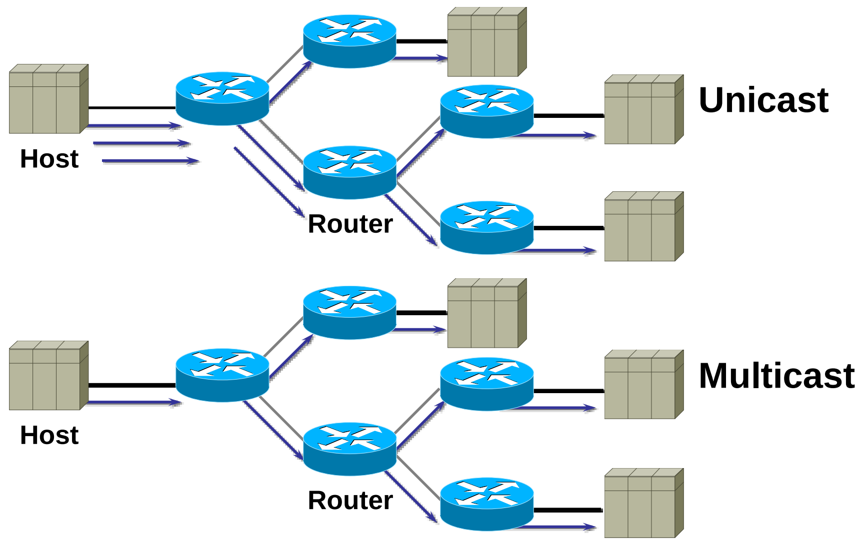
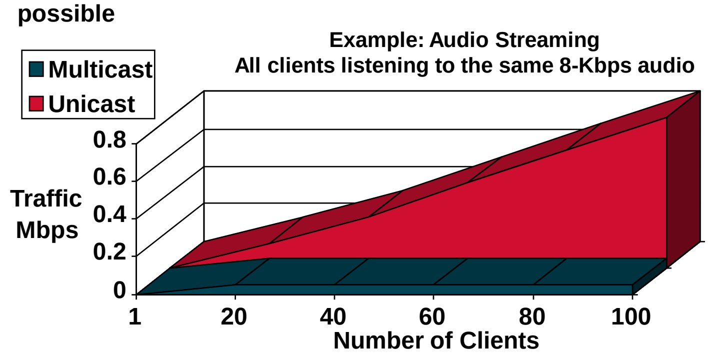
Unicast versus Multicast



Multicast Advantages

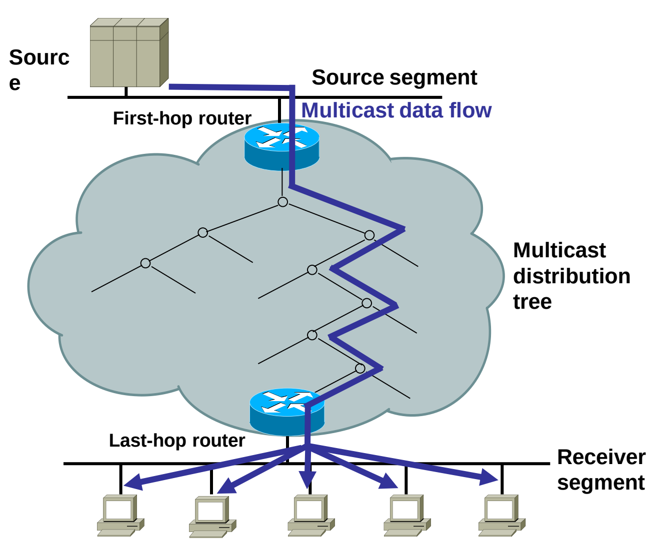
- **Enhanced Efficiency**: Controls network traffic and reduces server and CPU loads
- **Optimized Performance**: Eliminates traffic redundancy
- **Distributed Applications**: Makes multipoint applications possible



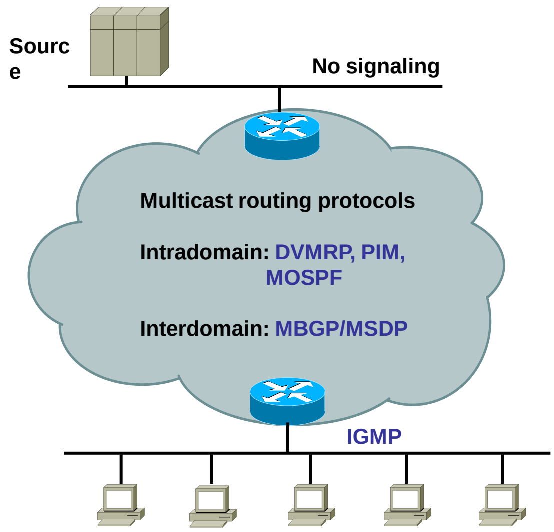
IP Multicast Service Model

- RFC 1112 “Host Extensions for Multicast Support”
- Each multicast group is identified by an IP address (224.0.0.0/4)
- Members join and leave the group and indicate this to the routers
- Routers listen to all multicast addresses and use multicast routing protocols to manage groups

Multicast Conceptual Model



Multicast Protocols (cont.)



IP Multicast Basic Addressing

- IP group addresses
  - High-order 4 bits are set as 1110
  - Range from 224.0.0.0 through 239.255.255.255
- Well-known addresses assigned by IANA
  - Reserved use: 224.0.0.0 through 224.0.1.255
    - 224.0.0.1 - all multicast systems on subnet
    - 224.0.0.2 - all routers on subnet
    - 224.0.0.4 - all DVMRP routers
    - 224.0.0.13 - all PIMv2 routers
    - 224.0.0.5, 224.0.0.6 - OSPF
    - 224.0.0.9 – RIPv2

IP Multicast
Basic Addressing (cont.)

- Transient addresses, assigned and reclaimed dynamically (within applications)
  - Global range: 224.0.2.0-238.255.255.255
    - 224.2.x.x usually used in Mbone applications
  - Limited (local) scope: 239.0.0.0/8 - “private IP multicast addresses” – RFC-2365
    - Site-local scope: 239.253.0.0/16
    - Organization-local scope: 239.192.0.0/14
- Part of a global scope recently used for new protocols and temporary usage

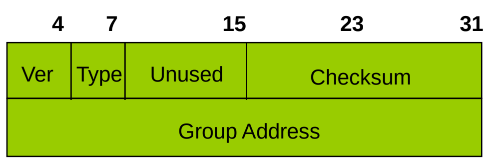
IGMP

- Internet Group Management Protocol
  - The way hosts tell routers about group membership
  - Routers solicit group membership from directly connected hosts
- RFC 1112 specifies the first version of IGMP
- RFC 2236 specifies the current version of IGMP
- IGMP v3 enhancements
- Supported on UNIX systems, PCs, and MACs

IGMPv1

- **RFC 1112, *Host extensions for IP Multicasting***
  - **Membership Queries**
    - **Querier sends IGMP query messages to 224.0.0.1 with TTL=1**
    - **One router on LAN is designated/elected to send queries**
    - **Query interval 60 to 120 seconds**
  - **Membership Reports**
    - **IGMP report sent by one host suppresses sending by others**
    - **Restrict to one report per group per LAN**
    - **Unsolicited reports sent by host when it first joins the group**

IGMPv1—Packet Format

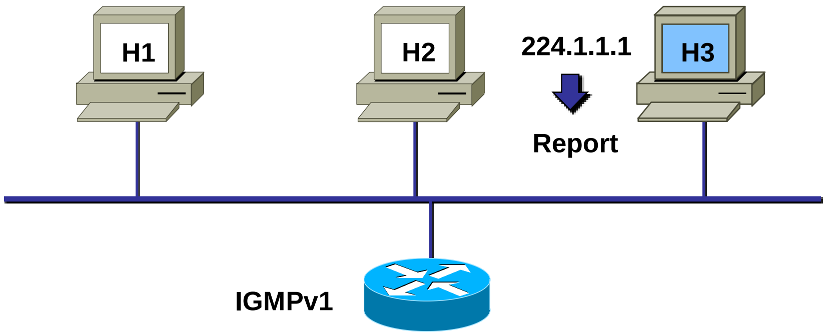


Ver:
Code Version = 1

Type:
1 = Host Membership Query
2 = Host Membership Report

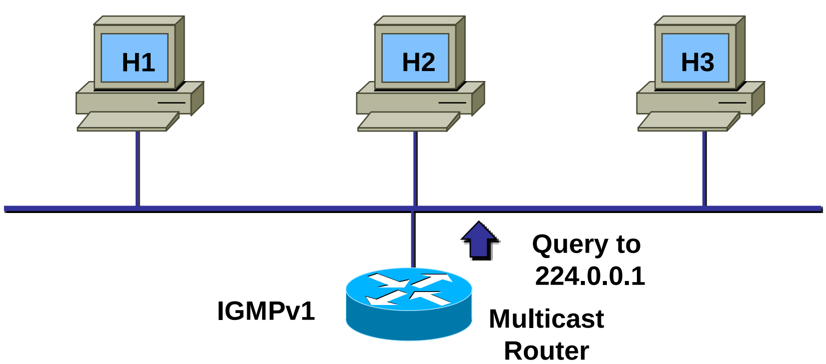
Group Address:
Multicast Group Address

IGMPv1—Joining a Group



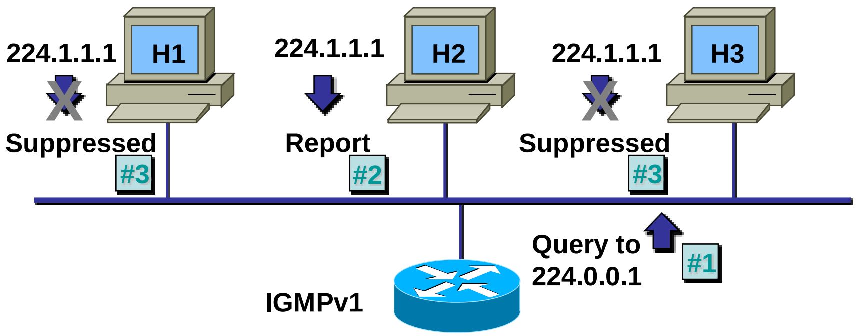
- Joining member sends report to 224.1.1.1 immediately upon joining

IGMPv1—General Queries



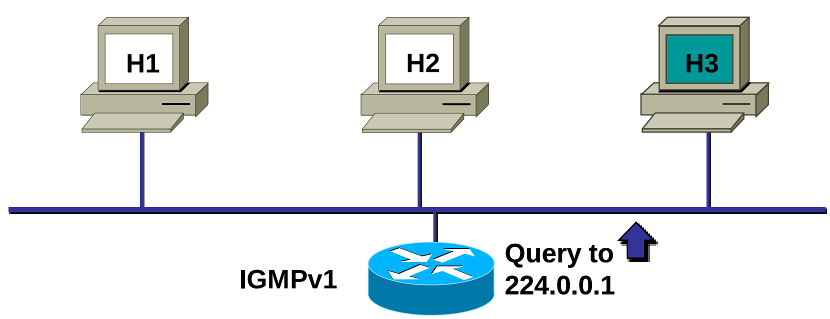
- Periodically sends General Queries to 224.0.0.1 to determine memberships

IGMPv1—Maintaining a Group



- #1 **Router sends periodic queries**
- #2 **One member per group per subnet report**
- #3 **Other members suppress reports**

IGMPv1—Leaving a Group



- **Router sends periodic queries**
- **Hosts silently leave group**
- **Router continues sending periodic queries**
- **No reports for group received by router**
- **Group times out**

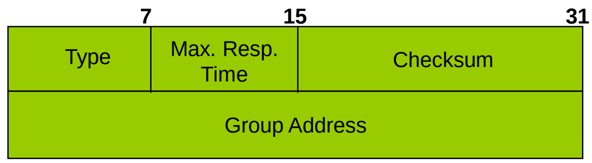
IGMPv2

- RFC 2236
  - Group-specific query
    - Router sends Group-Specific Query to make sure there are no members present before stopping to forward data for the group for that subnet
  - Leave Group message
    - Host sends Leave message if it leaves the group and is the last member (reduces leave latency in comparison to v1)

IGMPv2 (cont.)

- Querier election mechanism
  - On multiaccess networks, an IGMP querier router is elected based on the lowest IP address. Only the querier router sends queries.
- Query-interval response time
  - General Queries specify “Max. Response Time,” which inform’s hosts of the maximum time within which a host must respond to a General Query. (Improves burstiness of the responses.)
- Backward compatible with IGMPv1

IGMPv2—Packet Format

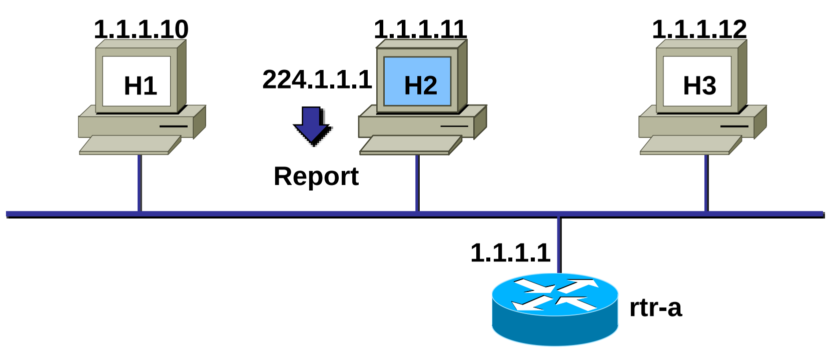


Type:
- **0x11 = Membership Query**
- **0x12 = Version 1 Membership Report**
- **0x16 = Version 2 Membership Report**
- **0x17 = Leave Group**

Max. Response Time
- **max. time before sending a responding report in 1/10 secs. (Default = 10 secs)**

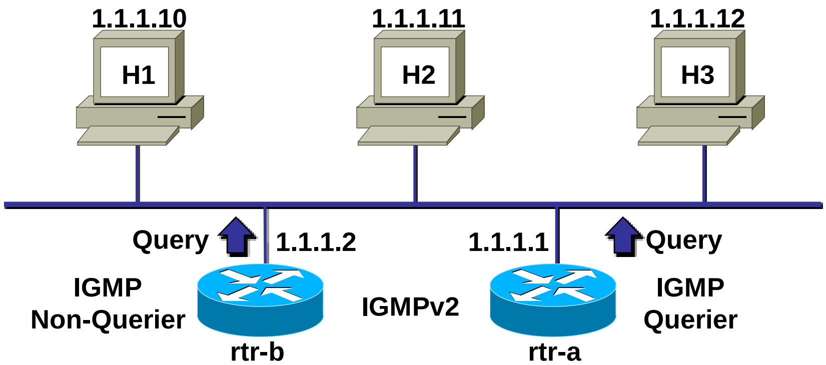
Group Address:
- **Multicast Group Address (0.0.0.0 for General Queries)**

IGMPv2—Joining a Group



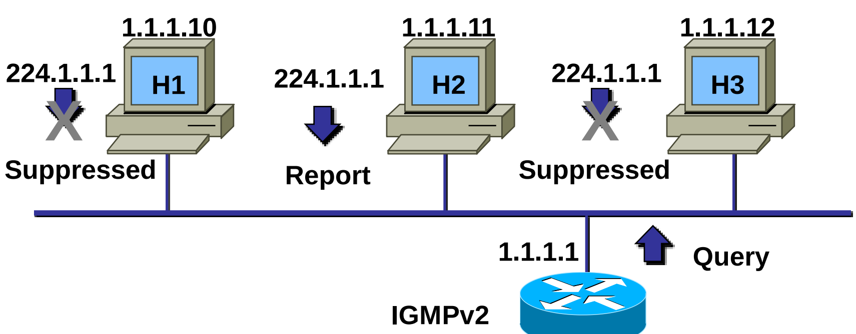
- Joining member sends report to 224.1.1.1 immediately upon joining (same as IGMPv1)

IGMPv2—Querier Election



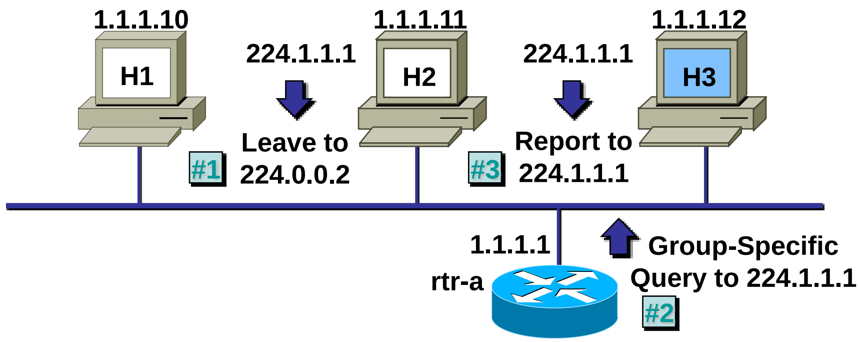
- **Initially, all routers send out a query**
- **Router with the lowest IP address is elected querier**
- **Other routers become non-queriers**

IGMPv2—Maintaining a Group



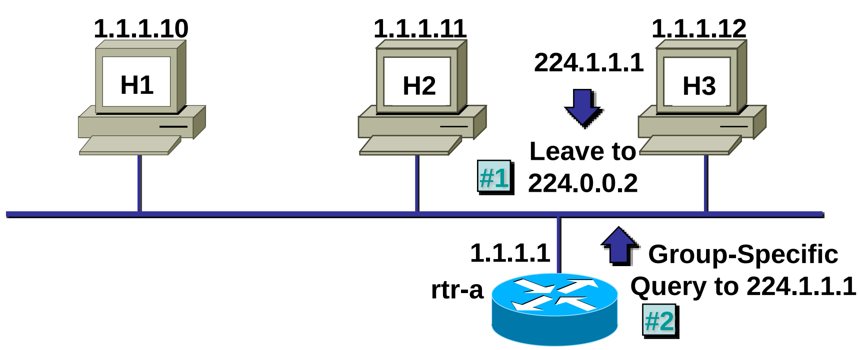
- Router sends periodic queries
- **One member per group per subnet reports**
- **Other members suppress reports**

IGMPv2—Leaving a Group (cont.)



- H2 leaves group; sends Leave message
- **Router sends Group-Specific Query**
- **A remaining member host sends report**
- **Group remains active**

IGMPv2—Leaving a Group (cont.)

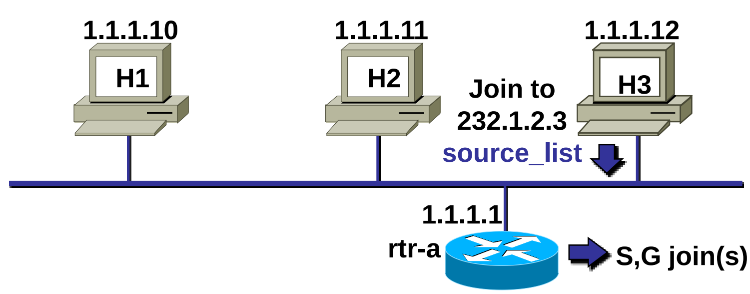


- **Last host leaves group; sends Leave message**
- **Router sends Group-Specific Query**
- **No report is received**
- **Group times out**

IGMPv3

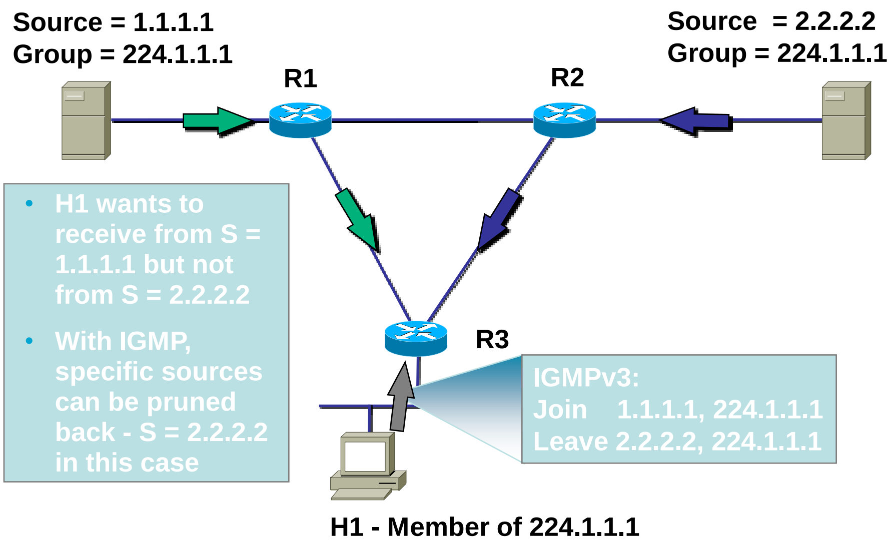
- Enables hosts to listen only to a specified subset of the hosts sending to the group
- Allows routers in sparse mode to build source distribution trees directly (avoiding RPs entirely)

IGMPv3 (cont.)



- **Host sends IGMPv3 join for group, which can specify a list of sources to be explicitly included.**
- **Router adds membership**
- **Router send (S,G) join directly to sources in the source_list, and is not required to send (*,G) join to RP (it must not be in the address range 232/8)**

IGMPv3 (cont.)

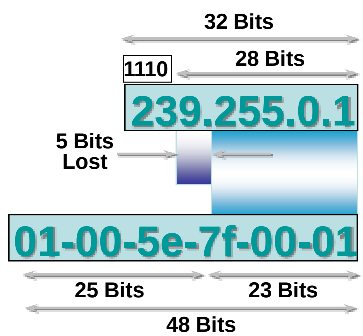


IGMPv3 – Who Needs it?

– Multicast receivers – hosts
– Last hop routers with directly attached receivers
– LAN switches doing IGMP snooping
– DSL aggregation point or other IGMP proxies passing on IGMP reports

Layer 2 Multicast Addressing

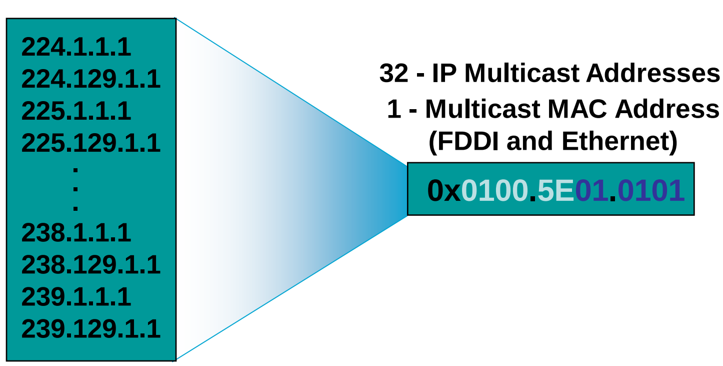
- IP Multicast MAC Address Mapping (WLAN and Ethernet)



Layer 2 Multicast Addressing (cont.)

- **IP Multicast MAC Address Mapping (WLAN and Ethernet)**

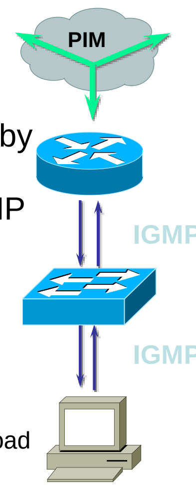
Be Aware of the 32:1 Address Overlap



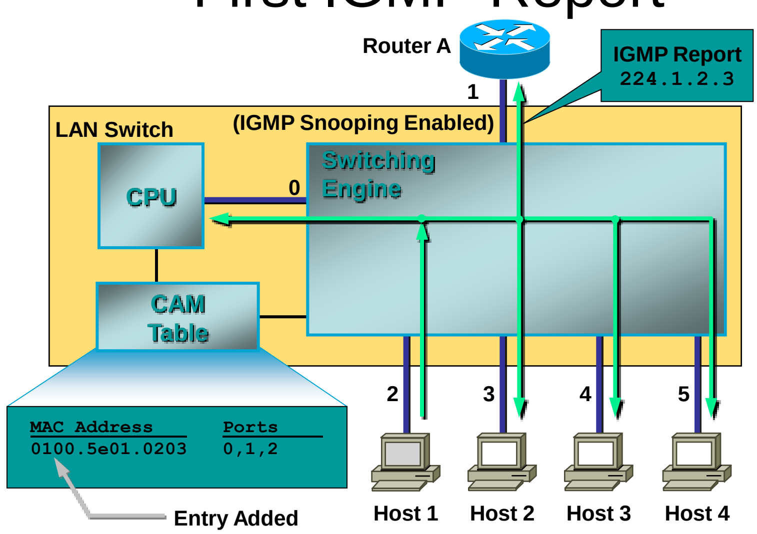
L2 Multicast Frame Switching IGMP Snooping

Solution : IGMP Snooping

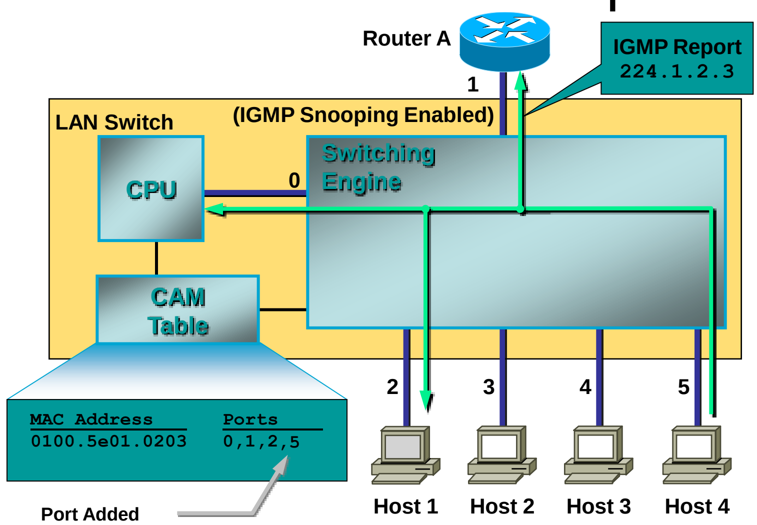
- Switches become “IGMP” aware
- IGMP packets intercepted by the NMP or by special hardware ASICs
- The switch must examine contents of IGMP messages to determine which ports want what traffic
  - IGMP membership reports
  - IGMP leave messages
- Effect on switch:
  - Must process all Layer 2 multicast packets
  - Administration load increases with multicast traffic load
  - Requires special hardware to maintain throughput



Typical L2 Switch First IGMP Report



Typical L2 Switch Second IGMP Report



Multicast Forwarding

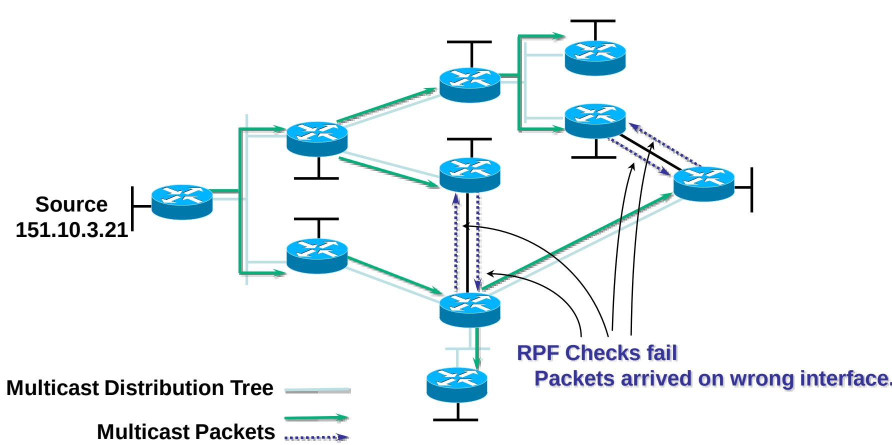
- Multicast routing works the opposite way of unicast routing
  - Unicast routing is concerned with where the packet is going
  - Multicast routing is concerned with where the packet comes from
- Multicast routing uses Reverse Path Forwarding to prevent forwarding loops

Reverse Path Forwarding (RPF)

- What is RPF?
  - A router forwards a multicast datagram only if received on the upstream interface to the source, i.e. it follows the distribution tree
- The RPF Check
  - The routing table for unicast is checked against the source address in the multicast datagram
  - If the datagram arrived on the interface specified in the routing table for the source address:
    - The RPF check succeeds
    - Otherwise, the RPF check fails

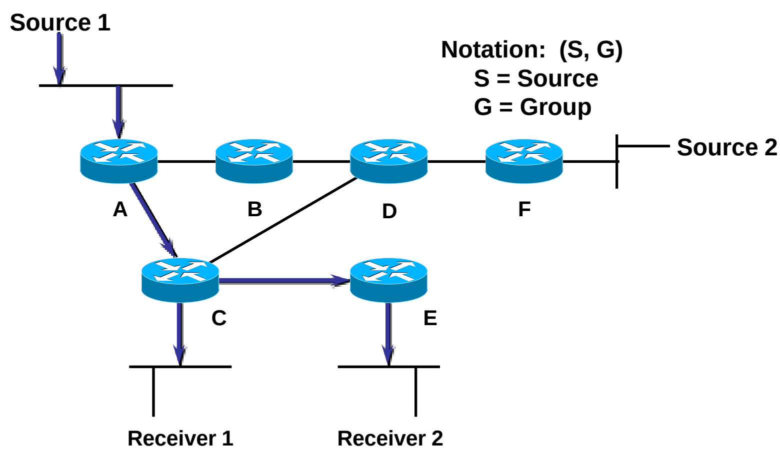
RPF Checking

- **Example: RPF Checking**



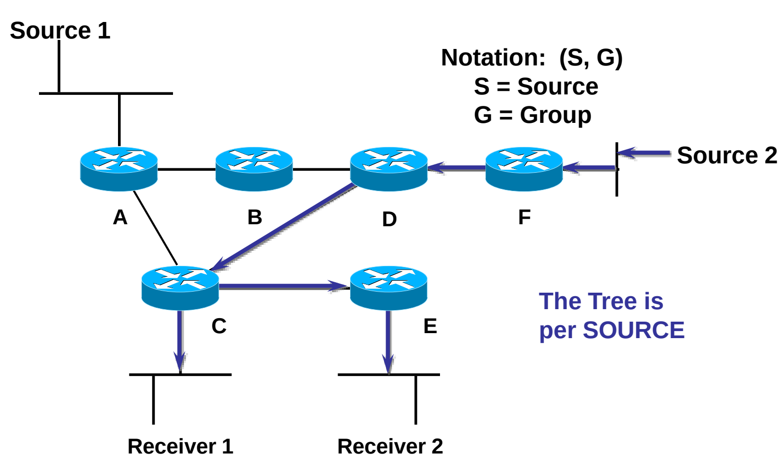
Shortest-Path Trees

- **Shortest-Path or Source Distribution Tree**



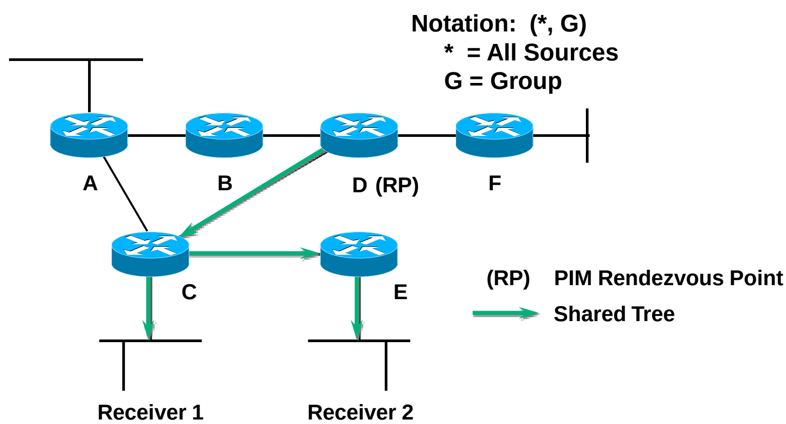
Shortest-Path Trees (cont.)

- **Shortest-Path or Source Distribution Tree**



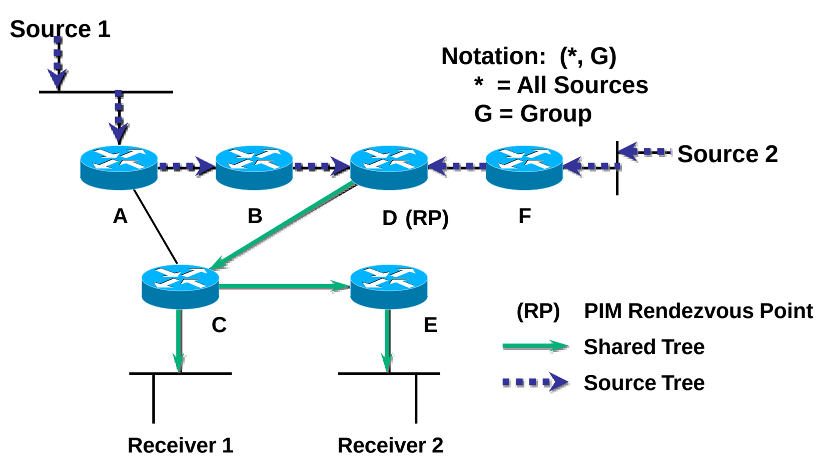
Shared Distribution Trees

- **Shared Distribution Tree**



Shared Distribution Trees (cont.)

- **Shared Distribution Tree**



Multicast Distribution Trees Identification

- (S,G) entries
  - For this particular Source sending to this particular Group
  - Traffic is forwarded via the shortest path from the Source
- (*,G) entries
  - For any (*) source sending to this Group
  - Traffic is forwarded via a meeting point for this Group

Dense Mode Protocols

- The push model is implemented
  - Referred to as flood and prune
  - Initial traffic flooded to all the branches of the distribution tree
  - Branches without receivers get pruned (for a limited time only)

Sparse Mode Protocols

- The pull model is implemented
  - An explicit join model
  - Last-hop routers “pull” the traffic from the meeting point or from the source
  - Branches without receivers never get the traffic

Multicast Protocol Review

- Intradomain multicast routing protocols:
  - PIM (Protocol Independent Multicast)
    - Sparse Mode (RFC 2362) Proposed Standard
    - SSM (Source Specific Mode)
    - Dense Mode (Internet-draft)
  - DVMRP (Distance Vector Multicast Routing Protocol) v2, v3 (Internet-Draft); v1 (RFC 1075) is obsolete
  - MOSPF (Multicast extensions to OSPF) (RFC 1584) Proposed Standard

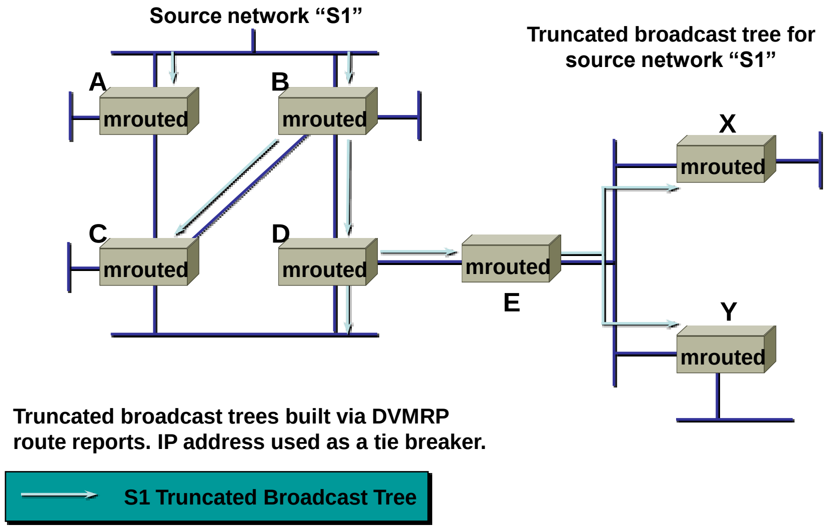
Multicast Distribution Trees Building

- How are distribution trees built?
  - PIM
    - Uses an existing unicast routing table plus a join/prune/graft mechanism to build the tree
  - DVMRP
    - Uses the DVMRP routing table plus a special Poison-Reverse mechanism to build the tree
  - MOSPF
    - Uses an extension of OSPFs link-state mechanism to build the tree

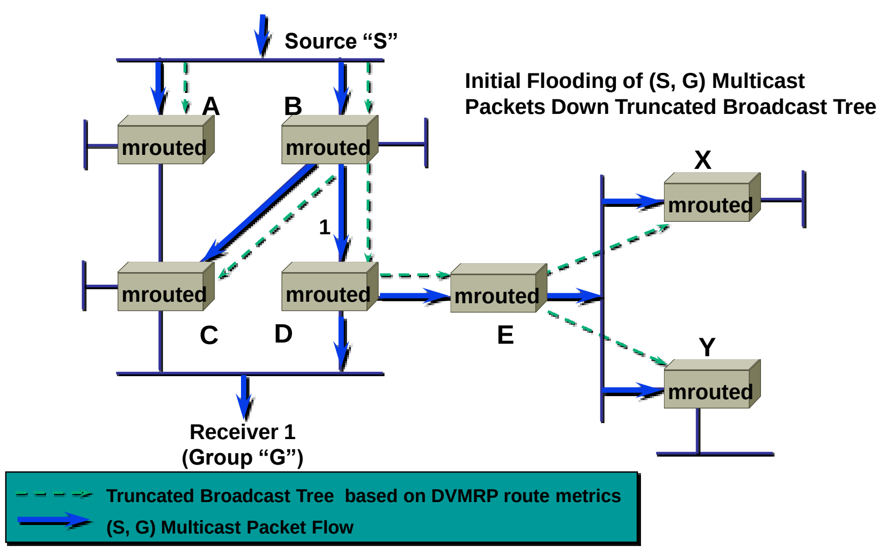
DVMRP Overview

- **Distance Vector Multicast Routing Protocol** (distance vector-based)
  - Similar to RIP
  - Infinity = 32 hops
  - Subnet masks in route advertisements
- Similar to PIM Dense Mode
  - Broadcast and prune operation (push model)
  - Uses DVMRP route table for RPF check

DVMRP - Source Trees



DVMRP — Pruning



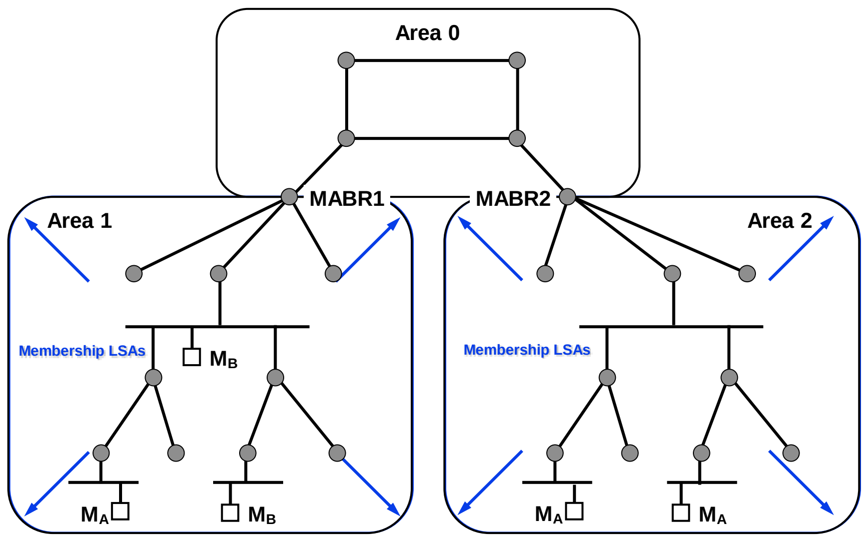
DVMRP — Evaluation

- Was widely used on the MBONE
- Significant scaling problems
  - Slow convergence—RIP-like behavior
  - Significant amount of multicast routing state information stored in routers—(S,G) everywhere
  - Maximum number of hops < 32
- Not appropriate for large scale production networks
  - Flood and prune behavior
  - Scaling problems (RIP-like)

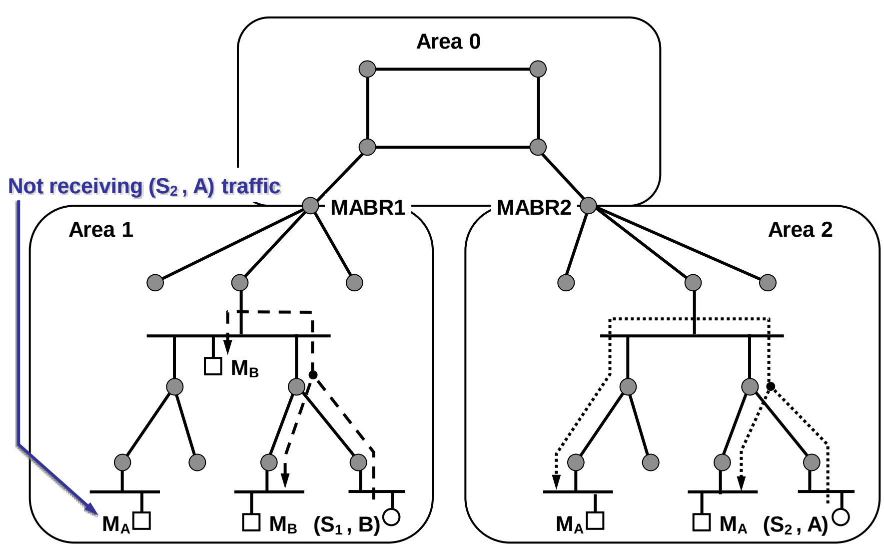
MOSPF (RFC 1584)

- **Multicast extensions of OSPF** routing protocol
  - Multicast information included in the OSPF link-state advertisements to construct distribution trees
  - Each router knows the topology of the entire network
- Group Membership LSAs flooded throughout the OSPF routing domain
- The Dijkstra's algorithm computes the shortest path
  - A separate calculation is required for each (SNet, G) pair

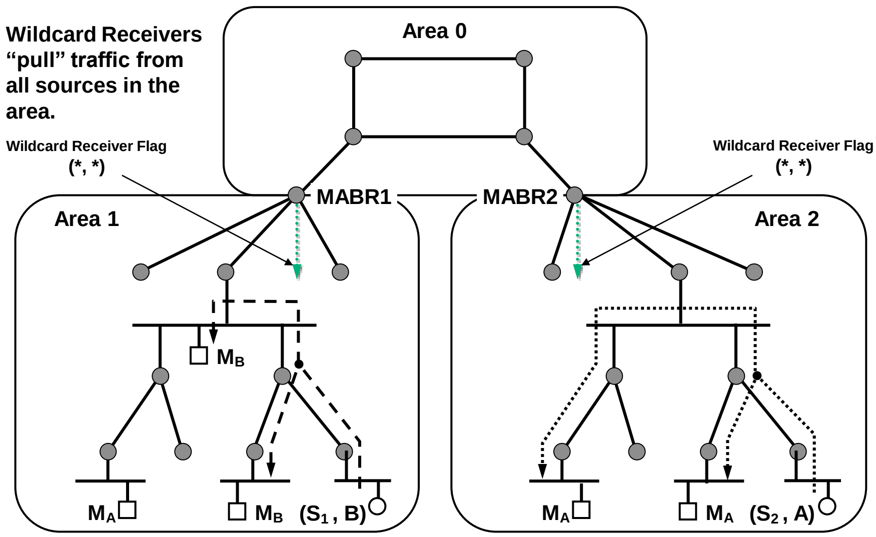
MOSPF Membership LSAs



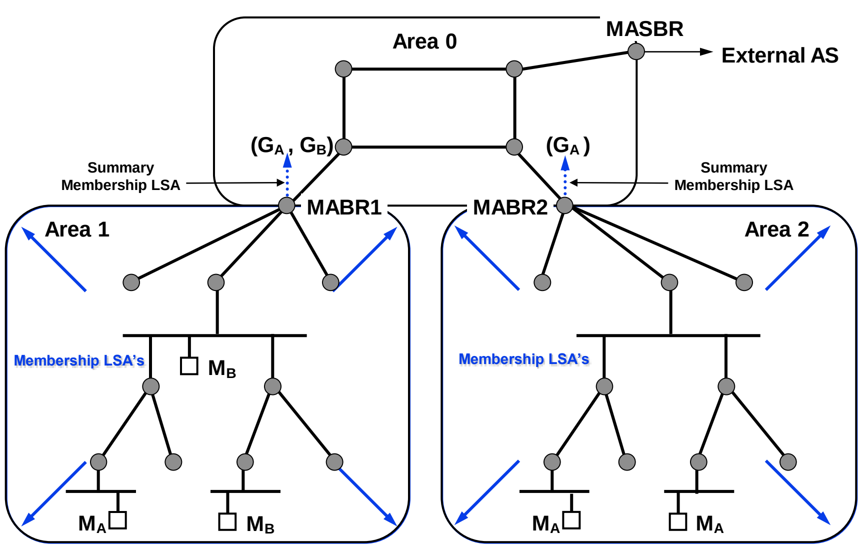
MOSPF Intra-Area Traffic



MOSPF Inter-Area Traffic



MOSPF Interdomain Traffic



MOSPF — Evaluation

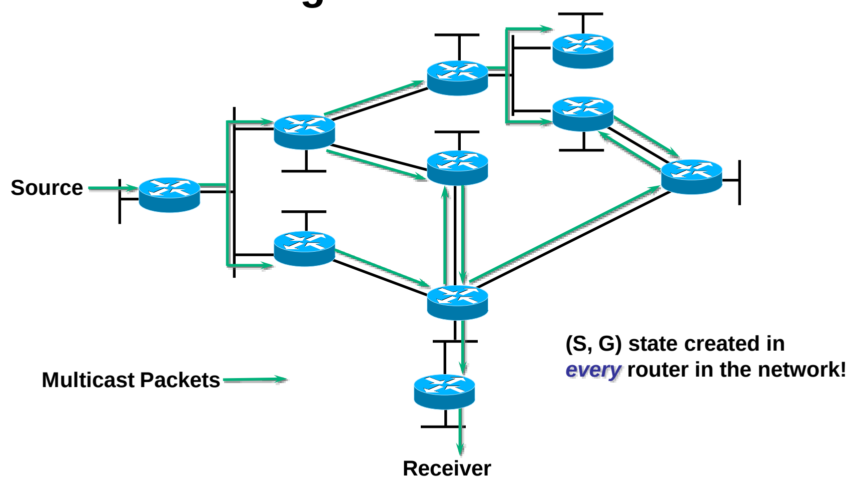
- If multicast traffic is not flooded everywhere, LSAs and the link-state database used
- Protocol dependent – OSPF only
- Significant scaling problems
  - Dijkstra's algorithm:
    - Run for every multicast (SNet, G)
    - Rerun on Group Membership or network state changes
- Not appropriate for networks with:
  - Large number of senders
  - Dynamic group membership

PIM - Dense Mode (PIM-DM)

- Protocol independent – supports all underlying unicast routing protocols: static, RIP, IGRP, EIGRP, IS-IS, OSPF, and BGP
- Uses flood and prune mechanism
  - Floods network and prunes back based on multicast group membership
  - Assert mechanism used to prune off redundant flows on multiaccess networks
- Appropriate for smaller implementations and pilot networks

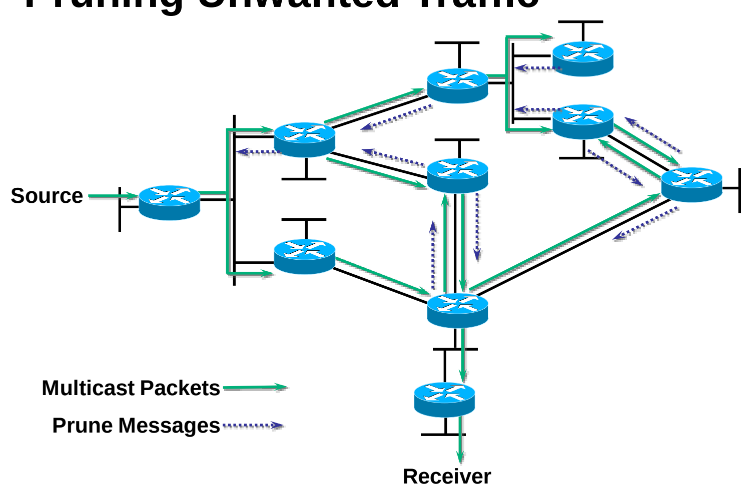
PIM-DM Flood and Prune

- **Initial Flooding**



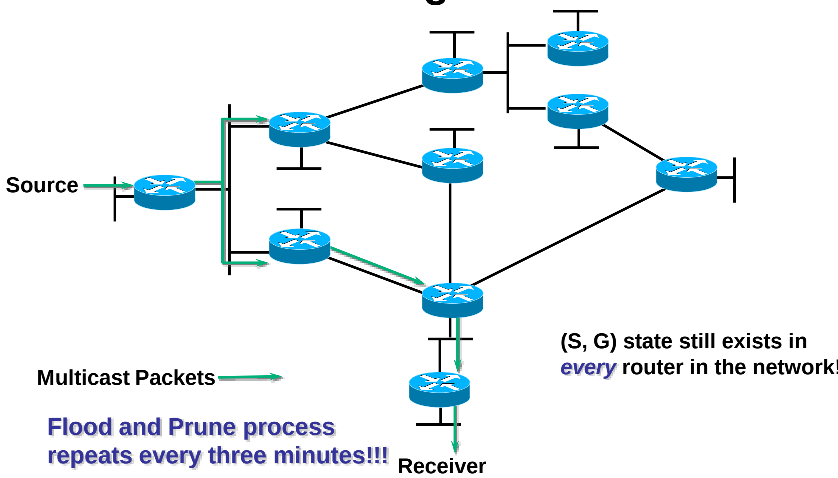
PIM-DM Flood and Prune (cont.)

- **Pruning Unwanted Traffic**



PIM-DM Flood and Prune (cont.)

- **Results after Pruning**



PIM-DM — Evaluation

- Most effective for small trial networks
- Advantages:
  - Easy to configure—two commands
  - Simple flood and prune mechanism
- Potential issues:
  - Inefficient flood and prune behavior
  - Complex assert mechanism
  - Mixed control and data planes
    - Results in (S, G) state in every router in the network
    - Can result in nondeterministic topological behaviors

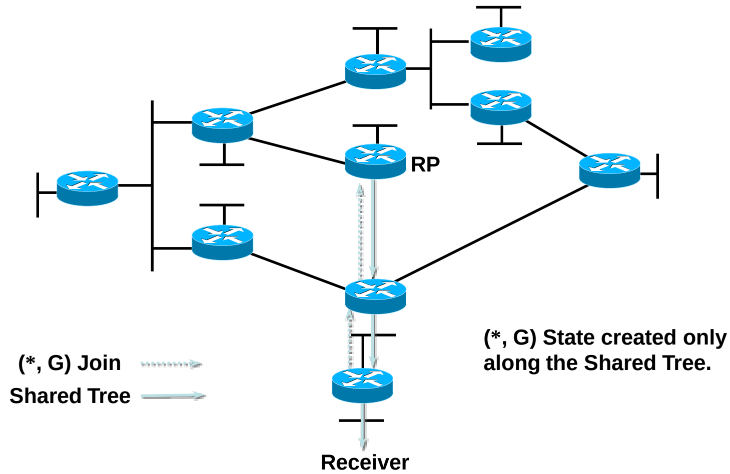
PIM Sparse Mode

- Protocol independent – works with any of the underlying unicast routing protocols
- Supports both **source and shared trees**
- Based on an explicit pull model
- Uses a rendezvous point (RP)
  - Senders and receivers "meet each other"
    - Senders are registered with RP by their first-hop router
    - Receivers are joined to the shared tree (rooted at the RP) by their local designated router (DR)

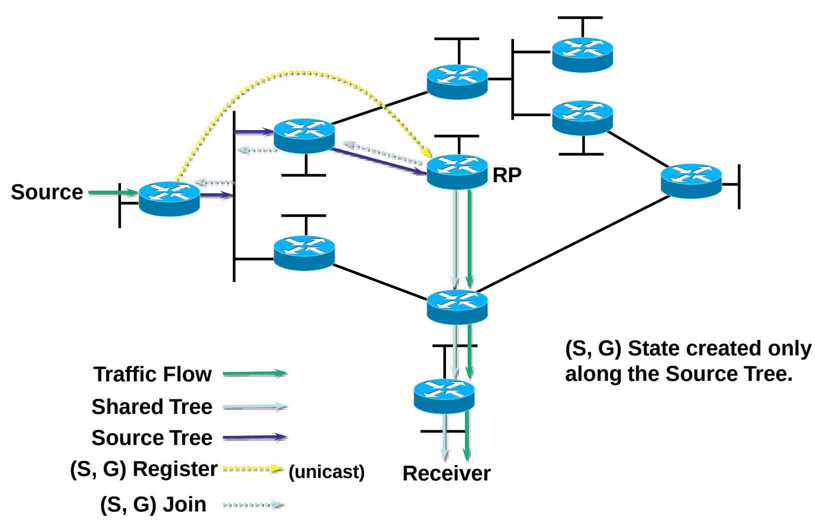
PIM Sparse Mode (cont.)

- Appropriate for:
  - Large-scale deployment for both densely and sparsely populated groups in the enterprise
  - Optimal choice for all production networks regardless of size and membership density
- Optimizations and derivatives:
  - Bidirectional mode
  - Source Specific Multicast

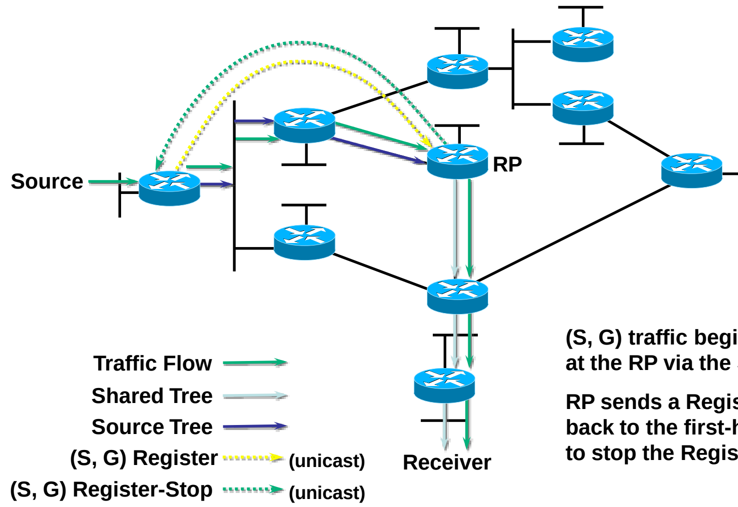
PIM-SM Shared Tree Join



PIM-SM Sender Registration



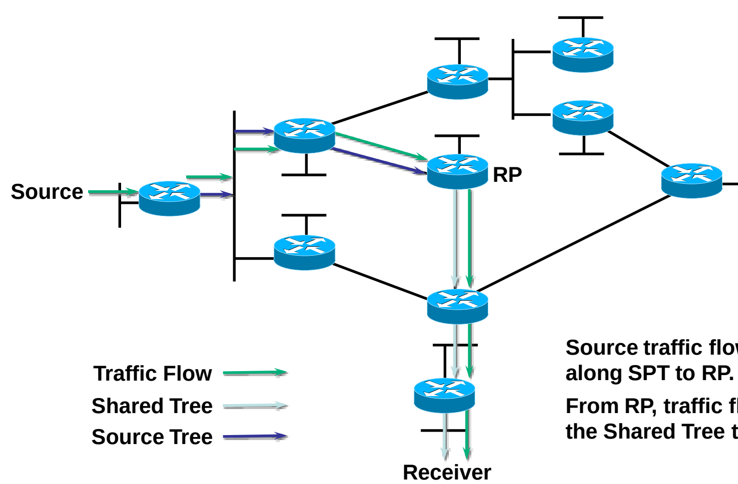
PIM-SM Sender Registration (cont.)



(S, G) traffic begins arriving at the RP via the Source tree.

RP sends a Register-Stop back to the first-hop router to stop the Register process.

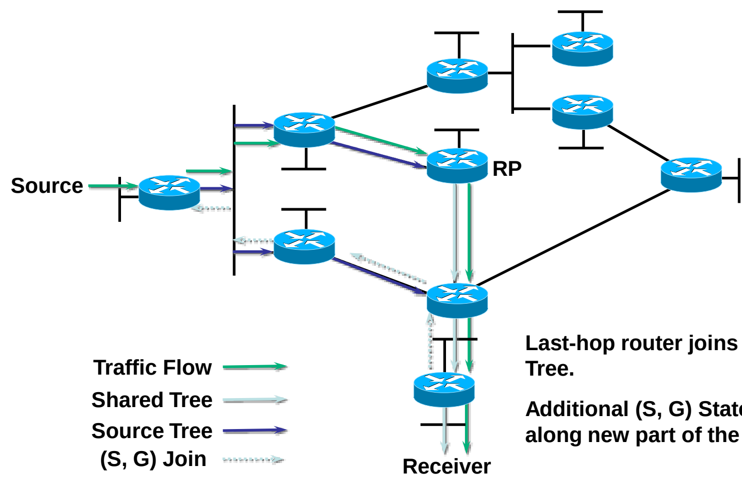
PIM-SM Sender Registration (cont.)



Source traffic flows natively along SPT to RP.

From RP, traffic flows down the Shared Tree to Receivers.

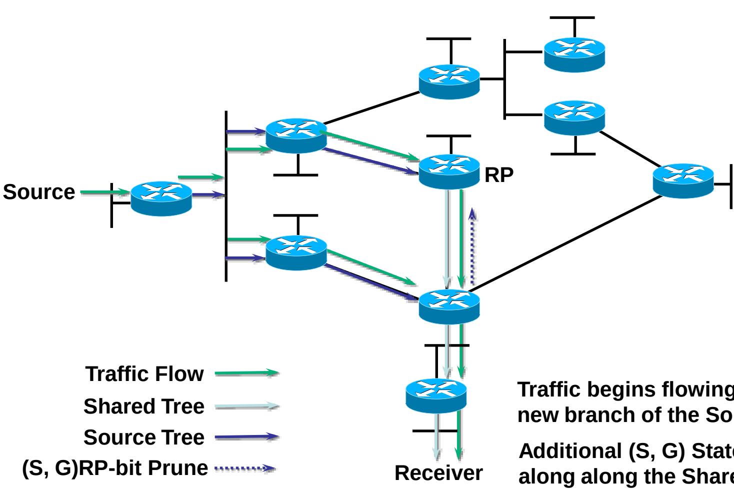
PIM-SM SPT Switchover



Last-hop router joins the Source Tree.

Additional (S, G) State is created along new part of the Source Tree.

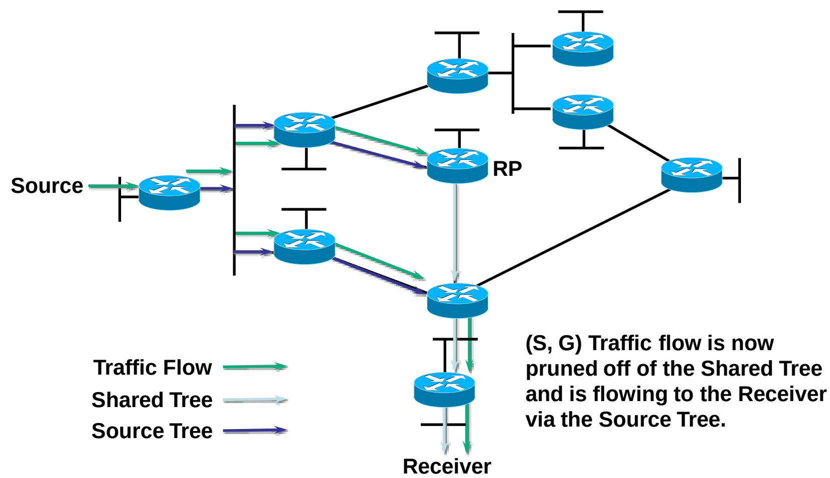
PIM-SM SPT Switchover (cont.)



Traffic begins flowing down the new branch of the Source Tree.

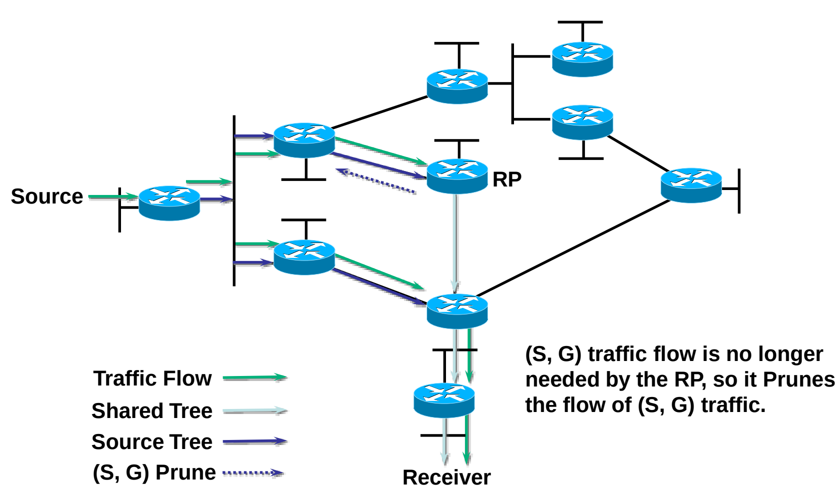
Additional (S, G) State is created along along the Shared Tree to prune off (S, G) traffic.

PIM-SM SPT Switchover (cont.)



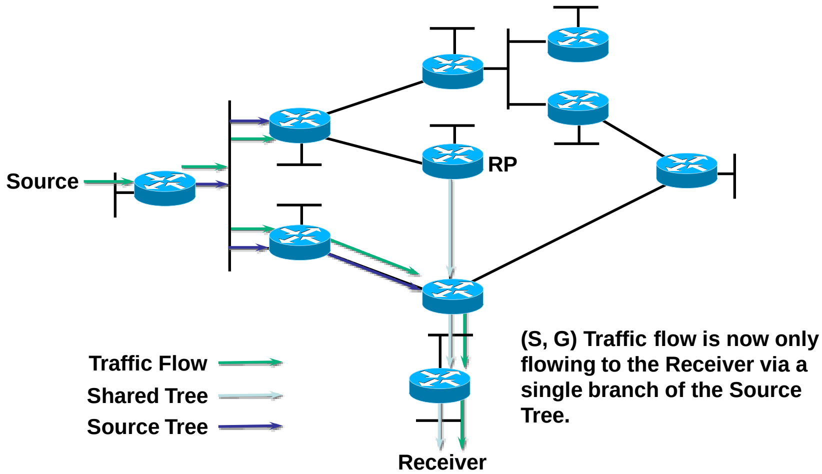
(S, G) Traffic flow is now pruned off of the Shared Tree and is flowing to the Receiver via the Source Tree.

PIM-SM SPT Switchover (cont.)



(S, G) traffic flow is no longer needed by the RP, so it Prunes the flow of (S, G) traffic.

PIM-SM SPT Switchover (cont.)



PIM-SM — Evaluation

- Effective for sparse or dense distribution of multicast receivers
- Advantages:
  - Traffic only sent down joined branches
  - Dynamically switches to optimal source trees for high traffic sources
  - Unicast routing protocol-independent
  - Basis for inter-domain multicast routing
    - When used with MBGP and MSDP

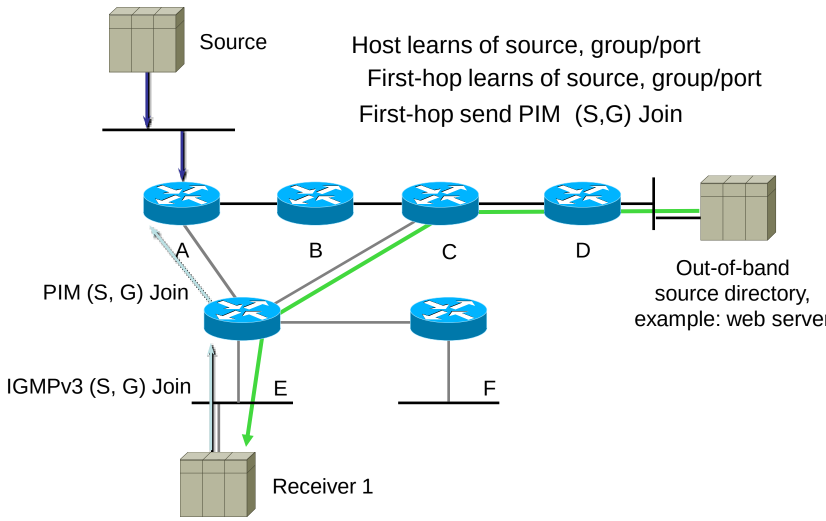
Source Specific Multicast (SSM)

- Uses Source Trees only.
- Assumes One-to-Many model.
  - Most Internet multicast fits this model.
  - IP/TV also fits this model.
- Hosts responsible for source discovery.
  - Typically via some out-of-band mechanism.
    - Web page, Content Server, etc.
  - Eliminates need for RP and Shared Trees.
  - Eliminates need for MSDP.

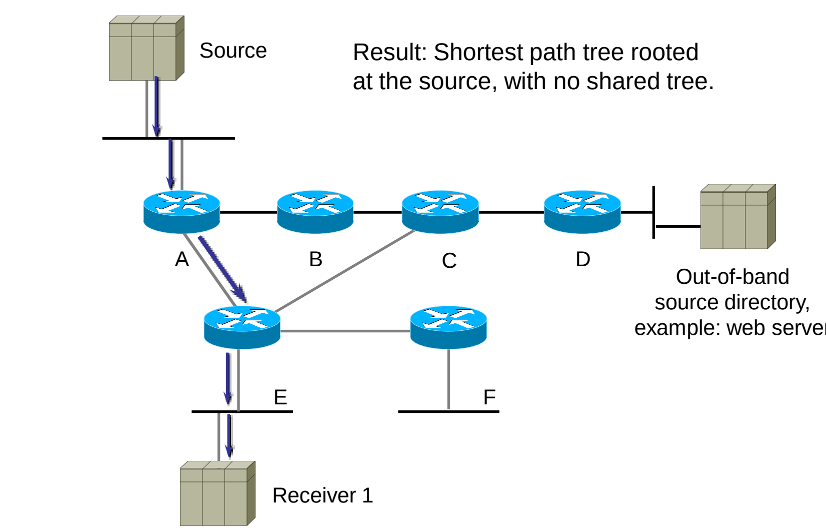
SSM Overview

- Hosts join a *specific* source within a group.
  - Content identified by specific (S,G) instead of (*,G).
  - Hosts responsible for learning (S,G) information.
- Last-hop router sends (S,G) join toward source
  - Shared Tree is never Joined or used.
  - Eliminates possibility of content Jammers.
  - Only specified (S,G) flow is delivered to host.
- Simplifies address allocation.
  - Dissimilar content sources can use same group without fear of interfering with each other.

SSM Example



SSM Example

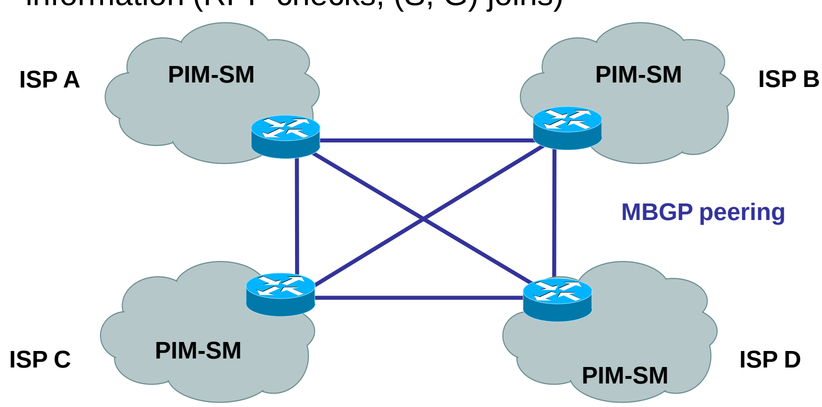


SSM — Host Signalling

- SSM Host Signalling: IGMPv3
  - Proposed for IP SSM
    - Also for filtering in RFC1112 style IP Multicast service.
  - IGMPv3 will only be active ...
    IF supported in last-hop routers
    AND IF supported in host operating systems
    AND IF supported in receiver applications

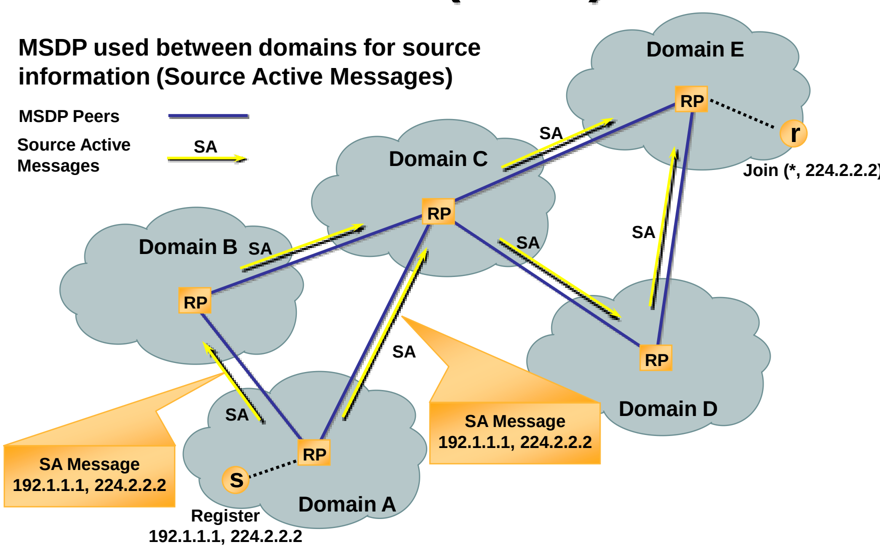
Interdomain Multicast Solution MBGP

- PIM-SM used within domains
- MBGP used between domains for source network information (RPF checks, (S, G) joins)



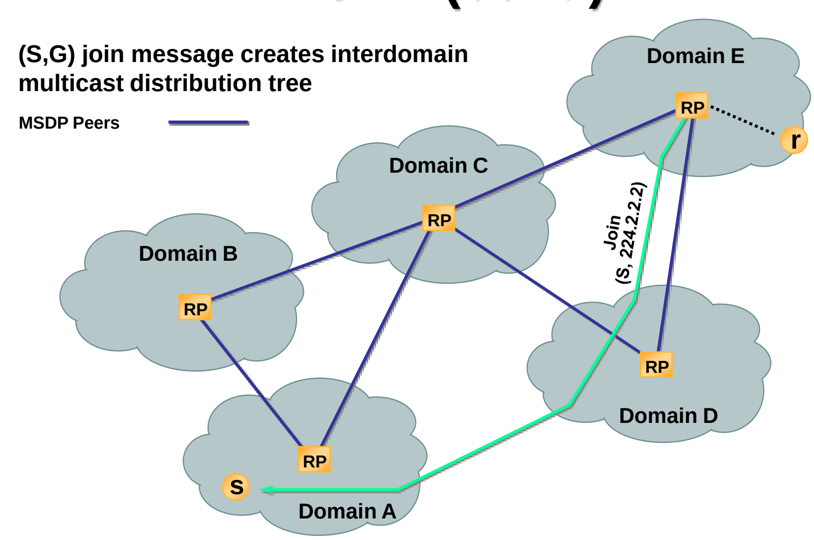
Interdomain Multicast Solution MSDP (cont.)

MSDP used between domains for source information (Source Active Messages)



Interdomain Multicast Solution MSDP (cont.)

(S,G) join message creates interdomain multicast distribution tree



Interdomain Multicast Solution MSDP (cont.)