IP Multicast Deepdive

Paul Borghese
Chesapeake NetCraftsmen

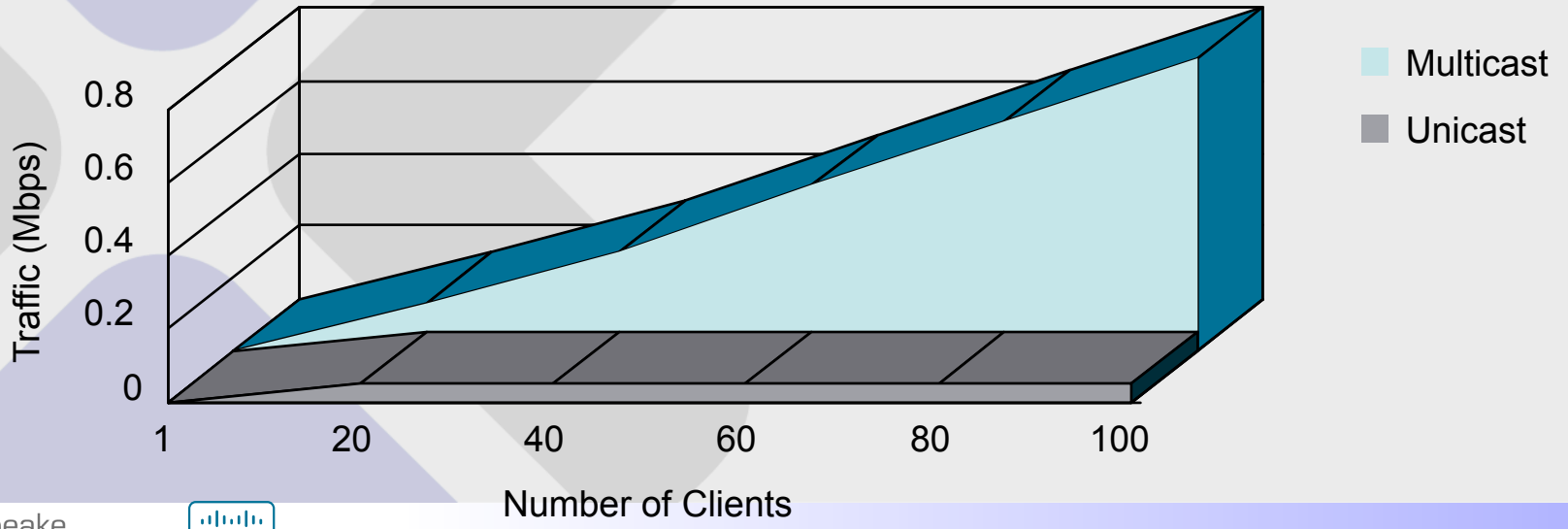
Cisco Mid-Atlantic Users Group

Multicast Background

Multicast Advantages

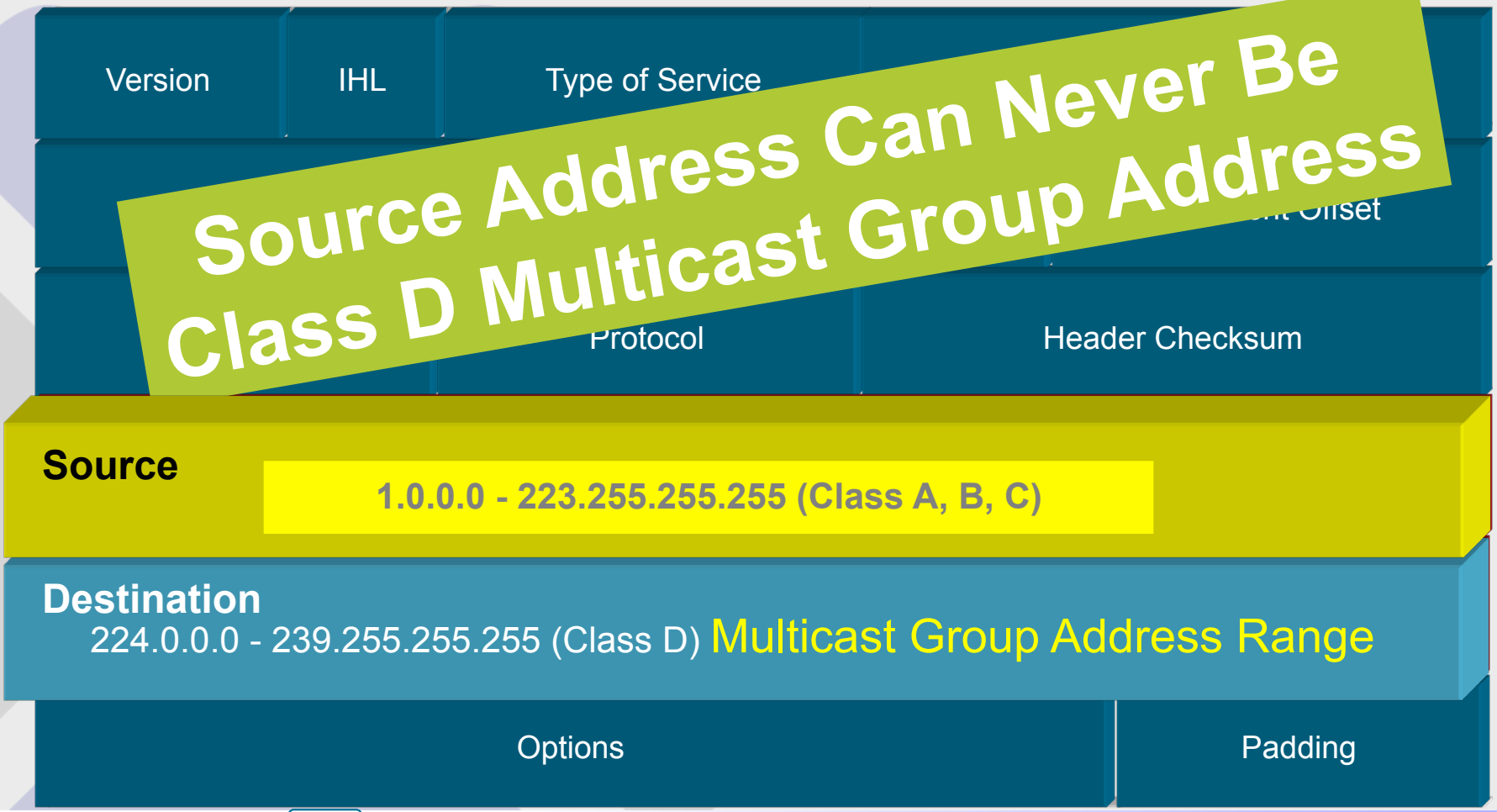
- Enhanced efficiency: **controls network traffic and reduces server and CPU loads**
- Optimized performance: **Eliminates traffic redundancy**
- Distributed applications: **Makes multipoint applications possible**

Example: Audio Streaming
All Clients Listening to the Same 8 Kbps Audio



Multicast Addressing

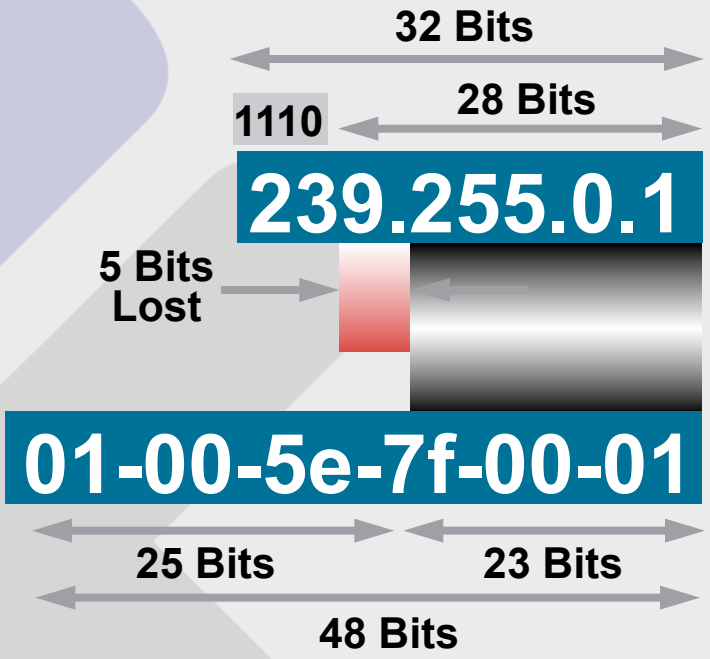
IPv4 Header



IP to Multicast Translation

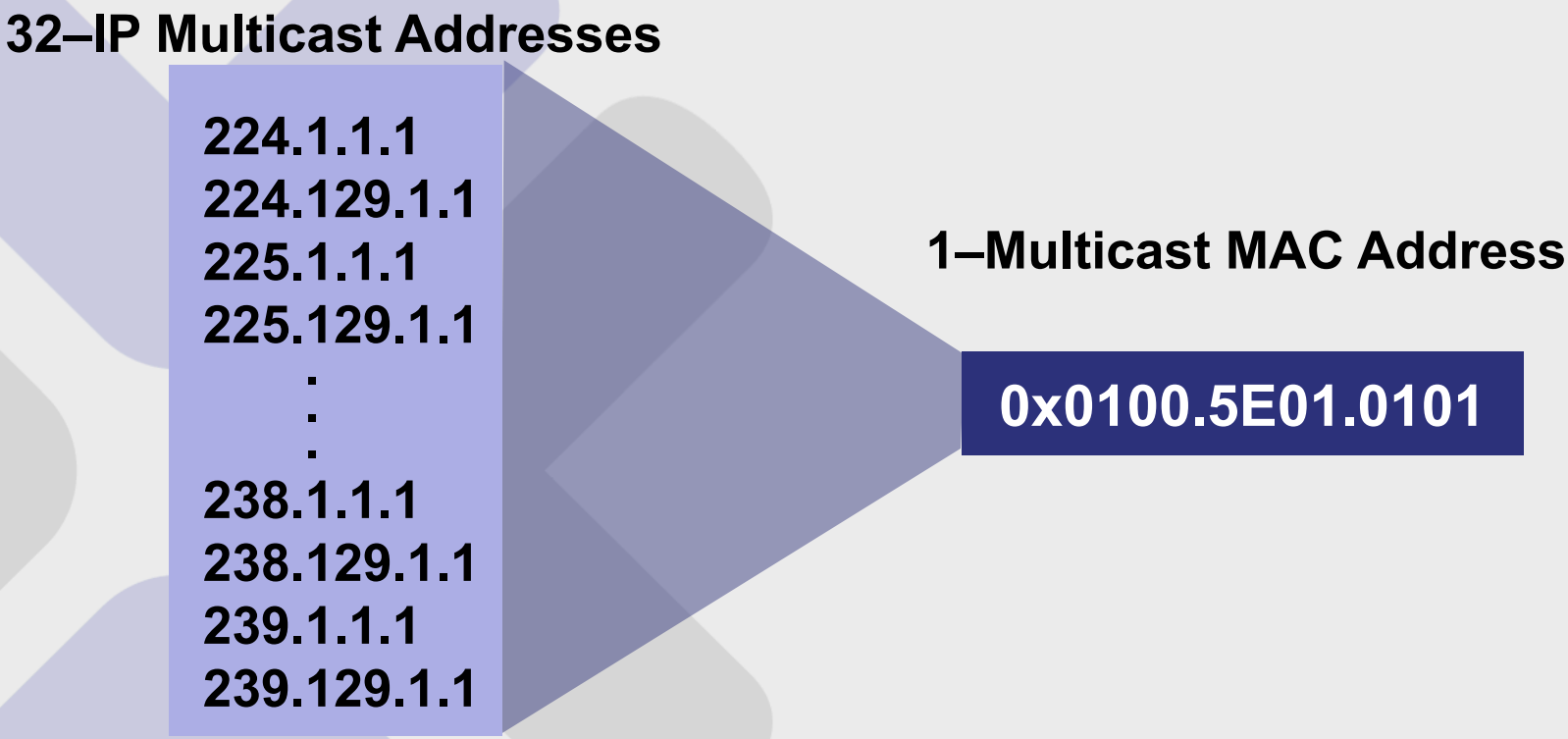
- The lower 23 bits of the IP address are mapped into the last 24 bits of the mac address 01-00-5e-xx-xx-xx
- Class D address of 224.x.x.x to 239.x.x.x
- There is some overlap between the IP address and the MAC address. So multiple IP addresses may lead to the same MAC address

IP Multicast MAC Address Mapping



IP Multicast MAC Address Mapping

Be Aware of the 32:1 Address Overlap



Multicast Addressing—224/4

- **Reserved link-local addresses**
  - 224.0.0.0–224.0.0.255
  - Transmitted with TTL = 1
  - Examples
    - 224.0.0.1 All systems on this subnet
    - 224.0.0.2 All routers on this subnet
    - 224.0.0.5 OSPF routers
    - 224.0.0.13 PIMv2 routers
    - 224.0.0.22 IGMPv3
- **Other reserved addresses**
  - 224.0.1.0–224.0.1.255
  - Not local in scope (transmitted with TTL > 1)
  - Examples
    - 224.0.1.1 NTP (Network Time Protocol)
    - 224.0.1.32 Mtrace routers
    - 224.0.1.78 Tibco Multicast1

Multicast Addressing—224/4

- Administratively scoped addresses
  - 239.0.0.0–239.255.255.255
  - Private address space
    - Similar to RFC1918 unicast addresses
    - Not used for global Internet traffic—scoped traffic
- SSM (Source Specific Multicast) range
  - 232.0.0.0–232.255.255.255
  - Primarily targeted for Internet-style broadcast
- GLOP (honest, it's not an acronym)
  - 233.0.0.0–233.255.255.255
  - Provides /24 group prefix per ASN

Host-Router Signaling: IGMP

- How hosts tell routers about group membership
- Routers solicit group membership from directly connected hosts
- RFC 1112 specifies version 1 of IGMP
  - Supported on Windows 95
- RFC 2236 specifies version 2 of IGMP
  - Supported on latest service pack for Windows and most UNIX systems
- RFC 3376 specifies version 3 of IGMP
  - Supported in Window XP and various UNIX systems

IGMP Features

- IGMP is enabled by default when multicast is applied to an interface.
- To make a router participate in a multicast group (for testing) use the command:
  - ip igmp join-group group-address

IGMP

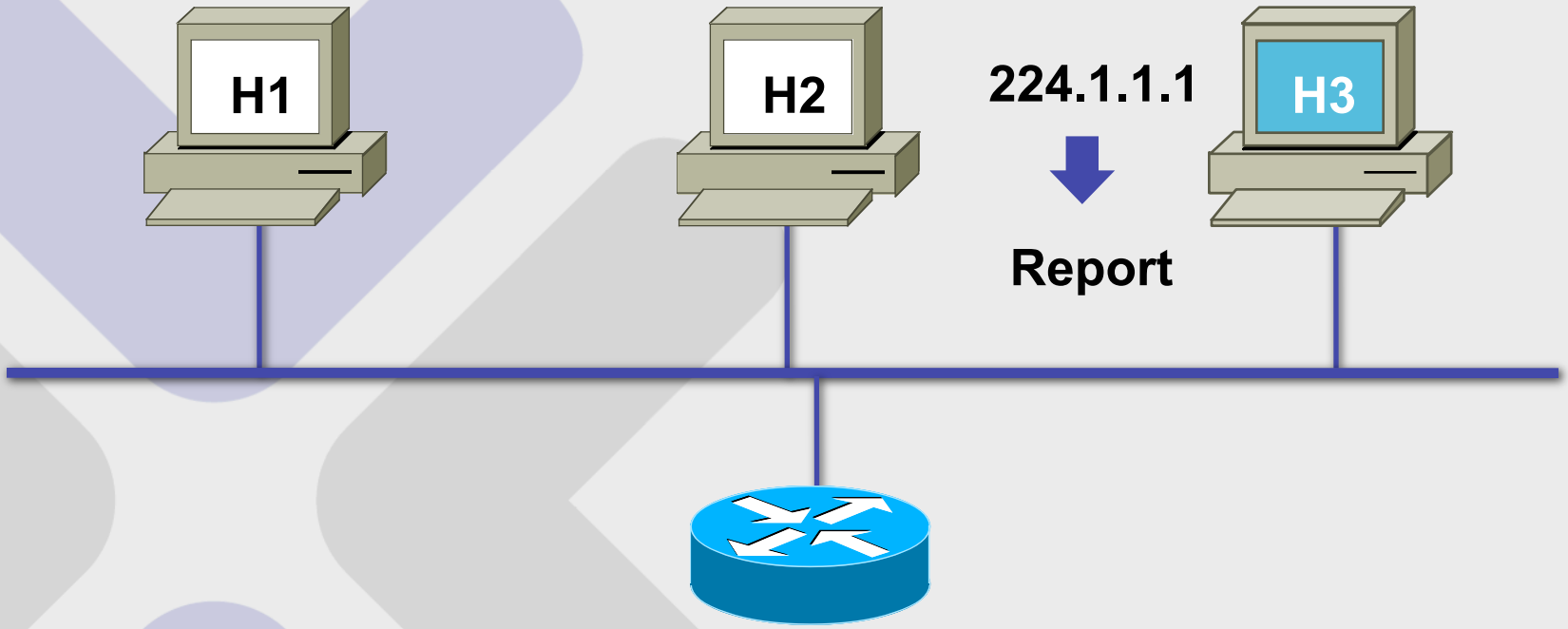
- **To filter which groups a client may join**
  - **ip igmp access-group *access-list-number***

IGMP

- **Version may be selected:**
  - **ip igmp version {3 | 2 | 1}**
- **To change the query interval for DR in a multiaccess network:**
  - **ip igmp query-interval *seconds***

Host-Router Signaling: IGMP

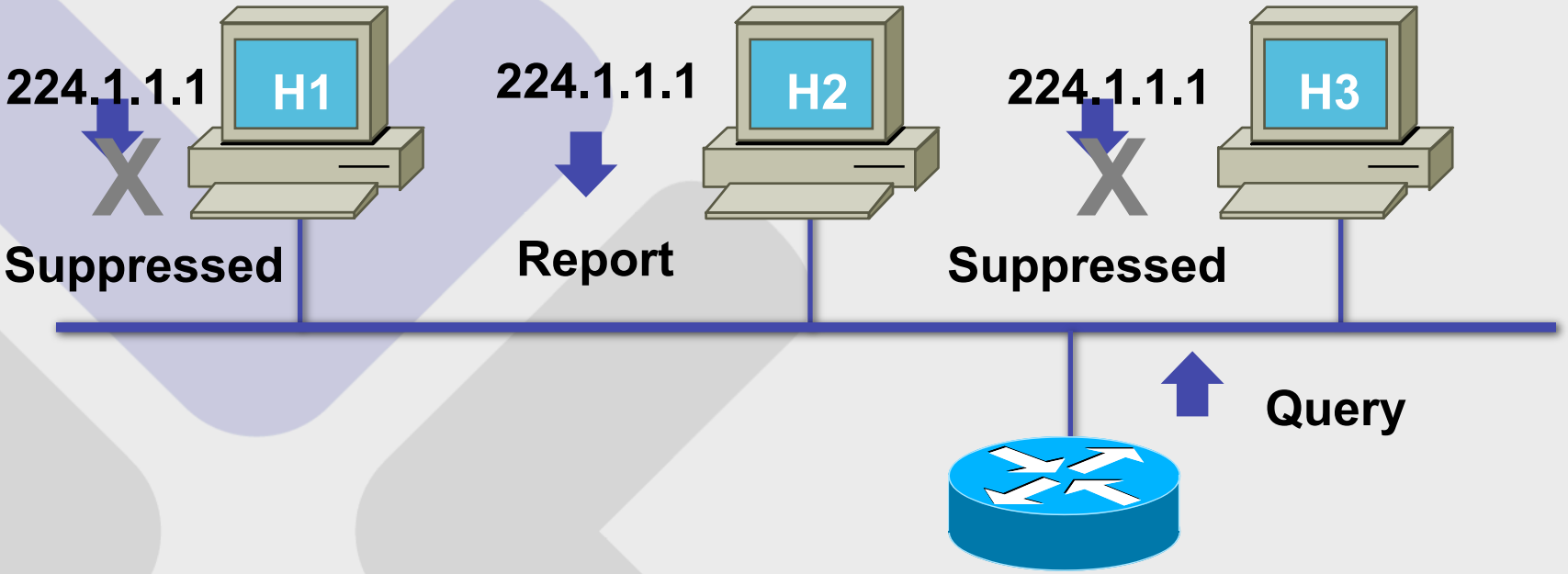
Joining a Group



- Host sends IGMP report to join group

Host-Router Signaling: IGMP

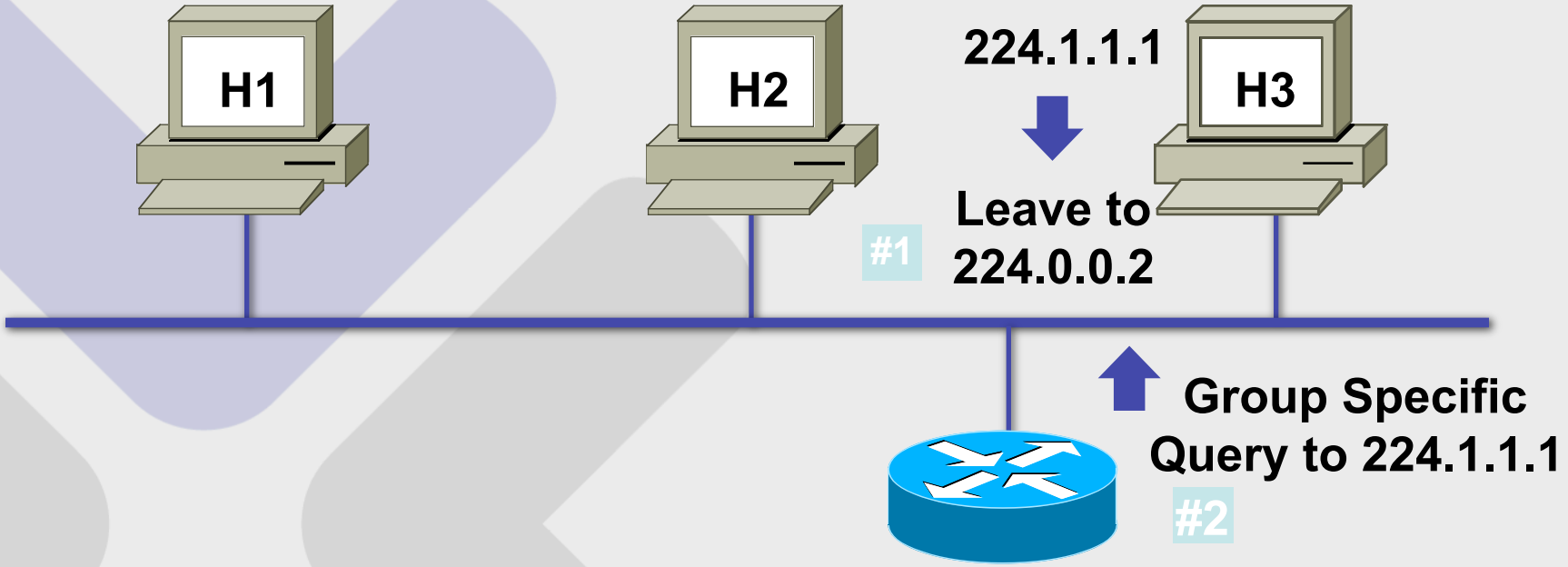
Maintaining a Group



- Router sends periodic queries to 224.0.0.1
- One member per group per subnet reports
- Other members suppress reports

Host-Router Signaling: IGMP

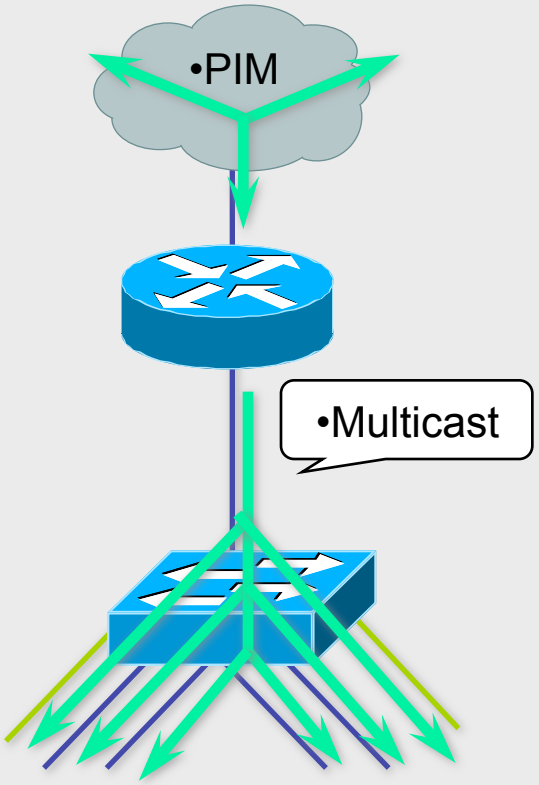
Leaving a Group (IGMPv2)



- Host sends leave message to 224.0.0.2
- Router sends group-specific query to 224.1.1.1
- No IGMP report is received within ~ 3 seconds
- Group 224.1.1.1 times out

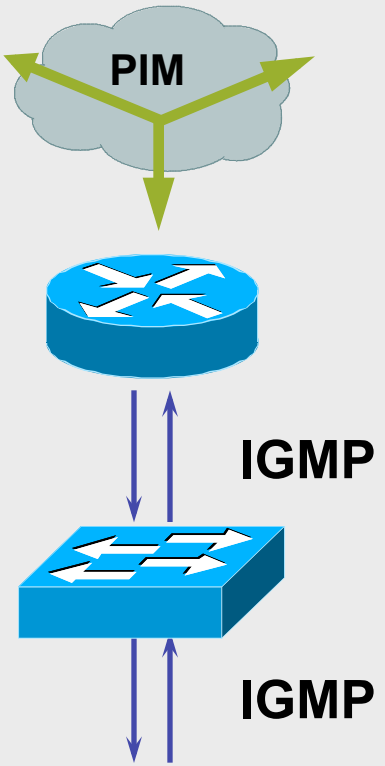
Issue: Layer 2 Multicast Frame Switching

- Layer 2 switches by default flood the frame to every port on the destination LAN.
- Static entries can sometimes be set to specify which ports should receive which group(s) of multicast traffic.
- Dynamic configuration of these entries cuts down on user administration.



Solution: IGMP Snooping

- **Switches snoop, are IGMP-aware**
  - **Switch must examine contents of IGMP messages to determine which ports want what traffic:**
    - **Uses IGMP membership reports.**
    - **Uses IGMP leave messages.**
  - **Switch can forward multicast traffic more efficiently.**
- **Some Cisco switches support IGMP snooping hardware**
  - **Catalyst 6500/4500 switches support multicast packet replication in hardware.**



L2 Multicast Frame Switching

Impact of IGMPv3 on IGMP Snooping

- IGMPv3 reports sent to separate group (224.0.0.22)
  - Switches listen to just this group
  - Only IGMP traffic—no data traffic
  - Substantially reduces load on switch CPU
  - Permits low-end switches to implement IGMPv3 snooping
- No report suppression in IGMPv3
  - Enables individual member tracking
- IGMPv3 supports source-specific includes/excludes

IGMP Snooping Summary

IGMP Snooping

- **Switches with Layer 3-aware hardware/ASICs**
  - **High-throughput performance maintained**
  - **Increases cost of switches**
- **Switches without Layer 3-aware hardware/ASICs**
  - **Suffer serious performance degradation or even meltdown**
  - **However, shouldn't be a problem when IGMPv3 is implemented**

Multicast Routing

Multicast Routing Is Backwards from Unicast Routing

- Unicast routing is concerned about where the packet is going
- Multicast routing is concerned about where the packet came from

Unicast vs. Multicast Forwarding

Unicast Forwarding

- Destination IP address directly indicates where to forward packet
- Forwarding is hop-by-hop
  - Unicast routing table determines interface and next-hop router to forward packet

Unicast vs. Multicast Forwarding

Multicast Forwarding

- Destination IP address (group) doesn't directly indicate where to forward packet
- Forwarding is connection-oriented
  - Receivers must first be "connected" to the tree before traffic begins to flow
  - Connection messages (PIM joins) follow unicast routing table toward multicast source
  - Build multicast distribution trees that determine where to forward packets
  - Distribution trees rebuilt dynamically in case of network topology changes

PIM Protocols

Enabling Multicasting

- **To enable multicast**
  - **ip multicast-routing**

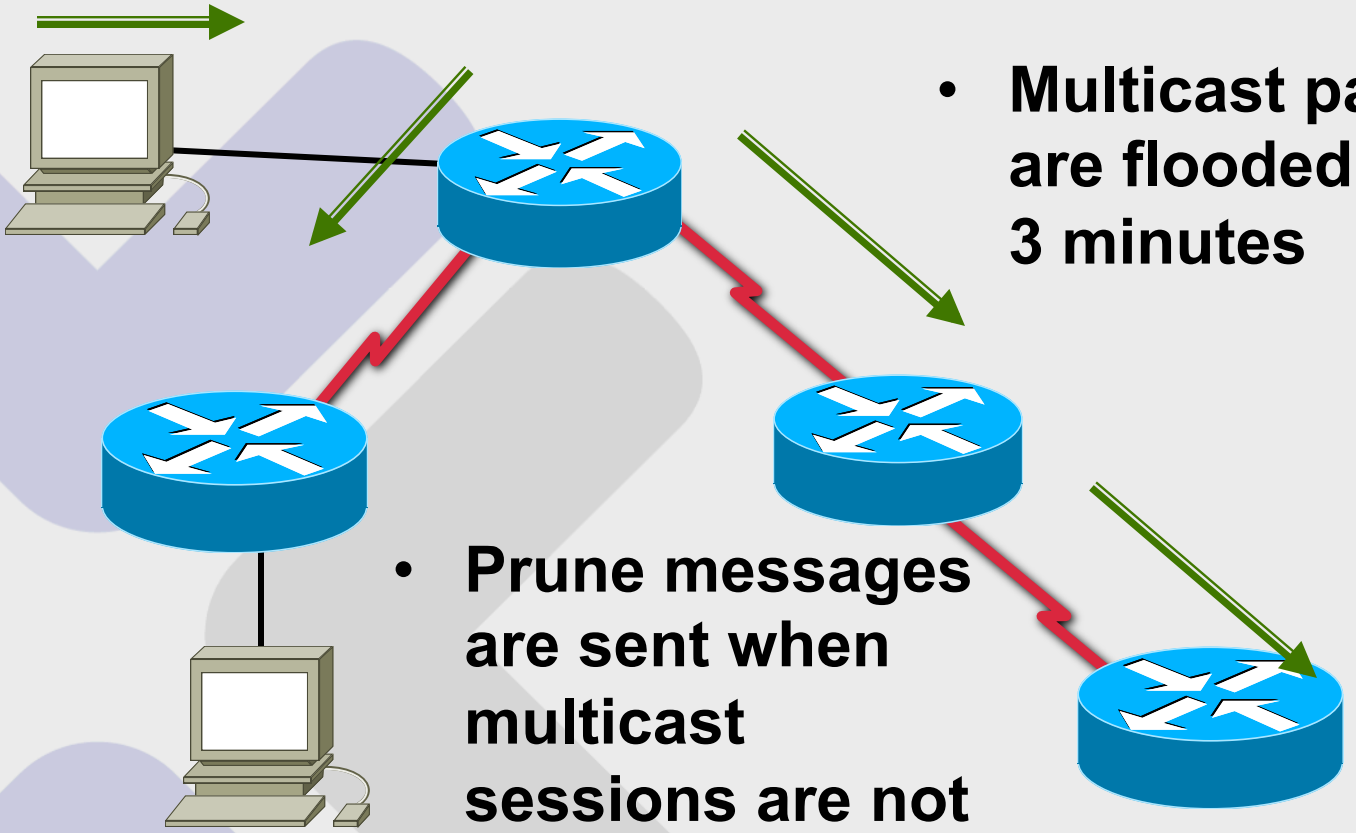
PIM

- There are two PIM modes, Dense and Sparse
  - ip pim dense-mode — enable dense mode on an interface
  - ip pim sparse-mode — enable sparse mode on an interface
  - ip pim sparse-dense-mode — dependent upon the group

PIM Modes

- Dense mode operates on a “flood and prune” basis.
- Multicast packets are initially flooded to all routers.
- Routers that do not have downstream clients who are interested participating in the session send prune messages.

PIM Dense Mode



- Multicast packets are flooded every 3 minutes
- Prune messages are sent when multicast sessions are not needed.

PIM Dense Mode

- PIM Dense Mode State Refresh prevents the reflooding and pruning by sending control messages which update the pruning state
- State Refresh also allow for recognition of typology changes before the three minute interval

State Refresh

- To disable in global configuration
  - ip pim state-refresh disable
- To change the state interval from the default value of 60 seconds
  - ip pim state-refresh origination-interval [*interval*]

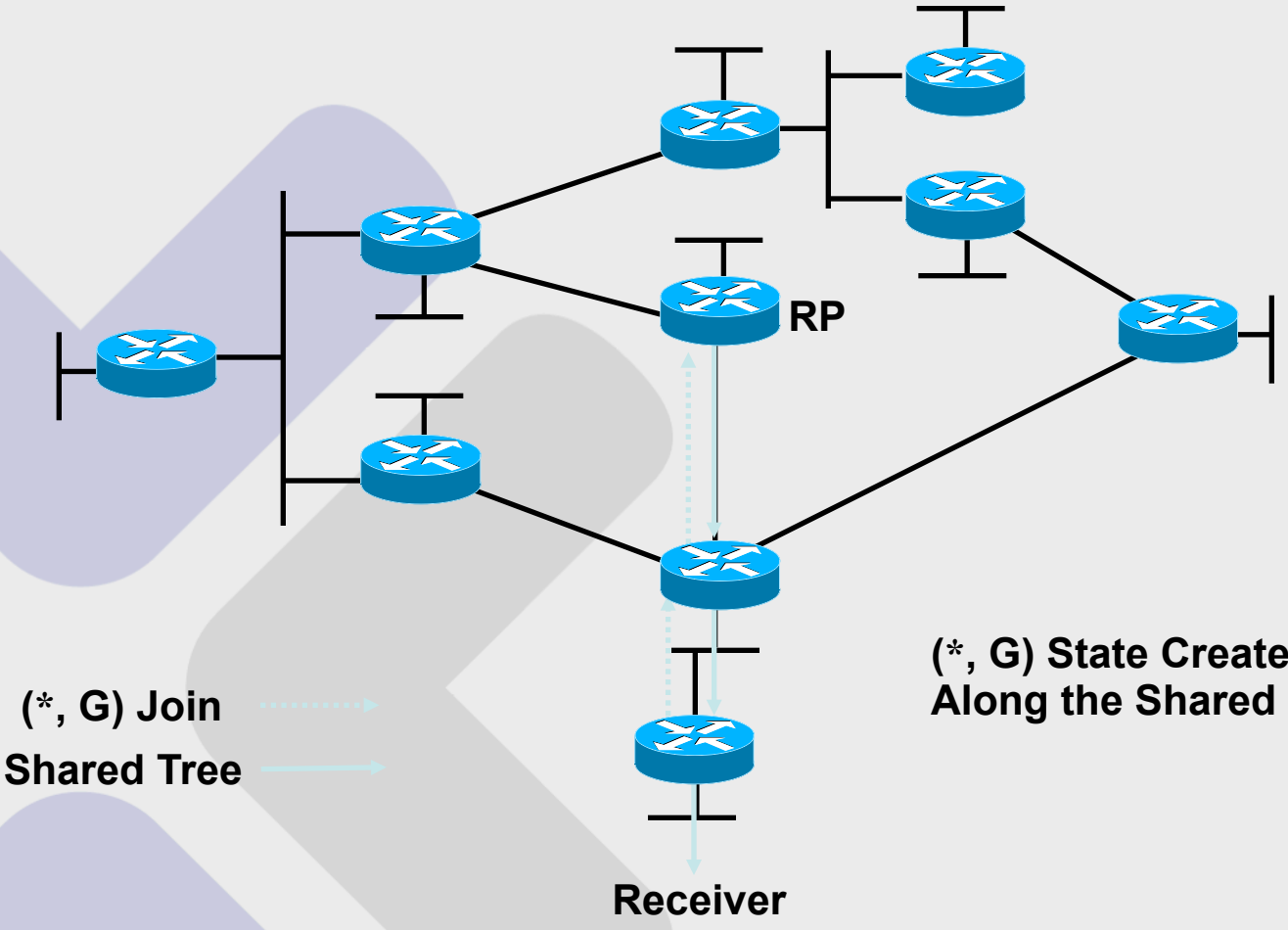
Lab Demonstration

PIM Dense Mode

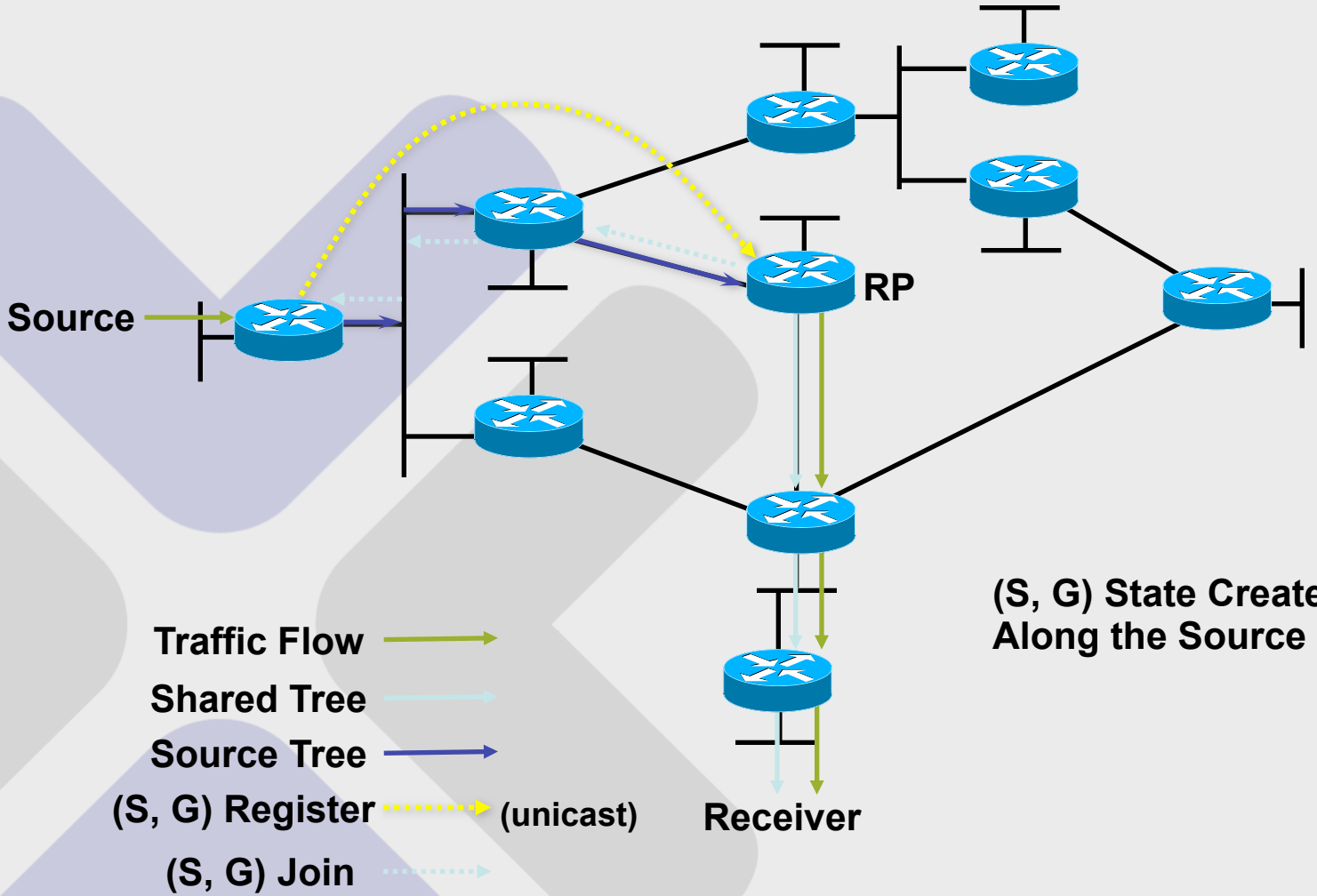
PIM Sparse Mode

- Sparse mode requires the use of a Rendezvous Point (RP)
- First hop routers send PIM register messages on behalf of the sending host to the group
- Last hop routers to send PIM join and prune messages to the RP
- No additional configuration is necessary on the RP

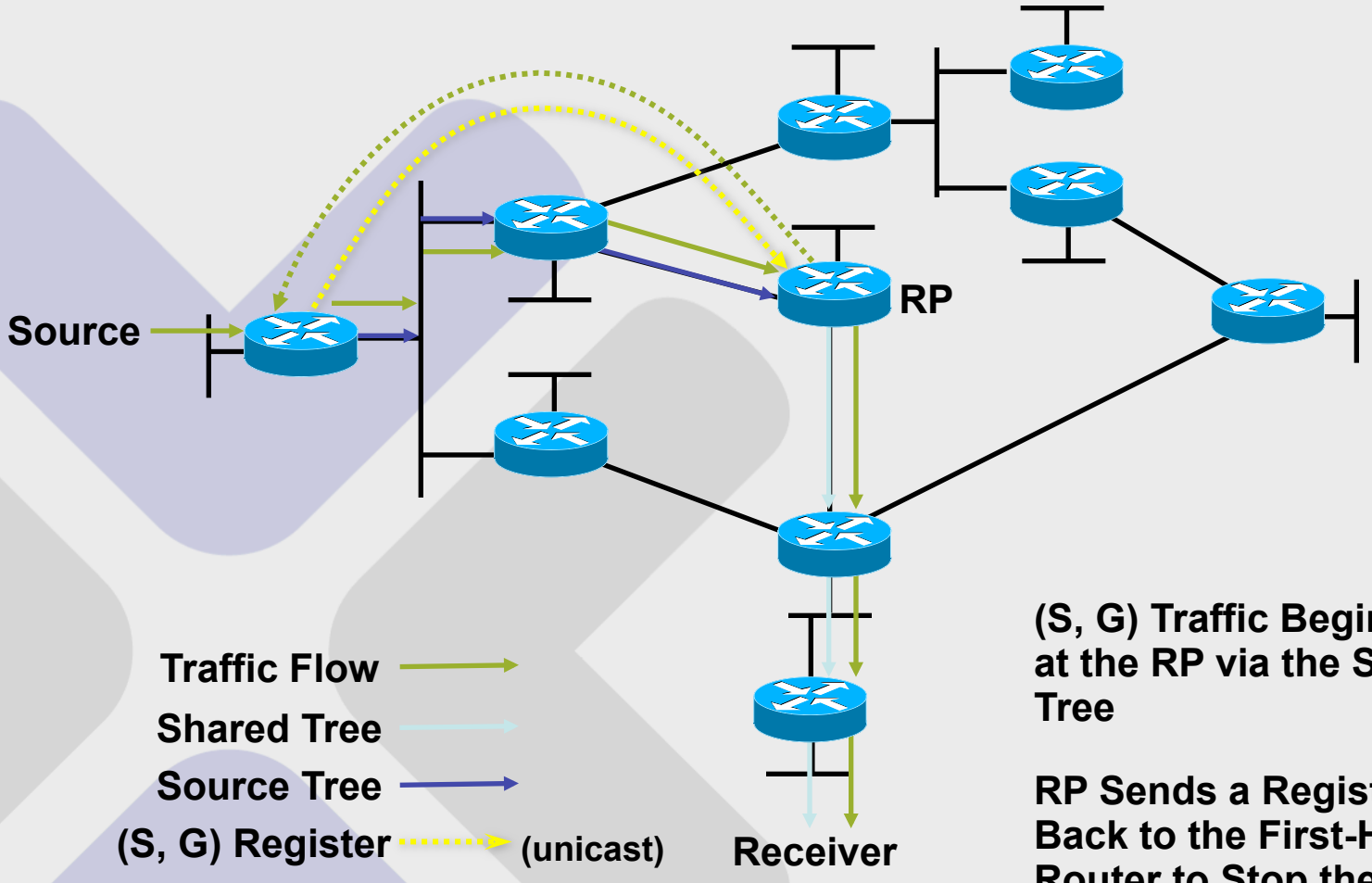
PIM-SM Shared Tree Join



PIM-SM Sender Registration



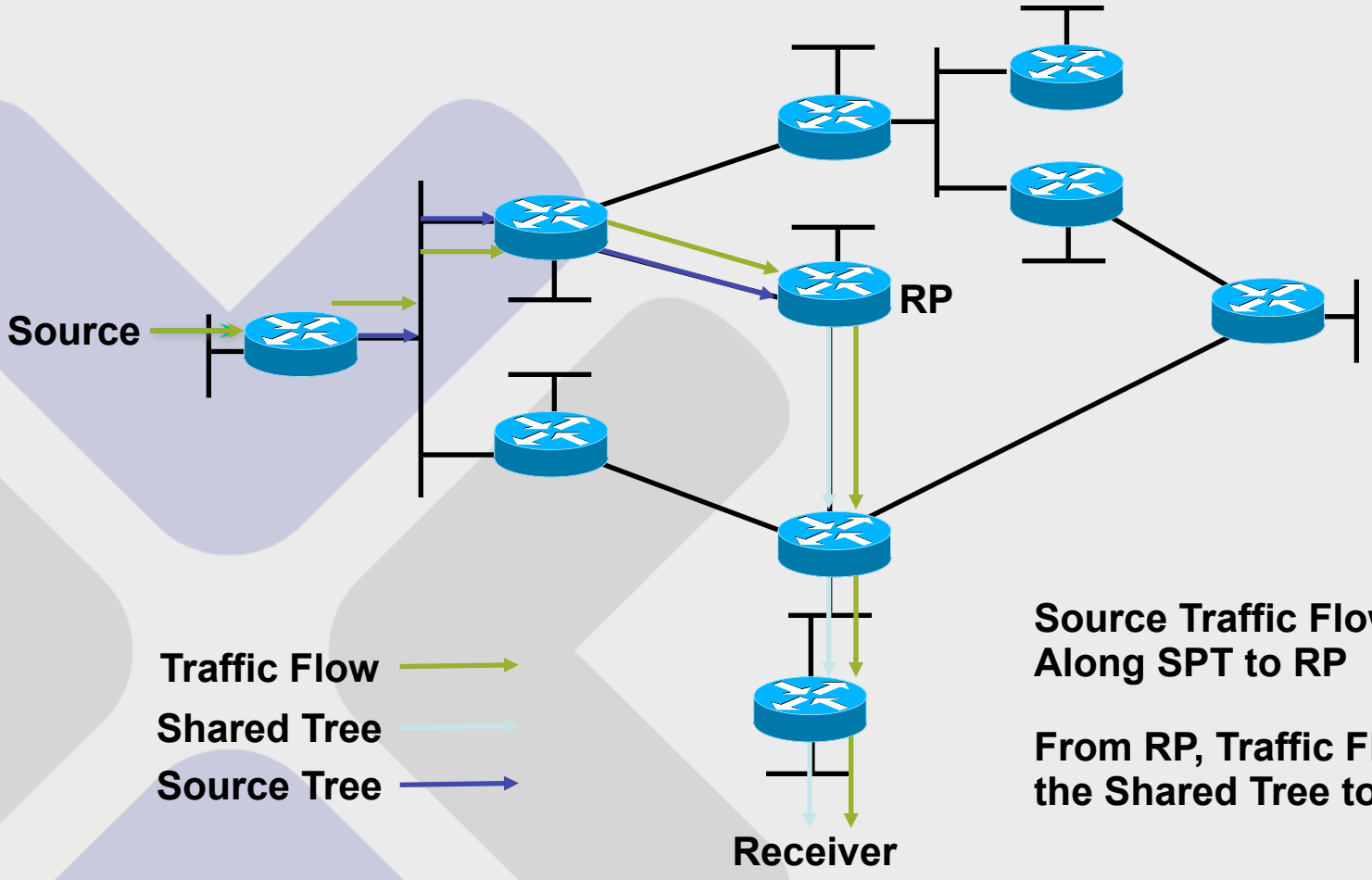
PIM-SM Sender Registration



(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

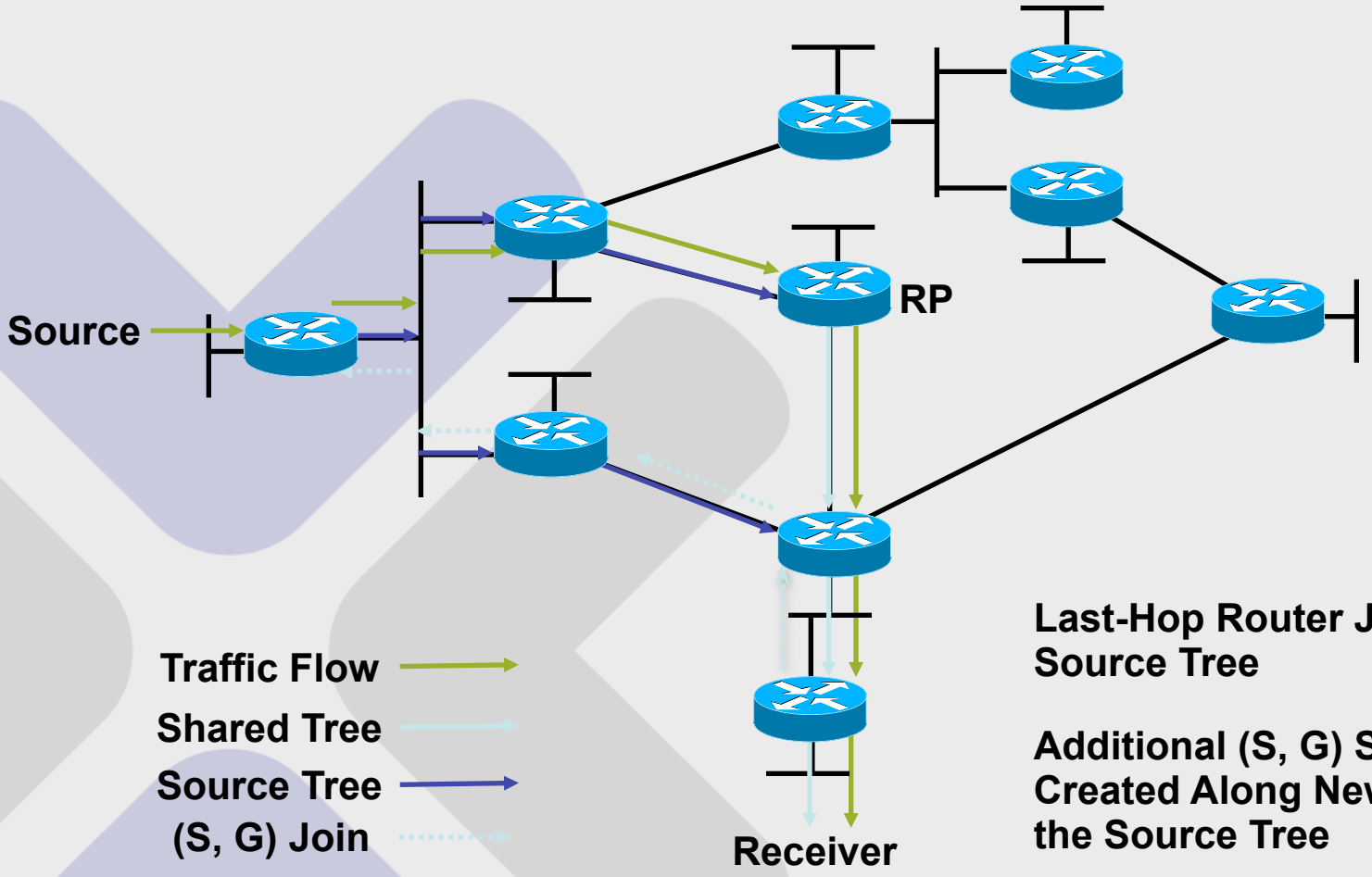
PIM-SM Sender Registration



Source Traffic Flows Natively Along SPT to RP

From RP, Traffic Flows Down the Shared Tree to Receivers

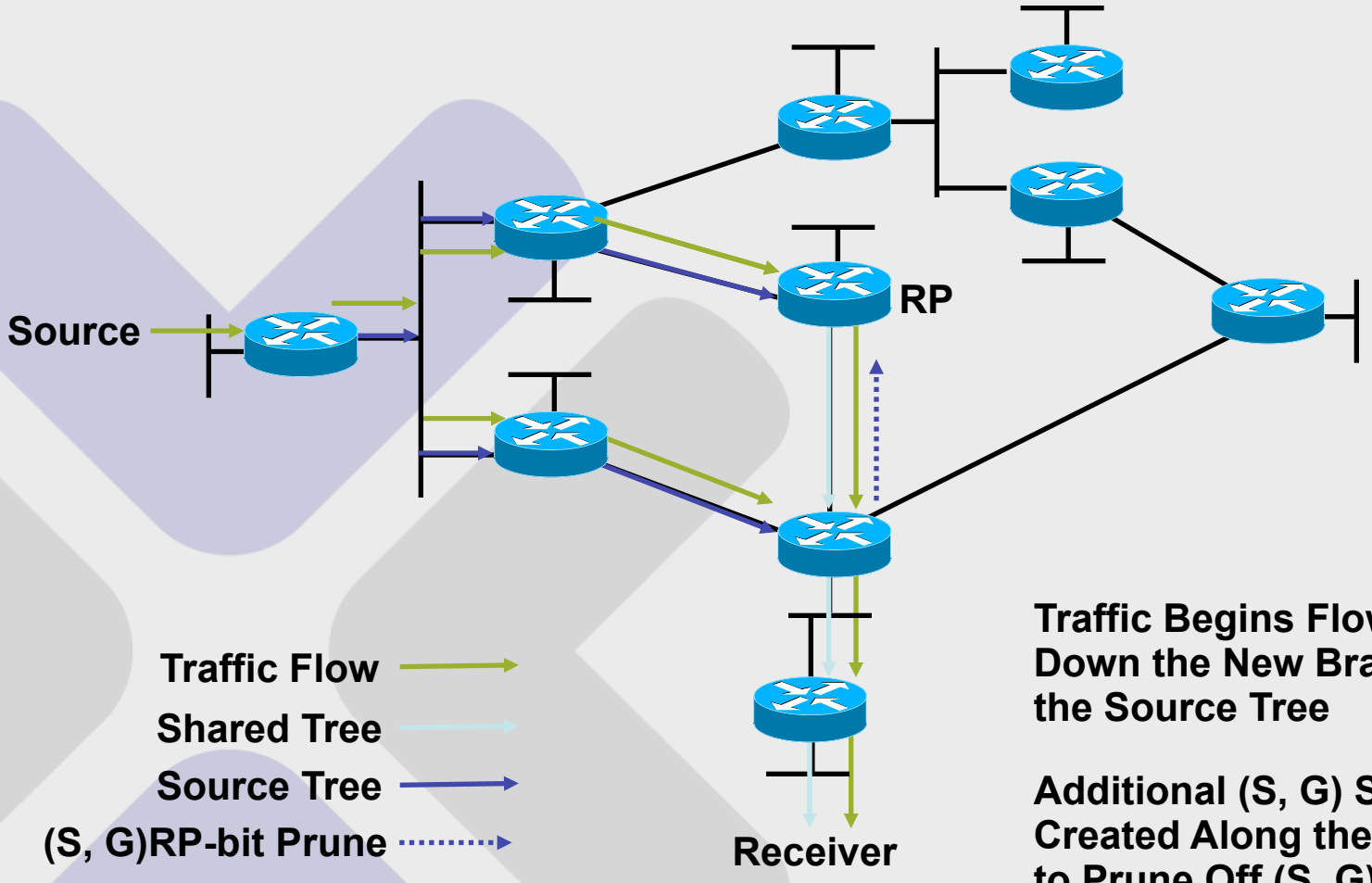
PIM-SM SPT Switchover



Last-Hop Router Joins the Source Tree

Additional (S, G) State Is Created Along New Part of the Source Tree

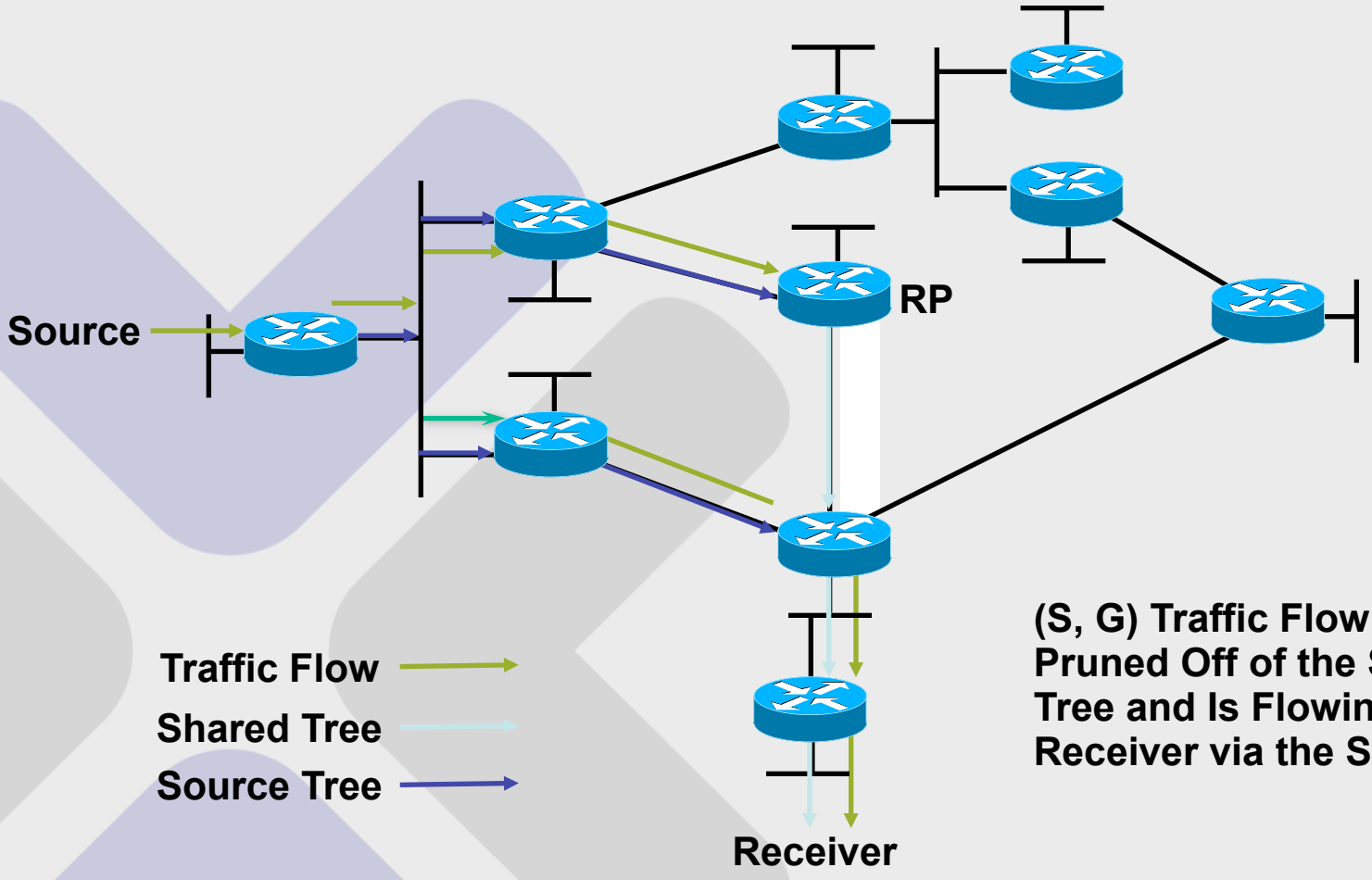
PIM-SM SPT Switchover



Traffic Begins Flowing Down the New Branch of the Source Tree

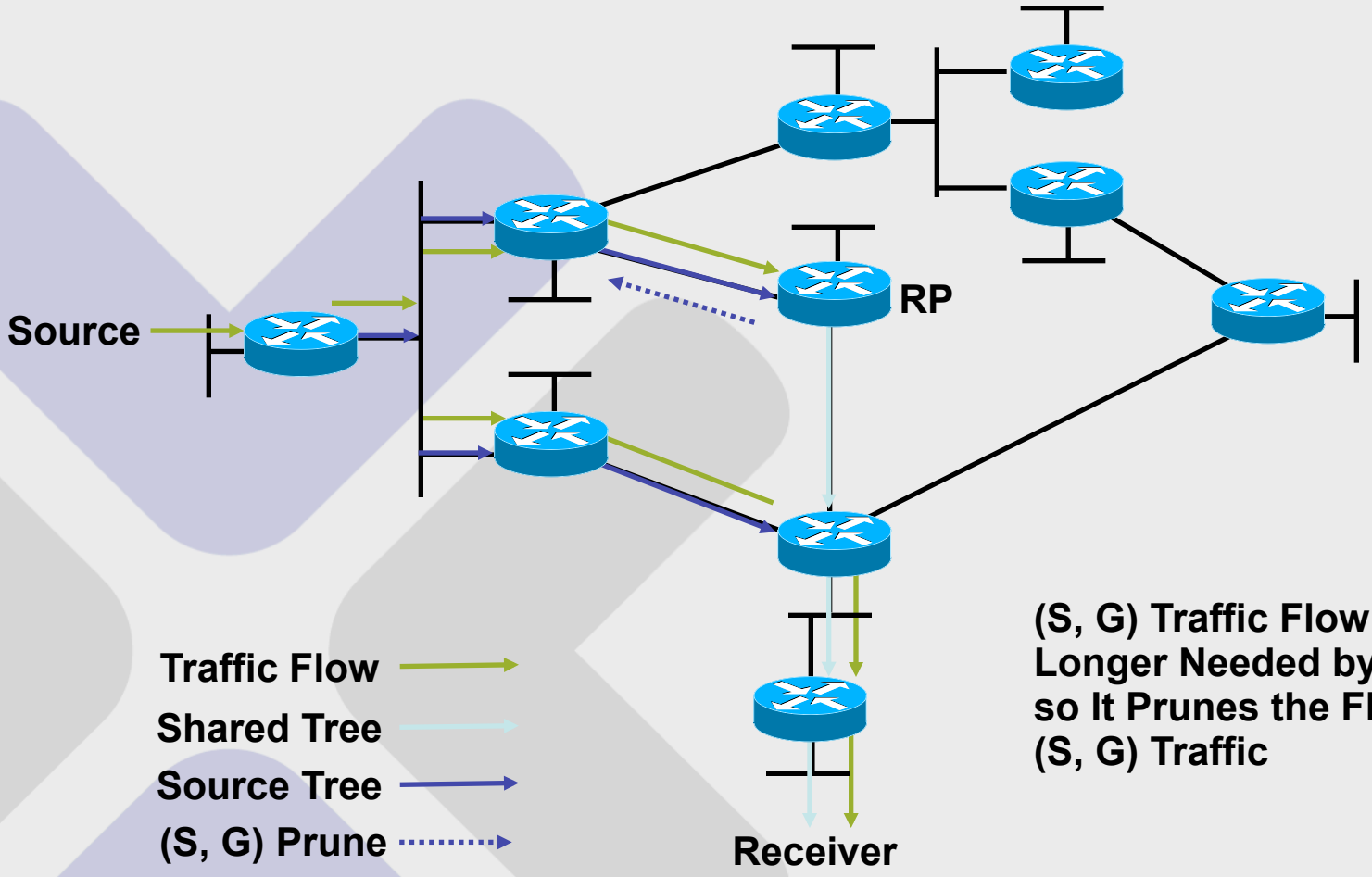
Additional (S, G) State Is Created Along the Shared Tree to Prune Off (S, G) Traffic

PIM-SM SPT Switchover

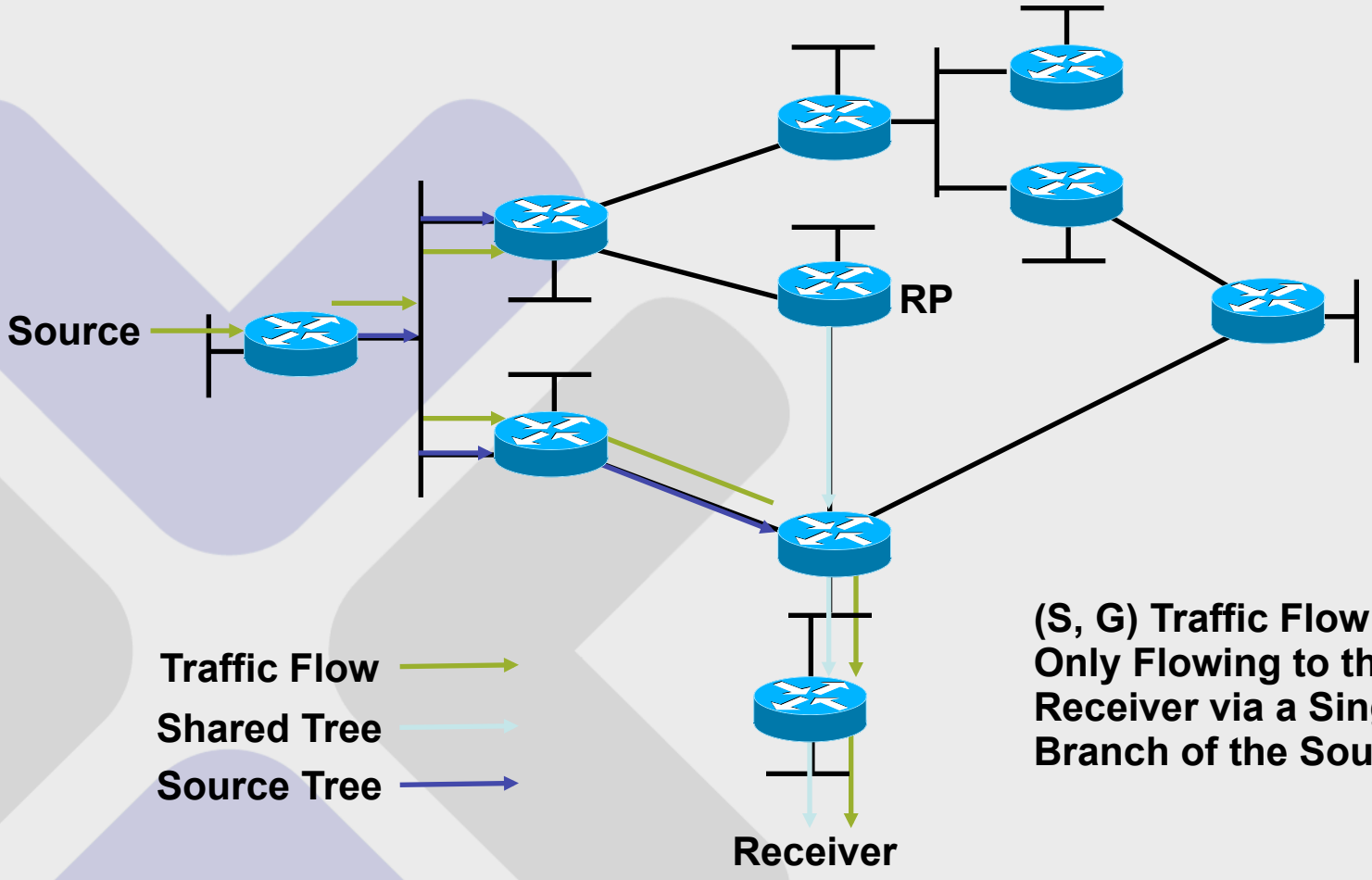


(S, G) Traffic Flow Is Now Pruned Off of the Shared Tree and Is Flowing to the Receiver via the Source Tree

PIM-SM SPT Switchover



PIM-SM SPT Switchover



(S, G) Traffic Flow Is Now Only Flowing to the Receiver via a Single Branch of the Source Tree

“The default behavior of PIM-SM is that routers with directly connected members will join the shortest path tree as soon as they detect a new multicast source.”

PIM-SM Frequently Forgotten Fact

PIM-SM Evaluation

- Effective for sparse or “dense” distribution of multicast receivers
- Advantages
  - Traffic only sent down “joined” branches
  - Can switch to optimal source-trees for high traffic sources dynamically
  - Unicast routing protocol-independent
  - Basis for interdomain, multicast routing
    - When used with MBGP, MSDP and/or SSM

PIM Sparse Mode — RP

- To specify the RP
  - ip pim rp-address *rp-address* [*access-list*] [override]
- The access-list specifies which groups the machine is acting as the RP

PIM-SM ASM RP Requirements

- Group to RP mapping
  - Consistent in all routers within the PIM domain
- RP redundancy requirements
  - Eliminate any single point of failure

Lab Demonstration

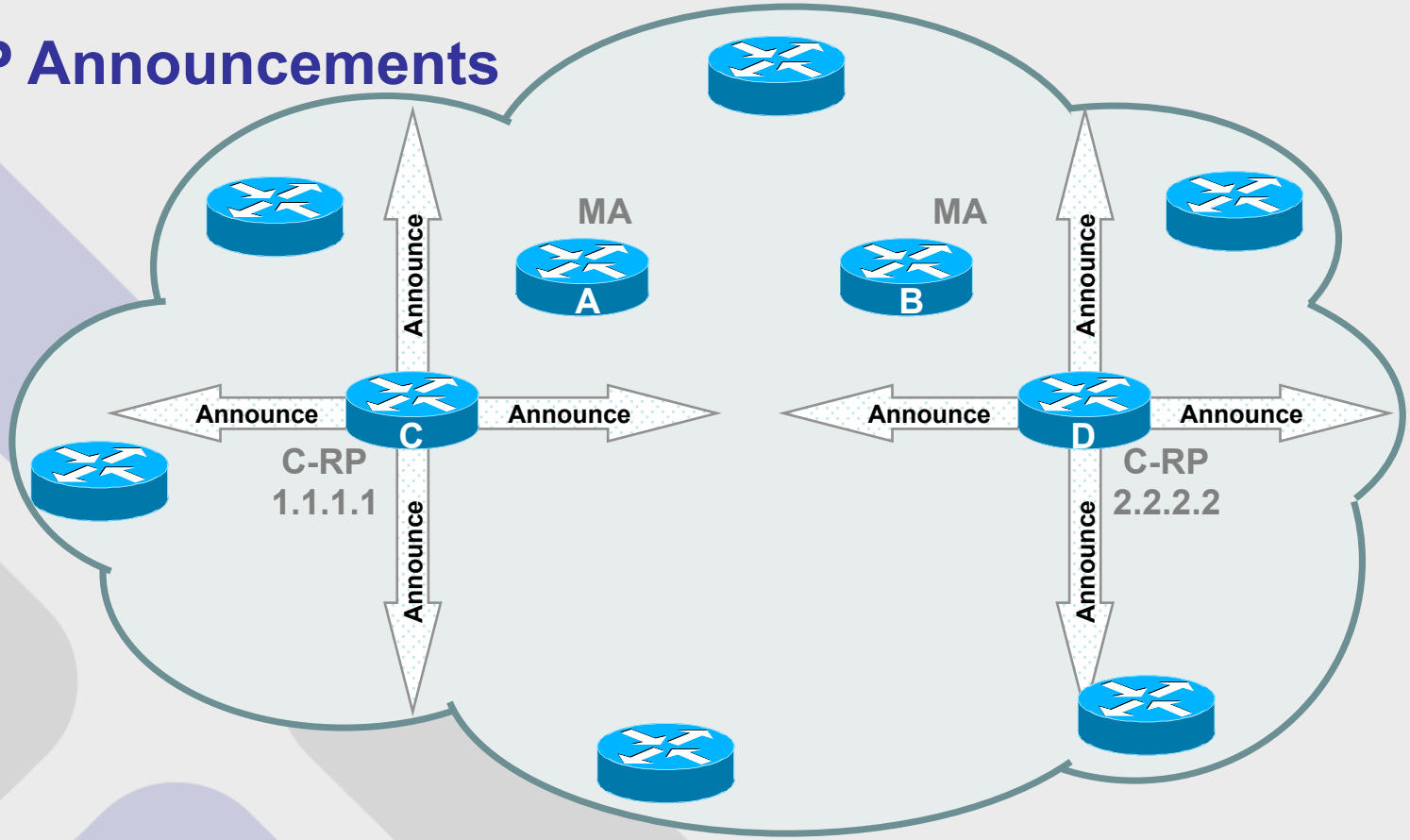
PIM Sparse Mode

How Does the Network Know About the RP?

- Static configuration
  - Manually on every router in the PIM domain
- AutoRP
  - Originally a Cisco® solution
  - Facilitated PIM-SM early transition
- BSR
  - draft-ietf-pim-sm-bsr
- Anycast RP with MSDP

Auto-RP Overview

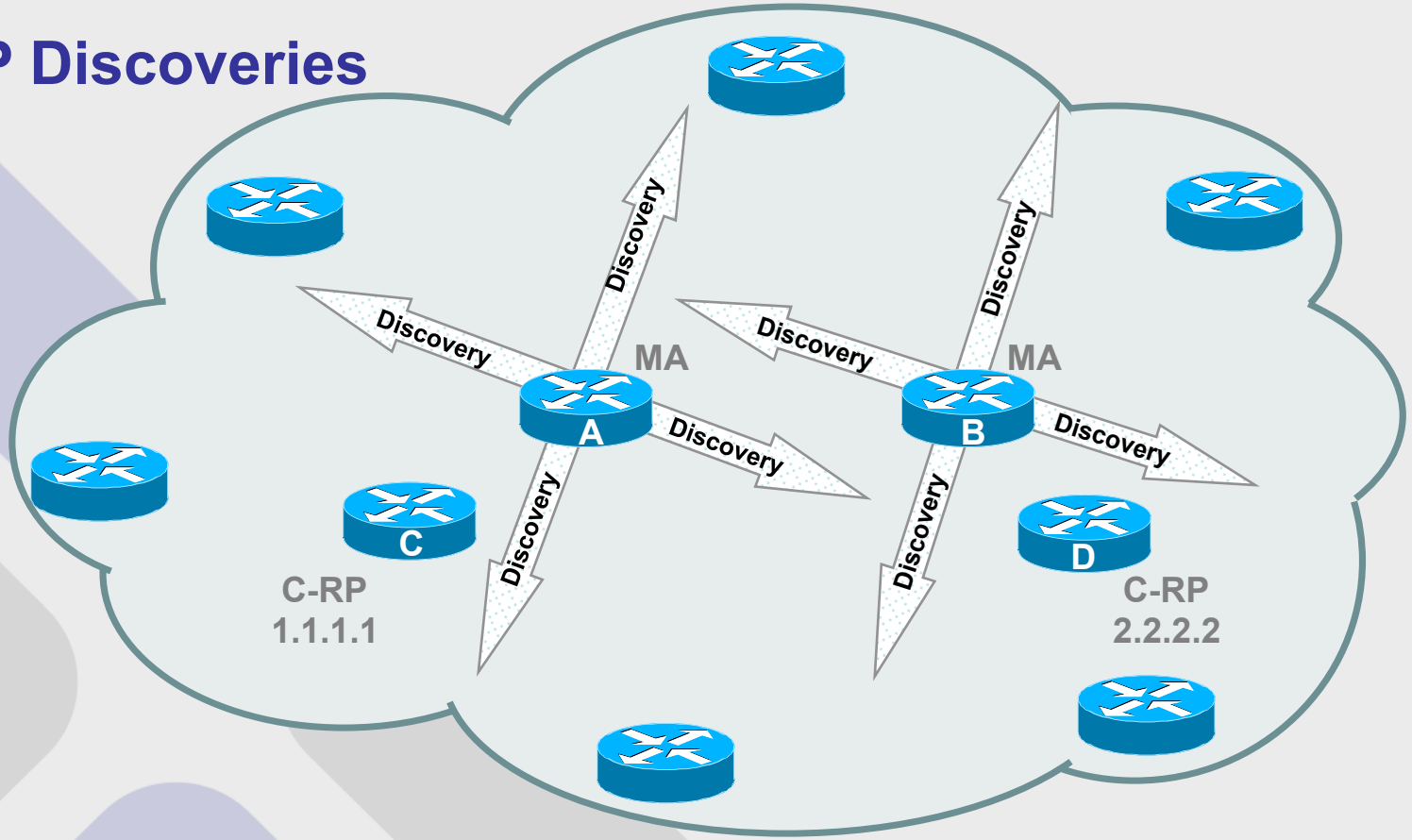
RP Announcements



- RP announcements are multicast to 224.0.1.39 group by C-RPs

Auto-RP Overview

RP Discoveries



- Mapping agent selects RP for each group
- Mapping agents multicast RP discoveries to 224.0.1.30 group

PIM Sparse Mode — Auto RP

- Auto RP allows the dynamic distribution of RP information
- Complex RP configurations are easy to configure
- Allows load splitting between RP
- Avoids configuration errors
- Only works in sparse-dense mode. Sparse mode routers muse use manual RP

PIM Sparse-Dense Mode — Auto RP

- In a network with a combination of Sparse-Dense and Sparse mode routers, use Auto RP for the Sparse-Dense routers and manually configure a default RP for the sparse mode interfaces.
- RP's discovered dynamically take precedence over statically configured RP's.

PIM Sparse-Dense Mode — Auto RP

- **To designate an RP**
  - **ip pim send-rp-announce *type number* scope *ttl-value* [group-list *access-list*] [interval*seconds*]**
- **A permit in the access-list specifies the group is serviced by the RP**

PIM Sparse-Dense — Auto RP

- The RP mapping agent sends the authoritative discovery packets informing other routers the group-to-RP mappings.
  - ip pim send-rp-discovery scope *ttl-value*
- To view the mappings
  - show ip pim rp [mapping | metric] [*rp-address*]

PIM Sparse-Dense — Auto RP

- To accept all RP's advertised via Auto-RP
  - ip pim accept-rp auto-rp
- To filter which auto-rp messages to accept
  - ip pim rp-announce-filter rp-list *access-list* group-list *access-list*

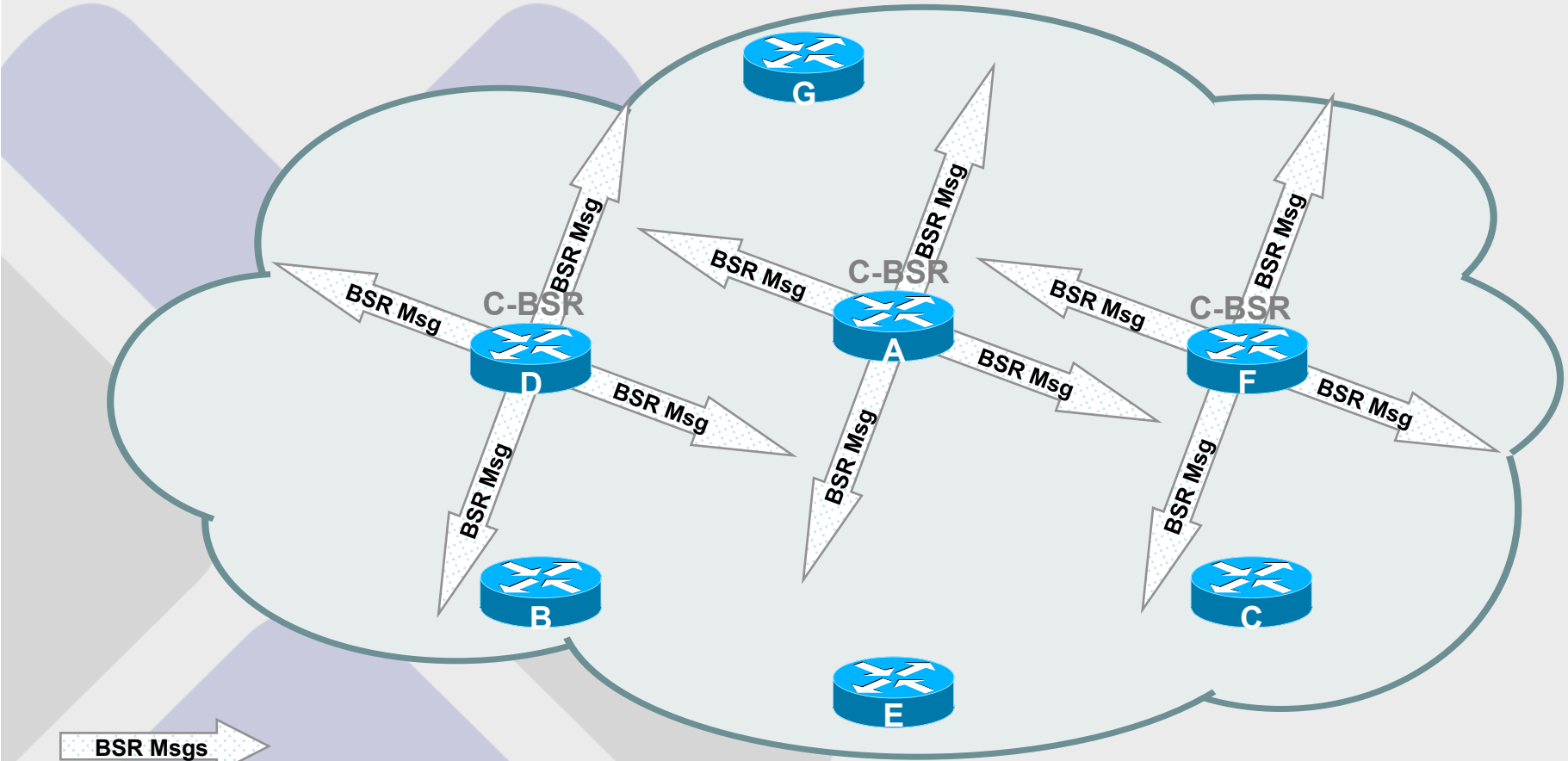
Lab Demonstration

Auto-RP

BootStrap Router (BSR)

BSR Overview

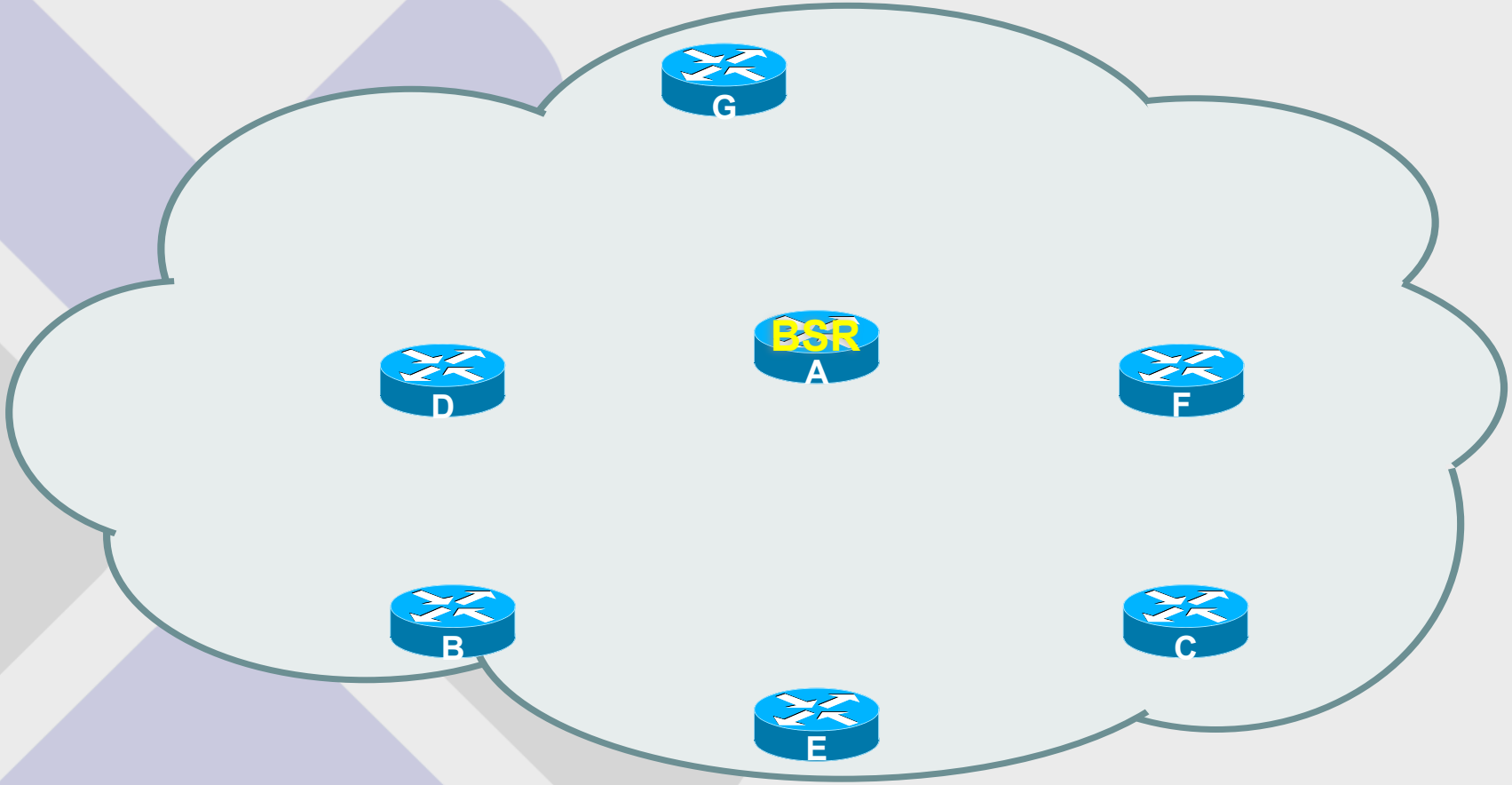
BSR Election Process



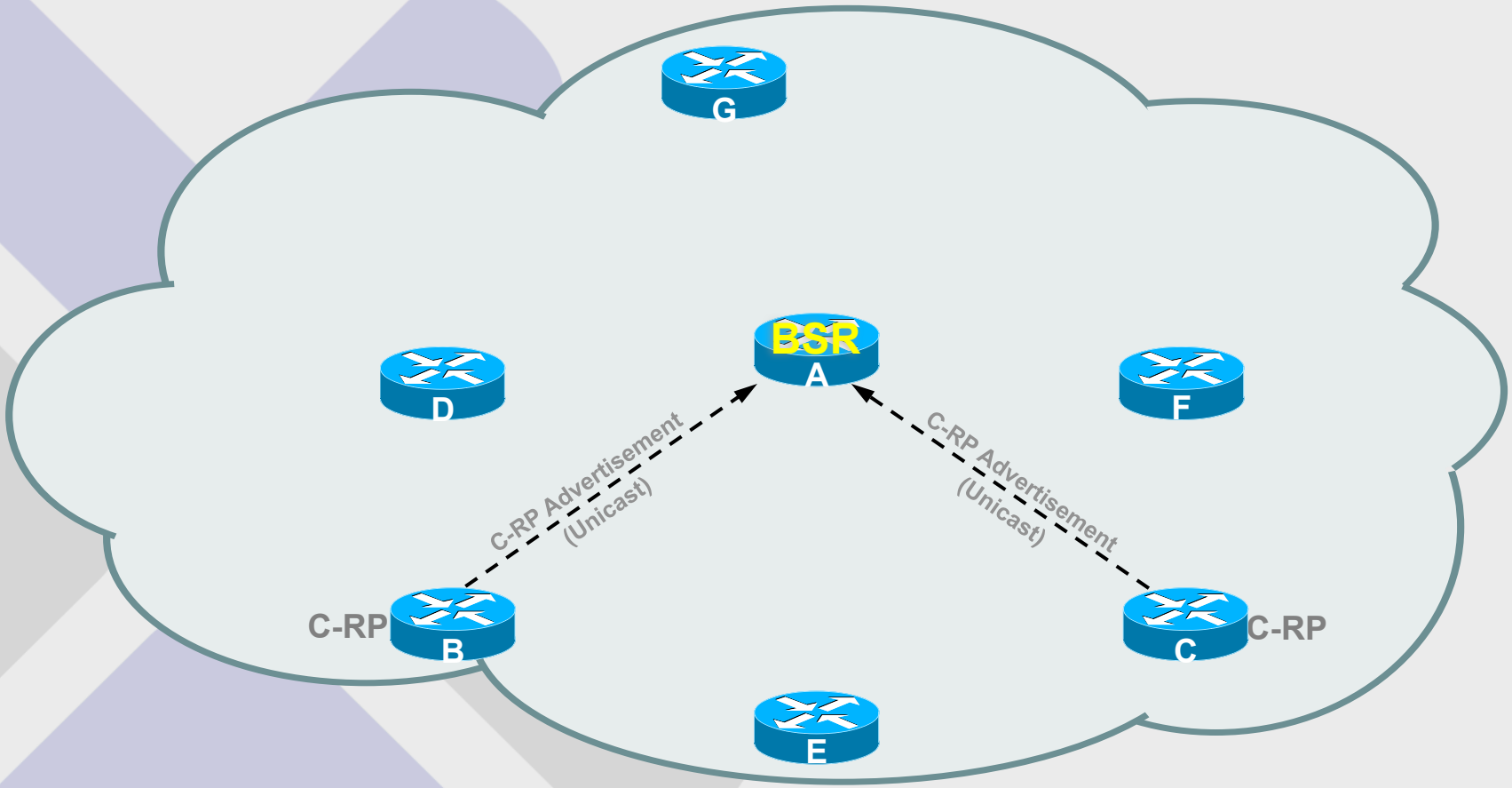
BSR Messages Flooded Hop-by-Hop

BSR Overview

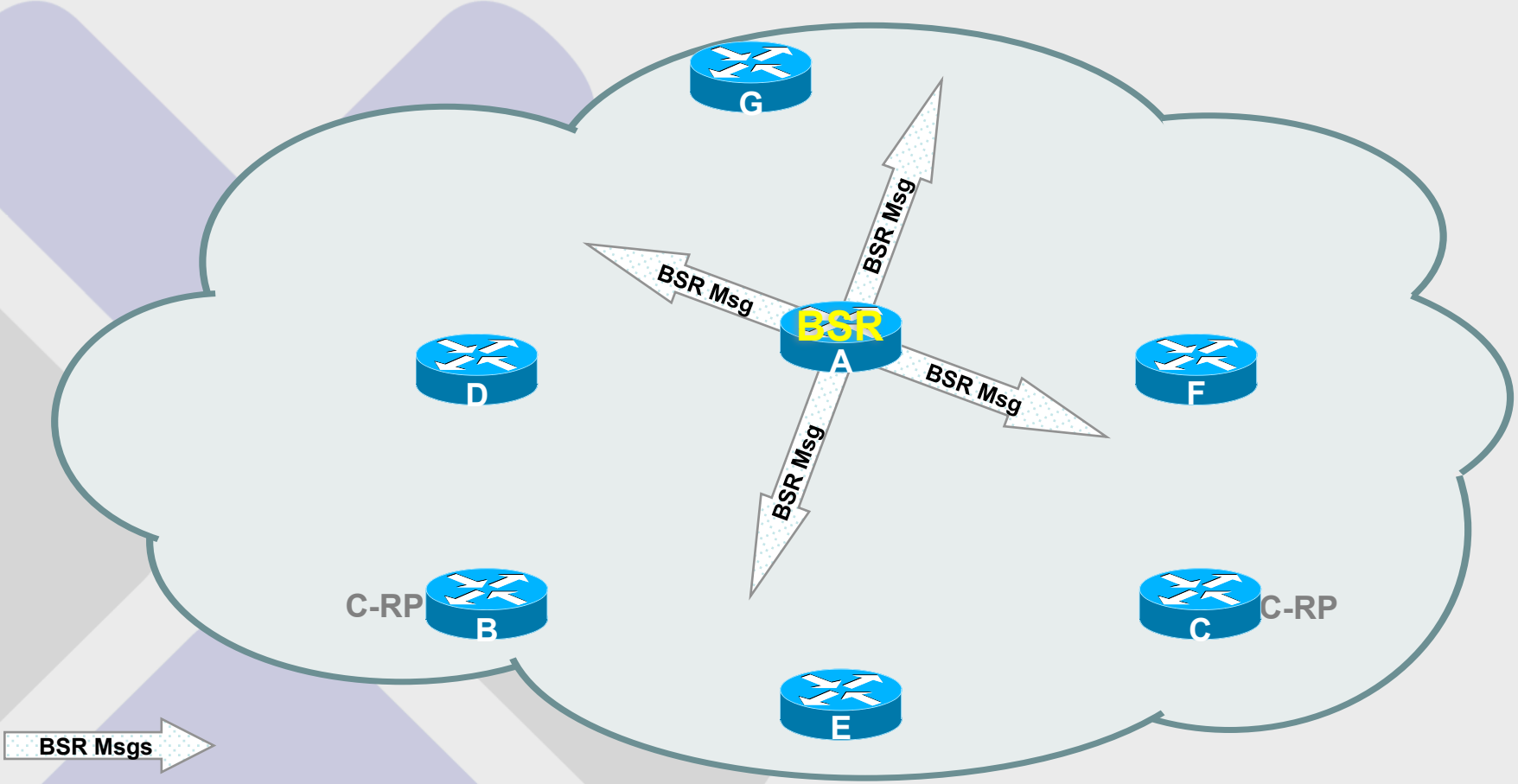
Highest Priority C-BSR Is Elected as BSR



BSR Overview



BSR Overview



BSR Messages Containing RP-Set Flooded Hop-by-Hop

Bootstrap Router

- **To define the bootstrap router**
  - **ip pim bsr-candidate type number *hash-mask-length* [*priority*]**

Defining the RP

- To specify a candidate RP for PIM v2
  - ip pim rp-candidate type number [group-list *access-list*]

Specify PIMv2 Border

- To prevent (bootstrap router) BSR message from traveling outside the domain, specify the BSR border on the interface
  - ip pim bsr-border

Lab Demonstration

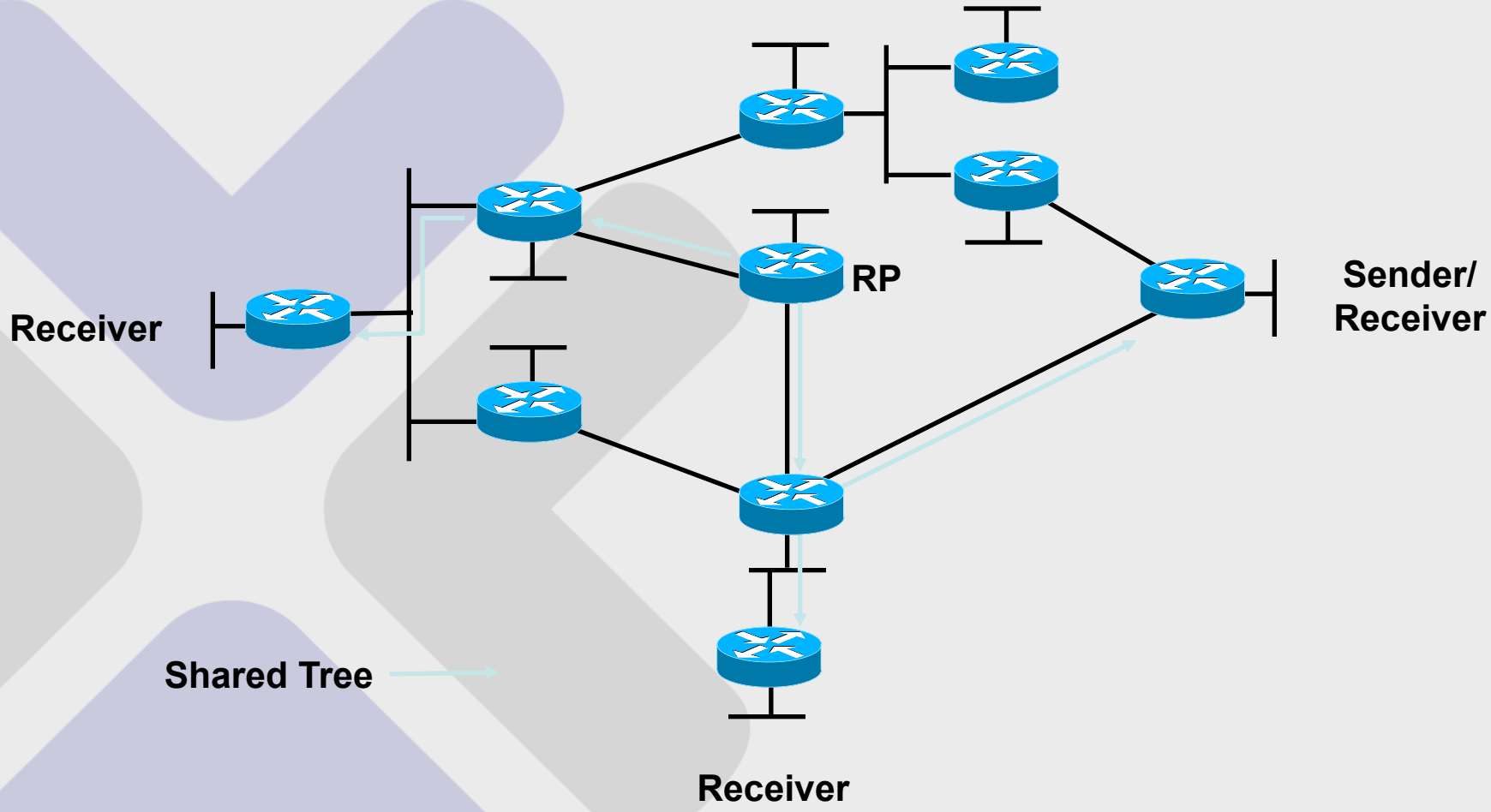
BSR

Bi-Directional PIM

Bidirectional PIM

- Bidirectional PIM does not use encapsulation to communicate with the RP
- Instead the traffic is sent upstream via the shared tree
- SPT are not allowed, all trees are shared
- Each link elects a Designated Forwarder (DF) to prevent loops

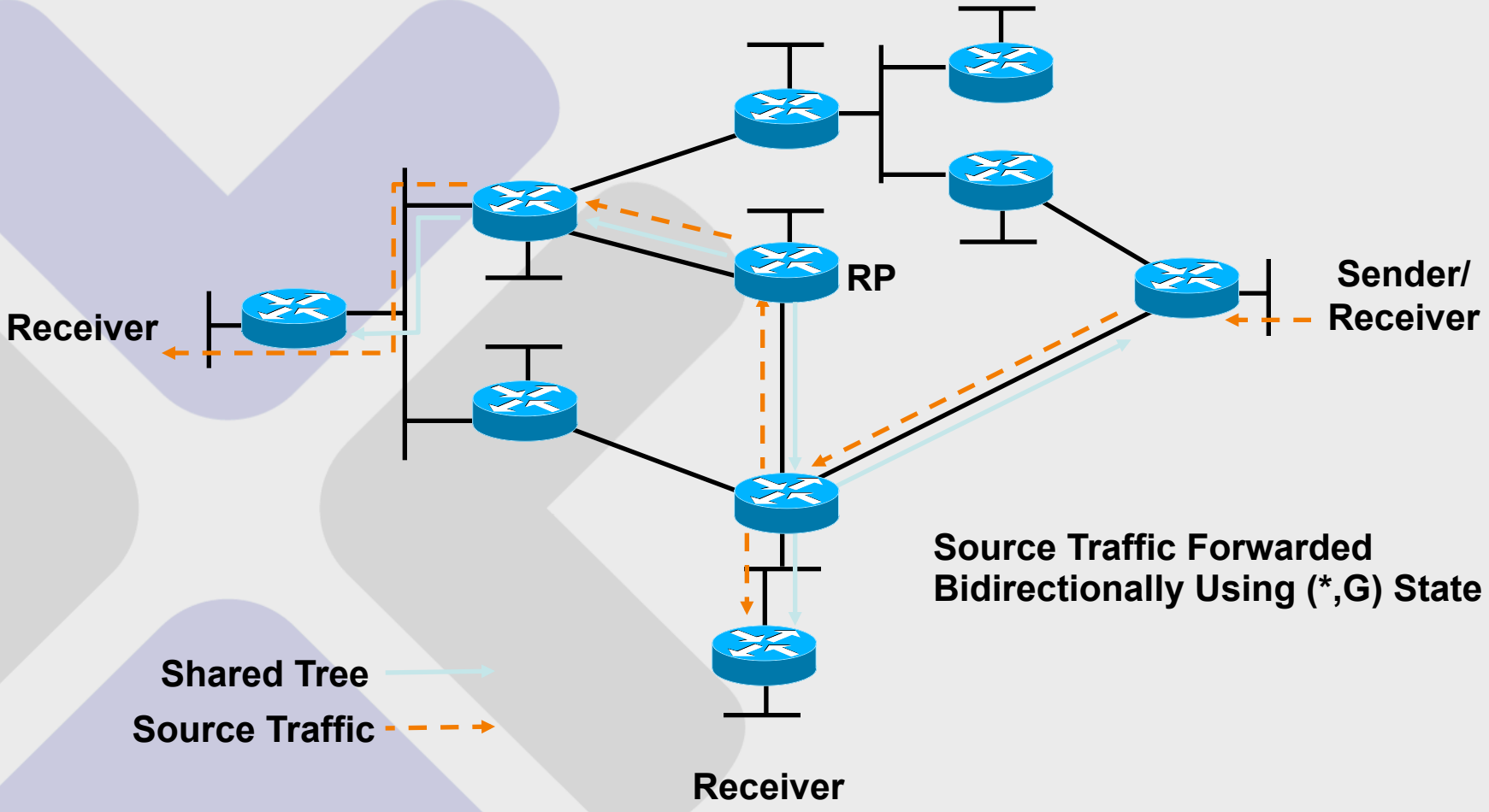
Bidirectional PIM Overview



BiDir PIM Evaluation

- Ideal for many to many applications
- Drastically reduces network mroute state
  - Eliminates all (S,G) state in the network
    - SPTs between sources to RP eliminated
    - Source traffic flows both up and down shared tree
  - Allows many-to-any applications to scale
    - Permits virtually an unlimited number of sources

Bidirectional PIM Overview



Bidirectional PIM

- Bidirectional PIM must be deployed on every router.
- SPT may not be mixed with Bi-PIM.
- DF is used to forward traffic upstream
- Bi-PIM uses normal PIM-SM mechanisms in a shared tree environment for downstream traffic. No switchover to SPT is permitted.

Bidir-PIM

- **Enable Bidir-PIM on the router**
  - **ip pim bidir-enable**
- **Configure RP (when not using auto-RP or BSR)**
  - **ip pim rp-address rp-address [access-list] [override] bidir**

Bidir-PIM

- **Configuring Auto-RP RP**
  - **ip pim rp-candidate type number [group-list access-list] bidir**
- **Configuring BSR RP**
  - **ip pim send-rp-announce *type number* scope *ttl-value* [group-list *access-list*] [interval *seconds*] bidir**

Bidir-PIM Verification

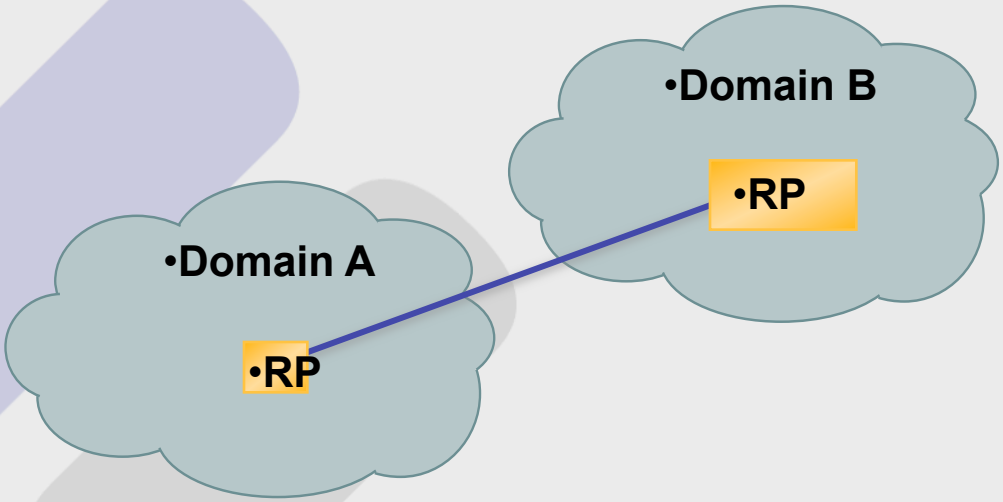
- **View the DF election**
  - **show ip pim interface [*type number*] [df | count] [*rp-address*]**
- **View Mapping**
  - **show ip pim rp [mapping | metric] [*rp-address*]**

Lab Demonstration

Bi-Directional PIM

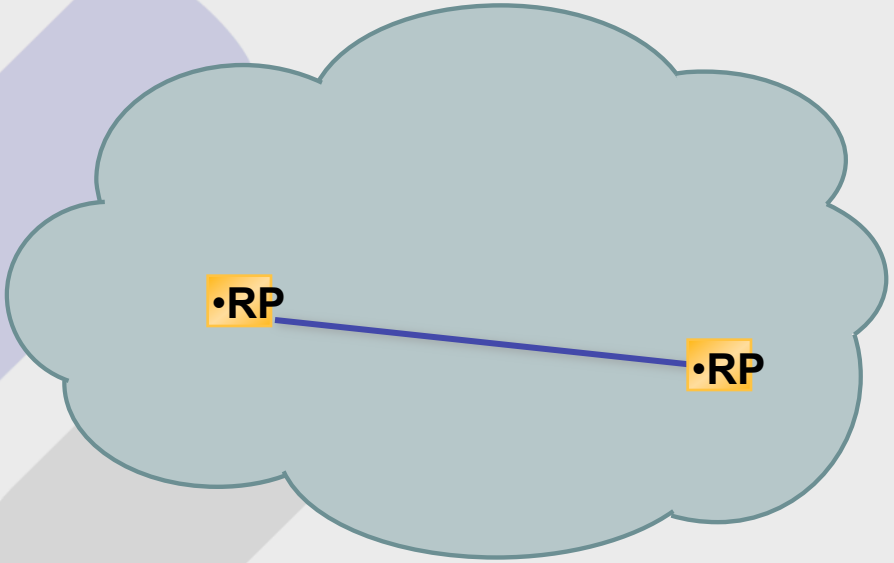
Multicast Source Discovery Protocol (MSDP)

MSDP



- Allows RPs to share information between domains

MSDP



- **Allows for backup RPs within an area**

MSDP Overview

- MSDP connects multiple Protocol Independent Multicast sparse mode (PIM-SM) domains
- A RP in a PIM-SM domain has an MSDP peering relationship with MSDP-enabled routers in another domain via TCP.
- Each domain is not depend on RPs in other domains

MSDP Overview

- When a source is registered with the RP the packet is reencapsulated in a Source-Active (SA) message that is forwarded to all MSDP peers.
- The SA message contains the source, the group the source is sending to, and the address or the originator ID of the RP.
- The peer forwards the SA down the shared-tree.

MSDP Configuration

- **Configure an MSDP peer**
  - **ip msdp peer {*peer-name* | *peer-address*} [connect-source type number] [remote-as as-number]**
- **To prevent delay, save SA state information**
  - **ip msdp cache-sa-state [list access-list]**
- **Allow other routers to query SA information from cache**
  - **ip msdp sa-request {*peer-address* | *peer-name*}**

MSDP Filtering

- To filter which (S,G) pairs received from a RP are forwarded via SA
  - ip msdp redistribute [list access-list] [asn as-access-list] [route-map map-name]

MSDP Filtering

- To filter SA Requests being received from the cache server
  - ip msdp filter-sa-request {*peer-address* | *peer-name*} list access-list
- Or per device
  - ip msdp filter-sa-request {*peer-address* | *peer-name*}

MSDP SA Filtering

- To filter which SA messages are forwarded
  - ip msdp sa-filter out {*peer--address* | *peer-name*} list access-list
- Or with a router map
  - ip msdp sa-filter out {*peer-address* | *peer-name*} route-map map-name

MSDP SA Inbound Filtering

- To block which SA messages are received from a peer
  - ip msdp sa-filter in {*peer-address* | *peer-name*} list access-list

Or

  - ip msdp sa-filter in {*peer-address* | *peer-name*} route-map map-name

Anycast

- Anycast offers the ability to have redundant RP in a network
- Both RP's are given the same IP address.
- Clients point to the common IP address
- RP's communicate via MSDP
- Allows for fault tolerance and load sharing

Anycast Example

- **Create a loopback address with a common IP address**
  - **ROUTER A:**

```
interface loopback 0
 ip address 10.1.1.1 255.255.255.255
interface loopback 1
  ip address 192.168.1.2 255.255.255.255
ip msdp peer 192.168.1.1 connect-source lo1
ip msdp originator-id lo1
```

Anycast Example

- **Router B contains a duplicate IP address for Loopback 0**
  - **ROUTER B:**

```
interface loopback 0
 ip address 10.1.1.1 255.255.255.255
interface loopback 1
  ip address 192.168.1.1 255.255.255.255
ip msdp peer 192.168.1.2 connect-source lo1
ip msdp originator-id lo1
```

Anycast Example

- **All other routers are configured with the RP as the duplicated IP address**
  - **ROUTER C:**

```
ip pim rp-address 10.1.1.1
```

Anycast Notes

WARNING

- OSPF does not allow duplicate Router-ID's. Manually set the router ID to be that of the peering IP address.
- BGP Router ID's must be the same as OSPF and the same as the MSDP peering relationship

Lab Demonstration

MSDP with Anycast

Source Specific Multicast

Source Specific Multicast

- Allows large scale multicast distributions without requiring the network to maintain a list of active sources for a multicast group
- Source information is obtained via out-of-band methods such as a website
- Cisco is developing “URD”, a protocol that allows the router to intercept SSM information and reconfigure without requiring a special SSM enabled application
- 232.0.0.0/8 is reserved for SSM
- Required IGMP v3

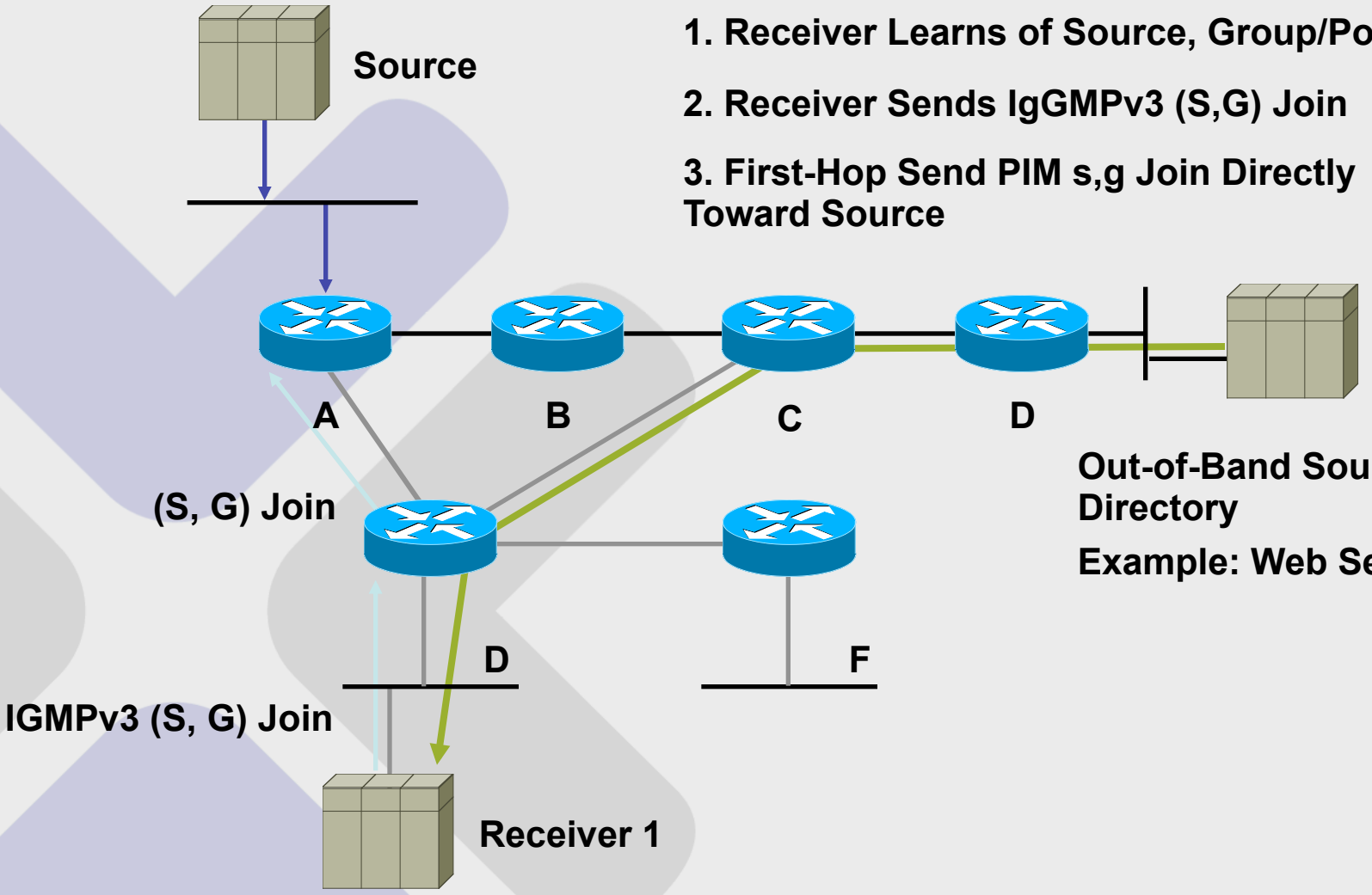
Source Specific Multicast

- Assume a one-to-any multicast model
  - Example: video/audio broadcasts, stock market data
- Why does ASM need a shared tree?
  - So that hosts and first hop routers can learn who the active source is for the group—source discovery
- What if this was already known?
  - Hosts could use IGMPv3 to signal exactly which (S, G) SPT to join
  - The shared tree and RP wouldn't be necessary
  - Different sources could share the same group address and not interfere with each other
- Result: Source Specific Multicast (SSM)

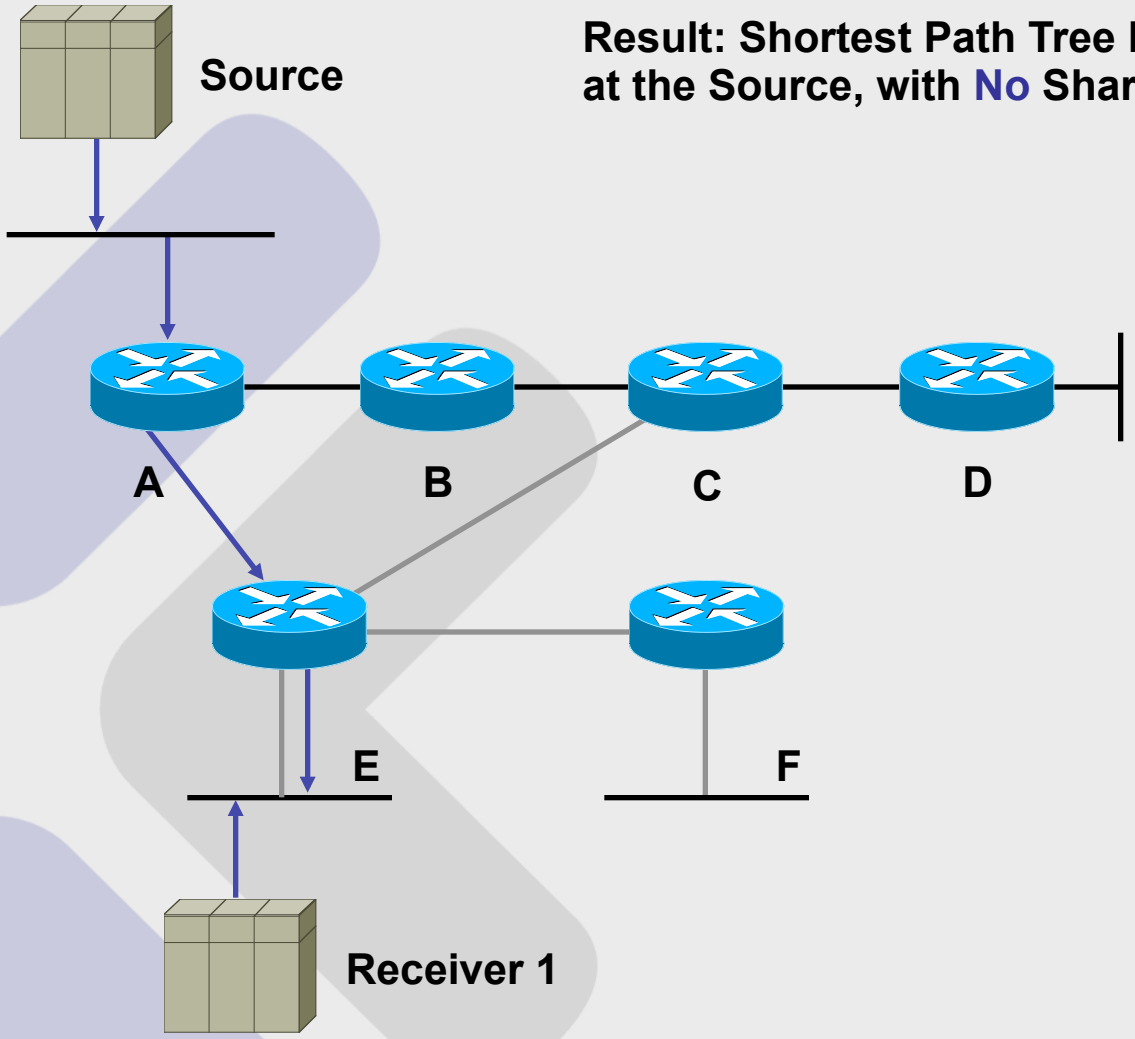
Source Specific Multicast

- SSM uses source trees only.
  - Receivers are responsible for source and group discovery.
  - Receivers select what traffic they want from a group.
  - Receivers use IGMPv3 to signal which (S,G) to join.
  - RP and shared trees are not needed in the network.
- SSM solves multicast address allocation problems.
  - Flows differentiated by both source and group.
  - Content providers can use same group ranges.
    - Each (S,G) flow is unique.
    - Only explicitly request flows are forwarded to receivers.
- RFC 3569: An Overview of Source Specific Multicast (SSM)

PIM Source Specific Mode



PIM Source Specific Mode



SSM Evaluation

- Ideal for applications with one source sending to many receivers
- Uses a simplified subset of the PIM-SM protocol
  - Simpler network operation
- Solves multicast address allocation problems
  - Flows differentiated by both source and group, not just by group
  - Content providers can use same group ranges since each (S,G) flow is unique
- Helps prevent certain DoS attacks
  - "Bogus" source traffic can't consume network bandwidth so not received by host application

Many-to-Many State Problem

- Creates huge amounts of (S,G) state
  - State maintenance workloads skyrocket
    - High OIL fan-out makes the problem worse
  - Router performance begins to suffer
- Using shared trees only
  - Provides some (S, G) state reduction
    - Results in (S, G) state only along SPT to RP
    - Frequently still too much (S, G) state
    - Need a solution that only uses (*, G) state

Configuring Source Specified Multicast

- Define which addresses are used for SSM (232.0.0.0/8 is default)
  - ip pim ssm [default | range access-list]
- Enable IGMP v3 on the interfaces
  - ip igmp version 3
  - Optionally enable URD on interface
  - ip urd

Questions?