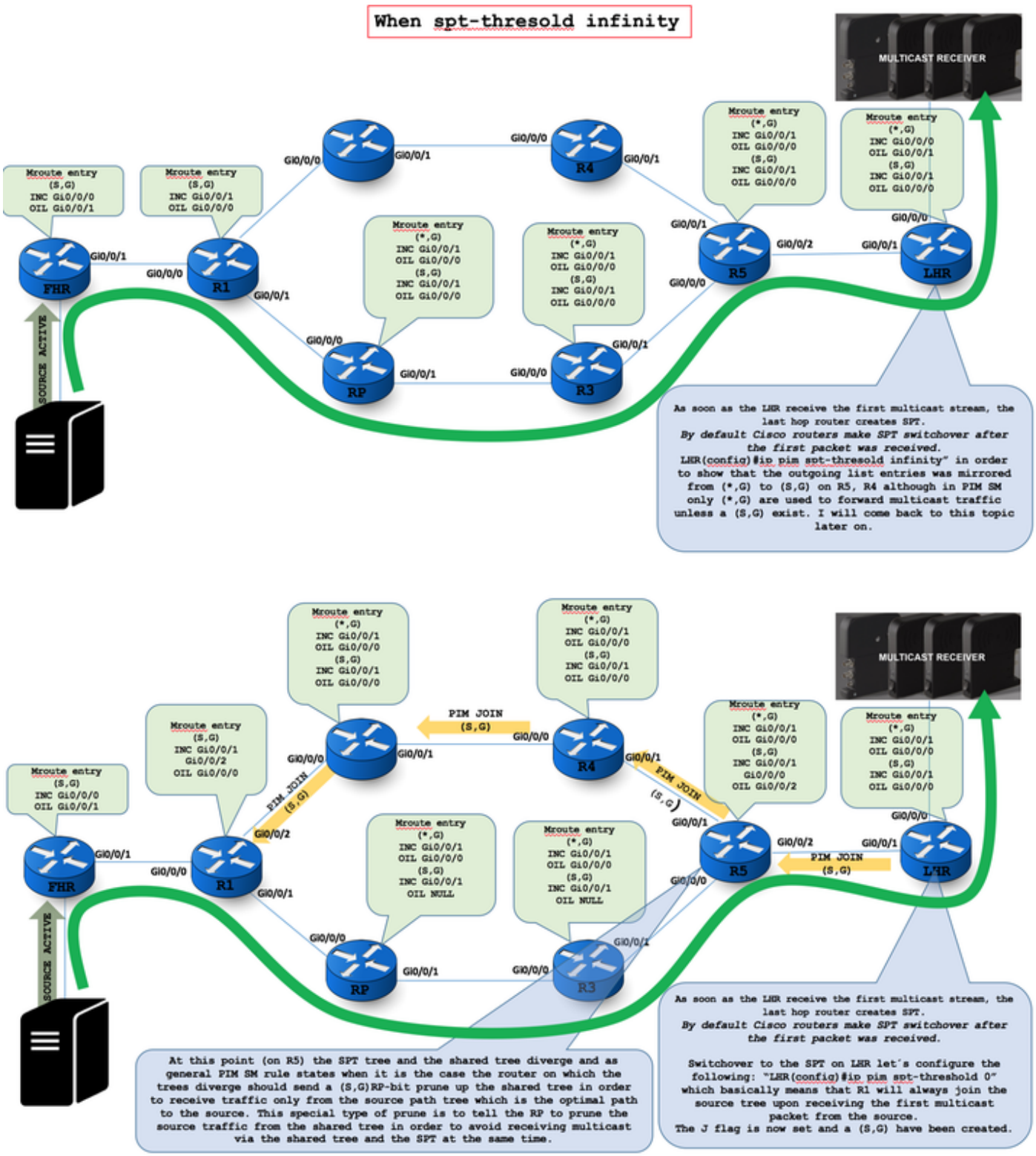


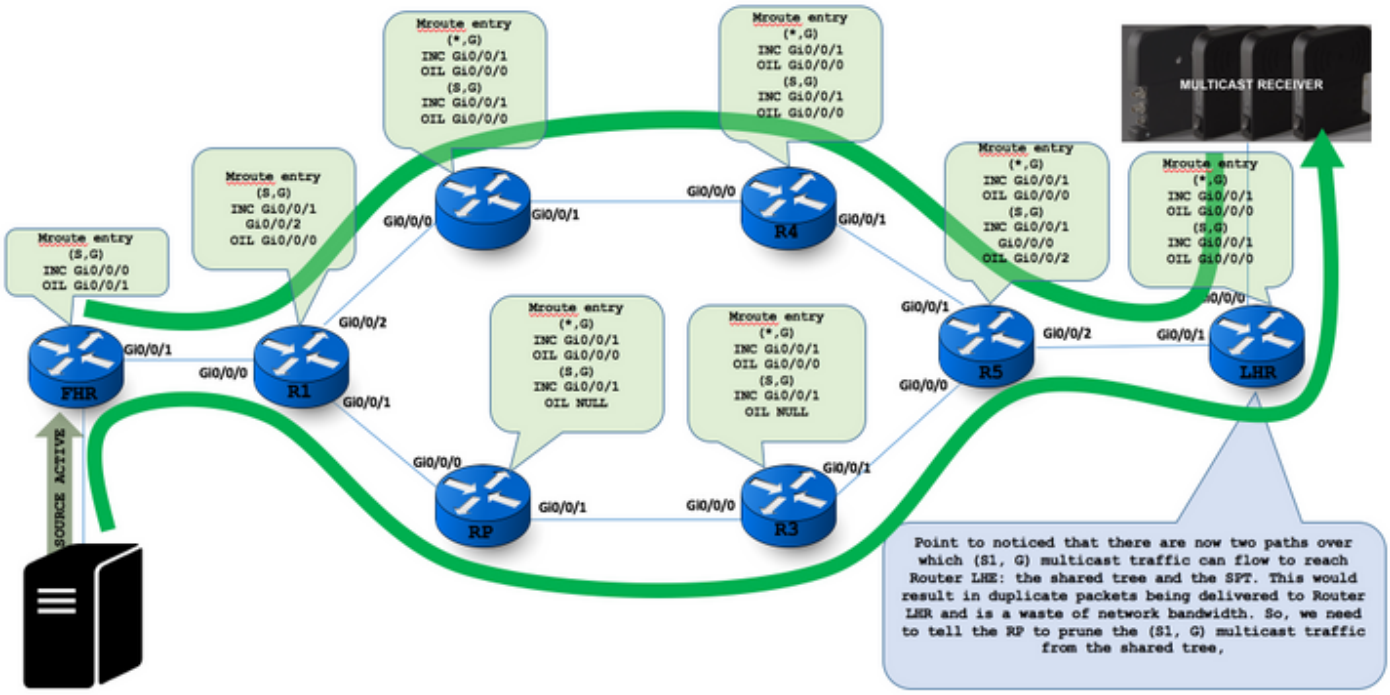
PIM-SM enables a last hop DR (that is, a DR with directly connected hosts that have joined a multicast group) to switch from the shared tree to the SPT for a specific source. This step is usually accomplished by specifying an SPT-Threshold in terms of bandwidth. If this threshold is exceeded, the last-hop 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.)

- Once the RP receive the source traffic via the source tree. Hardware source path is establish. Set the ”T” Flag ad the tree is built.
- RP sends REGISTER STOP to torn down the tunnel from the FHR to RP after receive the source traffic.
- To stop this register process the RP sends a PIM register stop message

When spt-thresold 0




When spt-thresold infinity
MULTICAST RECEIVER
Mroute entry
(S,G)
INC Gi0/0/0
OIL Gi0/0/1
Mroute entry
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/0
OIL Gi0/0/1
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
FHR
R1
RP
R3
R4
R5
LHR
Gi0/0/0
Gi0/0/1
Gi0/0/2
SOURCE ACTIVE
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.
MULTICAST RECEIVER
Mroute entry
(S,G)
INC Gi0/0/0
OIL Gi0/0/1
Mroute entry
(S,G)
INC Gi0/0/1
Gi0/0/2
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
PIM JOIN
(S,G)
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
Gi0/0/0
OIL Gi0/0/2
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL Gi0/0/0
Mroute entry
(*,G)
INC Gi0/0/1
OIL NULL
Mroute entry
(*,G)
INC Gi0/0/1
OIL Gi0/0/0
(S,G)
INC Gi0/0/1
OIL NULL
FHR
R1
RP
R3
R4
R5
L'HR
PIM JOIN
(S,G)
SOURCE ACTIVE
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.
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 R1 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.



Step 6. LHR Receives Traffic from SPT and Sends Prune Message towards the Shared Tree

After receipt of two streams of Multicast Traffic, the LHR starts to receive the traffic from SPT and sends prune message towards the Shared tree.

The J flag means the respective (*,G) state is to switch 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 that are registered with the RP. If the “F” flag persists, then your router is most likely unable to receive the PIM Register-Stop messages back from the RP, and thus there are sources that have not switched to the SPT.

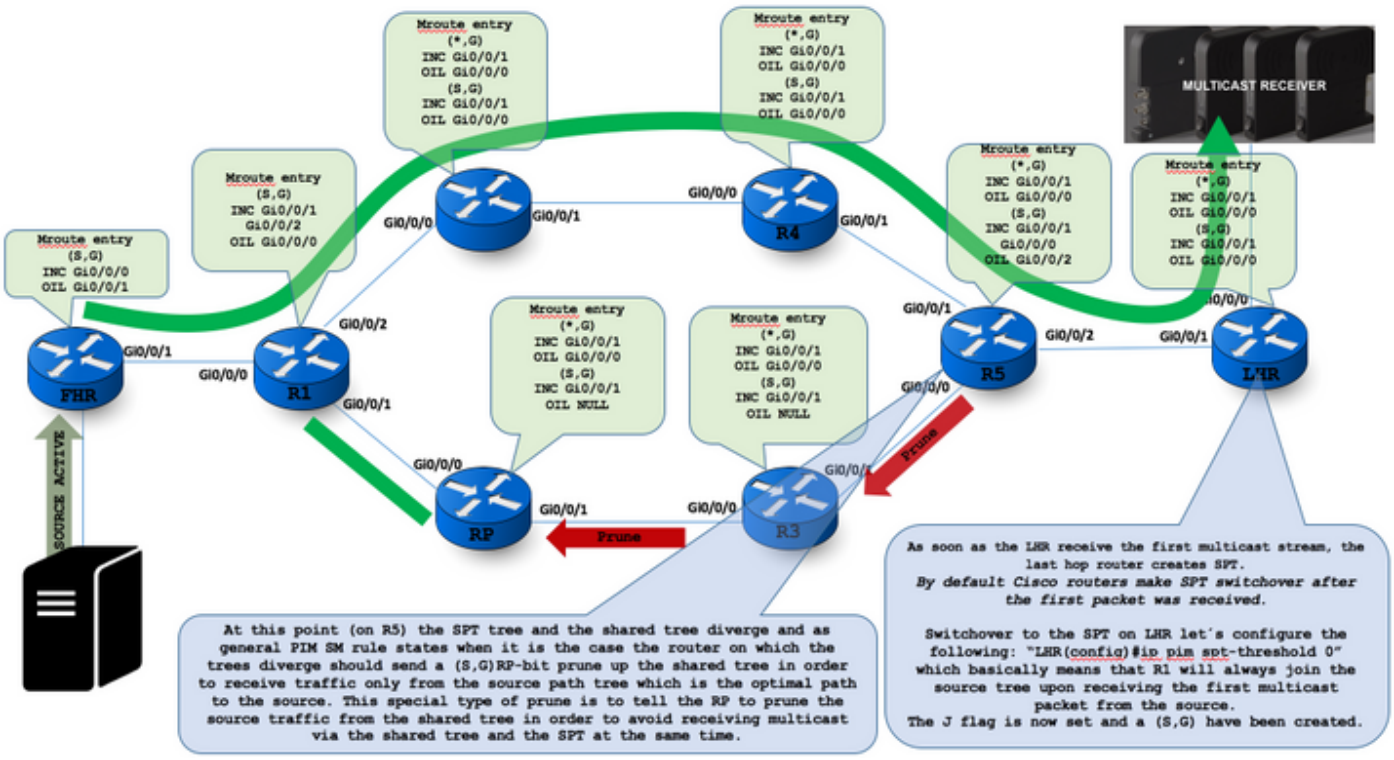
The J flag means the respective (*,G) state is to be switched the SPT by the leaf router.

```
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.

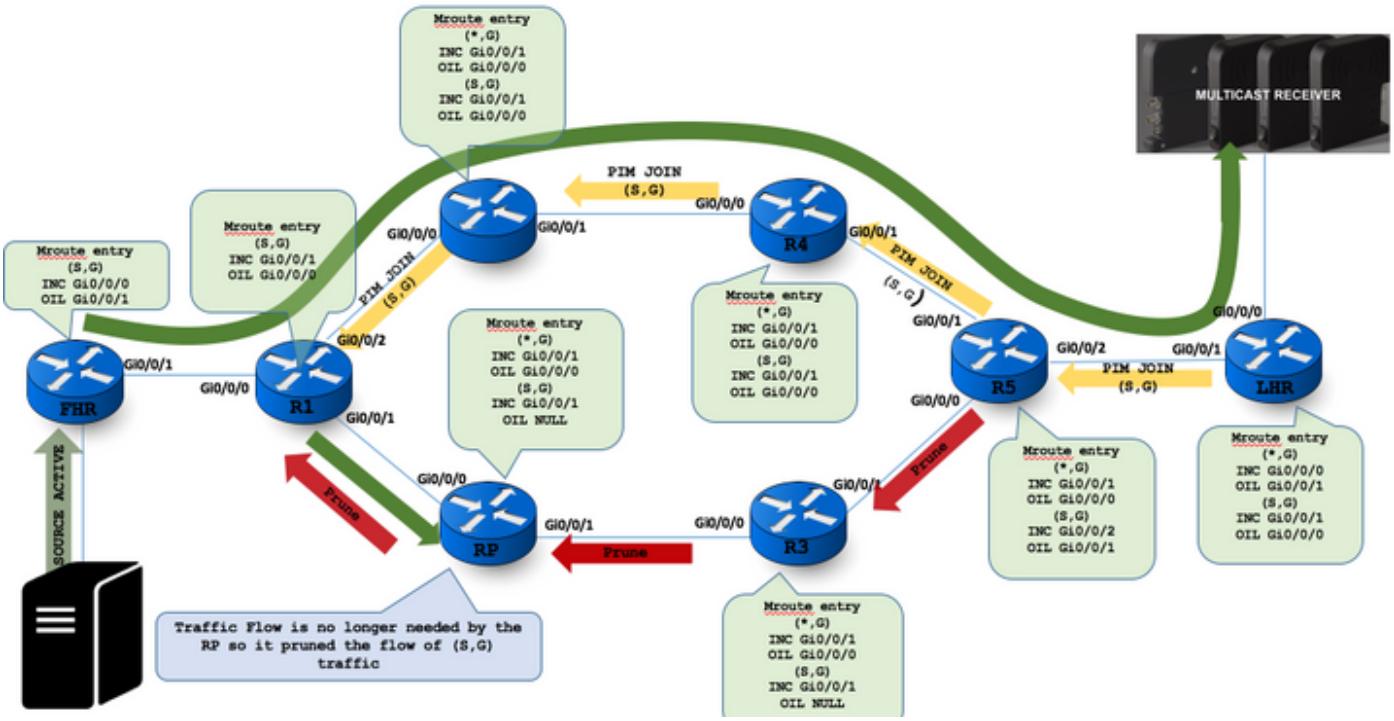
```
RP #
(10.0.12.1, 224.1.1.1), 00:00:10/00:02:53, flags: PTX
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.24.2
  Outgoing interface list: Null
```

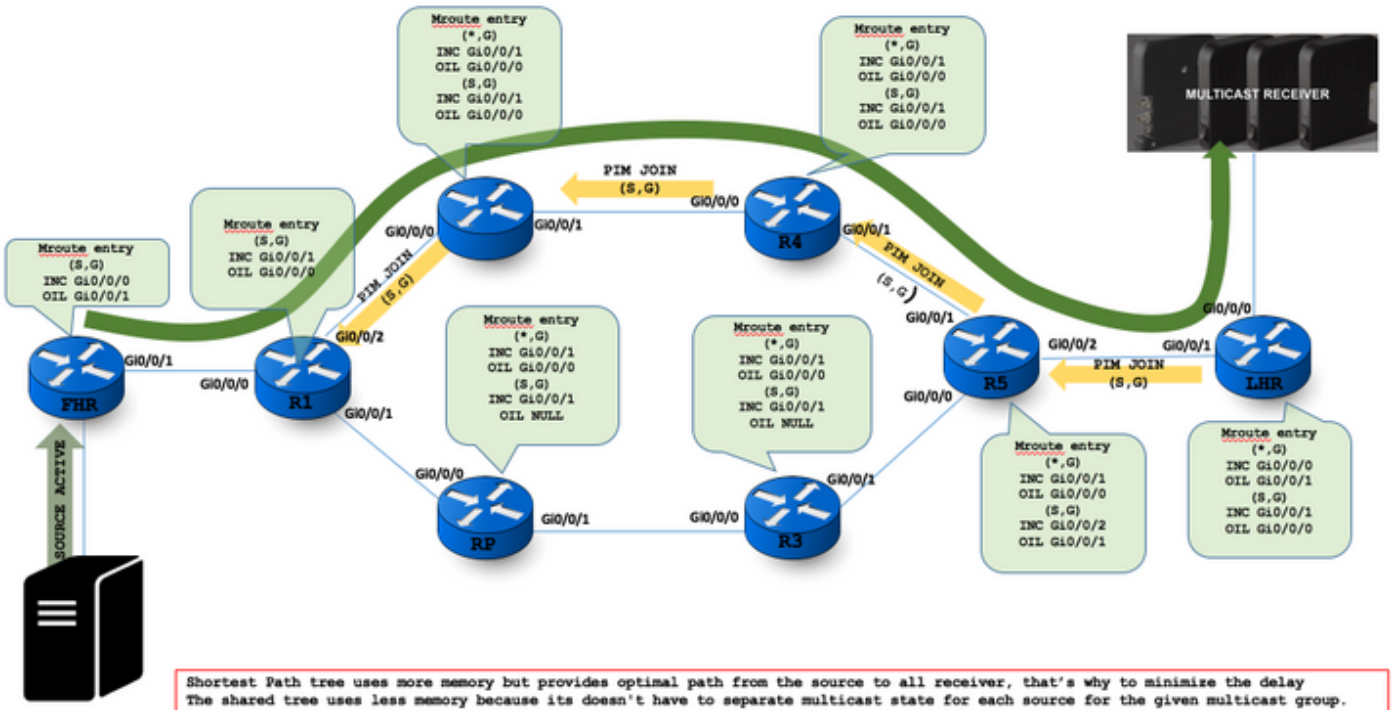
P Bit (Prune Flag) received from the diverge point.

```
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.

"Incoming interface" is set to Null, which means there is no incoming traffic for this group. If any physical interface the traffic is their.

"C" means there is a group-member directly connected

"L" means the router itself joined the group.

```
R5#sh ip mroute

(*, 239.1.1.1), 00:27:32/00:02:08, RP 4.4.4.4, flags: SJCL
  Incoming interface: FastEthernet0/0, RPF nbr 10.0.78.7
  Outgoing interface list:
    GigabitEthernet1/0, Forward/Sparse, 00:27:32/00:02:08
```

possibly the next-hop router

Expire times (How soon the group will expired if no refreshed

Uptime (How long this state has been created)

Incoming interface: Null, RPF nbr 155.29.0.5

If the incoming interface is null and the RPF neighbor is IP address, then there is a RPF failure. Mtrace will confirm the issue.