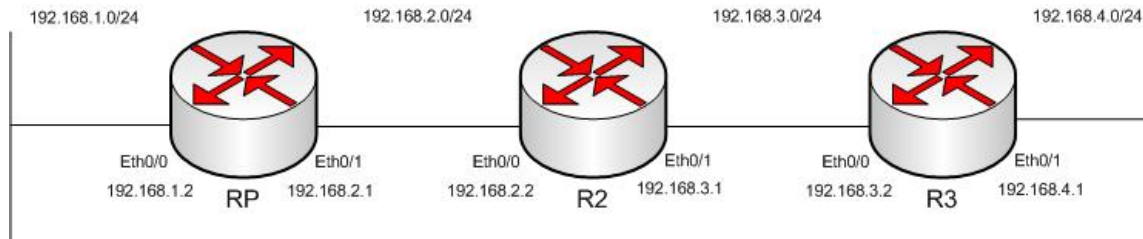


# **Protocol Independent Multicast: Sparse Mode**

**November 1, 2004  
Prepared for: William Farkas  
Prepared by: Hoon Song  
John Tewfik  
Leif Madsen**

## Summary

In this lab we will build a PIM-SM multicast enabled network to study the operation of the protocol in different scenarios. Cisco routers and switches will be used as the routing core. Laptops using the Mcast and TFGen software to receive and source multicast traffic will be used as the hosts. The figure below describes the basic topology that all procedures were based on.



Each subnet had a port monitor to monitor the traffic on the subnet. There was also a place to plug in a host on any of the subnets. If a host needed to be a receiver, the mcast program was then run and configured to listen to group address 239.255.5.1. If a host was a source of multicast traffic, it would run the tfggen program which was configured to source to the same group. Both programs were set to use port 5000 as their receiving and sourcing ports respectively. Hosts are always configured to use 192.168.x.100 where x is the subnet.

## Procedure 1

In this procedure, details of how Host A joins a multicast group will be outlined with respect to IGMP and CGMP messages. An IGMP message is used to indicate to an edge router of a multicast group member, so the router will forward traffic out the necessary interface. It is important to note that the IGMP message generated by a perspective multicast group member is not intercepted by a switch along the way to an edge router. Switches by default are designed to deliver multicast traffic to all ports, which can have undesirable affects for hosts not participating in multicast and wanting contention for the medium. For this reason, Cisco Systems introduced CGMP. This layer 2 message, understood by the switch, will allow the switch to forward group specific multicast traffic to only the ports where participating hosts are present.

### IGMP

The host wishing to inform an edge router of multicast group membership will issue an IGMP Report (type 2) destined for the group address. In our lab, the group address was 239.255.5.1. However, after dissecting captured data, we noticed that this address was evident in our IGMP Reports from Host A and a rather different multicast address of 239.255.255.250. We later learned that this was a UPnP feature of WinXP machines to join group multicasts.

```
▷ Internet Protocol, Src Addr: 192.168.4.100 (192.168.4.100), Dst Addr: 239.255.5.1 (239.255.5.1)
▽ Internet Group Management Protocol
  IGMP Version: 2
  Type: Membership Report (0x16)
  Max Response Time: 0.0 sec (0x00)
  Header checksum: 0xf4fe (correct)
  Multicast Address: 239.255.5.1 (239.255.5.1)
```

**Figure 1-1: IGMP Membership Report**

### CGMP

Now that our edge router is aware of a participating multicast group member, it will forward relative traffic out the interface from which it received the IGMP message. This is enough to ensure delivery of the multicast traffic, however, for the sake of the L2 switch to which host A is attached; a new message will be generated by the router. A CGMP message is a Layer 2 logical link control message that is forwarded out the interface from which the IGMP message was received. The switch is enabled to interpret this message, and makes necessary configuration changes. There are join/leave messages to inform the switch as to whether forwarding multicast traffic out of desired ports is necessary.

```
▽ Logical-Link Control
  DSAP: SNAP (0xaa)
  IG Bit: Individual
  SSAP: SNAP (0xaa)
  CR Bit: Command
  ▷ Control field: u, func=UI (0x03)
    organization Code: cisco (0x00000c)
    PID: CGMP (0x2001)
  ▽ Cisco Group Management Protocol
    0001 .... = Version: 1
    .... 0000 = Type: Join (0)
    Count: 1
    Group Destination Address: 01:00:5e:7f:ff:fa (01:00:5e:7f:ff:fa)
    Unicast Source Address: 00:0d:60:75:08:46 (192.168.4.101)
```

**Figure 1-2: CGMP Join message**

The above message was observed coming from our edge router to the switch. This message informs the switch as to which MAC address is a member of a multicast group. The switch's cam table illustrates which port contains information on Host A's MAC address. The switch will now forward necessary group multicast traffic to the port on which Host A is situated.

Note the count field of this message is set to 1, this indicates the number of group, and destination pairs there are present in the message. Host A could be a member of multiple groups, or, multiple hosts may reside on a given port.

## Procedure 2

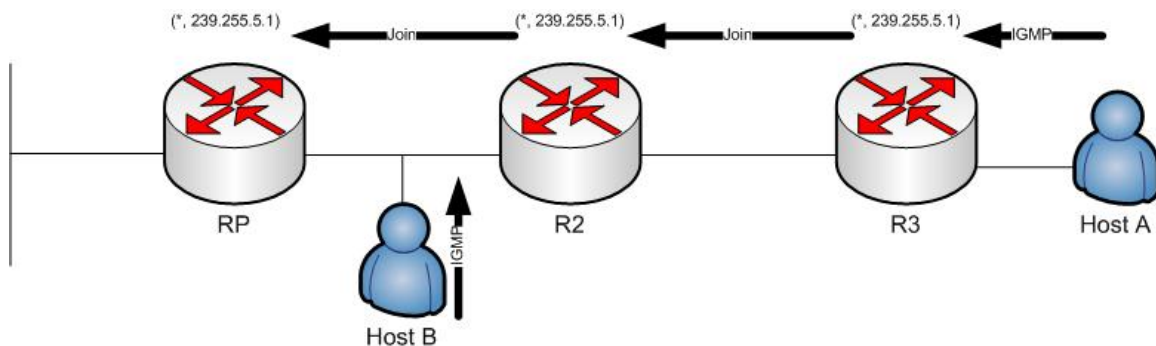
The following is an outline of the steps taken by participating routers in a multicast session. These are the “rules” which can be applied to procedure two, and gives a general overview of how PIM-SM works.

- 1) Host A sends an IGMP Report (also known as an IGMP-Join) message to R3.

```
▷ Internet Protocol, Src Addr: 192.168.4.100 (192.168.4.100), Dst Addr: 239.255.5.1 (239.255.5.1)
▽ Internet Group Management Protocol
  IGMP Version: 2
  Type: Membership Report (0x16)
  Max Response Time: 0.0 sec (0x00)
  Header checksum: 0xf4fe (correct)
  Multicast Address: 239.255.5.1 (239.255.5.1)
```

- 2) R3 then sends a PIM Join message towards the RP. The destination of the packet will be 229.0.0.13 (All PIM Routers). The routers along the path know that the RP is located at 10.10.1.1 due a field within the PIM parameters (Join: 1, IP Address: 10.10.1.1). The next hop (upstream-neighbor) is calculated at each router based on the routers routing table.

Each router along the path examines this PIM-Join message which creates a (\*, G) entry in each routers multicast table. This (\*, G) entry is referred to as the Shared Tree. It must be noted that the TTL of this PIM-Join message is set to one (1) in the IP datagram header. This is illustrated in Figure 2-1.



**Figure 2-1: PIM Join Message**

Below is a capture of a PIM-Join message. We can see that the Time to Live (TTL) of this PIM-Join message is set to 1. This is because the message only needs to go as

```
Ethernet II, Src: 00:02:4b:61:f3:21, Dst: 01:00:5e:00:00:0d
Internet Protocol, Src Addr: 192.168.3.2 (192.168.3.2), Dst Addr: 224.0.0.13 (224.0.0.13)
  Version: 4
  Header length: 20 bytes
  ▷ Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
  Total Length: 54
  Identification: 0x6be3 (27619)
  ▷ Flags: 0x00
  Fragment offset: 0
  Time to live: 1
  Protocol: PIM (0x67)
  Header checksum: 0xa906 (correct)
  Source: 192.168.3.2 (192.168.3.2)
  Destination: 224.0.0.13 (224.0.0.13)
Protocol Independent Multicast
  Version: 2
  Type: Join/Prune (3)
  Checksum: 0x220e (correct)
  ▷ PIM parameters
    Upstream-neighbor: 192.168.3.1
    Groups: 1
    Holdtime: 210
  ▷ Group 0: 224.0.1.40/32
    ▷ Join: 1
      IP address: 10.10.1.1/32 (SWR)
    Prune: 0
```

far as the next hop. The destination address is 239.0.0.13 (All PIM Routers) and the next hop is contained within the Upstream-neighbor field. Once that router receives the PIM-Join message, it inserts a (\*, G) entry into its multicast routing table. Within this routing table is included the incoming and outgoing physical interface for which the multicast traffic will pass through. This can be illustrated in Figure 2-2 below.

We can see that the entry contains the IP address of the Rendezvous Point (RP 10.10.1.1), the interface that incoming traffic will be accepted on (FastEthernet 0/0) and the interface that this incoming traffic will be forwarded out of (FastEthernet 0/1). Also of interest is the RPF (reverse path forwarding) entry of 192.168.3.1. The RPF algorithm allows a router to receive multicast traffic only on the same interface for which it would send unicast traffic to the source. This verifies that the source is coming in on the correct port (ie. to avoid spoofed traffic) and to make sure that we don't send traffic out of a port that we just received it on.

```
<*, 239.255.5.1>, 00:01:47:00:02:37, RP 10.10.1.1, flags: SJC
Incoming interface: FastEthernet0/0, RPF nbr 192.168.3.1
Outgoing interface list:
FastEthernet0/1, Forward/Sparse, 00:01:47:00:02:37
```

**Figure 2-2: Shared Tree Entry in Mcast Routing Table (R3)**

- 3) Host B is then brought up into listening mode. An IGMP Report message is sent to R2. Since this router is already a part of the Shared Tree, a PIM-Join message is not sent to the RP.
- 4) Host A then starts to send multicast traffic destined for 239.255.5.1. R3 then encapsulates this traffic in unicast messages destined for the RP (10.10.1.1). These encapsulated messages follow basic IP routing rules towards the RP. These messages are referred to as Register packets.

```

> Internet Protocol, Src Addr: 192.168.3.2 (192.168.3.2), Dst Addr: 10.10.1.1 (10.10.1.1)
< Protocol Independent Multicast
  Version: 2
  Type: Register (1)
  Checksum: 0x9927 (incorrect, should be 0xdef)
  < PIM parameters
    > Flags: 0x00000000
    > Internet Protocol, Src Addr: 192.168.4.100 (192.168.4.100), Dst Addr: 239.255.5.1 (239.255.5.1)
    > User Datagram Protocol, Src Port: 1420 (1420), Dst Port: 5000 (5000)

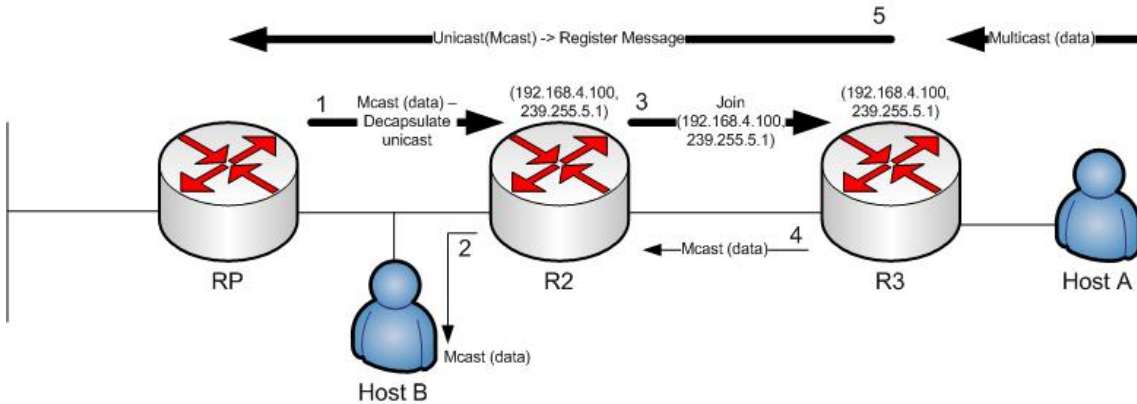
```

Above we can clearly see unicast encapsulating the multicast packet. The first line contains the source address of the designated router (R3 – 192.168.3.2) for Host A and a unicast destination address for the RP (10.10.1.1). Within that IP datagram is contained the PIM message which encapsulates the multicast IP header and data. The encapsulated multicast packet has a source address of the sending host (source, Host A – 192.168.4.100) and a multicast destination address (group 239.255.5.1).

- 5) Once the Register packets are received at the RP, the unicast portion of the message is decapsulated, leaving the pure multicast packet.
- 6) The multicast packets are then forwarded towards Host B along the Shared Tree (step 1 in Figure 2-3). Multicast packets will be broadcast onto the subnet where router R2 and Host B will see the traffic. Router R2 will then forward this traffic on towards R3. RPF is not applied here due to R2 not

having seen the multicast traffic in the Register packet. RPF would be applied at R3 and the traffic simply dropped as duplicate.

- 7) When the first multicast message is received by R2, the source address (192.168.4.100) is observed. Once this address is learned by the router, a PIM-Join message is sent towards R3 (step 3). When R3 sees this Join message, it will start to send multicast packets destined for the group (step 4). This effectively creates a (Source, Group) path between R2 and R3. Register packets continue to be sent to the RP from R3 (step 5). This is illustrated in Figure 2-3 below.

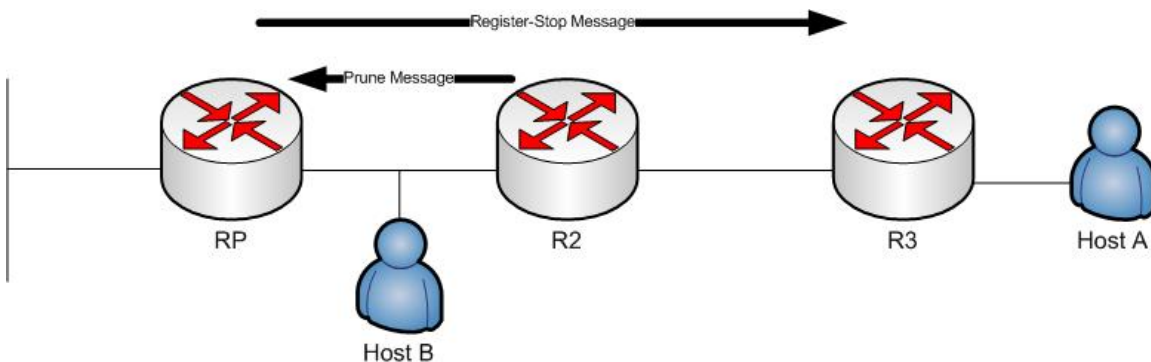


**Figure 2-3: SBT Switching Procedure**

Below is a sample of the (Source, Group) entry in R2.

```
<192.168.4.100, 239.255.5.1>, 00:01:20/00:01:50, flags: PT
Incoming interface: FastEthernet0/1, RPF nbr 192.168.2.2
Outgoing interface list: Null
```

- 8) Once R2 sees the first multicast packet coming from R3, it will send a PIM Prune message to the RP. This is saying, "I am receiving my traffic from the source directly, so you no longer need to send to me".
- 9) After the Prune message is received from R2, the RP sees that there are no longer any listeners that it needs to source to. The RP then sends a Register-Stop to R3 which essentially says, "there are no more listeners that I need to send to, so you can stop sending me Register packets". This is illustrated in Figure 2-4.



**Figure 2-4: Stopping the Register packets from source**

10) Router B then stops sending Register packets to the RP and continues to send multicast traffic to R2 which then broadcasts it on Host B's subnet.

### Procedure 3

- 1) Host A sends an IGMP Report to R3. This establishes Host A as a listener in group 239.255.5.1.
- 2) R3 then sends a PIM-Join with a destination of 229.0.0.13 (All PIM Routers). The PIM-Join is destined for the RP. This works the same as step 2 in Procedure 2.
- 3) Host A then starts to send multicast data for group 239.255.5.1 towards R3. R3 then encapsulates the multicast data inside a unicast datagram. This is then sent to the RP as a Register packet.

```
▶ Frame 108 (75 bytes on wire, 75 bytes captured)
▶ Ethernet II, Src: 00:02:4b:61:f3:21, Dst: 00:0c:85:0d:a7:21
▶ Internet Protocol, Src Addr: 192.168.3.2 (192.168.3.2), Dst Addr: 10.10.1.1 (10.10.1.1)
▼ Protocol Independent Multicast
  Version: 2
  Type: Register (1)
  Checksum: 0x992b (incorrect, should be 0xdef)
▼ PIM parameters
  ▶ Flags: 0x00000000
  ▶ Internet Protocol, Src Addr: 192.168.4.100 (192.168.4.100), Dst Addr: 239.255.5.1 (239.255.5.1)
  ▶ User Datagram Protocol, Src Port: 1121 (1121), Dst Port: 5000 (5000)
```

- 4) Once the RP receives the Register packet, it immediately sends back a Register-Stop. This is due to a lack of listeners for the group address on the network.

```
.....
▶ Frame 109 (60 bytes on wire, 60 bytes captured)
▶ Ethernet II, Src: 00:0c:85:0d:a7:21, Dst: 00:02:4b:61:f3:21
▶ Internet Protocol, Src Addr: 10.10.1.1 (10.10.1.1), Dst Addr: 192.168.3.2 (192.168.3.2)
▼ Protocol Independent Multicast
  Version: 2
  Type: Register-stop (2)
  Checksum: 0x21d2 (correct)
▼ PIM parameters
  Group: 239.255.5.1/32
  Source: 192.168.4.100
```

- 5) For a period of time, Register and Register-Stop messages are exchanged between the RP and R3. This is essentially performing a keep-alive function to allow the routers to maintain adjacency. It was observed that our routers exchanged the Register / Register-Stop messages every 120 seconds (2 minutes). There was no indication within the packets as to where this hold time was stored. Based on observations, an educated guess of the time being statically defined within the Cisco implementation of PIM-SM would explain why the hold time is 120 seconds. Further, it can be observed that Register/Register-Stop pairs are sent every 60 seconds, but the group that they describe is rotated. Because we see only two groups being rotated, this would explain why this happens every 120 seconds. It should be fair to say that as long as Register/Register-Stop messages are exchanged every 60 seconds, and that the group for which it is describing does not matter, but that the state being refreshed is what counts.

Seq	Time	Src	Dst	Protocol	Message
386	169.03700	192.168.3.2	10.10.1.1	PIMv2	Register
387	169.03888	10.10.1.1	192.168.3.2	PIMv2	Register-stop
396	172.75948	192.168.3.1	224.0.0.13	PIMv2	Hello
417	180.03630	192.168.3.2	224.0.0.13	PIMv2	Hello
424	183.72380	10.10.1.1	224.0.0.2	PIMv1	RP-Reachable
433	186.62381	10.10.1.1	224.0.0.2	PIMv1	RP-Reachable
465	202.03690	192.168.3.2	224.0.0.13	PIMv2	Join/Prune
466	202.55107	192.168.3.1	224.0.0.13	PIMv2	Hello
483	210.03625	192.168.3.2	224.0.0.13	PIMv2	Hello
493	215.06839	192.168.3.2	10.10.1.1	PIMv2	Register
494	215.07014	10.10.1.1	192.168.3.2	PIMv2	Register-stop
498	217.06861	192.168.3.2	224.0.0.13	PIMv2	Join/Prune
532	232.01923	192.168.3.1	224.0.0.13	PIMv2	Hello
552	240.06768	192.168.3.2	224.0.0.13	PIMv2	Hello
594	261.56723	192.168.3.1	224.0.0.13	PIMv2	Hello
598	262.06853	192.168.3.2	224.0.0.13	PIMv2	Join/Prune
615	270.06752	192.168.3.2	224.0.0.13	PIMv2	Hello
622	273.76039	10.10.1.1	224.0.0.2	PIMv1	RP-Reachable
626	275.09953	192.168.3.2	10.10.1.1	PIMv2	Register
627	275.10115	10.10.1.1	192.168.3.2	PIMv2	Register-stop
631	276.10026	192.168.3.2	224.0.0.13	PIMv2	Join/Prune
632	276.68839	10.10.1.1	224.0.0.2	PIMv1	RP-Reachable
658	289.09953	192.168.3.2	10.10.1.1	PIMv2	Register
659	289.10115	10.10.1.1	192.168.3.2	PIMv2	Register-stop

- 6) At this point Host B joins the group by sending an IGMP Report to the RP (the first-hop-router of the host).

```

> Internet Protocol, Src Addr: 192.168.1.100 (192.168.1.100), Dst Addr: 239.255.5.1 (239.255.5.1)
  Internet Group Management Protocol
    IGMP Version: 2
    Type: Membership Report (0x16)
    Max Response Time: 0.0 sec (0x00)
    Header checksum: 0xf4fe (correct)
    Multicast Address: 239.255.5.1 (239.255.5.1)
  
```

- 7) A PIM-Join message is sent from the RP acting as a DR for Host B towards R3 with a destination of 229.0.0.13 (All PIM Routers). This PIM-Join message is essentially telling R3 that it may start sending traffic again since there is now a listener.

```

Internet Protocol, Src Addr: 192.168.3.1 (192.168.3.1), Dst Addr: 224.0.0.13 (224.0.0.13)
Protocol Independent Multicast
  Version: 2
  Type: Join/Prune (3)
  Checksum: 0x5733 (correct)
  PIM parameters
    Upstream-neighbor: 192.168.3.2
    Groups: 1
    Holdtime: 210
    Group 0: 239.255.5.1/32
      Join: 1
        IP address: 192.168.4.100/32 (S)
      Prune: 0

```

Above we see the PIM-Join message as observed between R2 and R3. The PIM parameters contain a field with the group name (239.255.5.1) and the IP address of our source which we want to hear data from. As soon as the packet arrives at R3, it opens the gates to allow multicast traffic to start flowing.

The result is that we send multicast data instead of unicast Register messages. This is because the RP router - acting as the DR router for Host B - has established bi-directional communication. This is because the original PIM-Join message sent in step 2 of this procedure left residual entries in the routers which established the Shared Tree Path – a (\*, G) entry. When the RP sent the PIM-Join message back to allow communications to flow again, this created a Source, Group pairing in the routers leading to a Source Based Tree. With this Source Based Tree established, we are able to send the information in pure multicast form, relieving R3 from the extra processing power required to encapsulate the data, and in turn the RP from decapsulating. This is seen in the following screenshot.

904	394.16245	192.168.3.2	224.0.0.13	PIMv2	Join/Prune
905	394.88849	192.168.3.1	224.0.0.13	PIMv2	Join/Prune
906	394.89498	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]
907	394.90472	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]
908	394.91471	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]
909	394.92483	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]
910	394.93483	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]
911	394.94469	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]
912	394.95481	192.168.4.100	239.255.5.1	UDP	Source port: 1121 Destination port: 5000 [Malformed Packet]

## Questions

- 1) A PIM-SM RPT join message is propagated towards the RP (Rendezvous Point) using information gathered from the IP portion of the message, as well as information included in the PIMv2 protocol portion of the message. From the IP portion of the join message, the TTL value of 1 is important to note as the packet is set to expire as soon as it reaches a neighboring router. From the PIMv2 portion of the message the upstream neighbor information is used to issue the (\*,G) pair to the neighboring router. The neighbor then generates a similar join message destined for the next router along the path to the RP.
- 2) There is media payload redundancy during the period where the source is sending register packets to the RP and multicast to the learned (S,G) pair. This continues until the RP issues a register stop to the source.
- 3) In routers participating in the shared RPT, there are entries in the multicast routing table indicating RPF (Reverse Path Forwarding) information. The RPF algorithm allows a router to accept a multicast datagram only on the interface from which the router would send a unicast datagram to the source of the multicast datagram. This entry indicates the network address of this interface.
- 4) A wildcard prune is used when no more listeners are active for a router along the RPT. When this happens, multicast is no longer needed from the (\*,G) path and a prune message is issued to the RP. A specific source prune may occur towards the RP if a listener has started receiving multicast traffic along SPT from source.
- 5) Under PIM-SMv2 dynamically configuring a RP router includes the usage of a BSR (Bootstrap Router). This router is elected based on priority, and in the case of a tie, on highest IP address. The BSR is used to propagate a RP set of routers to ALL-PIM-ROUTERS in a domain. It is common for a BSR to also be a candidate to become a RP. The BSR is responsible for flooding this information throughout the domain.
- 6) Sender 2 will start to source multicast packets. R2 only aware of the (\*,G) pairing between it and the RP will encapsulate the packets in unicast messages destined for the RP. The RP will then decapsulate these messages and place them back onto the subnet. This does not disregard the RPF rule as the packets were received via unicast, so it does not apply here. However it will be applied at R2 so as not to forward the traffic back to Sender 2. R3 will then see the multicast message from RP. It will see the source address of Sender 2 and send a PIM-Join message to R2 (creating a S,G entry in the mcast routing tables) to receive its multicast traffic from the source directly (instead of from RP). R2 will continue to send Register packets to the RP while at the same time sourcing multicast traffic onto the subnet. RP will continue to decapsulate the Register packets and place them onto the subnet as well. At this point R3 will see duplicate data arriving. R3 will then send a PIM-Prune message to RP to stop this duplicate traffic. It will be pruning the S,G,RPT entries in the mcast routing table and not the S,G pairing that it requires to continue receiving traffic from Sender 1. RP will then notice it no longer needs to decapsulate and source the multicast traffic and will send a Register-Stop to R2. At this point the network will have stabilized with R2

sourcing multicast data traffic directly onto the subnet which R3 will then forward to the Receiver.