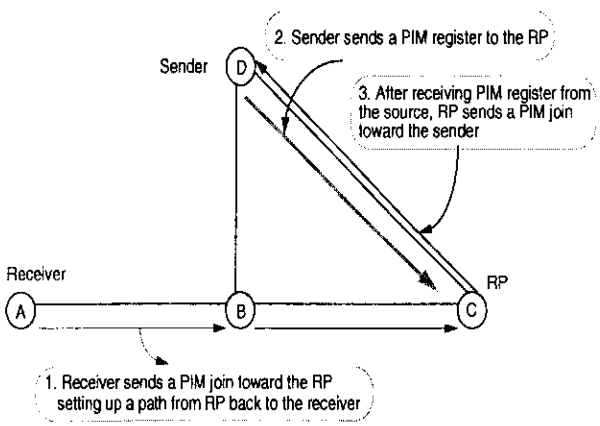

Figure 3: How senders rendezvous with receivers

packets are sent until routers without local (or downstream) members send explicit prune messages to remove themselves from the distribution tree.

(b) whereas dense mode IP multicast tree construction is all data driven, PIM must use per-group *Rendezvous Point(s)* for receivers to "meet" new sources. Rendezvous Points (RP) are used by senders to announce their existence and by receivers to learn about new senders of a group. Source-specific trees in PIM are in fact data driven, however the RP-tree is receiver-join driven in anticipation of data.

The shortest path tree state maintained in routers is of the same order as the forwarding information that is currently maintained by routers running existing IP multicast protocols such as MOSPF, i.e., source (S), multicast address (G), outgoing interface set (oif), incoming interface (iif). We refer to this forwarding information as the *multicast forwarding entry* for **(S,G)**. The oif's and iif's of (S, G) entries in all routers together form a shortest path tree rooted at S.

An entry for a shared tree can match packets from any source for its associated group if the packets come through the right incoming interface, we denote such an entry (*,G). A (*,G) entry keeps the same information an (S,G) entry keeps, except that it saves the RP address in place of the source address. There is an RP-flag indicating that this is a shared tree entry.

Figure 3 shows a simple scenario of a receiver and a sender joining a multicast group via an RP. When the receiver signals that it wants to join a PIM multicast group (i.e., by sending an IGMP message[8]), its first hop PIM router (*A* in fig 3) sends a PIM-Join message toward one of the RPs advertised for the group. (We explain how routers identify the appropriate RP for a group in later sections.) Processing of this message by intermediate routers sets up the multicast tree branch from the RP to the receiver. When sources start sending to the multicast group, the first hop PIM-router (D in fig 3) sends a PIM-Register message, piggybacked on the data packet, to the RP(s) for that group. The RP responds by sending a join toward the source. Processing of these messages by intermediate routers (there are no intermediate routers between the RP and the source in fig 3) sets up a packet delivery path from the source to the RP(s).

If source-specific distribution trees are desired, the first hop PIM router for each member eventually joins the source-rooted distribution tree for each source by sending a PIM-Join message towards the source. After data packets are received on the new path, router B in fig 3 sends a PIM-prune message toward the RP B knows, by checking the incoming interface in its routing table, that it is at a point where the shortest path tree and the RP tree branches diverge. A flag, called SPT bit, is included in (S, G) entries to indicate whether the transition from shared tree to shortest path tree has finished. This makes a smooth transition, e.g. there is no loss of data packets.

One or more Rendezvous Points (RPs) are used *initially* to propagate data packets from sources to receivers. An RP may be any PIM-speaking router that is close to one of the members of the group, or it may be some other PIM-speaking router in the network. A sparse mode group, i.e., one that the receiver's directly connected PIM router will join using PIM, is identified by the presence of RP address(es) associated with the group in question. The mapping information may be configured or may be learned through another protocol mechanism (e.g., a new IGMP message used by hosts to distribute information about RPs to their local routers).

PIM avoids explicit enumeration of receivers, but does require enumeration of sources. If there are very large numbers of sources sending to a group but the sources' average data rates are low, then one possibility is to support the group with a shared tree instead which has less per-source overhead. If shortest path trees are desired then when the number of sources grows very large, some form of aggregation or proxy mechanism will be needed; see section 4. We selected this tradeoff because in many existing and anticipated applications, the number of receivers is much larger than the number of sources. And when the number of sources is very large, the average data rate tends to be lower (e.g. resource discovery).

The remainder of this section describes the protocol design in more detail.

3.1 Local hosts joining a group

A host sends an IGMP-Report message identifying a particular group, G, in response to a directly-connected router's IGMP-Query message, as shown in figurev 4. From this point on we refer to such a host as a receiver, R, (or member) of the group G.

When a *designated router* (DR) receives a report for a new group G it checks to see if it has RP address(es) associated with G. The mechanism for learning this mapping of G to RP(s) is somewhat orthogonal to the specification of this protocol; however, we require some mechanism in order for the protocol to work. At the very least this information