IP Multicasting

Applications with multiple receivers

MCAST services

- Many applications like radio and video broadcasting, videoconferencing, etc. transmit the same data at one time to multiple receivers. A network must have mechanisms to support such applications in an efficient manner. Traditionally, multicast services are deployed in IP networks using well-known protocols like IGMP for L2 and PIM-SM/DM for IP

MCAST VPNs

- Service providers who have an installed base of Layer 3 VPN for unicast services are looking for the technology which enables transparent multicast services within customer's VPN.

Multicast Groups

- The set of receivers for a multicast transmission is called a **multicast group**
  - A multicast group is identified by a **multicast address**
  - A user that wants to receive multicast transmissions **joins** the corresponding multicast group, and becomes a **member** of that multicast group
- After a user joins, the network builds the necessary routing paths so that the user receives the data sent to the multicast group

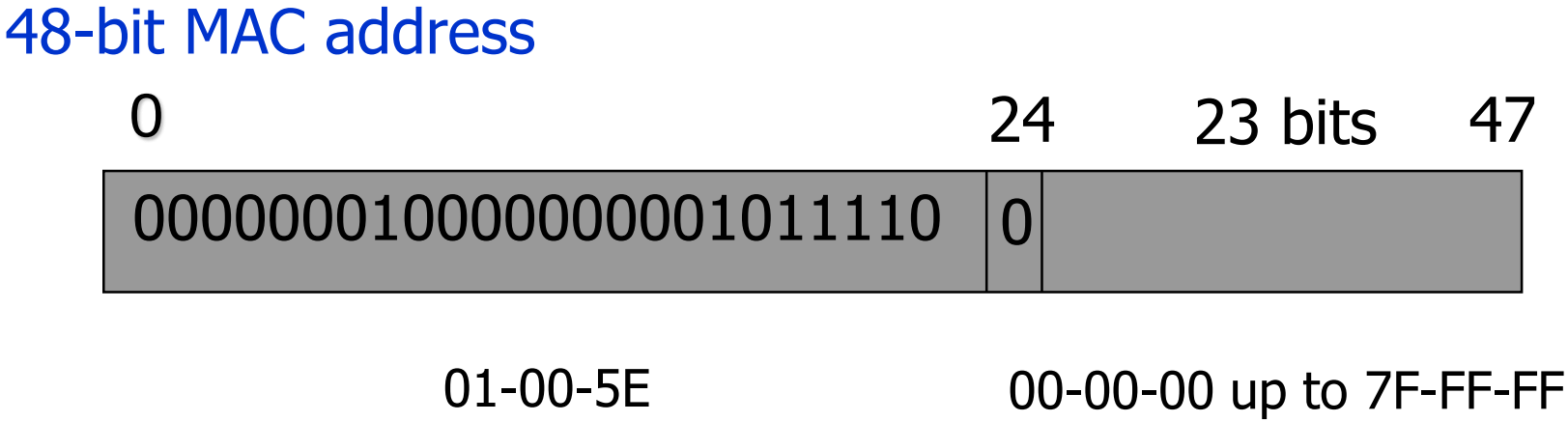
Semantics of IP Multicast

- Multicast groups are identified by IP addresses in the range 224.0.0.0 - 239.255.255.255 (class D address)
- Every host (interface) can join and leave a multicast group dynamically
  - no access control
- Every IP datagram send to a multicast group is transmitted to all members of the group
  - no security, no flood control
  - Sender does not need to be a member of the group
- The IP Multicast service is unreliable

IP Multicasting

- There are three components of the IP Multicast service
  - IP Multicast Addressing
  - Internet Group Management
  - Multicast Routing

Multicast Addressing



One multicast MAC address maps to 32 multicast IP addresses

Multicast IP - MAC Translation Example

224.0.0.13 – IANA - All PIM Routers

Multicast IP **224.0.0.13** converts to

- MAC address 01:00:5e:**00:00:0d**

- Host on LAN can receive MAC-layer multicast packets for groups to which it does not belong
- These packets are dropped

224.0.0.13	224.128.0.13	225.0.0.13
225.128.0.13	226.0.0.13	226.128.0.13
227.0.0.13	227.128.0.13	228.0.0.13
228.128.0.13	229.0.0.13	229.128.0.13
230.0.0.13	230.128.0.13	231.0.0.13
231.128.0.13		

Matched multicast
IP group addresses

232.0.0.13	232.128.0.13	233.0.0.13
233.128.0.13	234.0.0.13	234.128.0.13
235.0.0.13	235.128.0.13	236.0.0.13
236.128.0.13	237.0.0.13	237.128.0.13
238.0.0.13	238.128.0.13	239.0.0.13
239.128.0.13		

Types of Multicast addresses

- The range of addresses between 224.0.0.0 and 224.0.0.255 is reserved for the use of routing protocols and other low-level topology discovery or maintenance protocols
- Multicast routers should not forward any multicast datagram with destination addresses in this range.

- Examples of special and reserved Class D addresses

  - 224.0.0.1 All Systems on this Subnet
  - 224.0.0.2 All Routers on this Subnet
  - 224.0.0.5 OSPFIGP OSPFIGP All Routers
  - 224.0.0.6 OSPFIGP OSPFIGP Designated Routers
  - 224.0.0.9 RIP2 Routers

IGMP

- The Internet Group Management Protocol (IGMP) is a simple protocol for the support of IP multicast
- Multicast management on IPv6 networks is handled by Multicast Listener Discovery (MLD) which uses ICMPv6 messaging

- IGMP is defined in RFC 1112, 2236, 3376&4604
- IGMP operates on a Ethernet Segment only
- IGMP is used by multicast routers to keep track of membership in a multicast group

- Supports:
  - Joining a multicast group
  - Query membership
  - Send membership reports

IGMPv2 Protocol

- A host sends an **IGMP report** when it joins a multicast group
- No report is typically sent when a process leaves a group
  - IGMPv1
- IGMPv2 enhancements
  - **Leave Group** Message when process leaves a group
  - **Group-Specific Query** – Added to permit queries form a router to a specific group

- A multicast router regularly multicasts an **IGMP query** to all hosts (group address is set to zero)
- A host responds to an IGMP query with an **IGMP report**

- Multicast router keeps a table on the multicast groups that have joined hosts. The router only forwards a packet, if there is a host still joined
- Router does not keep track which host is joined

IGMPv3 and other features

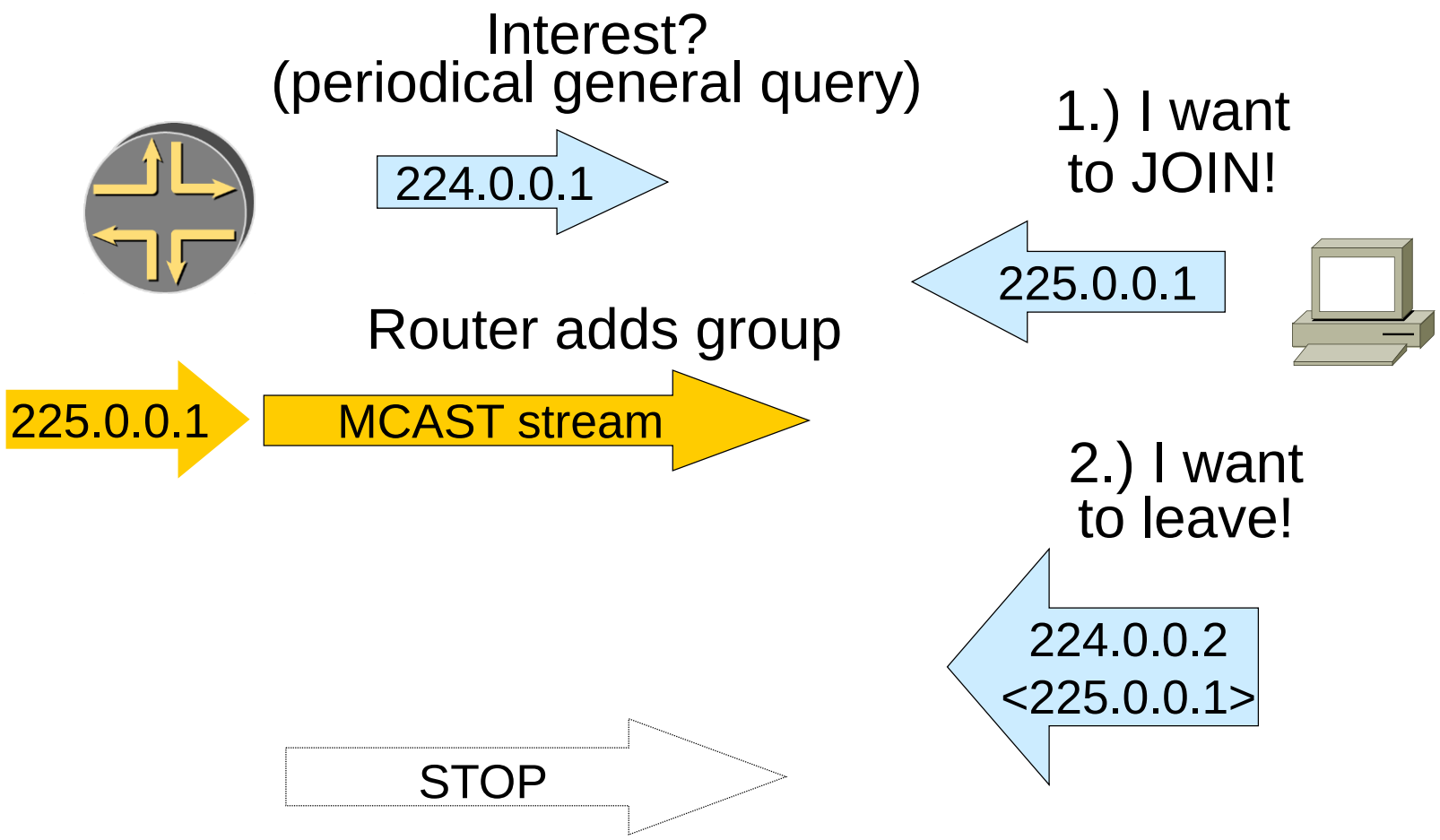
- v3 must be interoperable with v1 and v2
- Source-filtering
  - Only from a source (list of sources)
  - All but a source (list of sources)
- IGMP Snooping & Proxy reporting

IGMP snooping (L2 switch functionality) is designed to prevent hosts on a local network from receiving traffic for a multicast group they have not explicitly joined. Proxy - Report suppression - actively filters IGMP packets in order to reduce load on the multicast router

- Fast leave

Enables the software to remove the group state when it receives an IGMP Leave report without sending an IGMP query message. This parameter is used for IGMPv2 hosts when no more than one host is present on each VLAN port.

IGMPv2 Protocol



IGMPv2 Protocol

```
⊞ Frame 1: 60 bytes on wire (480 bits), 60 bytes captured (480 bits)
⊞ Ethernet II, Src: Cisco_6f:c8:00 (00:01:63:6f:c8:00), Dst: IPv4mcast_00:00:01 (01:00:5e:00:00:01)
⊞ Internet Protocol Version 4, Src: 10.60.0.189 (10.60.0.189), Dst: 224.0.0.1 (224.0.0.1)
⊟ Internet Group Management Protocol
    [IGMP Version: 2]
    Type: Membership Query (0x11)
    Max Response Time: 10.0 sec (0x64)
    Header checksum: 0xee9b [correct]
    Multicast Address: 0.0.0.0 (0.0.0.0)
```

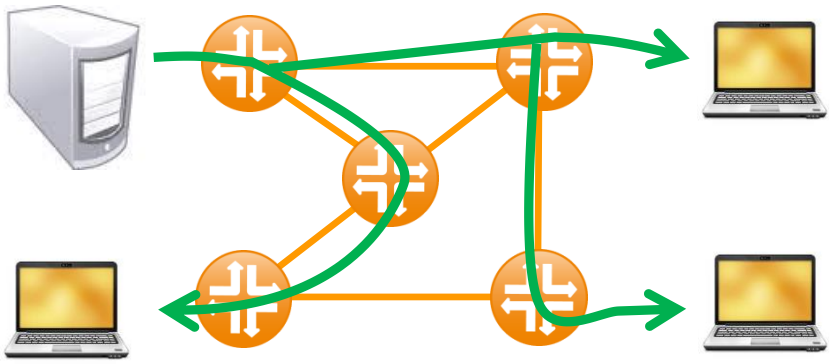
```
⊞ Frame 2: 60 bytes on wire (480 bits), 60 bytes captured (480 bits)
⊞ Ethernet II, Src: Hewlett-_e6:47:c6 (00:14:38:e6:47:c6), Dst: IPv4mcast_00:01:3c (01:00:5e:00:01:3c)
⊞ Internet Protocol Version 4, Src: 10.60.0.20 (10.60.0.20), Dst: 224.0.1.60 (224.0.1.60)
⊟ Internet Group Management Protocol
    [IGMP Version: 2]
    Type: Membership Report (0x16)
    Max Response Time: 0.0 sec (0x00)
    Header checksum: 0x08c3 [correct]
    Multicast Address: 224.0.1.60 (224.0.1.60)
```

IGMPv2 destination address

Message Type	Multicast Address
General Query	All hosts (224.0.0.1)
Group-Specific Query	The group being queried
Membership Report	The group being reported
Leave Group	All routers (224.0.0.2)

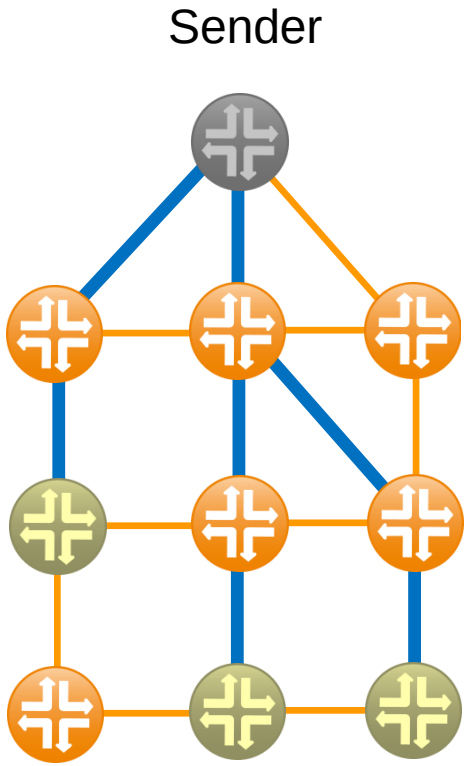
Multicast Routing Protocols

- **Goal**
  - Build a spanning tree between all members of a multicast group



Multicast routing as a graph problem

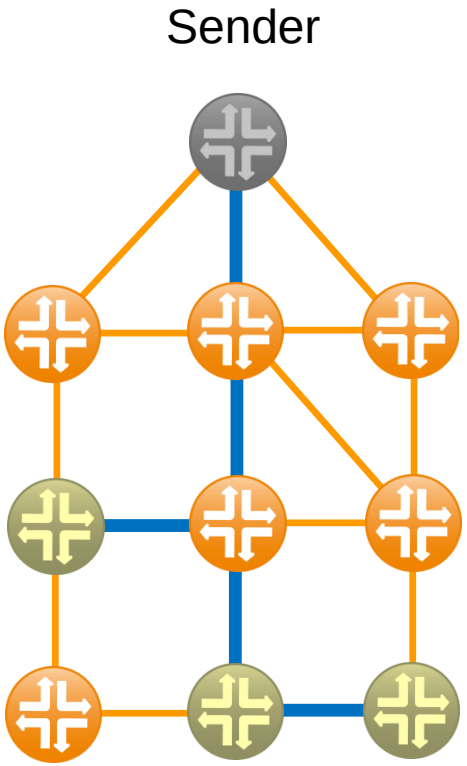
- Goal: Embed a tree such that all multicast group members are connected by the tree
- Solution 1: Shortest Path Tree
  - Build a tree that minimizes the path cost from the source to each receiver
  - Easy to compute
  - Good for one sender



Multicast routing as a graph problem

- Goal: Embed a tree such that all multicast group members are connected by the tree

- Solution 2: Minimum-Cost Tree
  - Build a tree that minimizes the total cost of the edges
  - Expensive to compute
  - Good solution if there are multiple senders



Multicast routing in practice

- A host sends an **IGMP report** when it joins a multicast group
- Routing Protocols implement two different trees:

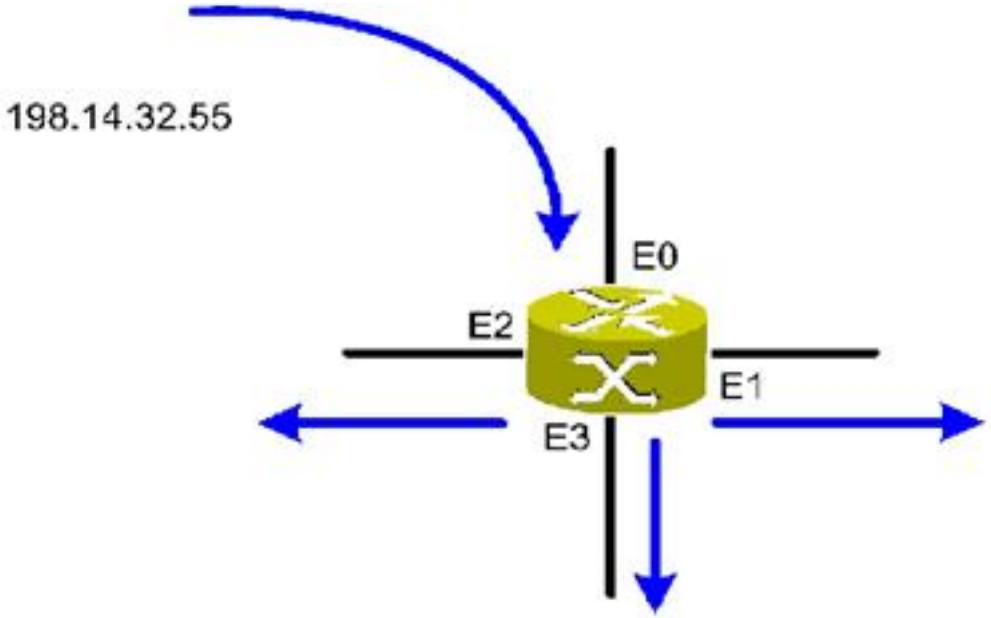
- Source Tree (S,G):
  - Implements Solution 1
  - Builds one SPT for each sender
  - Tree is built from receiver to the sender
    - → reverse shortest path / reverse path forwarding

- Shared Tree (*,G):
  - Implements Solution 1
  - Build a single distribution tree that is shared by all senders
  - Selects one router as a central point also called "**Rendezvous point**"
  - All receivers build a shortest path to the RP
    - → reverse shortest path / reverse path forwarding
  -

RPF – Reverse Path Forwarding

- A technique to ensure loop-free forwarding of data multicast packets
- Builds a Shared/RP Tree and Source Tree in a distributed fashion by taking advantage of the unicast routing tables
- Forward a packet only if it is receives from an RPF neighbor

Sender's IP address



Packet has been received on the correct port E0 according to the routing table and is replicated to the remaining ports

Routing Table

151.10.0.0/16	E2
198.14.32.0/24	E0
105.1.10.0/23	E1

Protocol Independent Multicast

- PIM-DM
  - The basic assumption behind PIM Dense Mode is that the multicast packet stream has receivers at most locations. Initial multicast packets are flooded everywhere, with **pruning** cutting off traffic to locations that do not need the multicast feed. Flood&Prune behavior every 3 minutes
- PIM-SM
  - PIM Sparse Mode uses an explicit request approach, periodically joins to the group
  - **ASM mode**: Any-Source Multicast. Traditional multicast – data and joins are forwarded to an RP
    - All receivers/routers in a PIM domain must have RP mapping
    - When receiver joins, a **Join** message is sent to RP on RPF
    - When load exceeds threshold, forwarding switched to Source Tree
    - Source Tree state is refreshed when data is forwarded and with **Join/Prune** control messages
  - **SSM mode**: Source-Specific Multicast. PIM-SM without RPs – the source is learned out-of-band, and the Source Tree is built directly

Pruning

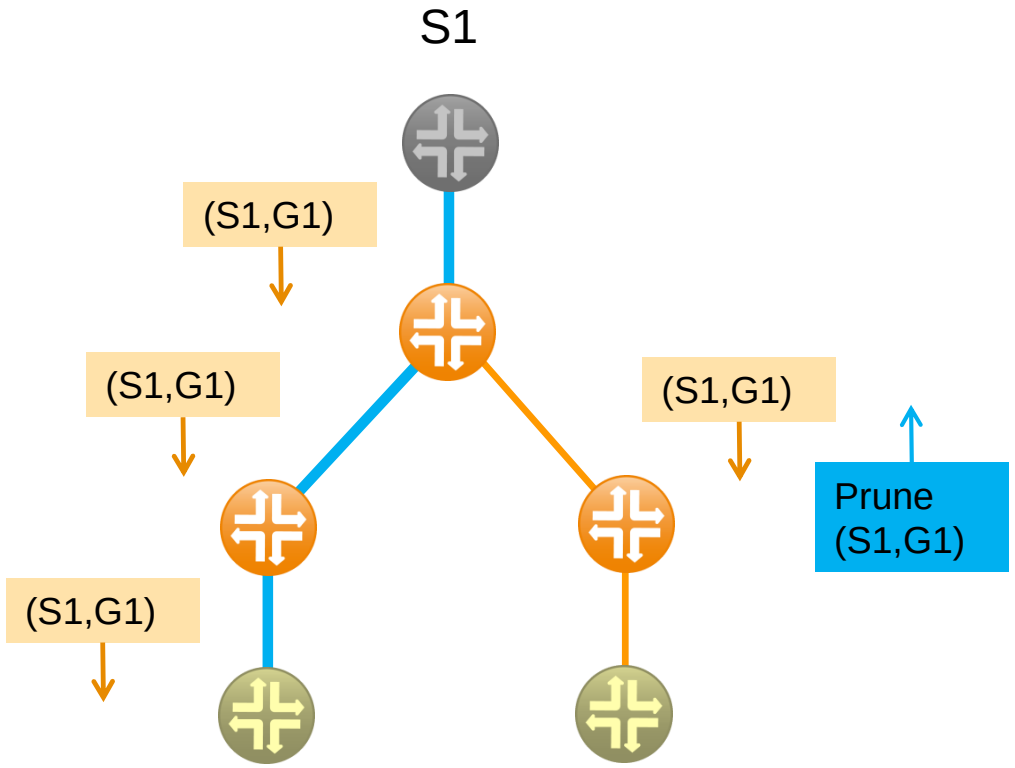
- Prune message disables a routing table entry
  - Effect: Removes a link from the multicast tree
  - No multicast messages are sent on a pruned link
  - Prune message is sent in response to a multicast packet
    - In case of PIM-DM – just temporarily disables mcasting
      - Question: Why temporarily?

- Who sends prune messages?
  - A router with no group members in its local network and no connection to other routers (sent on RPF interface)
  - A router with no group members in its local network which has received a prune message on all non-RPF interfaces (sent on RPF interface)
  - A router with group members which has received a mcast packet from a non-RPF neighbor (sent to non-RPF neighbor)

PIM Dense Mode

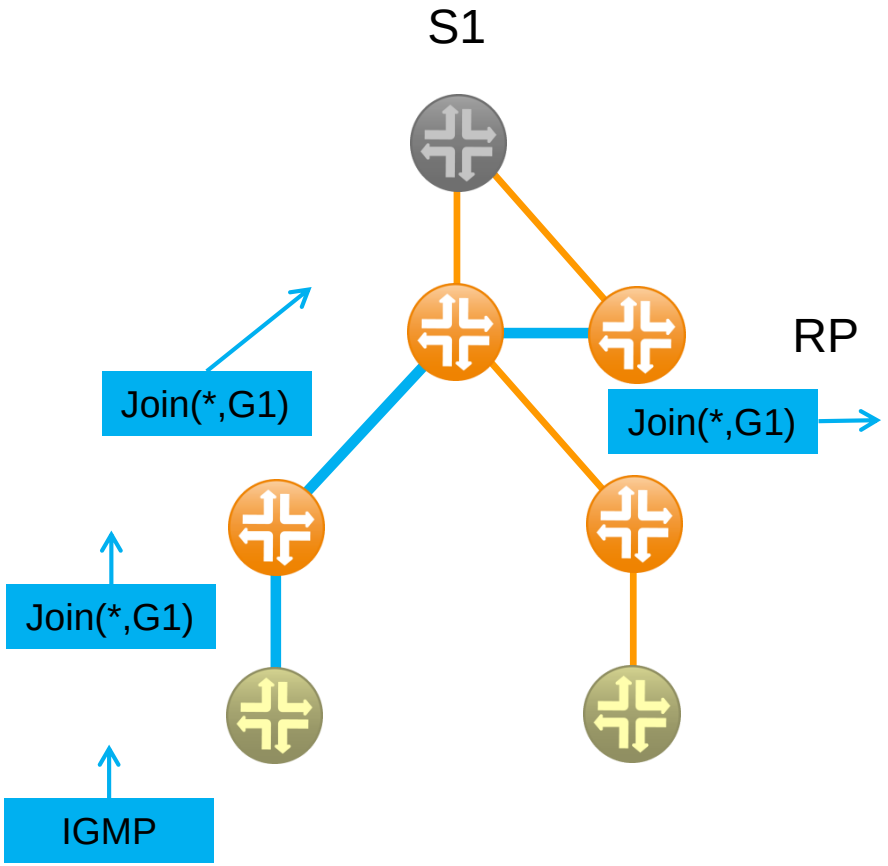
- Prune message disables a routing table entry
- PIM-DM implements flood&prune

- Orange packet
  - Multicast data
- Blue packet
  - PIM control message



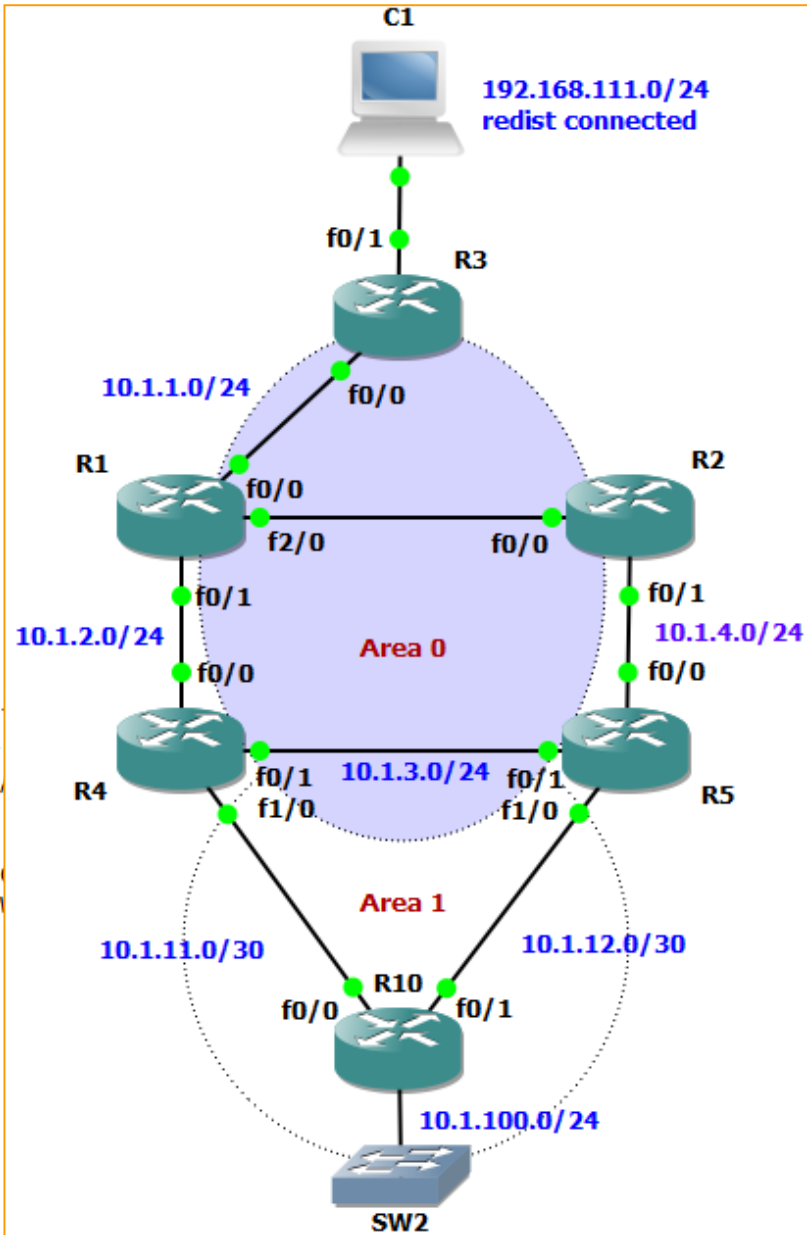
PIM Sparse Mode

- Receivers know RP (statically configured or dynamically elected)
- When receiver joins, a Join message is sent to RP on RPF
- Orange packet
  - Multicast data
- Blue packet
  - IGMP/PIM control message



PIM Sparse Mode

```
R2#sh ip pim rp 224.1.1.1
Group: 224.1.1.1, RP: 10.1.255.2, next RP-reachable in 00:00

--
R2#sh ip pim neighbor
PIM Neighbor Table
Mode: B - Bidir Capable, DR - Designated Router, N - Default
      S - State Refresh Capable
Neighbor          Interface                Uptime/Expires
Address
10.1.5.11         FastEthernet0/0          00:11:55/00:01:40
10.1.4.15         FastEthernet0/1          00:12:25/00:01:39
R2#
--

R2#sh ip mroute 224.1.1.1
Group 224.1.1.1 not found
R2#

--
R2#sh ip mroute 224.1.1.1
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group
       L - Local, P - Pruned, R - RP-bit set, F - Register f
       T - SPT-bit set, J - Join SPT, M - MSDP created entry
       X - Proxy Join Timer Running, A - Candidate for MSDP
       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 gr
Outgoing interface flags: H - Hardware switched, A - Assert
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 00:00:16/00:03:13, RP 10.1.255.2, flags: S
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/1, Forward/Sparse, 00:00:16/00:03:13
```



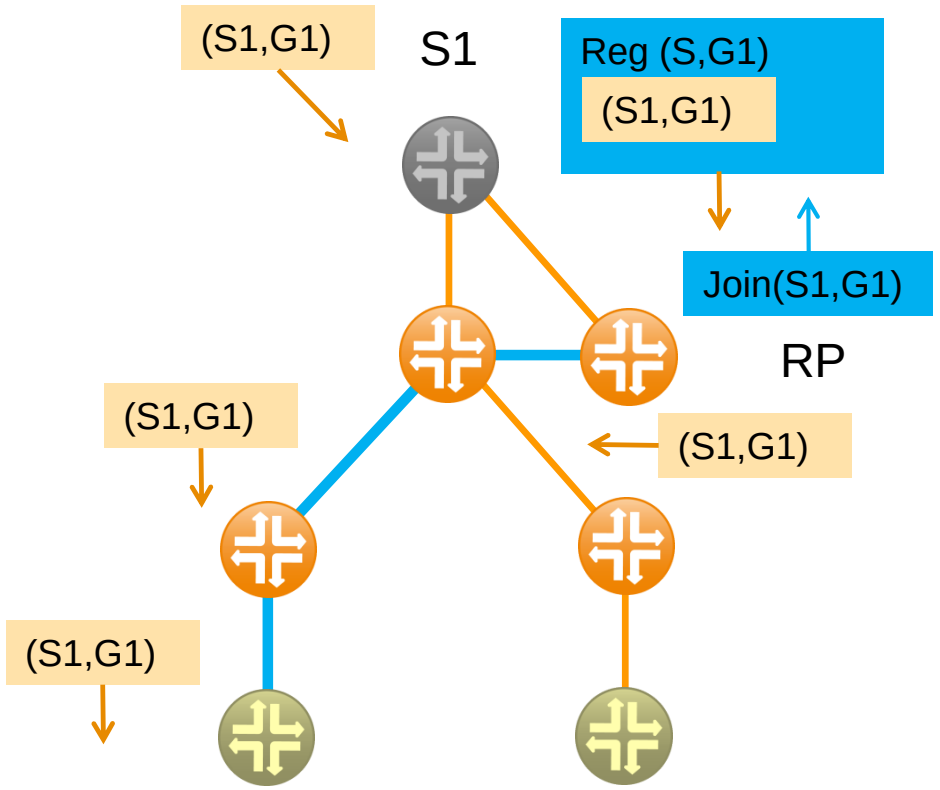
PIM Sparse Mode

No.	Time	Source	Destination	Protocol	Length	Info
123	198.789000	10.1.4.15	224.0.0.13	PIMv2	68	Hello
129	204.006000	10.1.4.15	224.0.0.13	PIMv2	68	Join/Prune
140	227.328000	10.1.4.12	224.0.0.13	PIMv2	68	Hello
141	228.581000	10.1.4.15	224.0.0.13	PIMv2	68	Hello

```
⊞ Frame 129: 68 bytes on wire (544 bits), 68 bytes captured (544 bits)
⊞ Ethernet II, Src: c2:01:1c:18:00:00 (c2:01:1c:18:00:00), Dst: IPv4mcast_00:00:0d (01:00:5e:00:00:0d)
⊟ Internet Protocol Version 4, Src: 10.1.4.15 (10.1.4.15), Dst: 224.0.0.13 (224.0.0.13)
    Version: 4
    Header length: 20 bytes
  ⊞ Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00: Not-ECT (Not ECN-Capab
    Total Length: 54
    Identification: 0x00d8 (216)
  ⊞ Flags: 0x00
    Fragment offset: 0
    Time to live: 1
    Protocol: PIM (103)
  ⊞ Header checksum: 0xc9ac [correct]
    Source: 10.1.4.15 (10.1.4.15)
    Destination: 224.0.0.13 (224.0.0.13)
⊟ Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0011 = Type: Join/Prune (3)
    Reserved byte(s): 00
    Checksum: 0xd9d7 [correct]
  ⊟ PIM options
      Upstream-neighbor: 10.1.4.12
      Reserved byte(s): 00
      Num Groups: 1
      Holdtime: 210s
    ⊟ Group 0: 224.1.1.1/32
      ⊟ Num Joins: 1
          IP address: 10.1.255.2/32 (SWR)
        Num Prunes: 0
```

PIM Sparse Mode Data Transmission

- Source sends multicast packet to RP
- Packet is attached to an RP **Register** message
- When packet reaches RP, it is forwarded in the tree
- RP sends a Join message on reverse path to S1, when RP Join messages reaches R1, it sends a native multicast packet to the RP, Register stops
- Orange packet
  - Multicast data
- Blue packet
  - PIM control message



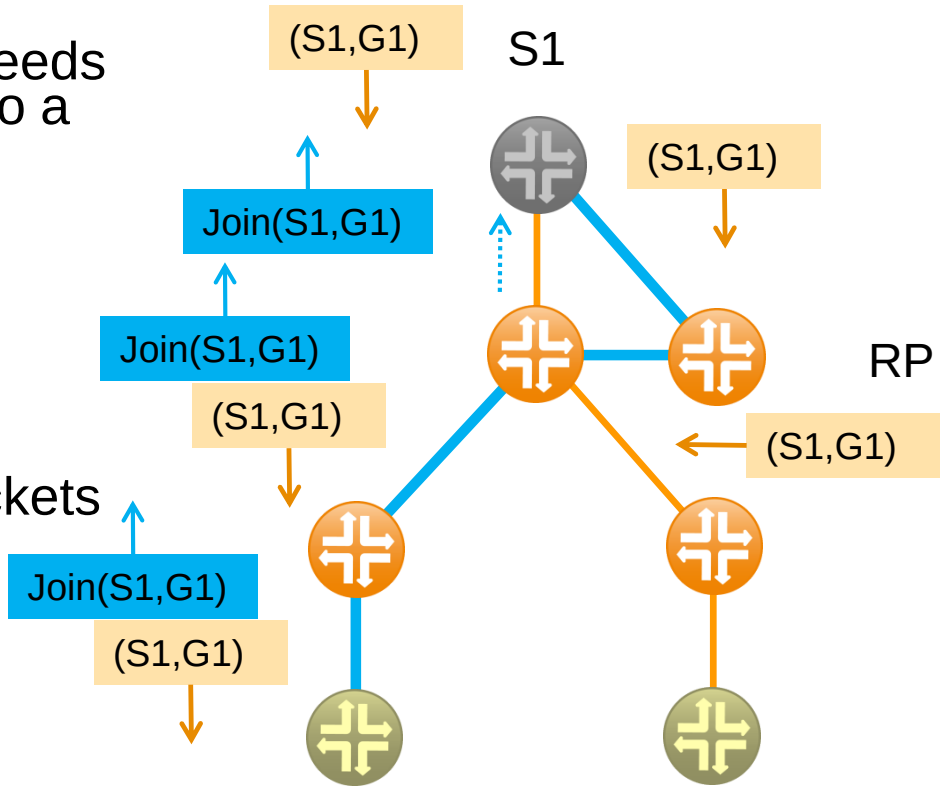
PIM Sparse Mode Data Transmission

```
R2#sh ip mroute 224.1.1.1
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,
       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
Outgoing interface flags: H - Hardware switched, A - Assert winner
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 00:24:30/00:03:29, RP 10.1.255.2, flags: S
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/1, Forward/Sparse, 00:24:30/00:03:29

(192.168.111.114, 224.1.1.1), 00:00:07/00:02:57, flags: PTX
  Incoming interface: FastEthernet0/1, RPF nbr 10.1.4.15
  Outgoing interface list: Null
```

```
⊞ Frame 41: 170 bytes on wire (1360 bits), 170 bytes captured (1360 bits)
⊞ Ethernet II, Src: c2:03:15:84:00:20 (c2:03:15:84:00:20), Dst: c2:02:0c:c8:00:00 (c2:02:0c:c8:00:00)
⊞ Internet Protocol Version 4, Src: 10.1.1.13 (10.1.1.13), Dst: 10.1.255.2 (10.1.255.2)
⊟ Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0001 = Type: Register (1)
    Reserved byte(s): 00
    Checksum: 0xf09a [incorrect, should be 0xdeff]
  ⊟ PIM options
    ⊟ Flags: 0x00000000
        0... .... .... .... .... .... .... .... = Not border
        .0.. .... .... .... .... .... .... .... = Not Null-Register
    ⊞ Internet Protocol Version 4, Src: 192.168.111.114 (192.168.111.114), Dst: 224.1.1.1 (224.1.1.1)
    ⊞ User Datagram Protocol, Src Port: 5001 (5001), Dst Port: 5001 (5001)
    ⊞ Data (100 bytes)
```

PIM Sparse Mode Switching to Source Tree

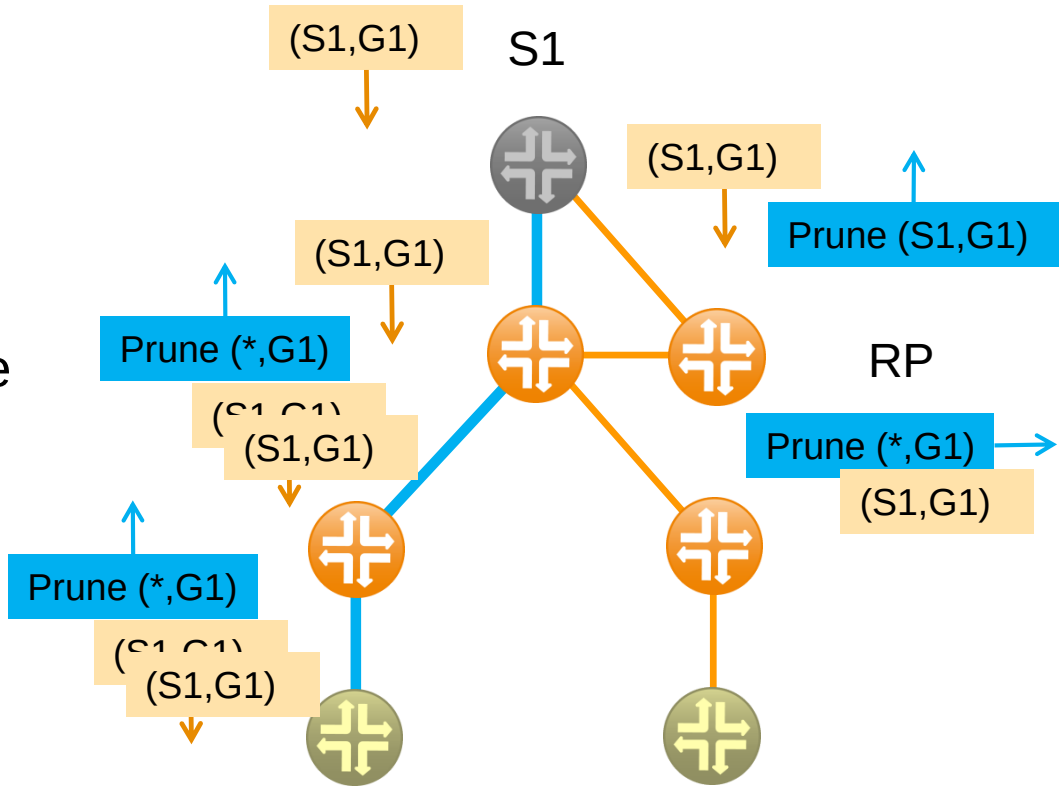
- Source sends multicast packet to RP
- When data to receivers exceeds a threshold, routers switch to a source-based tree
- This is done by sending an explicit join message to the source
- There may be duplicate packets being sent for some time



- Orange packet
  - Multicast data
- Blue packet
  - PIM control message

PIM Sparse Mode Switching to Source Tree

- When data to receivers exceeds a threshold, routers switch to a source-based tree
- This is done by sending an explicit join message to the source
- There may be duplicate packets being sent for some time



- Orange packet
  - Multicast data
- Blue packet
  - PIM control message

Rendezvous Point

- Rendezvous Point (RP) is a distribution point and used as a temporary way to connect a multicast receiver to an existing shared multicast tree passing through the rendezvous point
- Scalability issues
  - The traffic is concentrated around RP
  - Intensive CPU load
  - Memory
- Solution
  - Different RPs for different groups
  - Anycast RP – more physical RPs with the same logical IP

- There are three ways to set up a RP

  - manual configuration in each leaf routers
  - auto-RP
  - bootstrap router with PIM version 2

Manual RP Configuration

- Simple but not scalable
- Single point of failure
  - Exception: Anycast-RP (RFC 3446)

- Supply the RP address on each router in the network. If you supply an access-list, it defines (permits) the multicast groups for this particular RP.

Auto-RP

- All routers dynamically learn RP's IP address
- Two specific roles
  - Candidate RP (C-RP)
  - Mapping Agent (MA)
- Two special IANA MCAST Groups are used
  - Cisco-Announce 224.0.1.39
  - Cisco-Discovery 224.0.1.40
- Dense Mode needed to forward these two groups
  - So called **Sparse-Dense Mode**

Auto-RP

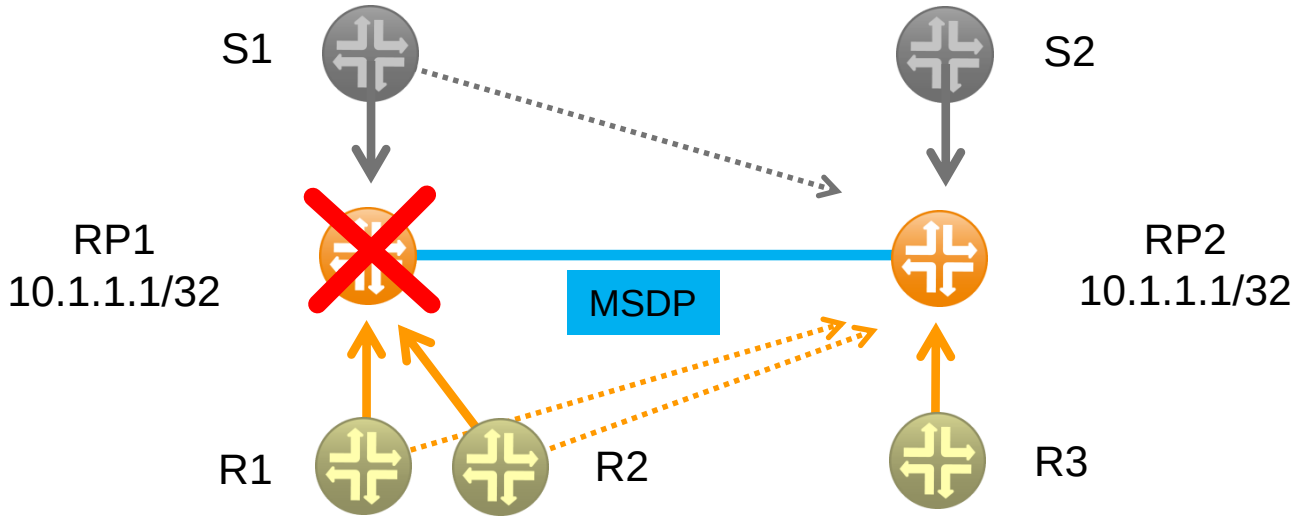
- C-RPs announce (224.0.1.39) to MAs its presence
  - MAs are joined to this mcast group using DM
- MAs elect the highest IP as RP
- MAs announce to other routers elected RP via Discovery message (224.0.1.40)
  - All other routers are joined to Discovery group via DM
- The same physical router can be configured for the C-RP and MA

BSR

- Multiple candidate BSRs (Bootstrap Router) can be configured
  - The highest priority C-BSR or IP
- A single BSR is elected
- C-RPs announce its presence to the new elected BSR
  - CRP-Advertisement is a unicast message
  - All C-RPs are stored in **RP-Set**
  - BSR message with RP-Set periodically flooded hop-by-hop to all routers (224.0.0.13, TTL=1)
- All routers select RP from RP-Set
  - The same selection algorithm – same RP for all routers

Anycast RP

- Used to provide RP load sharing and redundancy for static RP configuration
- More RPs with the same logical IP address for the same group ranges
- Sources and Receivers use the closest RP based on unicast routing
- Uses TCP based MSDP (Multicast Source Discovery Protocol) to communicate existence between sources
  - Also used for inter-domain multicasting



Ďakujem za pozornosť

roman dot kaloc at gmail dot com