



PIM Dense Mode

Peter J. Welcher

Introduction

This article continues the series on IP Multicast. We'll take a look at how basic IP multicast works. We'll then look at how PIM Dense Mode (PIM-DM) operates, how to configure it, and how to troubleshoot it. This should warm us up before taking on the slightly more complex and more scalable PIM Sparse Mode in a later article.

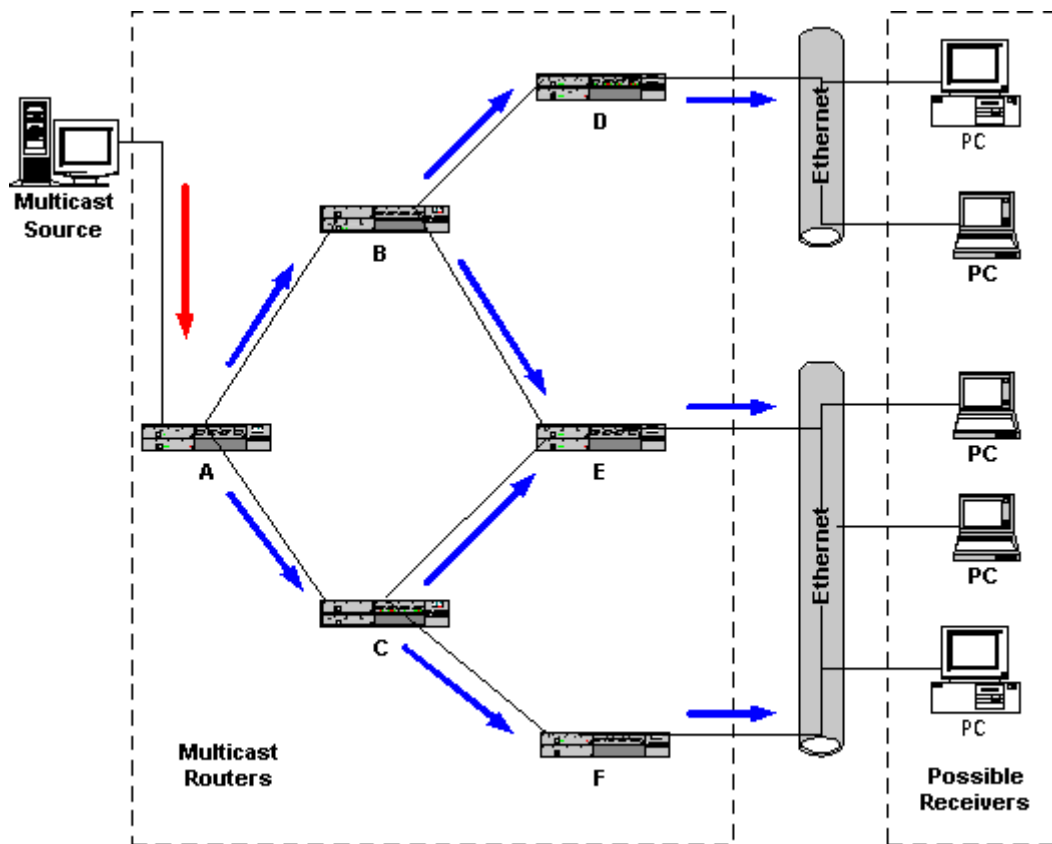
The initial multicast article contains links to the key IETF working groups at its end. It can be found at <http://www.netcraftsmen.net/welcher/papers/multicast01.html>.

Understanding IP Multicast Forwarding

The purpose of a multicast routing protocol is to allow routers to work together to efficiently deliver copies of multicast packets to interested receivers. In the process of doing this, the multicast routing protocol probably also provides a mechanism for neighbors to discover and track neighboring routers also using the multicast routing protocol.

As we saw in the last article, those computers interested in receiving a multicast packet stream use IGMP to notify adjacent router(s). The routers then use the multicast protocol to arrange for a copy of the multicast packet stream to be sent to them so they can forward it onto the LAN containing the receiver. We made no mention of what happens at the **source** end of the packet stream. In IP multicast, there is no protocol for the source to communicate or register or notify the routers. The source just starts sending IP multicast packets, and it is up to the neighboring router(s) to do the right thing. (What the "right thing" happens to be depends on the multicast routing protocol).

The following picture shows a source sending a multicast packet in red, and downstream routers duplicating the packet and flooding it to other routers and ultimately all LAN segments. As we'll see shortly, this is the initial (and periodic) behavior of Cisco's Protocol Independent Multicast (PIM) multicast routing protocol, when acting in Dense Mode.



Notice in the above picture that the blue arrows forward multicast packet copies in an organized way. There may be two copies of each packet coming into some of the routers or LAN segments. But the routers do not forward packets "backwards". This is a good thing: think about what might happen if router E were to forward a copy of the packet it received from C to router B. Would B then forward the packet to A, which might forward it to C, and so on, in a forwarding loop? This would be a sort of Layer 3 equivalent of what happens at Layer 2 when Spanning Tree is disabled in a loop topology: a good way to waste a lot of bandwidth and router capacity.

To prevent multicast forwarding loops, IP multicast always performs an RPF check, which we'll talk about shortly.

In addition to the RPF check, multicast routing protocols such as PIM may also work to prevent inefficiency. For example, router E in the picture does not need to receive two copies of each multicast packet (one from B, one from C).

The multicast routing protocol determines which interfaces to send copies out (or **not** send copies out). As the above picture suggests, multicast forwarding occurs along logical trees, branching paths through the network. All the multicast forwarding information is stored in the multicast state table, which some people call the multicast routing table. This information can be viewed with the very useful command, `show ip mroute`. Let's take a brief look at sample output from the `show ip mroute` command (with PIM Dense Mode running).

```
Router# show ip mroute 233.1.1.1
```

```
IP Multicast Routing Table
```

```

Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned
      R - RP-bit set, F - Register flag, T - SPT-bit set
Timers: Uptime/Expires
Interface state: Interface, Next-Hop, State/Mode

(*, 233.1.1.1), uptime 0:57:31, expires 0:02:59, RP is 0.0.0.0, flags: DC
  Incoming interface: Null, RPF neighbor 0.0.0.0
  Outgoing interface list:
    Ethernet0, Forward/Dense, 0:57:31/0:02:52
    FastEthernet1, Forward/Dense, 0:56:55/0:01:28
    FastEthernet2, Forward/Dense, 0:56:45/0:01:22

(172.16.16.1/32, 233.1.1.1), uptime 20:20:00, expires 0:02:55, flags: C
  Incoming interface: FastEthernet1, RPF neighbor 10.20.30.1
  Outgoing interface list:
    Ethernet0, Forward/Dense, 20:20:00/0:02:52

```

To understand this, note that addresses starting with 224-239 are IP multicast addresses or groups. (The group refers to the group of receivers for that multicast destination address).

The entry starting with (*, 233.1.1.1) is a shared multicast tree entry, sometimes referred to as a (*, G) entry. (G here is just 233.1.1.1). PIM-DM doesn't use these for packet forwarding, but does list interfaces with a role in multicast (known IGMP receiver or PIM neighbor) as outgoing interfaces under such entries.

The entry starting with (172.16.16.1, 233.1.1.1) is referred to as an (S, G) entry. S is source, G is group. If you prefer, think of this as source and destination in the IP header (since that's where they actually appear in the packets). This entry is a source-specific multicast tree for a particular multicast group. There will generally be one such entry for each source and group. Note that the unicast routing next hop shows up as the RPF neighbor, and the incoming interface is the RPF interface, the interface used by unicast routing towards the source 172.16.16.1. The outgoing interface list shows that any packet from 172.16.16.1 with destination 233.1.1.1 received on FastEthernet1 will be copied out Ethernet0.

You can draw the multicast forwarding tree for a particular (S, G) for troubleshooting purposes. To do this, run the `show ip mroute` command on each router. Take a copy of your network map and draw an outbound arrow for each interface in the outgoing interface list ("OIL" or "OILIST"). You'll end up with a diagram somewhat like the above picture.

State flags you might see in PIM-DM `show ip mroute` output:

- 1 D = Dense Mode. Appears on (*, G) entries only. Group is operating in Dense mode.
- 1 C = Directly Connected Host (IGMP!)
- 1 L = Local (Router is configured to be a member of the multicast group).
- 1 P = Pruned (All OILIST interfaces set to Prune). The router generally send Prune to its RPF neighbor when this occurs.
- 1 T = Forwarding via Shortest Path Tree (SPT), indicates at least one packet received / forwarded.
- 1 J = Join SPT. Always on in (*,G) entry in PIM-DM, doesn't mean much.

What is the RPF Check, and Why?

IP multicast forwarding always performs an RPF check. RPF stands for Reverse Path Forwarding. The goal of the RPF check is to try to prevent a multicast packet forwarding loop in a simple way. For each multicast stream, the multicast router checks the source address, what device sent the multicast. It then looks the sender up in the **unicast** routing table, and determines the interface it would use to send unicast packets to the multicast source. That interface is the RPF interface, the one on which the router "expects" to receive multicasts. Think of it as the "officially approved" interface for receiving multicasts from that particular source. The router stores the RPF interface as part of the multicast state information for that particular source and that particular multicast destination (group).

When a multicast packet is received by the router, the router tracks which interface the packet came in. If the packet is the first packet from a new source, the RPF interface is determined and stored in the mroute table, as just discussed. Otherwise, the router looks up the source and multicast group in the mroute table. If the packet was received on the RPF interface, the packet is copied and forwarded on each outgoing interface listed in the mroute table. If the packet was received on a non-RPF interface, it is discarded. If the router were a person, this would be the equivalent of "What's that person sending me this for? They must be confused, I'll ignore what they just sent me."

The router also does not send a copy of a packet back out the interface it came in (the RPF interface). In other words, even if the neighboring router on the RPF interface somehow were to request to be sent copies of a multicast stream, the router will not add the RPF interface to the list of outbound interfaces which receive copies of the multicast stream. This protects against any sort of protocol error getting two neighbors into a tight multicast forwarding loop.

Router E is also probably ignoring one of the two packet streams, based on whether B or C is connected to the RPF interface. Even with equal cost routes, normally just one interface is chosen as the RPF interface. So if you have two links connecting two routers, only one will typically be used for multicast from any one source subnet.

See however the command `ip multicast multipath` (new in Cisco IOS Version 12.0(8) T, 12.0(5) S). With this command, used properly, there can be load balancing for different sources for a particular multicast group. Since this is per-source and not per-stream, it really becomes more like load splitting. It may not end up being very balanced.

Load balancing traffic for one packet stream (one source and group) over two links between two routers can also be done using GRE tunnels between loopback interfaces, although I might worry about performance implications of doing this with a high bandwidth flow.

By the way, this also tells us how to direct IP multicast over links of our choosing. We control where multicast traffic goes by controlling the RPF check. If a router doesn't learn routes back to a multicast source on some interface, then that interface will not be the RPF interface. So with distance vector protocols and "route starvation" (not advertising a route to a neighbor), we can steer or direct multicast.

We'll see later that any PIM grafts or joins are sent out the RPF interface, back towards the source (or RP, in PIM Sparse Mode). By controlling where this activity takes place, we control which links the multicast is forwarded on. Static mroute entries (static routes back to the source for multicast RPF check purposes) are another way of directing multicast traffic. If / when we talk about Multicast MBGP (Multiprotocol BGP for Multicast), we'll see that MBGP also provides us with a way to direct or control the links used by multicast traffic. Also note that if the RPF interface does not have IP multicast enabled, then in effect the router is expecting packets where it will not receive them. The RPF interface should always be one where IP multicast is enabled.

PIM Dense Mode -- Overview

Protocol Independent Multicast has two modes, Dense Mode and Sparse Mode. This article only has space to cover the former. We'll get to PIM-SM in the next article.

PIM Dense Mode (PIM-DM) uses a fairly simple approach to handle IP multicast routing. The basic assumption behind PIM-DM is that the multicast packet stream has receivers at most locations. An example of this might be a company presentation by the CEO or President of a company. By way of contrast, PIM Sparse Mode (PIM-SM) assumes relatively fewer receivers. An example would be the initial orientation video for new employees.

This difference shows up in the initial behavior and mechanisms of the two protocols. PIM-SM only sends multicasts when requested to do so. Whereas PIM-DM starts by flooding the multicast traffic, and then stopping it each link where it is not needed, using a Prune message. I think of the Prune message as one router telling another "we don't need that multicast over here right now".

In older Cisco IOS releases, PIM-DM would re-flood all the multicast traffic every 3 minutes. This is fine for low volume multicast, but not higher bandwidth multicast packet streams. More recent Cisco IOS versions support a new feature called PIM Dense Mode State Refresh, since 12.1(5)T. This feature uses a PIM state refresh messages to refresh the Prune state on outgoing interfaces. Another benefit is that topology changes are recognized more quickly. By default, the PIM state refresh messages are sent every 60 seconds.

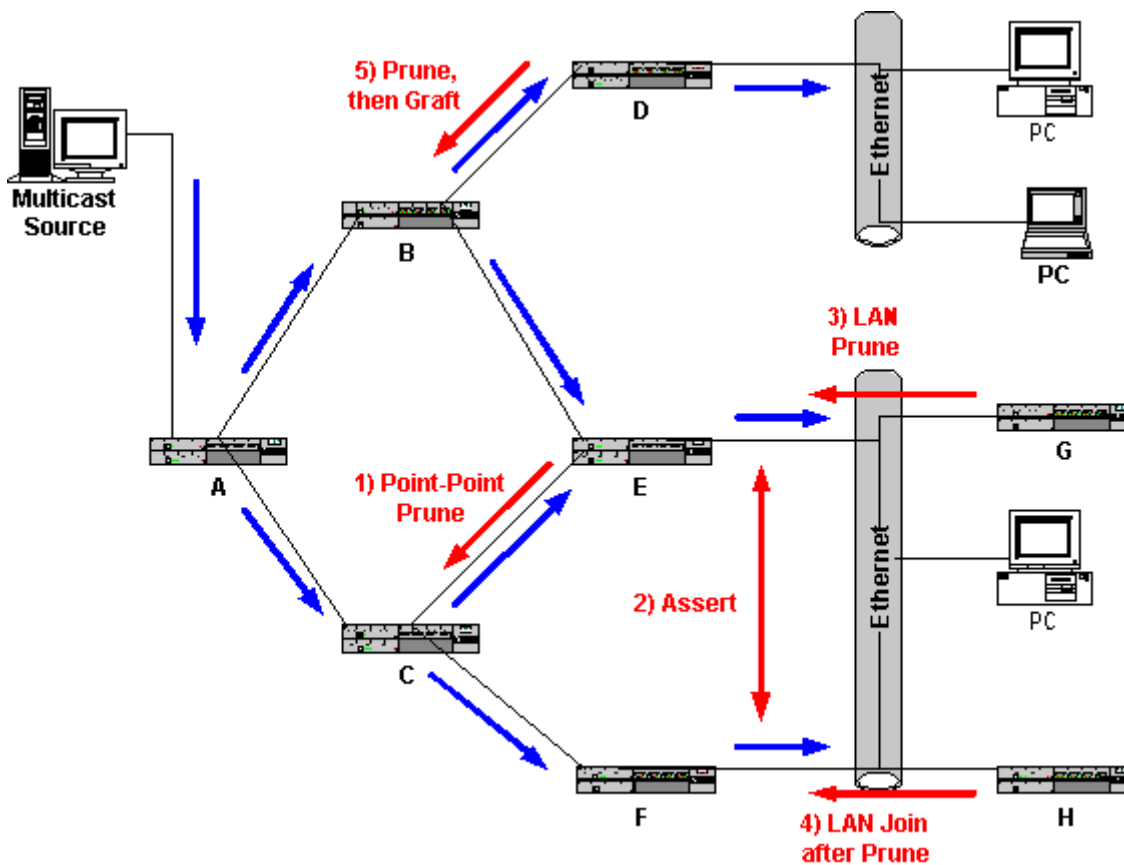
Consider routers E and F in the above picture. When two PIM-DM routers connect to a LAN, they will see the multicast packets from each other. One should forward packets to the LAN, and the other not. They both send Assert messages. Best routing metric wins, with higher IP address as a tie-breaker. If they are using different routing protocols, a weighted routing metric scheme, somewhat like administrative distance, settles which router is to be the Forwarder (forwarding the multicast packets onto the LAN). The Forwarder may be silenced by a Prune from a downstream router with no receivers, if there are also no receivers on the LAN segment. Downstream routers may have to adjust their RPF neighbor, to reflect the winner of the Assert process.

To repeat that in full detail: when multicast traffic is received on a non-RPF interface, a Prune message is sent, provided the interface is point-to-point. These Prune messages are rate-limited, to make sure the volume of them (potentially, one per multicast received) doesn't cause further problems.

If the non-RPF interface is a LAN, an Assert message is sent. Non-Forwarder routers then send a Prune on their RPF interface if they don't need the multicast stream. Only one such Prune is sent, at the time of the transition to having no interfaces in the Outgoing Interface List (OILIST). The LAN Prune receiver delays acting on it for 3 seconds, so that if another LAN router still needs the multicast stream, it can send a PIM Join message to counteract (cancel) the Prune. ("Yo, that router doesn't need it, but I still do!")

Suppose a router has Pruned, and some time later a receiver requests the multicast stream with an IGMP message. The router then sends a Graft message. In effect, "hey, I need that multicast stream over here now".

The following picture illustrates this, admittedly in somewhat compressed form.



Explanation of the picture:

(1), perhaps the router E chooses B as its RPF neighbor, based on unicast routing back to the source. Then E receives a multicast packet on the point-to-point interface from C. It sends a rate-limited Prune to C.

(2), the routers E and F on the LAN exchange Assert packets, when E or F sees the multicast forwarded by the other of the two. Suppose E wins, based on unicast routing metric or address. Then F knows not to forward multicasts on the LAN. Note that G and H are not involved, since the Ethernet is their RPF interface.

(3), suppose router G has no receivers downstream. It can then send a LAN Prune to the Forwarder for the LAN, router E.

(4), if router H has local or downstream receiver(s), it counters this with a LAN Join.

(5), suppose router D had no downstream or local receivers and sent a Prune to B. Suppose sometime later one of the PC's to its right sends it an IGMP message for the same multicast group. Router D can then send a PIM Graft to B, asking B to resume sending it the specified multicast group.

PIM Protocol and Packet Types

PIM is a full routing protocol, with various kinds of messages. I'm not about to drone on and on about the different messages. But I do think it's worth taking a brief look, since it tells us a little about the

protocol.

PIM uses Hello messages to discover neighbors and form adjacencies. The Hello is sent to the All-PIM-Routers local multicast address, 224.0.0.13, every 30 seconds (PIMv1 uses AllRouters, 224.0.0.2). Each LAN has a PIM Designated Router (DR), used in PIM Sparse Mode. It is **also** the IGMPv1 Querier: the highest IP address on the LAN. The `show ip pim neighbor` command shows neighbors and adjacency and timer information.

Clarification added 12/11/2004: IGMP querier and Designated Router are the two roles in question. See http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/ip_c/ipcprt3/1cdmulti.htm

In IGMP version 1,"The DR is responsible for the following tasks:

- 1 Sending PIM register and PIM join and prune messages toward the RP to inform it about host group membership.**
- 1 Sending IGMP host-query messages."**

In IGMP version 2, they are decoupled. The IGMP querier is elected by lowest IP on the LAN. PIM selects the DR by highest IP on the LAN, to forward multicasts to the LAN. (This offloads work).

PIM also has a Join/Prune message, used as described above. There are also Graft and Graft ACK messages, which tells us that Graft is done reliably (unlike the real world?).

And there is the PIM Assert message.

PIM-SM has 3 more message types: Register, Register-Stop, and RP-Reachability (not in PIMv2).

Configuring PIM Dense Mode

Configuring PIM-DM is downright easy compared to all the above.

Globally, enable multicast routing with the command:

```
ip multicast-routing
```

Then on each interface you wish to participate in multicasting, enable IP multicast and PIM with the interface command:

```
interface ...
 ip pim dense-mode
```

Actually, it is better practice to configure sparse-dense mode:

```
interface ...
 ip pim sparse-dense-mode
```

The reason is that this allows you to simply migrate some or all multicast groups to Sparse Mode, by letting the router know about a Rendezvous Point. And you can even do this without reconfiguring a lot of routers or router interfaces.

You can configure stub networks for simple IP multicast. The idea is to **not** run PIM to stub parts of the network, like small remote sites, for simplicity. This is a particularly good idea with routers that are not under your control: you don't want them sharing multicast routing with your routers. It also eliminates PIM-DM flooding to such routers (with older Cisco IOS releases).

To configure stub multicast, configure

```
ip igmp helper-address a.b.c.d
```

on any stub router LAN interfaces with potential multicast receivers. The address *a.b.c.d* is the address of the central PIM-speaking router. On the central router, you configure a filter to tune out any PIM messages the stub neighbor might send, with:

```
ip pim neighbor-filter access-list
```

See also the Command Guide at the URL:

http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122cgcr/fipr_c/ipept3/1cfmulti.htm .

There are several show commands to help troubleshoot IP multicast. The ones I like best:

```
1 show ip mroute
1 show ip pim interface
1 show ip pim neighbor
1 show ip rpf
```

Since space is tight, I'll refer you to the Command Reference for examples. See the following URL:

http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122cgcr/fiprmc_r/mult/1rfmult3.htm .

One troubleshooting note: if you do have high bandwidth multicast in your network, make sure you use the Prune State Refresh in the newer Cisco IOS releases if you're using sparse-dense mode. I worry about routers "forgetting" they have a RP and reverting to Dense Mode, with periodic flooding. This might be a rare accidental occurrence, but it could really ruin your morning! You may also wish to consider RP robustness techniques as well, a topic for our later PIM-SM or Rendezvous Point article.

In Conclusion

The recommended book for a lot of multicast topics: *Developing IP Multicast Networks: The Definitive Guide to Designing and Deploying CISCO IP Multicast Networks*, Beau Williamson. See also <http://www.amazon.com/exec/obidos/ASIN/1578700779/> .

Another extremely good Cisco multicast link: <http://www.cisco.com/warp/public/732/Tech/multicast> .

Next article: scaling with PIM Sparse Mode.

Dr. Peter J. Welcher (CCIE #1773, CCSI #94014) is a Senior Consultant with Chesapeake NetCraftsmen. NetCraftsmen is a high-end consulting firm and Cisco Premier Partner dedicated to quality consulting and knowledge transfer. NetCraftsmen has nine CCIE's, with expertise including large network high-availability routing/switching and design, VoIP, QoS, MPLS, network management, security, IP multicast, and other areas. See <http://www.netcraftsmen.net> for more information about NetCraftsmen. Pete's links start at <http://www.netcraftsmen.net/welcher> . New articles will be posted under the Articles link. Questions, suggestions for articles, etc. can be sent to pjw@netcraftsmen.net .

10/1/2001, updated 12/17/2004

Copyright (C) 2001, Peter J. Welcher