Troubleshoot HSRP Aware PIM

Contents

Introduction

This document describes how to troubleshoot the Hot Standby Router Protocol (HSRP)-aware Protocol Independent Multicast (PIM) feature and scenarios in which it can be used.

Explanation

In environments that require redundancy for you, HSRP runs normally. HSRP is a proven protocol and it works, but how do you handle when you have clients that need multicast? What triggers multicast to converge when the Active Router (AR) goes down? In this case, Topology 1 is used:
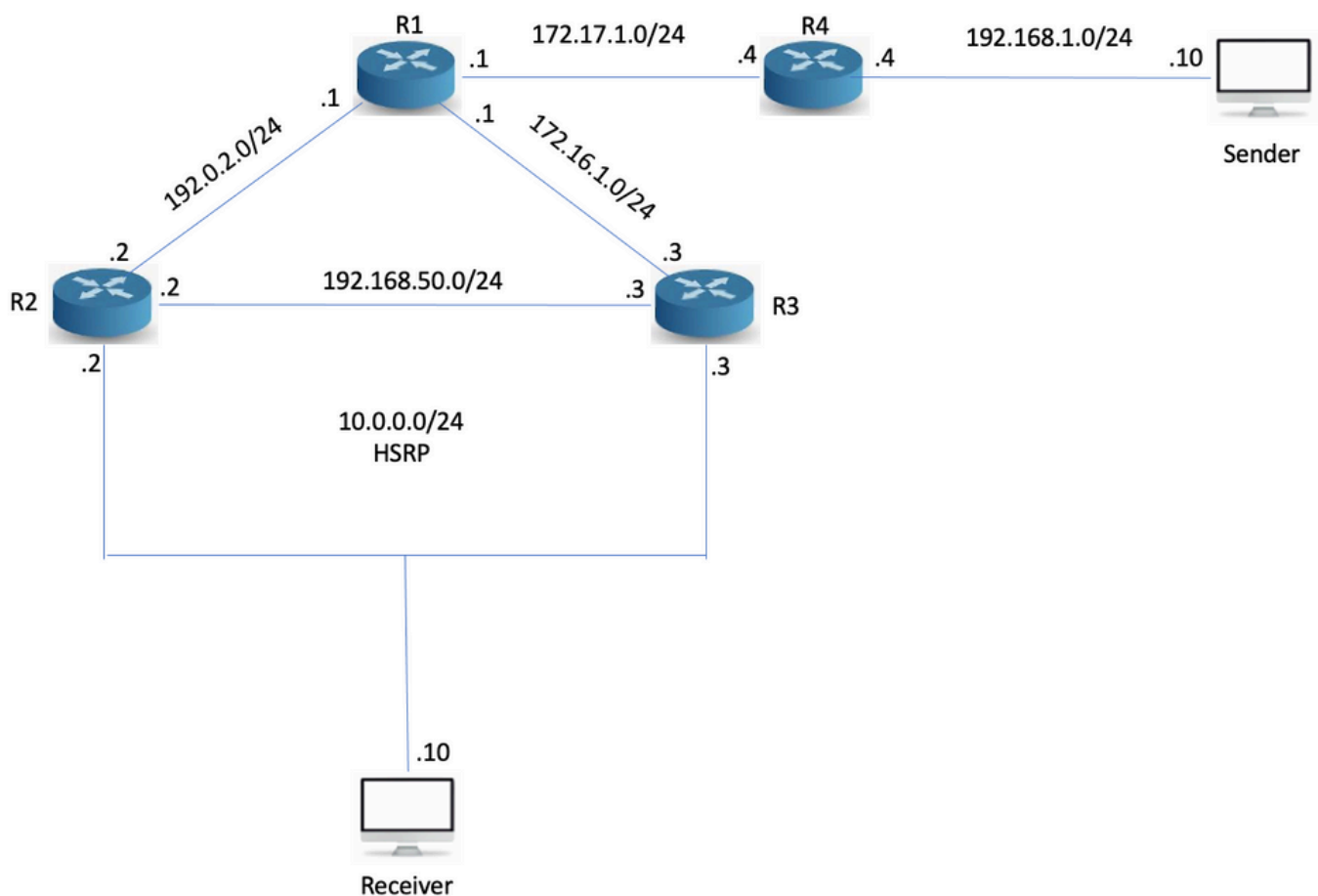
Topology 1

One thing to notice here is that R3 is the PIM Designated Router (DR) even though R2 is the HSRP AR. The network has been set up with Open Shortest Path First (OSPF), PIM and R1 are the Rendezvous Point (RP) with a 10.1.1.1 IP address. Both R2 and R3 receive Internet Group Management Protocol (IGMP) reports but only R3 sends the PIM Join since it is the PIM DR. R3 builds the '*,G' towards the RP:

```
R3#sh ip mroute 239.0.0.1 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, E - Extranet, 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, G - Received BGP C-Mroute, g - Sent BGP C-Mroute, N - Received BGP Shared-Tree Prune, n -
BGP C-Mroute suppressed, Q - Received BGP S-A Route, q - Sent BGP S-A Route, V - RD & Vector, v
- Vector, p - PIM Joins on route Outgoing interface flags: H - Hardware switched, A - Assert
winner, p - PIM Join Timers: Uptime/Expires Interface state: Interface, Next-Hop or VCD,
State/Mode (*, 239.0.0.1), 02:54:15/00:02:20, RP 10.1.1.1, flags: SJC Incoming interface:
Ethernet0/0, RPF nbr 172.16.1.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:25:59/00:02:20
```

You then ping 239.0.0.1 from the multicast source to build the S,G:

```
Sender#ping 239.0.0.1 re 3 Type escape sequence to abort. Sending 3, 100-byte ICMP Echos to
239.0.0.1, timeout is 2 seconds: Reply to request 0 from 10.0.0.10, 35 ms Reply to request 1
from 10.0.0.10, 1 ms Reply to request 2 from 10.0.0.10, 2 ms
```

The S,G has been built:

```
R3#sh ip mroute 239.0.0.1 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, E - Extranet, 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, G - Received BGP C-Mroute, g - Sent BGP C-Mroute, N - Received BGP Shared-Tree Prune, n -
BGP C-Mroute suppressed, Q - Received BGP S-A Route, q - Sent BGP S-A Route, V - RD & Vector, v
- Vector, p - PIM Joins on route Outgoing interface flags: H - Hardware switched, A - Assert
winner, p - PIM Join Timers: Uptime/Expires Interface state: Interface, Next-Hop or VCD,
State/Mode (*, 239.0.0.1), 02:57:14/stopped, RP 10.1.1.1, flags: SJC Incoming interface:
Ethernet0/0, RPF nbr 172.16.1.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:28:58/00:02:50 (192.168.1.10, 239.0.0.1), 00:02:03/00:00:56, flags: JT Incoming interface:
Ethernet0/0, RPF nbr 172.16.1.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:02:03/00:02:50
```

The unicast and multicast topology is not currently congruent. This can or cannot be important. What happens when R3 fails?

```
R3(config)#int e0/2 R3(config-if)#sh R3(config-if)#
```

No replies to the pings come in until PIM on R2 detects that R3 is gone and takes over the DR role. This takes between 60-90 seconds with the default timers in use.

```
Sender#ping 239.0.0.1 re 100 ti 1 Type escape sequence to abort. Sending 100, 100-byte ICMP
Echos to 239.0.0.1, timeout is 1 seconds: Reply to request 0 from 10.0.0.10, 18 ms Reply to
request 1 from 10.0.0.10, 2
ms............................................................... ....... Reply to request
77 from 10.0.0.10, 10 ms Reply to request 78 from 10.0.0.10, 1 ms Reply to request 79 from
10.0.0.10, 1 ms Reply to request 80 from 10.0.0.10, 1 ms
```

You can increase the DR priority on R2 to make it become the DR.

```
R2(config-if)#ip pim dr-priority 50 *May 30 12:42:45.900: %PIM-5-DRCHG: DR change from neighbor
10.0.0.3 to 10.0.0.2 on interface Ethernet0/2
```

HSRP-aware PIM is a feature that makes the HSRP AR the PIM DR. It also sends the PIM messages from the virtual IP which is useful in situations where you have a router with a static route towards a Virtual IP (VIP). This is how Cisco describes the feature:

HSRP Aware PIM enables multicast traffic to be forwarded through the HSRP AR, allows PIM to leverage HSRP redundancy, avoids potential duplicate traffic, and enables failover, which depends on the HSRP states in the device. The PIM-DR runs on the same gateway as the HSRP AR and maintains mroute states.

In Topology 1, HSRP runs towards the clients, so even though this feature sounds like a perfect fit, it cannot help in the multicast convergence. Configure this feature on R2:

```
R2(config-if)#ip pim redundancy HSRP1 hsrp dr-priority 100 R2(config-if)# *May 30 12:48:20.024:
%PIM-5-DRCHG: DR change from neighbor 10.0.0.3 to 10.0.0.2 on interface Ethernet0/2
```

R2 is now the PIM DR, and R3 now sees two PIM neighbors on interface E0/2:

```
R3#sh ip pim nei e0/2 PIM Neighbor Table Mode: B - Bidir Capable, DR - Designated Router, N -
Default DR Priority, P - Proxy Capable, S - State Refresh Capable, G - GenID Capable Neighbor
Interface Uptime/Expires Ver DR Address Prio/Mode 10.0.0.1 Ethernet0/2 00:00:51/00:01:23 v2 0 /
S P G 10.0.0.2 Ethernet0/2 00:07:24/00:01:23 v2 100/ DR S P G
```

R2 now has the S,G and you can see that it was the Assert winner because R3 was previously the multicast forwarder to the LAN segment.

```
R2#sh ip mroute 239.0.0.1 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, E - Extranet, 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, G - Received BGP C-Mroute, g - Sent BGP C-Mroute, N - Received BGP Shared-Tree Prune, n -
BGP C-Mroute suppressed, Q - Received BGP S-A Route, q - Sent BGP S-A Route, V - RD & Vector, v
- Vector, p - PIM Joins on route Outgoing interface flags: H - Hardware switched, A - Assert
winner, p - PIM Join Timers: Uptime/Expires Interface state: Interface, Next-Hop or VCD,
State/Mode (*, 239.0.0.1), 00:20:31/stopped, RP 10.1.1.1, flags: SJC Incoming interface:
Ethernet0/0, RPF nbr 192.0.2.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:16:21/00:02:35 (192.168.1.10, 239.0.0.1), 00:00:19/00:02:40, flags: JT Incoming interface:
Ethernet0/0, RPF nbr 192.0.2.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:00:19/00:02:40, A
```

What happens when the R2s LAN interface goes down? Can R3 become the DR? And how fast can it converge?

```
R2(config)#int e0/2 R2(config-if)#sh
```

HSRP changes to active on R3 but the PIM DR role does not converge until the PIM query interval has expired (3x hellos).

```
*May 30 12:51:44.204: HSRP: Et0/2 Grp 1 Redundancy "hsrp-Et0/2-1" state Standby -> Active R3#sh
ip pim nei e0/2 PIM Neighbor Table Mode: B - Bidir Capable, DR - Designated Router, N - Default
DR Priority, P - Proxy Capable, S - State Refresh Capable, G - GenID Capable Neighbor Interface
Uptime/Expires Ver DR Address Prio/Mode 10.0.0.1 Ethernet0/2 00:04:05/00:00:36 v2 0 / S P G
10.0.0.2 Ethernet0/2 00:10:39/00:00:36 v2 100/ DR S P G R3# *May 30 12:53:02.013: %PIM-5-NBRCHG:
```

```
neighbor 10.0.0.2 DOWN on interface Ethernet0/2 DR *May 30 12:53:02.013: %PIM-5-DRCHG: DR change
from neighbor 10.0.0.2 to 10.0.0.3 on interface Ethernet0/2 *May 30 12:53:02.013: %PIM-5-NBRCHG:
neighbor 10.0.0.1 DOWN on interface Ethernet0/2 non DR
```

You lose a lot of packets while PIM convergence happens:

```
Sender#ping 239.0.0.1 re 100 time 1 Type escape sequence to abort. Sending 100, 100-byte ICMP
Echos to 239.0.0.1, timeout is 1 seconds: Reply to request 0 from 10.0.0.10, 5 ms Reply to
request 0 from 10.0.0.10, 14
ms.................................................................. Reply to request 68 from
10.0.0.10, 10 ms Reply to request 69 from 10.0.0.10, 2 ms Reply to request 70 from 10.0.0.10, 1
ms
```

HSRP is aware PIM did not really help here. It is useful if you use the Topology 2 instead:
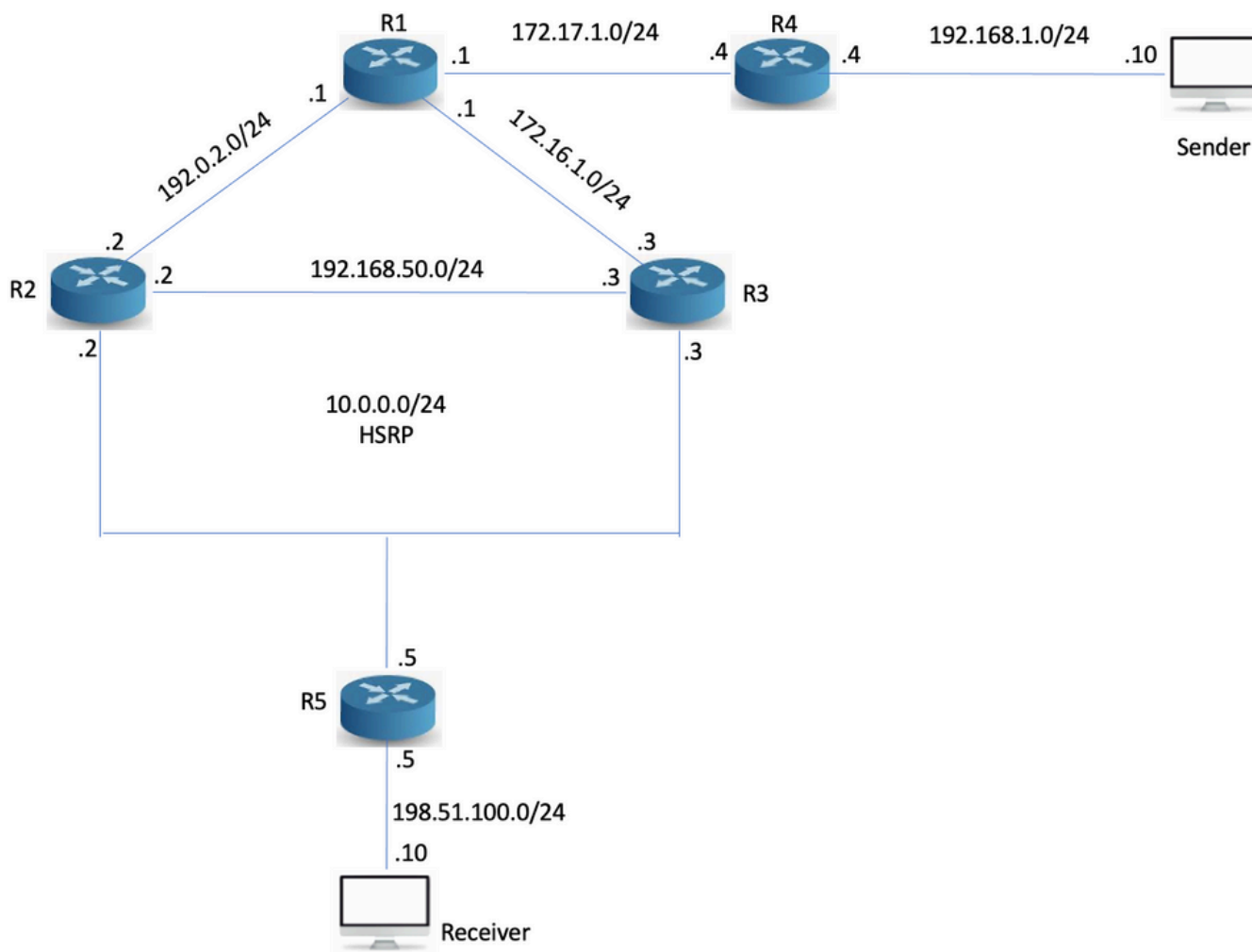
Topology 2

The router R5 has been added and the receiver sits behind R5 instead. R5 does not run routing
with R2 and R3, only with static route points at the RP and the multicast source:

```
R5(config)#ip route 10.1.1.1 255.255.255.255 10.0.0.1 R5(config)#ip route 192.168.1.0
255.255.255.0 10.0.0.1
```

Without HSRP aware PIM, the Reverse Path Forwarding (RPF) check fails because PIM peers
with the physical address, but R5 sees three neighbors on the segment, where one is the VIP:

```
R5#sh ip pim nei PIM Neighbor Table Mode: B - Bidir Capable, DR - Designated Router, N - Default
```

```
DR Priority, P - Proxy Capable, S - State Refresh Capable, G - GenID Capable Neighbor Interface
Uptime/Expires Ver DR Address Prio/Mode 10.0.0.2 Ethernet0/0 00:03:00/00:01:41 v2 100/ DR S P G
10.0.0.1 Ethernet0/0 00:03:00/00:01:41 v2 0 / S P G 10.0.0.3 Ethernet0/0 00:03:00/00:01:41 v2 1
/ S P G
```

R2 is the one that forwards multicast at the time of normal conditions since it is the PIM DR via HSRP state of the active router:

```
R2#sh ip mroute 239.0.0.1 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, E - Extranet, 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, G - Received BGP C-Mroute, g - Sent BGP C-Mroute, N - Received BGP Shared-Tree Prune, n -
BGP C-Mroute suppressed, Q - Received BGP S-A Route, q - Sent BGP S-A Route, V - RD & Vector, v
- Vector, p - PIM Joins on route Outgoing interface flags: H - Hardware switched, A - Assert
winner, p - PIM Join Timers: Uptime/Expires Interface state: Interface, Next-Hop or VCD,
State/Mode (*, 239.0.0.1), 00:02:12/00:02:39, RP 10.1.1.1, flags: S Incoming interface:
Ethernet0/0, RPF nbr 192.0.2.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:02:12/00:02:39
```

Try a ping from the source:

```
Sender#ping 239.0.0.1 re 3 Type escape sequence to abort. Sending 3, 100-byte ICMP Echos to
239.0.0.1, timeout is 2 seconds: Reply to request 0 from 198.51.100.10, 1 ms Reply to request 1
from 198.51.100.10, 2 ms Reply to request 2 from 198.51.100.10, 2 ms
```

The ping works and R2 has the S,G:

```
R2#sh ip mroute 239.0.0.1 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, E - Extranet, 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, G - Received BGP C-Mroute, g - Sent BGP C-Mroute, N - Received BGP Shared-Tree Prune, n -
BGP C-Mroute suppressed, Q - Received BGP S-A Route, q - Sent BGP S-A Route, V - RD & Vector, v
- Vector, p - PIM Joins on route Outgoing interface flags: H - Hardware switched, A - Assert
winner, p - PIM Join Timers: Uptime/Expires Interface state: Interface, Next-Hop or VCD,
State/Mode (*, 239.0.0.1), 00:04:18/00:03:29, RP 10.1.1.1, flags: S Incoming interface:
Ethernet0/0, RPF nbr 192.0.2.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:04:18/00:03:29 (192.168.1.10, 239.0.0.1), 00:01:35/00:01:24, flags: T Incoming interface:
Ethernet0/0, RPF nbr 192.0.2.1 Outgoing interface list: Ethernet0/2, Forward/Sparse,
00:01:35/00:03:29
```

What happens when R2 fails?

```
R2#conf t Enter configuration commands, one per line. End with CNTL/Z. R2(config)#int e0/2
R2(config-if)#sh R2(config-if)#
```

```
Sender#ping 239.0.0.1 re 200 ti 1 Type escape sequence to abort. Sending 200, 100-byte ICMP
Echos to 239.0.0.1, timeout is 1 seconds: Reply to request 0 from 198.51.100.10, 9 ms Reply to
request 1 from 198.51.100.10, 2 ms Reply to request 1 from 198.51.100.10, 11
ms............................................................
............................................................
.........................................................
```

The pings time out because when the PIM Join from R5 comes in, R3 does not realize that it must process the Join.

```
*May 30 13:20:13.236: PIM(0): Received v2 Join/Prune on Ethernet0/2 from 10.0.0.5, not to us
*May 30 13:20:32.183: PIM(0): Generation ID changed from neighbor 10.0.0.2
```

As it turns out, the PIM redundancy command must be configured on the secondary router as well, for it to process PIM Joins to the VIP.

```
R3(config-if)#ip pim redundancy HSRP1 hsrp dr-priority 10
```

After this has been configured, the incoming Join is processed. R3 triggers R5 to send a new Join because the GenID is set in the PIM hello to a new value.

```
*May 30 13:59:19.333: PIM(0): Matched redundancy group VIP 10.0.0.1 on Ethernet0/2 Active,
processing the Join/Prune, to us *May 30 13:40:34.043: PIM(0): Generation ID changed from
neighbor 10.0.0.1
```

After this configuration, the PIM DR role converges as fast as HSRP allows. The Bidirectional Forwarding Detection (BFD) is used in this scenario.

Conclusion

The key concept to understand the HSRP aware PIM here is that:

- Initially, the PIM redundancy configuration on the AR makes it the DR.
- PIM redundancy must be configured on the secondary router as well, otherwise, it cannot process PIM Joins to the VIP.
- The PIM DR role does not converge until PIM hellos have timed out. The secondary router processes the Joins, so the multicast converges.

Key Takeaway

This feature does not work when you have a receiver on an HSRP LAN, because the DR role is not moved until PIM adjacency expires.

Related Information

- https://www.cisco.com/en/US/docs/ios-xml/ios/ipmulti_pim/configuration/15-2s/imc_hsrp_aware.html#GUID-1294B212-466A-4D8D-AB20-D8DE0B3645CD
- Technical Support & Documentation - Cisco Systems