1 Routing Multicast

- Overview
- Different kinds of multicast routing
- Multicast within an Autonomous System (AS)
- Multicast between ASs
- Protocol overview
- Protocol Independent Multicast - Sparse Mode (PIM-SM)
- Multiprotocol Border Gateway Protocol/Multicast Source Discovery Protocol (MBGP/MSDP)

2 Overview

For multicasting to work, the routers need some basic information:

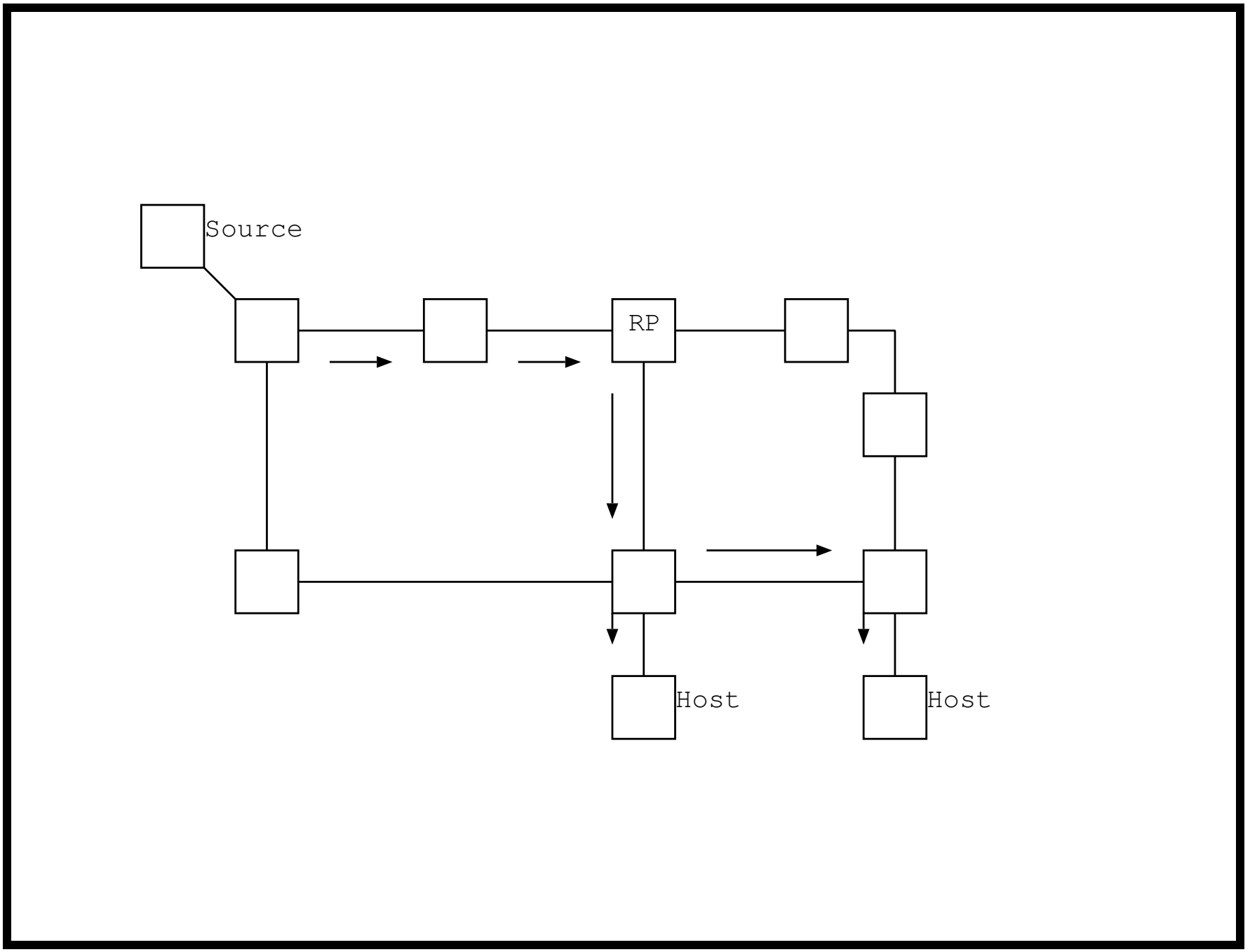
- Where is the source of a multicast group X?
- When a multicast packet from a certain source comes in on an interface, which interface(s) should the packet be forwarded to (where are the listeners)

3 Multicast within an AS

- Just as in IGP and EGP, there are different multicast protocols for routing within and between ASs.

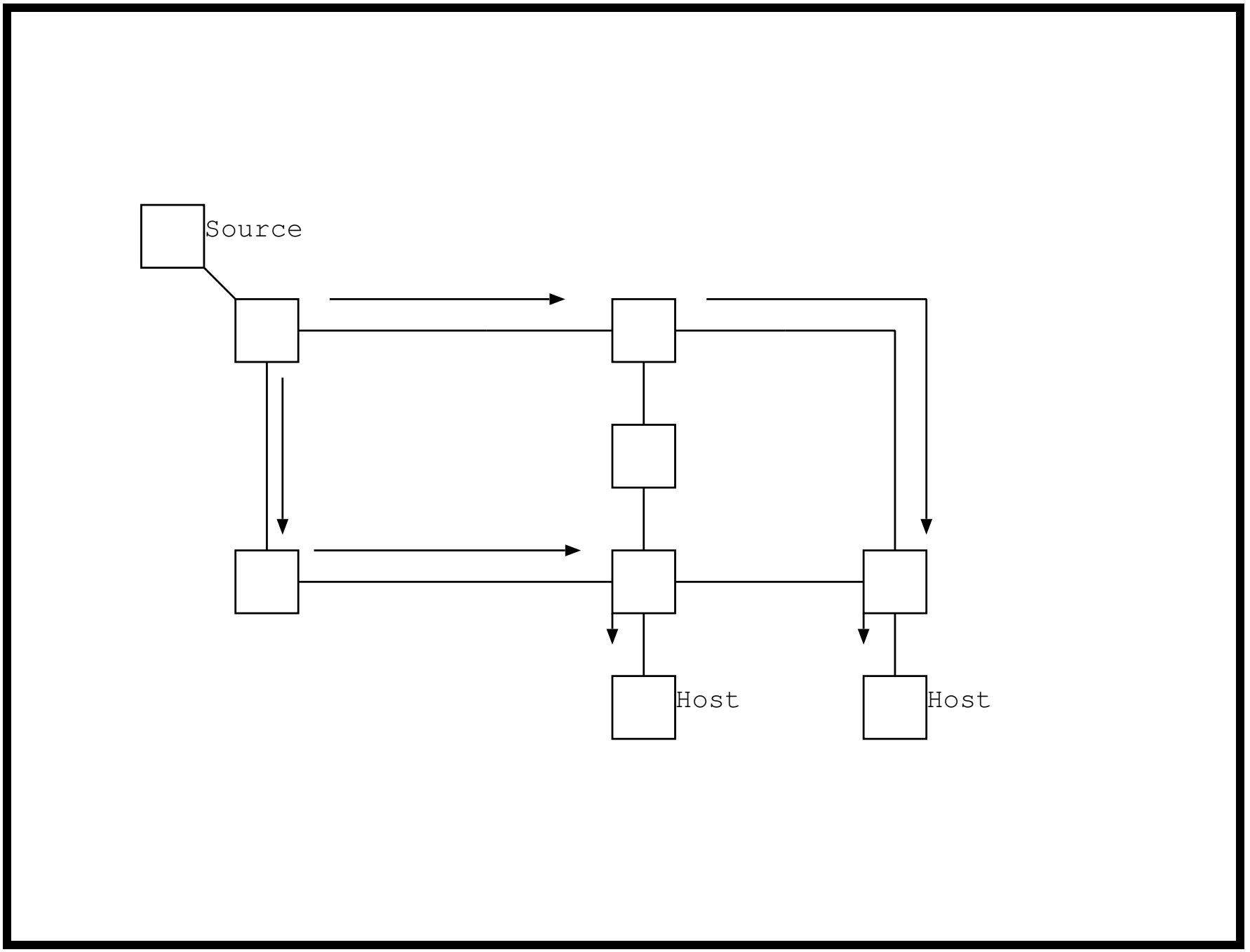
4 Shared tree

- The term **shared** in this context means to try to gather the listeners before sending a *join* to the source to have the tree built towards the source.
- The source always sends packets to a rendezvouz point (RP) in the net which then spreads the information to the nodes where there are listeners.
- If someone wants to listen to a multicast group, it tries to register itself to an RP.


Source
RP
Host
Host

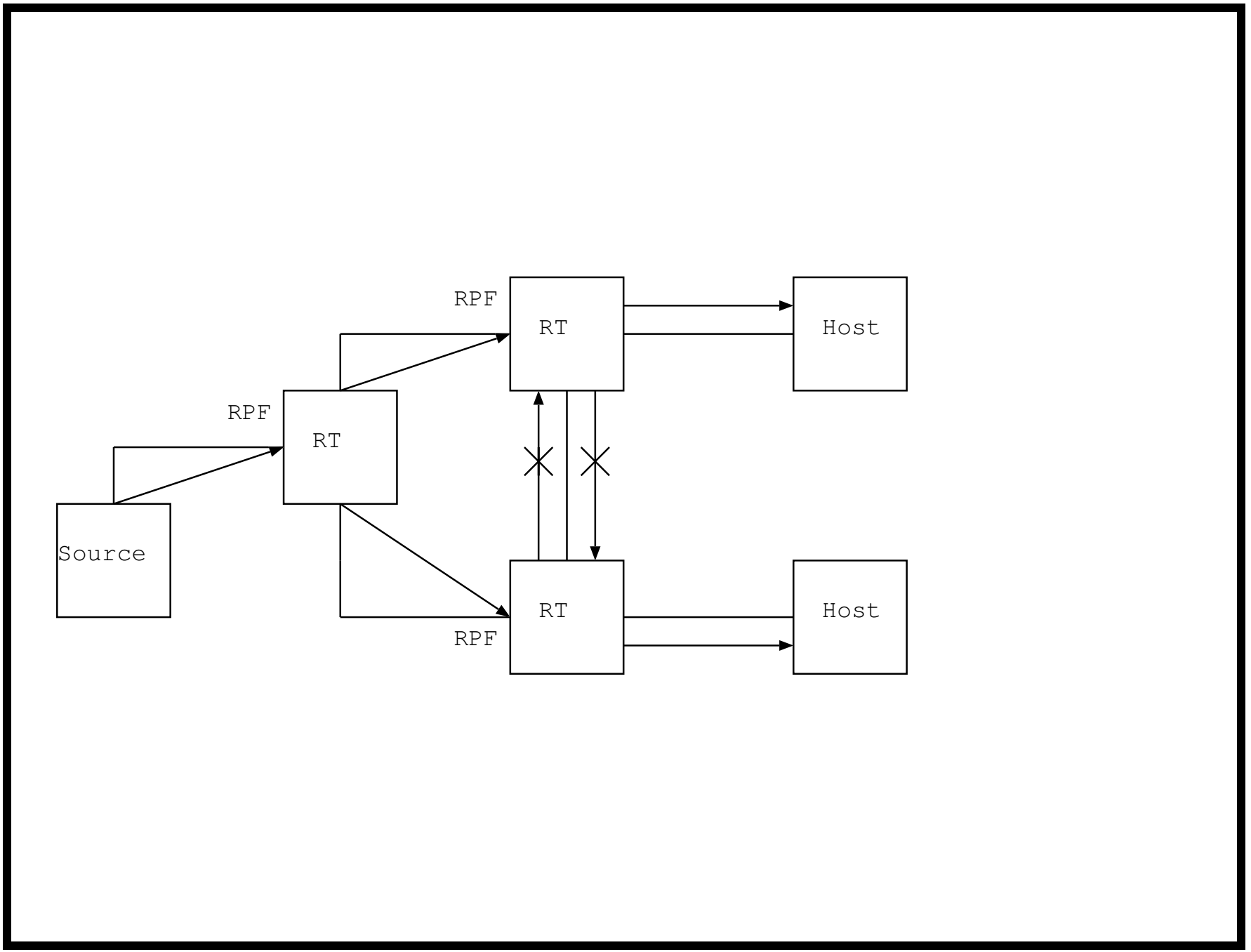
5 Source based tree

- The term **source based** means that one chooses the shortest path to the source.
- If someone wants to listen to a multicast group, the routers sends this information to the source directly.
- No RP.
- Also known as the shortest path tree (SPT)


Source
Host
Host

6 Reverse Path Forwarding - RPF

- What is RPF?
  The packet is sent on if it came in on an RPF-interface, and a reply packet would been sent through the same interface. Ie the source address of the packet is looked at.
- Why is RPF needed?
  So that the system stays loop-free.
- Which interface becomes the RPF-interface?
  When building a tree, the RPF-interfaces are set. (Some protocols use other techniques)


RPF
RT
Host
RPF
RT
Source
RT
Host
RPF

7 Different types of Multicast

Dense mode

- RPF - Reverse Path Forwardig
- Source based Tree
- Builds the tree in advance (but not PIM-DM)
- Prune/Graft
- Protocols:
  - DVMRP (Distance Vector Multicast Routing Protocol)
  - MOSPF (Multicast OSPF)
  - PIM-DM (Protocol Independant Multicast - Dense mode)

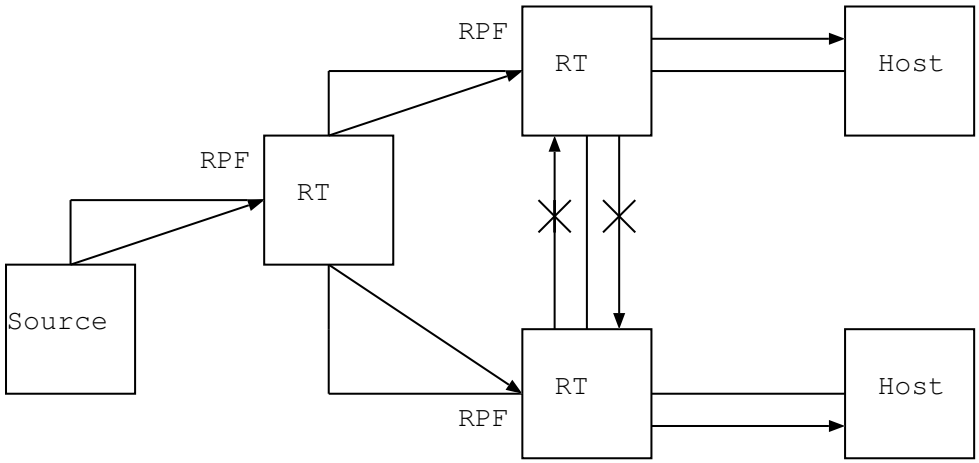
7.1 Different types of Multicast

Sparse mode

- RPF - Reverse Path Forwarding
- RP - Rendezvous Point
- Core based tree
- Builds the tree when it is needed
- Source based tree
- Protocols:
  - PIM-SM (Protocol Independant Multicast - Sparse mode)

8 Dense mode multicast

- Useful if many listeners are located near each other.
- Assumes that all hosts want to listen to the multicast group.



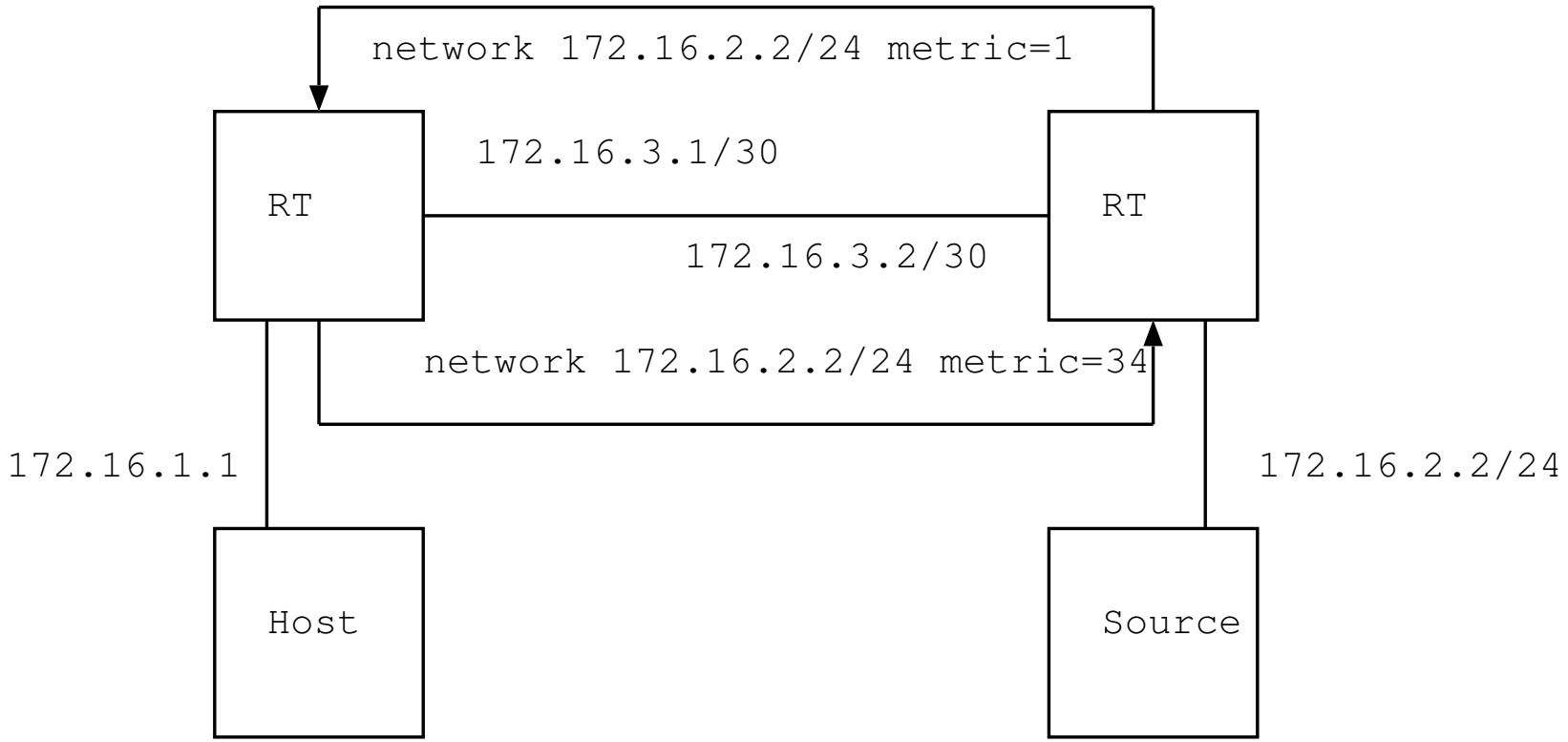
8.1 DVMRP

- DVMRP v1 has never been used (RFC 1075)
- DVMRP v2 is an old draft
- DVMRP v3 is the newest draft
- Similar to RIP
  Same problems as RIP (count to infinity ....)
  Uses poisonous reverse
  Builds a source tree in advance, and floods the multicast group
  Afterwards sends Graft (join) and Prune (leave)

8.2 DVMRP (cont'd)

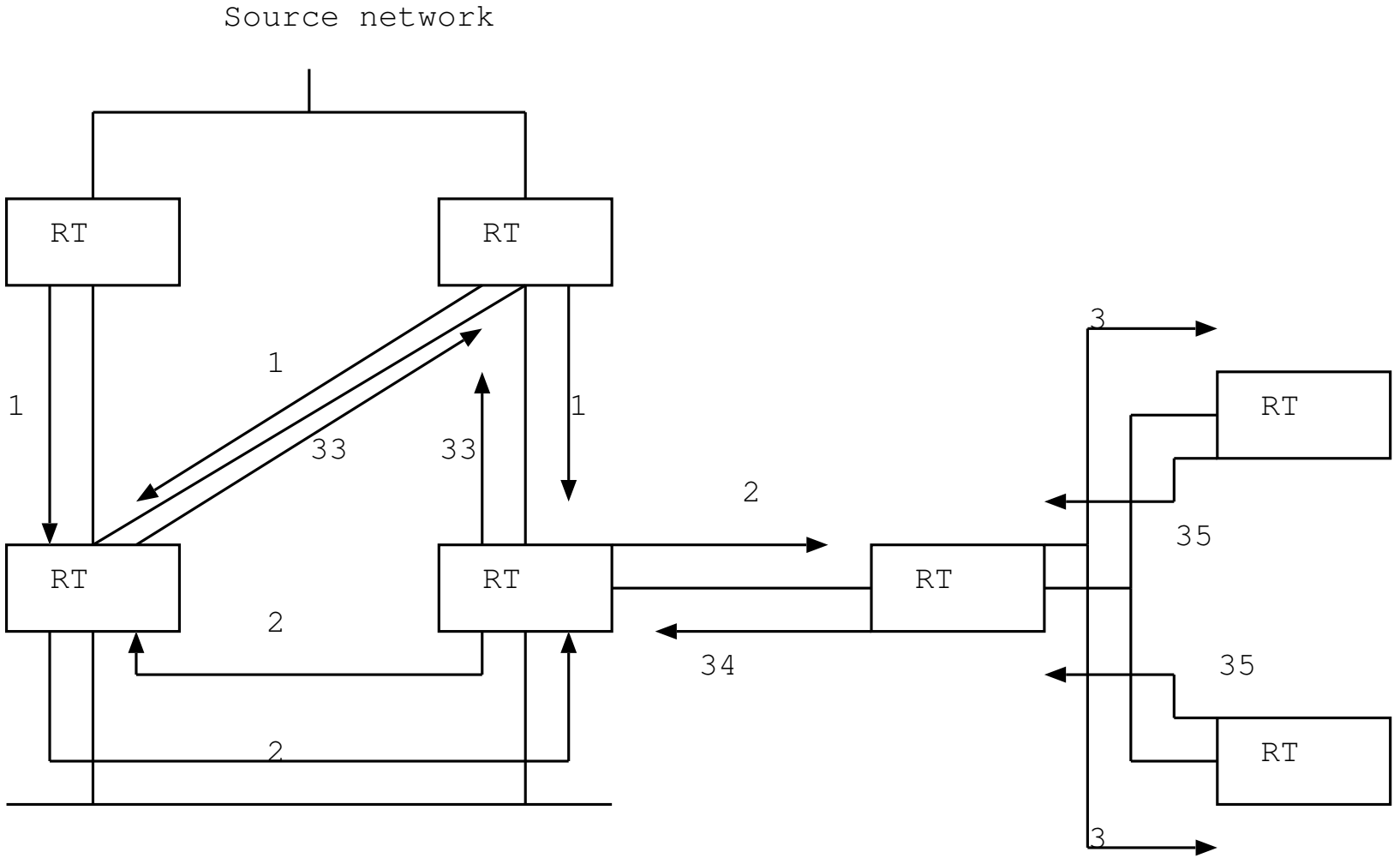
- Neighbor discovery
- multicast every tenth second
- After 35 seconds of silence, a neighbour is considered down.
- Route exchange - Same rules as RIP, but with an important exception
  By allowing values larger than infinity (32), one can see if routers are "upstream" towards the source. This helps when building a tree (one can keep track of RPF interfaces)

8.3 DVMRP (cont'd)

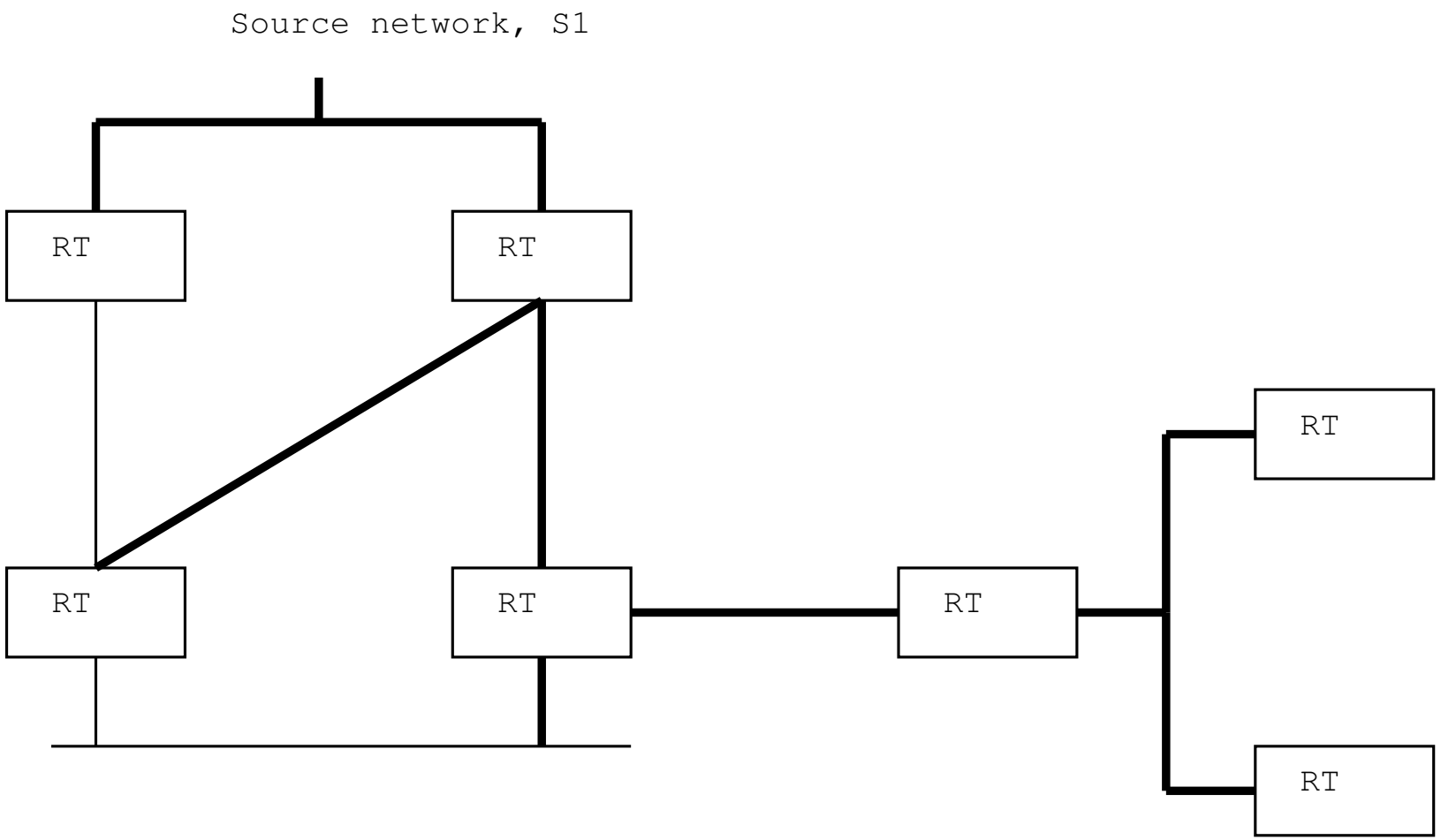


8.4 DVMRP (cont'd)

How to build a source based tree with DVMRP



8.5 DVMRP (cont'd)



8.6 DVMRP - Flooding, Pruning and Grafting

Flood

If a multicast group source begins to send packets, flood the packets everywhere in the tree, since it is assumed that everyone wants them.

Prune

Remove this leaf/branch from the multicast group tree.

Graft

Add this leaf/branch to the multicast group tree.

8.7 MOSPF

- Neighbour discovery is used here as usual (multicast every tenth second). A neighbour is assumed to be down after four missing HELLOs.
- Builds a tree based on the OSPF unicast table. Does not need to send separate "tree building messages".
- RPF, Flooding, Pruning and Graft are used in the same way as in DVMRP, albeit slightly more complicated if one uses areas.

8.8 PIM-DM

- Neighbour discovery is used here as usual (multicast every tenth second). A neighbour is assumed to be down after four missing HELLOs.
- Instead of building a tree, the Unicast table is used to choose the RPF-interface. PIM-DM supports OSPF, RIP, EIGRP, ...
- Floods a multicast group on all non-RPF interfaces. Receiving routers send a Prune back to the sending interface if they do not want this multicast group.

9 Sparse mode multicast

- Uses an RP (Rendezvous Point)
- Uses both source based and shared trees.
- PIM-SM v2 (Protocol Independent Multicast - Sparse Mode) is the only protocol used. (There is a version 1, but its very different and uses IGMP)

9.1 PIM-SM v1

- Never became an RFC.
- Used IGMP all the way, even for report routers to the shared tree and to RP.
- Not used.

9.2 PIM-SM v2

- RFC 2362
- IP protocol #103
- The only IGP multicast which is actually used today.

9.3 PIM-SM v2 (cont'd)

- Finds neighbours by periodically sending out HELLOs over multicast (224.0.0.13, ALL-PIM-ROUTERS)
- When a host reports interest for a group to a router, the router sends a join-message up toward the RP. Every router on the way toward the RP will process the packet and add a (*, G) state. If the packet hits a shared tree on the way, it stops there.
- When a source begins to multicast, its closest router sends a registration message to the RP with the multicast packet encapsuled in a unicast packet.

9.4 PIM-SM v2 (cont'd)

- When the RP gets a registration message, it sends a join message towards the source. The RP unpacks the multicast packet and sends it downstream.
- The source's first router continually encapsules the packets and sends them to the RP.
- When the tree between the RP and the source is built, multicast flows through the net naturally.
- When an interested host's first router recieves its first multicast packet, it can choose to connect directly to the source and send a PRUNE to the RP.

9.5 RP location

How do the routers know where the Rendezvous Points are located?

- It could be statically configured
- Or it could be dynamic:
  - Auto-RP (PIM-SM v1)
  - Bootstrap Routers (BSR)

9.6 Auto-RP

- Learns the location of the RP automatically.
- Multicast is used to distribute information
  Announce `224.0.1.39`
  Discovery `224.0.1.40`

9.7 BSR

- BSR election
- Only one router may be a BSR.
- If a BSR fails, another election is started.
- All c-rp's *(Candidate RP)* send information to the BSR

10 Multicast between ASs

Why don't we use PIM everywhere across the whole world?
This is for the same reasons why we don't use an IGP over the whole world.
The protocols used to multicast between ASs are:

- MSDP (*Multicast Source discovery Protocol*)
  Used so that the RP can find out whether there are sources in another AS.
- MBGP (*Multi Proticol Border Gateway Protocol*)
  Makes it possible do a *join* to the source.
- BGMP *Border Gateway Multicast Protocol*
  Not used yet

10.1 MSDP (Multicast Source Discovery Protocol)

- draft-ietf-msdp-spec-08.txt
- Uses MBGP to spread routes between ASs.
- A mechanism to tie together several PIM-SM domains
- Each PIM-SM domain uses its own independent RP(s) and does not have to depend on RPs in other.

10.2 MSDP (cont'd)

- TCP to port 639, connection between RPs in different domains.
- Spreads control information, for example:
  A PIM DR router which is directly connected to the source sends data to the RP encapsulated in a registration message. The RP then creates a "source active" message (SA) and sends it to its MSDP peers.
- Every MSDP peer floods the SA messages to its peers, in something called *peer RPF flooding*.
- Uses regular PIM-SM (MBGP) mechanisms to build the tree toward to source.
- The RP continues to send SA messages periodically to its peers as long as the source is sending.

10.3 Source Active (SA) filtering

- Only an RP in the same domain as the source may filter away SA messages.
- MBGP policy is used to control access directives (who may send to which group)

10.4 Anycast RP

- Used for redundancy for the RP
- All RP's have the same IP-address
- All RP's are in a MSDP full mesh