# Introduction to IP Multicast

RST-1261

Cisco Networkers poweredbycisco.
2006

IP Multicast at Networkers 2006

- **RST-1261 Introduction to IP Multicast**
- **RST-2261 Deploying IP Multicast**
- **RST-2262 Multicast Security**
- **RST-2263 Multicast Network Management**
- **RST-3261 Advanced IP Multicast**
- **RST-3262 IP Multicast Architecture & Troubleshooting for the Catalyst 6500**
- **TECRST-1008 Enterprise IP Multicast**

Cisco Public

Session Goal

To provide you with a thorough understanding of the concepts, mechanics and protocols used to build IP Multicast networks
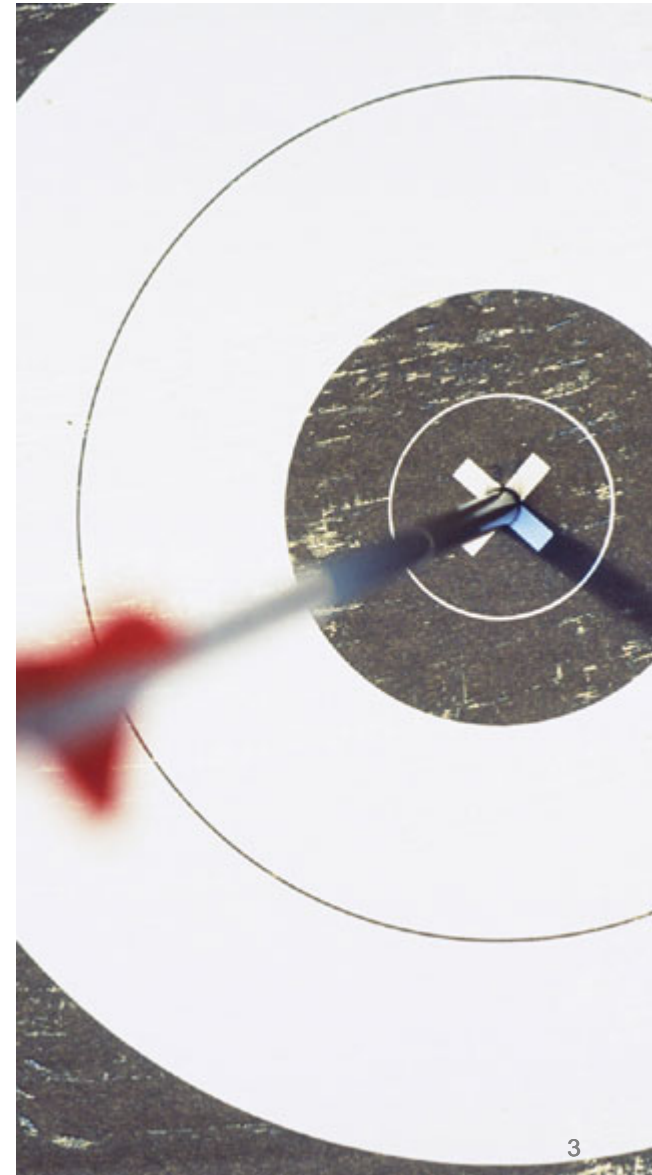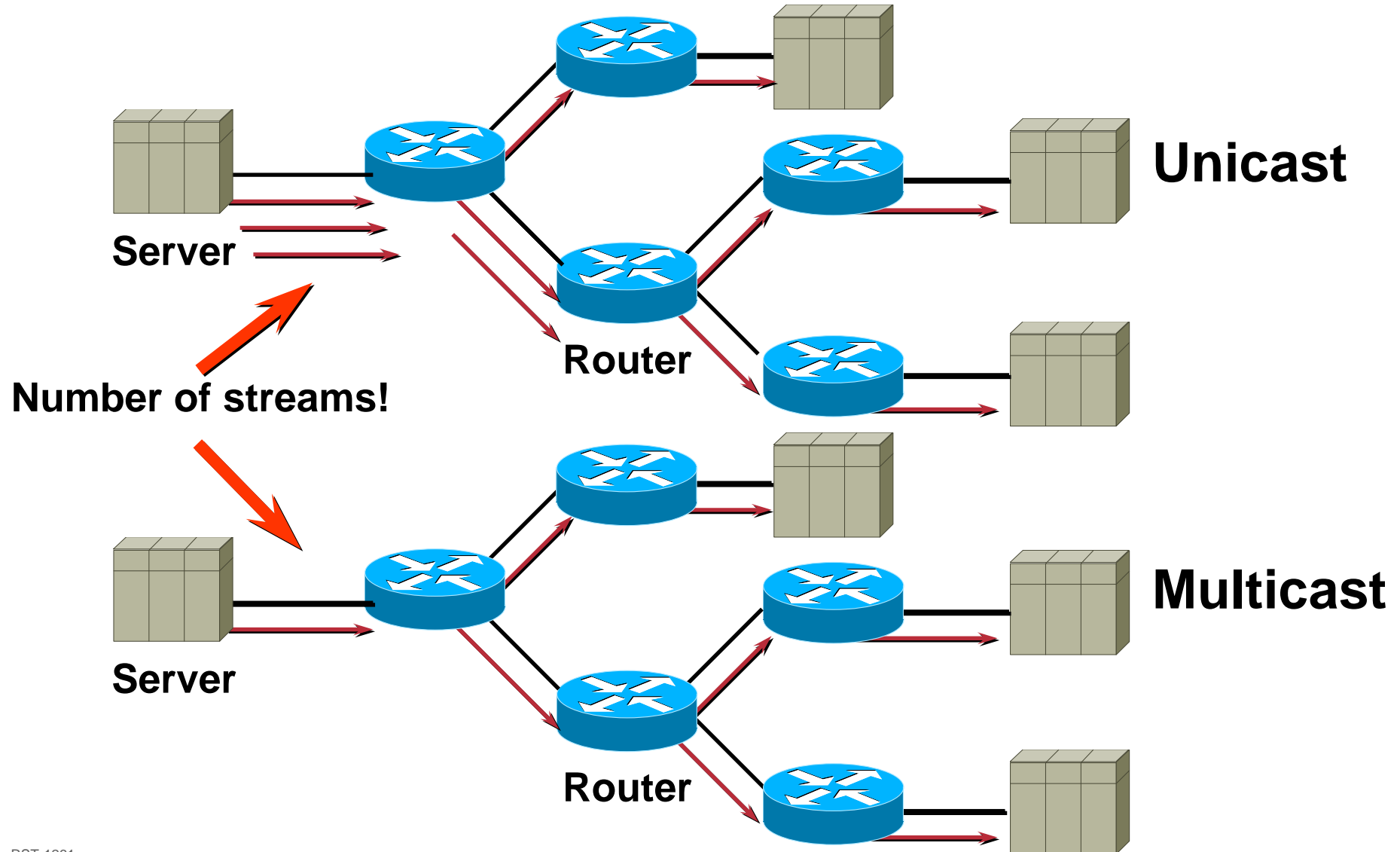
Agenda

- **Why Multicast?**

- **Multicast Fundamentals**

- **PIM Protocols**

- **RP choices**

- **Multicast at Layer 2**

- **Interdomain IP Multicast**

- **Latest Additions**

Why Multicast?

Cisco Public

Unicast vs. Multicast

Unicast

Server

Number of streams!

Router

Multicast

Server

Router

Cisco Public

Unicast vs. Multicast

- **TCP Unicast but NOT Multicast**
  - TCP is connection orientated protocol
  - Requires 3 way Handshake
  - Reliable due to sequence numbers + Ack
  - Flow control
- **UDP Unicast and Multicast**
  - Connectionless
  - Unreliable (application layer awareness)
- **Unicast Protocols**
  - ARP not applicable
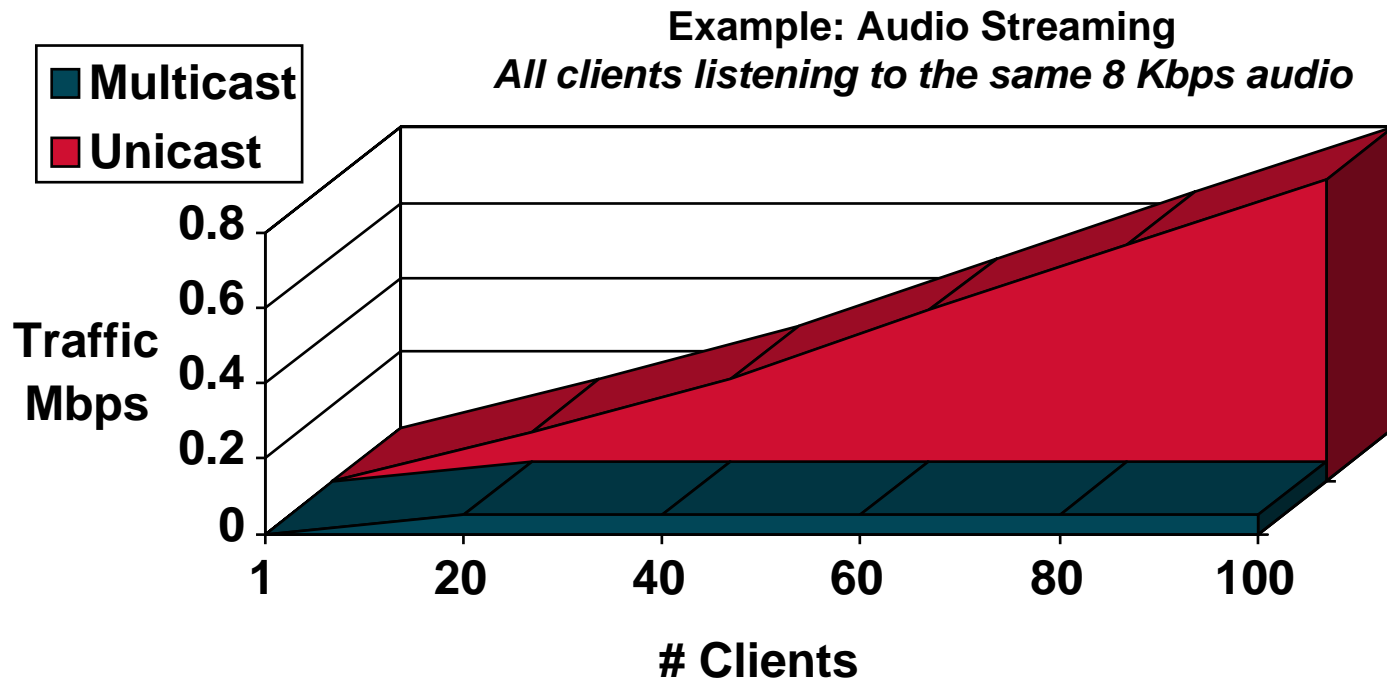 - HSRP etc are not applicable

Multicast Disadvantages

Multicast Is UDP Based!!!

- *Best Effort Delivery*: **Drops are to be expected. Multicast applications should not expect reliable delivery of data and should be designed accordingly. Reliable Multicast is still an area for much research. Expect to see more developments in this area. PGM offers reliability!**

- *No Congestion Avoidance*: **Lack of TCP windowing and "slow-start" mechanisms can result in network congestion. If possible, Multicast applications should attempt to detect and avoid congestion conditions.**

- *Duplicates*: **Some multicast protocol mechanisms (e.g. Asserts, Registers and SPT Transitions) result in the occasional generation of duplicate packets. Multicast applications should be designed to expect occasional duplicate packets.**

- *Out of Order Delivery* : **Some protocol mechanisms may also result in out of order delivery of packets.**
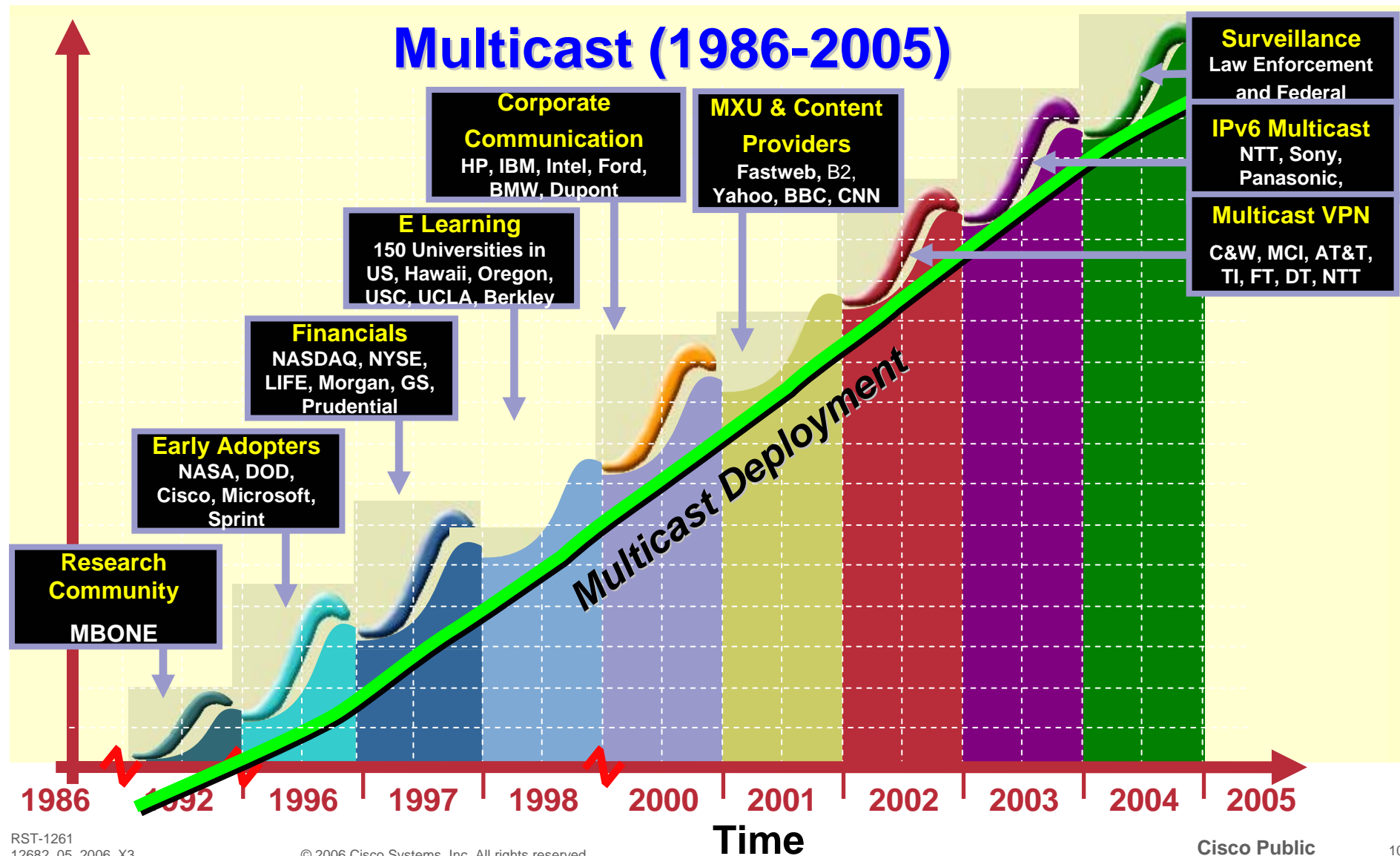
Multicast Advantages

- *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**
■ **Unicast**

Traffic Mbps

0.8
0.6
0.4
0.2
0

 # Clients

1   20   40   60   80   100

Cisco Public   9

Multicast Adoption
Past, Present, and Future

Multicast (1986-2005)

Research Community
MBONE

Early Adopters
NASA, DOD, Cisco, Microsoft, Sprint

Financials
NASDAQ, NYSE, LIFE, Morgan, GS, Prudential

E Learning
150 Universities in US, Hawaii, Oregon, USC, UCLA, Berkley

Corporate Communication
HP, IBM, Intel, Ford, BMW, Dupont

MXU & Content Providers
Fastweb, B2, Yahoo, BBC, CNN

Surveillance
Law Enforcement and Federal

IPv6 Multicast
NTT, Sony, Panasonic,

Multicast VPN
C&W, MCI, AT&T, TI, FT, DT, NTT

Multicast Deployment

1986   1992   1996   1997   1998   2000   2001   2002   2003   2004   2005

Time

Multicast Fundamentals

Cisco Public   11

Multicast Components
Cisco End-to-End Architecture

MSDP

ISP A

ISP B

IGMP Snooping PIM snooping

MBGP

IGMP

**PIM-SM
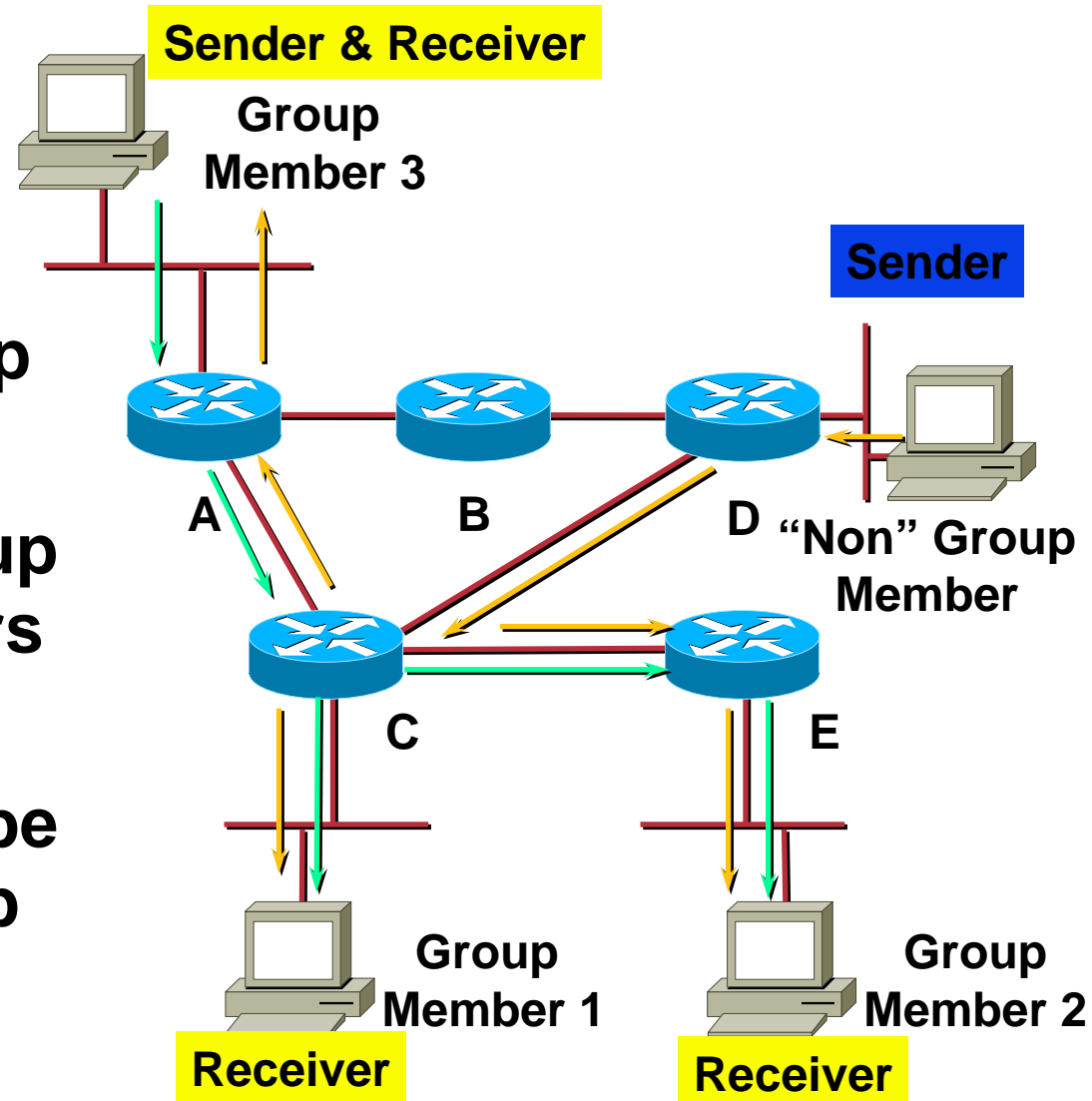Bidir PIM
PIM-SSM
MVPN**
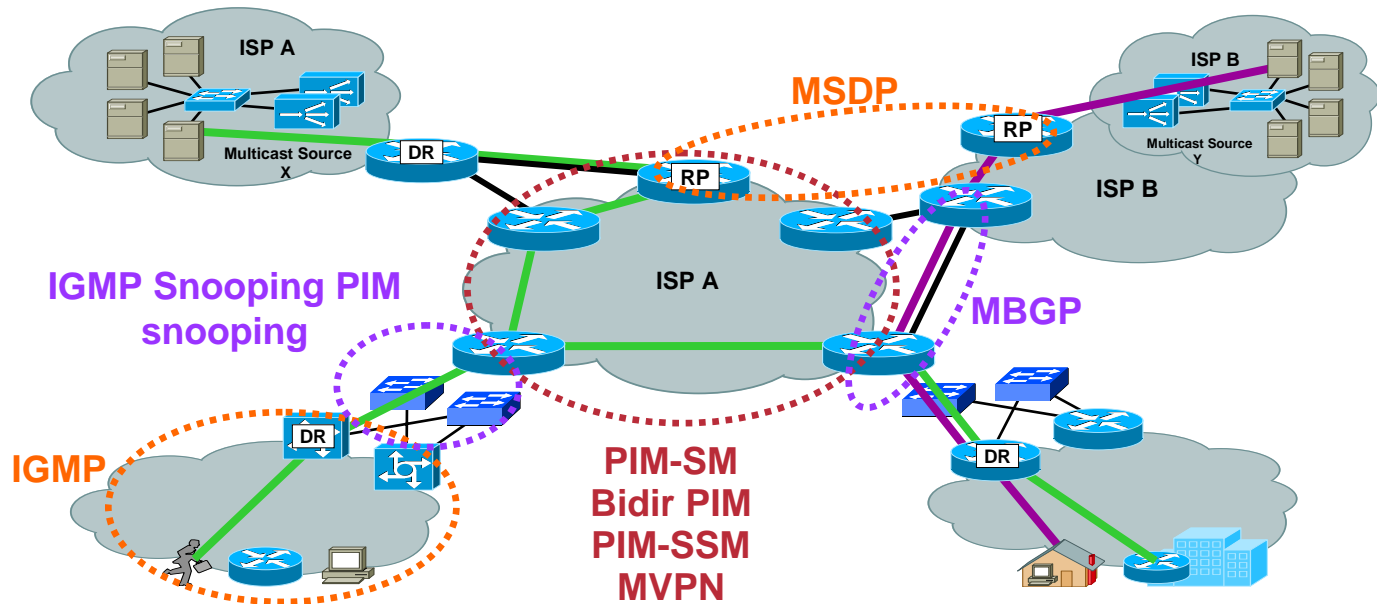
Campus Multicast

- **End Stations (hosts-to-routers):**
  - **IGMP**
- **Switches (Layer 2 Optimization):**
  - **IGMP Snooping PIM snooping**
- **Routers (Multicast Forwarding Protocol):**
  - **PIM Sparse Mode or Bidirectional PIM**

Interdomain Multicast

- **Multicast routing across domains**
  - **MBGP**
- **Multicast Source Discovery**
  - **MSDP with PIM-SM**
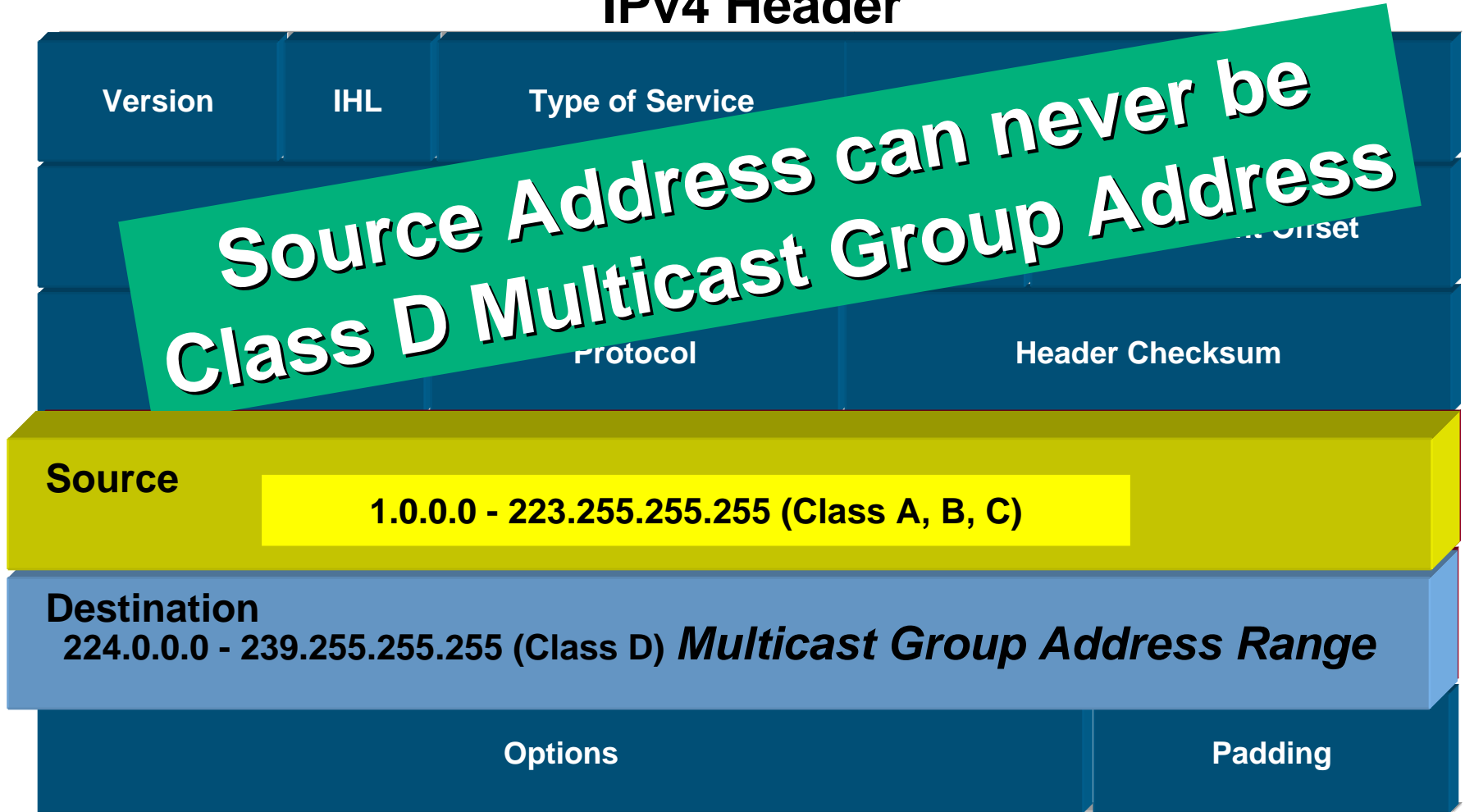- **Source Specific Multicast**
  - **PIM-SSM**

IP Multicast Group Concept

1. **You must be a "member" of a group to receive its data**

2. **If you send to a group address, all members receive it**

3. **You do not have to be a member of a group to send to a group**

Multicast Addressing

IPv4 Header

Version	IHL	Type of Service		
			Offset	
		Protocol	Header Checksum	

Source Address can never be Class D Multicast Group Address

Source	1.0.0.0 - 223.255.255.255 (Class A, B, C)

Destination 224.0.0.0 - 239.255.255.255 (Class D) *Multicast Group Address Range*

Options	Padding

Multicast Group Address range

224.0.0.0 - 239.255.255.255 (Class D)

- **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

- **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**
    - **Used to limit "scope" of multicast traffic**
    - **Same addresses may be in use at different locations for different multicast sessions**
  - **Examples**
    - **Site-local scope: 239.255.0.0/16**
    - **Organization-local scope: 239.192.0.0/14**
- **SSM (Source Specific Multicast) Range**
  - **232.0.0.0 – 232.255.255.255**
  - **Primarily targeted for Internet style Broadcast**

Multicast Addressing

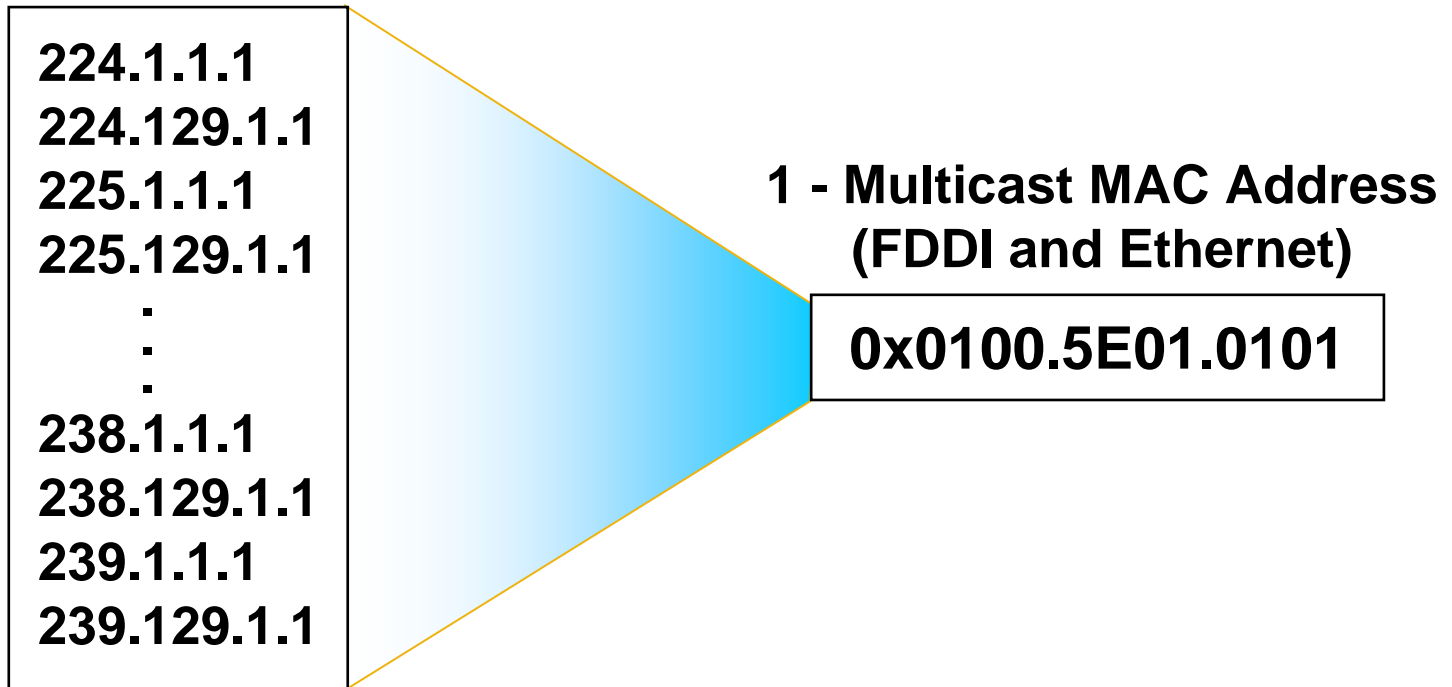## IP Multicast MAC Address Mapping
(FDDI and Ethernet)

32 Bits
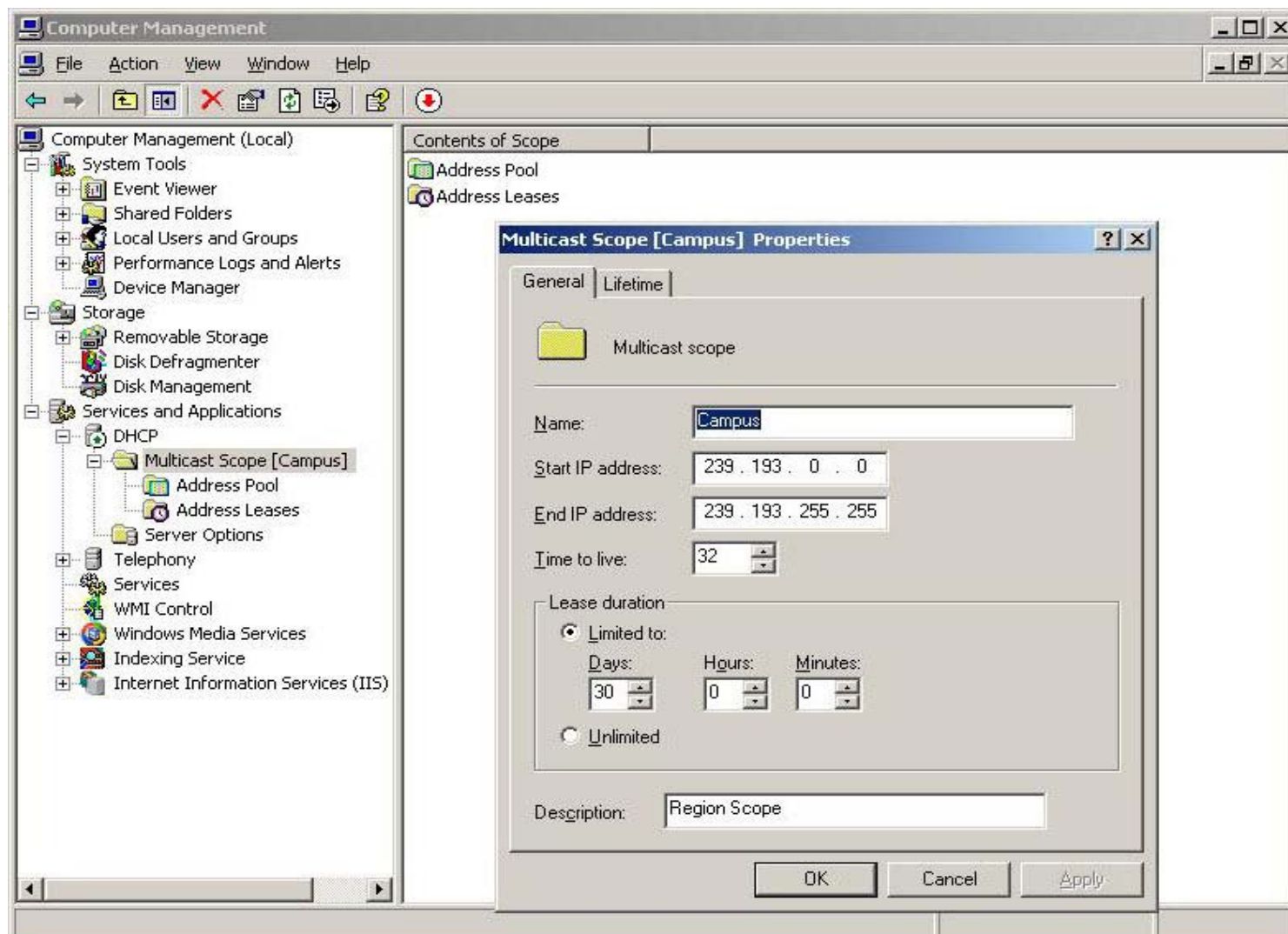
1110   28 Bits

239.255.0.1

5 Bits Lost

01-00-5e-7f-00-01

25 Bits   23 Bits

48 Bits
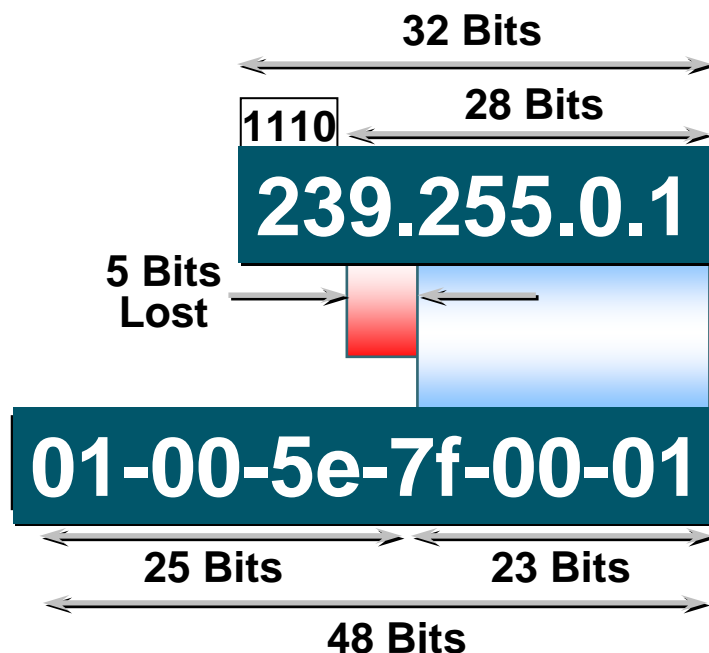
Multicast Addressing

IP Multicast MAC Address Mapping
(FDDI & Ethernet)

Be Aware of the 32:1 Address Overlap

32 - IP Multicast Addresses

```
224.1.1.1
224.129.1.1
225.1.1.1
225.129.1.1
    .
    .
    .
238.1.1.1
238.129.1.1
239.1.1.1
239.129.1.1
```

1 - Multicast MAC Address (FDDI and Ethernet)

0x0100.5E01.0101

Cisco Public   18

Madcap in MS Server

Cisco Public

# How are Multicast Addresses Assigned?

- **Static Global Group Address Assignment**
  - **Temporary method to meet immediate needs**
  - **Group range: 233.0.0.0 – 233.255.255.255**
    - **Your AS number is inserted in middle two octets**
    - **Remaining low-order octet used for group assignment**
  - **Defined in RFC 2770**
    - **"GLOP Addressing in 233/8"**

- **Manual Address Allocation by the Admin !!**
  - **Is still the most common practice**

Cisco Public

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**
  -

Cisco Public
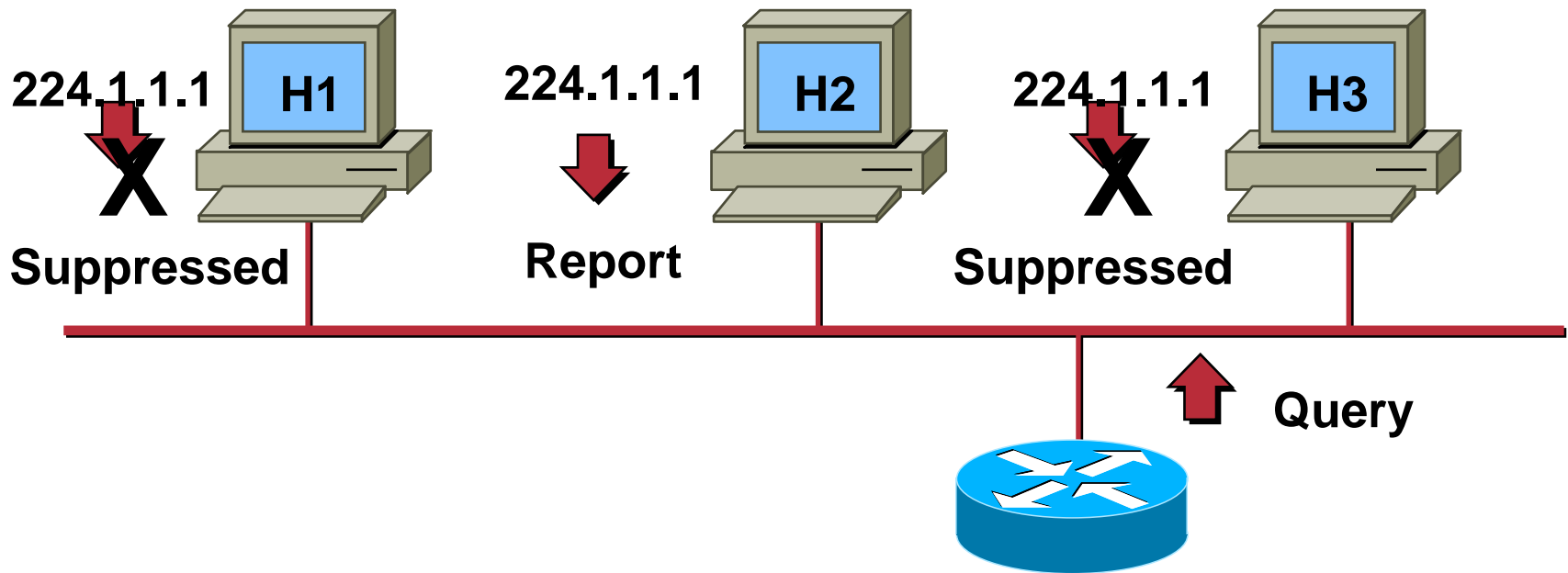
Host-Router Signaling: IGMP

Joining a Group

H1

H2

224.1.1.1

H3
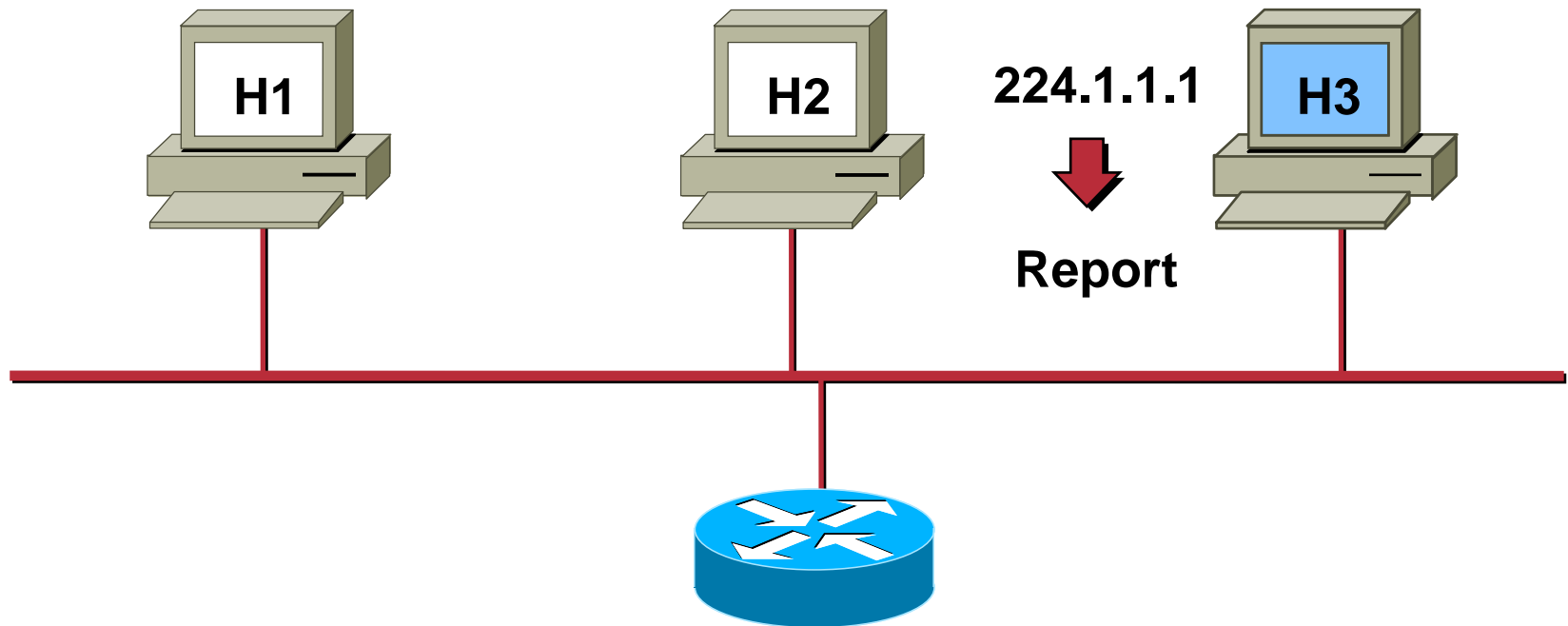
Report

- **Host sends IGMP Report to join group**

Host-Router Signaling: IGMP

Maintaining a Group

224.1.1.1 **H1**

224.1.1.1 **H2**

224.1.1.1 **H3**
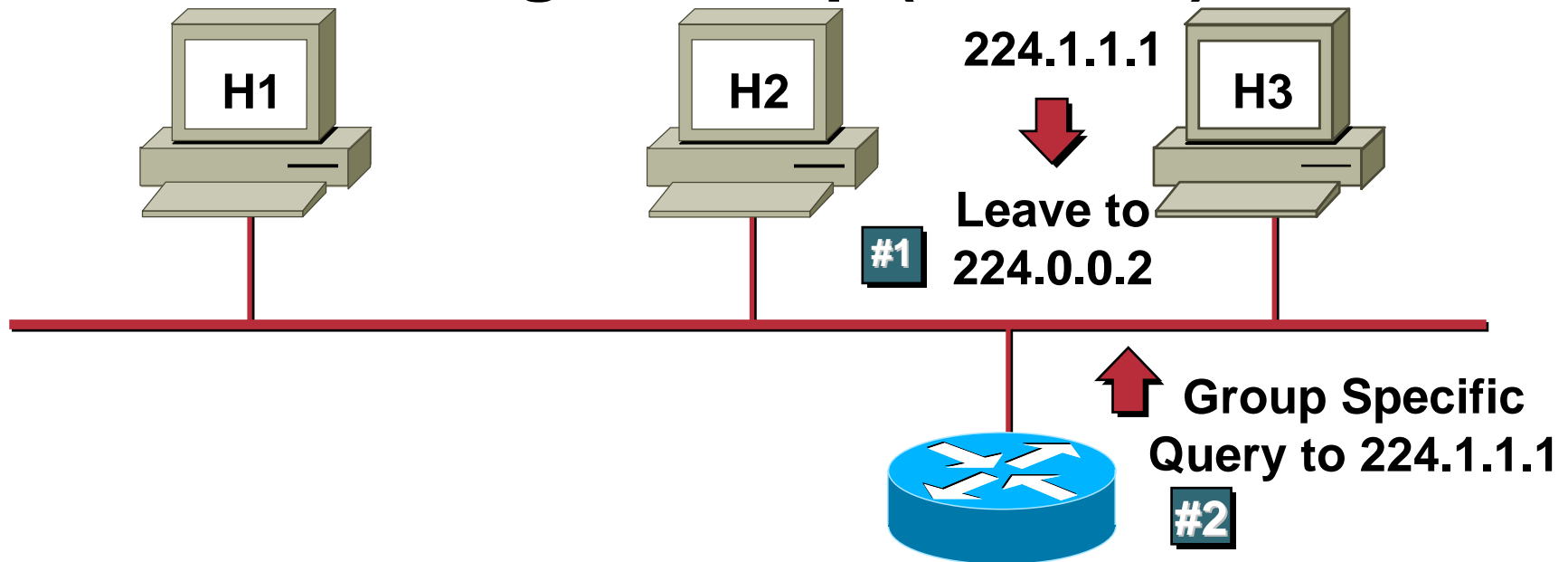
Suppressed

Report

Suppressed

Query

- **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)

H1   H2   224.1.1.1   H3

#1 Leave to 224.0.0.2

#2 Group Specific Query to 224.1.1.1

1. **Host sends Leave message to 224.0.0.2**
2. **Router sends Group specific query to 224.1.1.1**
3. **No IGMP Report is received within ~3 seconds**
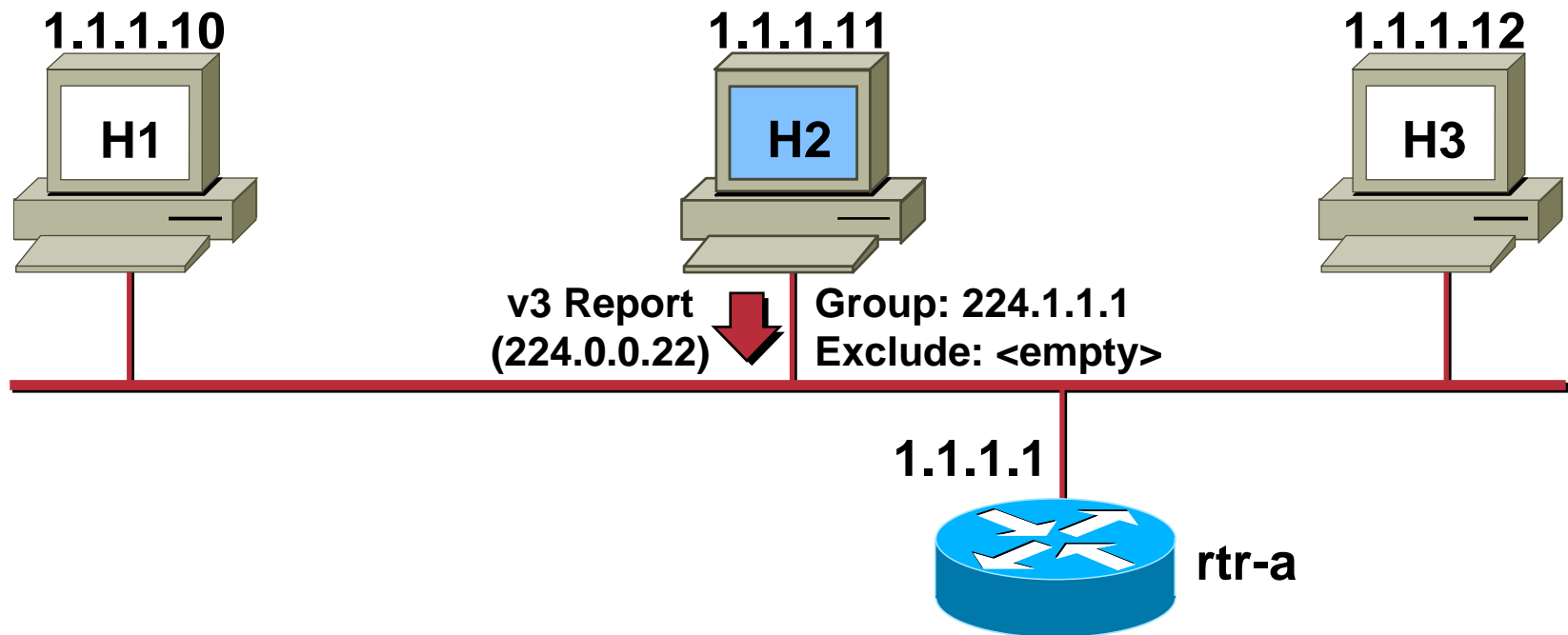4. **Group 224.1.1.1 times out**

Host-Router Signaling: IGMPv3

- **RFC 3376**
  - **Adds Include/Exclude Source Lists**
  - **Enables hosts to listen only to a specified subset of the hosts sending to the group**
  - **Requires new 'IPMulticastListen' API**
    - **New IGMPv3 stack required in the O/S.**
  - **Apps must be rewritten to use IGMPv3 Include/Exclude features**

Cisco Public

Host-Router Signaling: IGMPv3
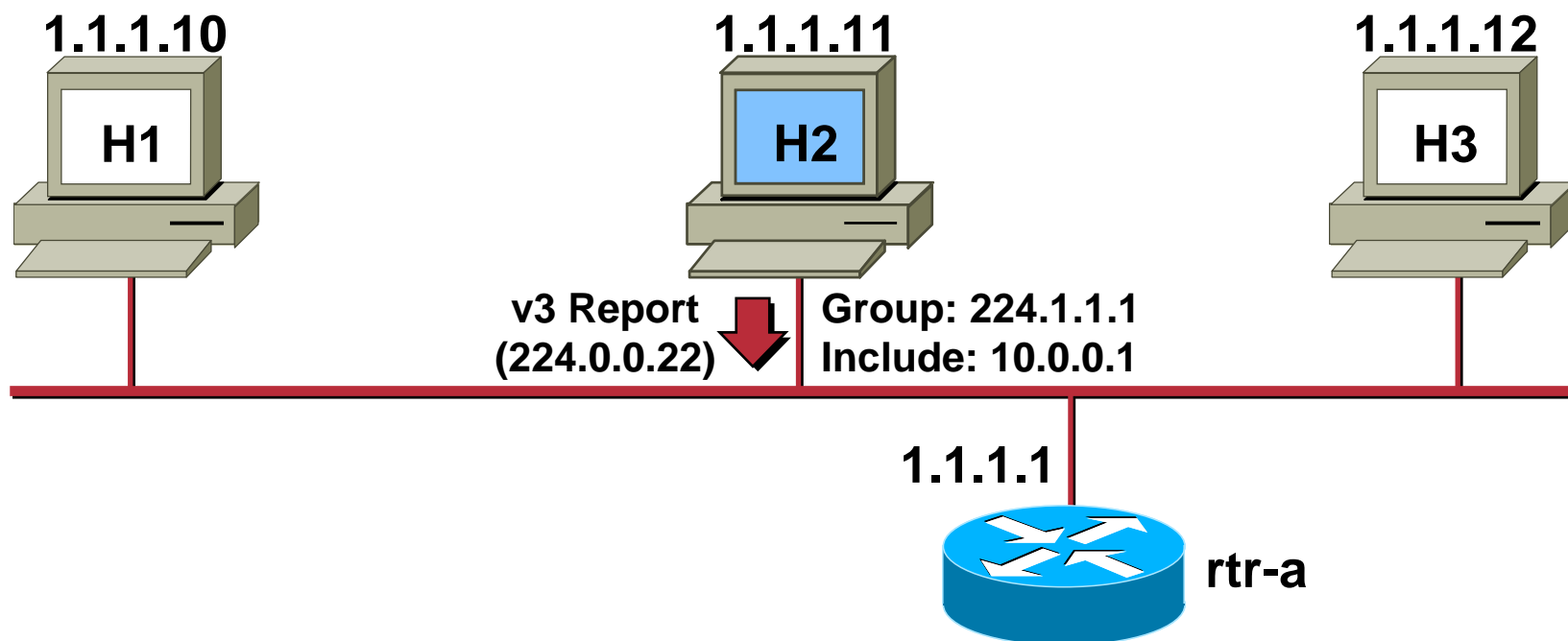
- **New Membership Report address**
  - **224.0.0.22 (IGMPv3 Routers)**
    - **All IGMPv3 Hosts send reports to this address**
      - » **Instead of the target group address as in IGMPv1/v2**
    - **All IGMPv3 Routers listen to this address**
    - **Hosts do not listen or respond to this address**
  - **No Report Suppression**
    - **All Hosts on wire respond to Queries**
      - » **Host's complete IGMP state sent in single response**
    - **Response Interval may be tuned over broad range**
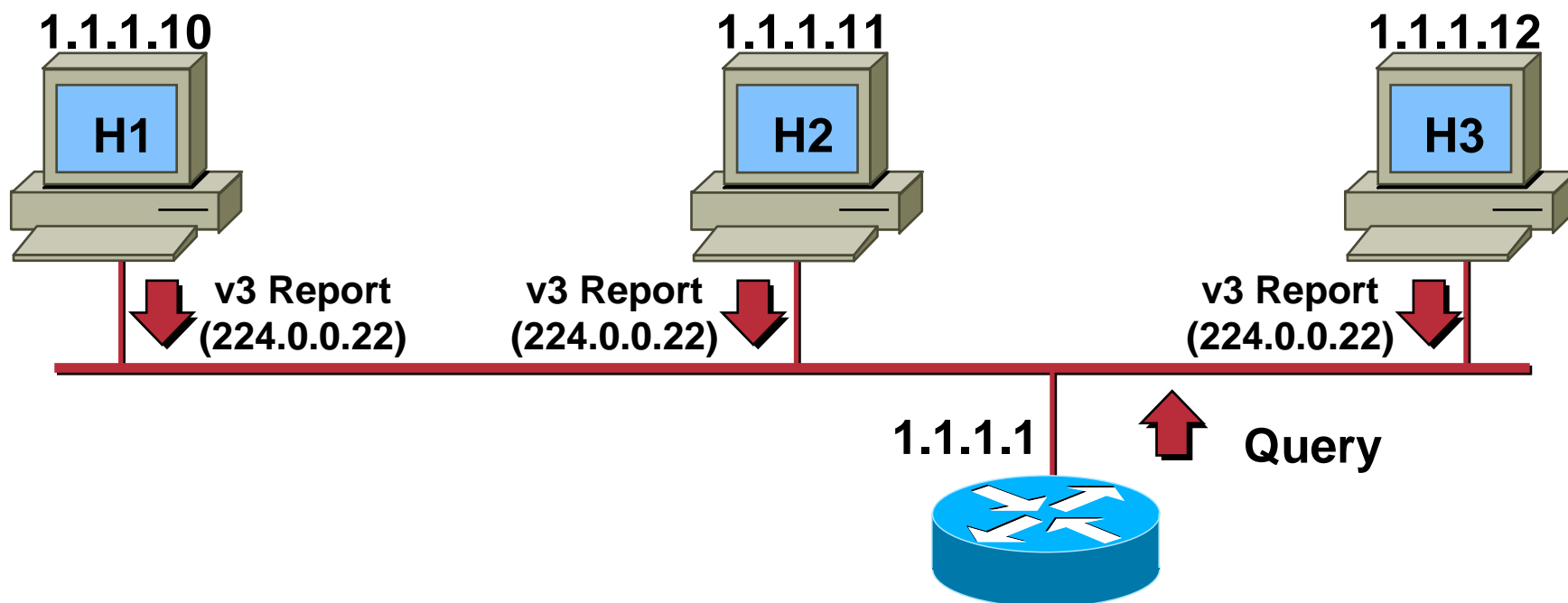   - » **Useful when large numbers of hosts reside on subnet**

IGMPv3—Joining a Group

1.1.1.10

H1

1.1.1.11

H2

1.1.1.12

H3

v3 Report
(224.0.0.22)

Group: 224.1.1.1
Exclude: <empty>

1.1.1.1

rtr-a

- **Joining member sends IGMPv3 Report to 224.0.0.22 immediately upon joining**

IGMPv3—Joining specific Source(s)

1.1.1.10

H1

1.1.1.11

H2

1.1.1.12

H3

v3 Report
(224.0.0.22)

Group: 224.1.1.1
Include: 10.0.0.1

1.1.1.1

rtr-a

- **IGMPv3 Report contains desired source(s) in the Include list.**

- **Only "Included" source(s) are joined.**

IGMPv3—Maintaining State

1.1.1.10

H1

**v3 Report
(224.0.0.22)**

1.1.1.11

H2

**v3 Report
(224.0.0.22)**

1.1.1.12

H3

**v3 Report
(224.0.0.22)**

1.1.1.1

Query

- **Router sends periodic queries**
- **All IGMPv3 members respond**

  - **Reports contain multiple Group state records**

Multicast L3 Forwarding

- **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.**

Cisco Public

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.**

Cisco Public

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 source 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.

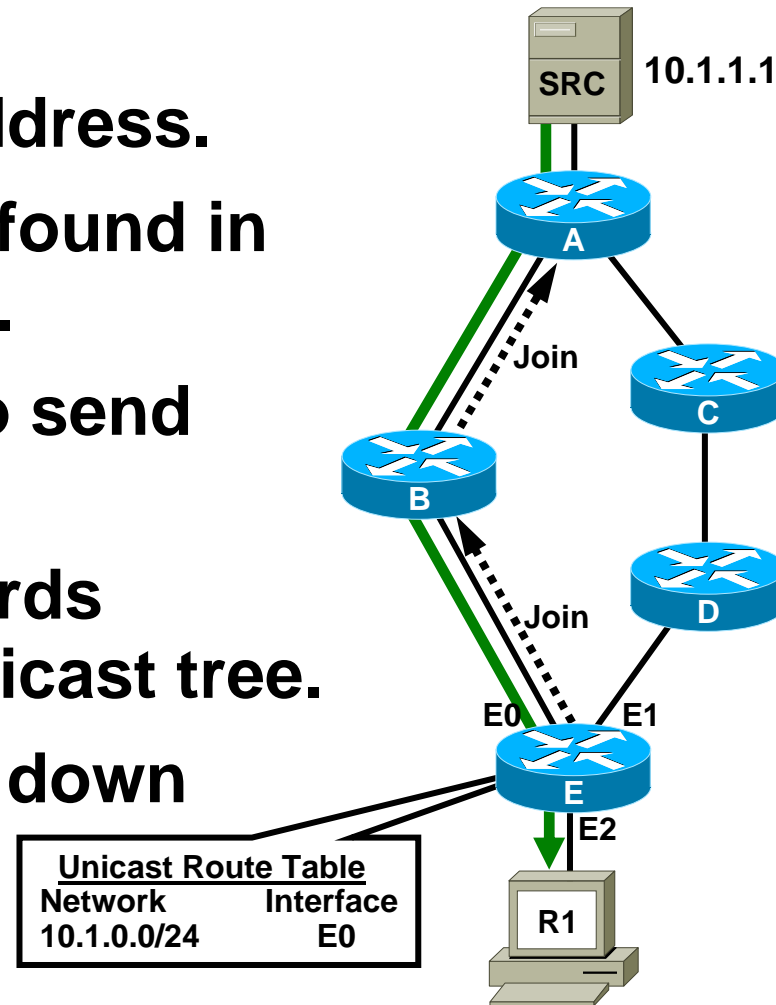# Reverse Path Forwarding (RPF)

- **The RPF Calculation**

  - **The multicast source address is checked against the unicast routing table.**

  - **This determines the interface and upstream router in the direction of the source to which PIM Joins are sent.**

  - **This interface becomes the "Incoming" or RPF interface.**

    - **A router forwards a multicast datagram only if received on the RPF interface.**

Cisco Public

Reverse Path Forwarding (RPF)
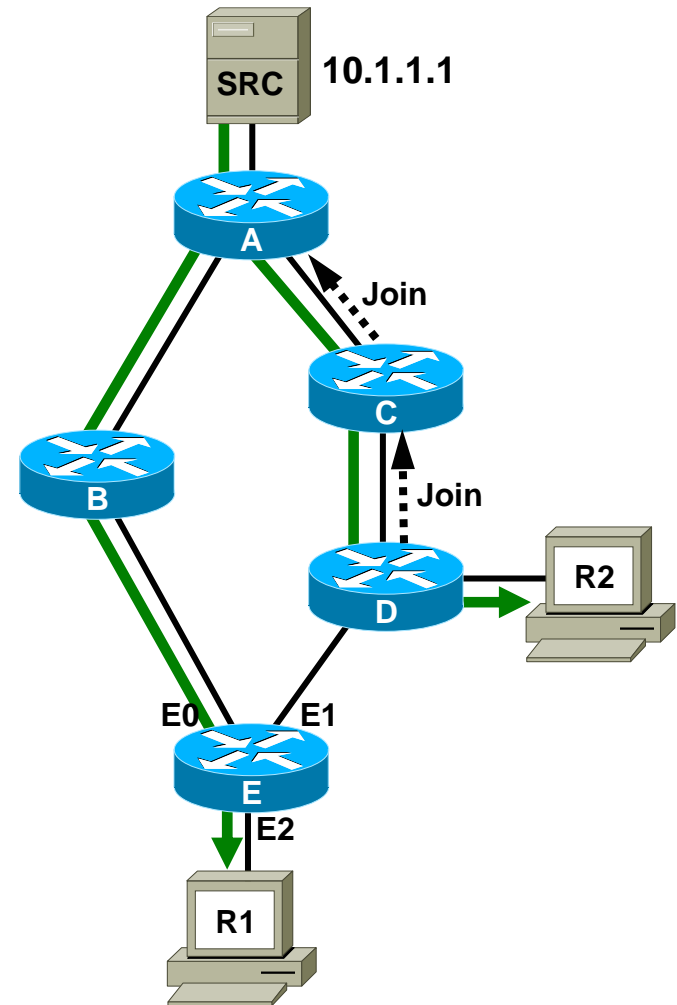
- **RPF Calculation**
  - **Based on Source Address.**
  - **Best path to source found in Unicast Route Table.**
  - **Determines where to send Join.**
  - **Joins continue towards Source to build multicast tree.**
  - **Multicast data flows down tree.**

SRC 10.1.1.1

A

Join

C

B

Join

D

E0   E1

E

E2

Unicast Route Table

Network	Interface
10.1.0.0/24	E0

R1

Reverse Path Forwarding (RPF)

- **RPF Calculation**
  - **Based on Source Address.**
  - **Best path to source found in Unicast Route Table.**
  - **Determines where to send Join.**
  - **Joins continue towards Source to build multicast tree.**
  - **Multicast data flows down tree.**
  - **Repeat for other receivers.**
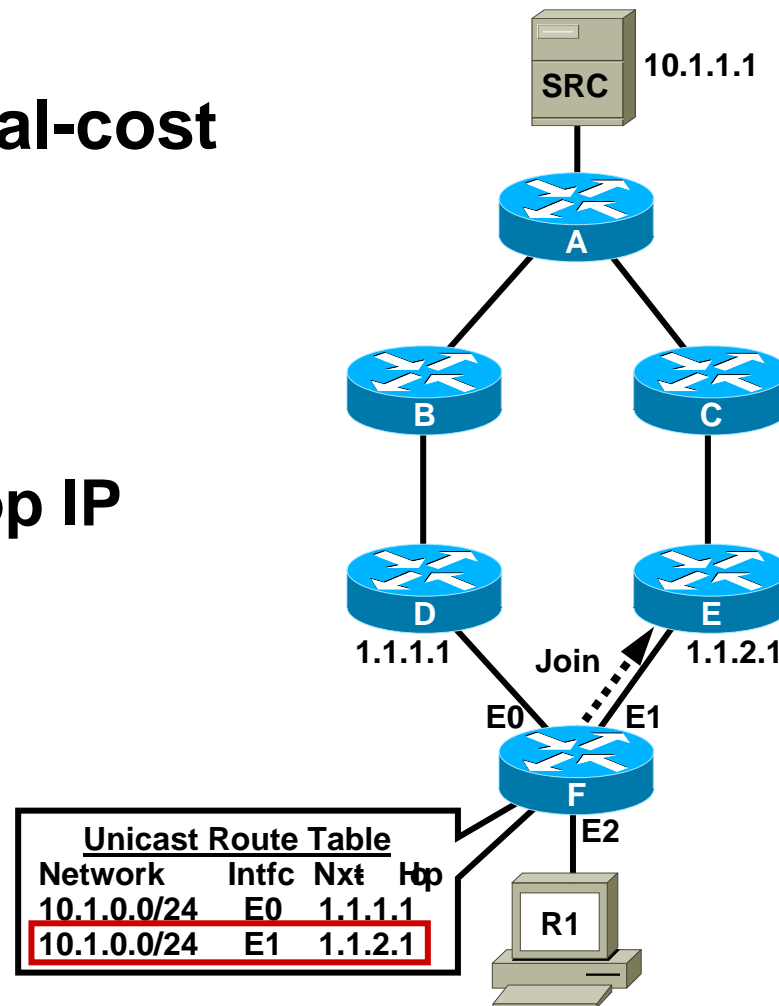
Cisco Public 35

Reverse Path Forwarding (RPF)

- **RPF Calculation**
  - **What if we have equal-cost paths?**
    - **We can't use both.**
  - **Tie-Breaker**
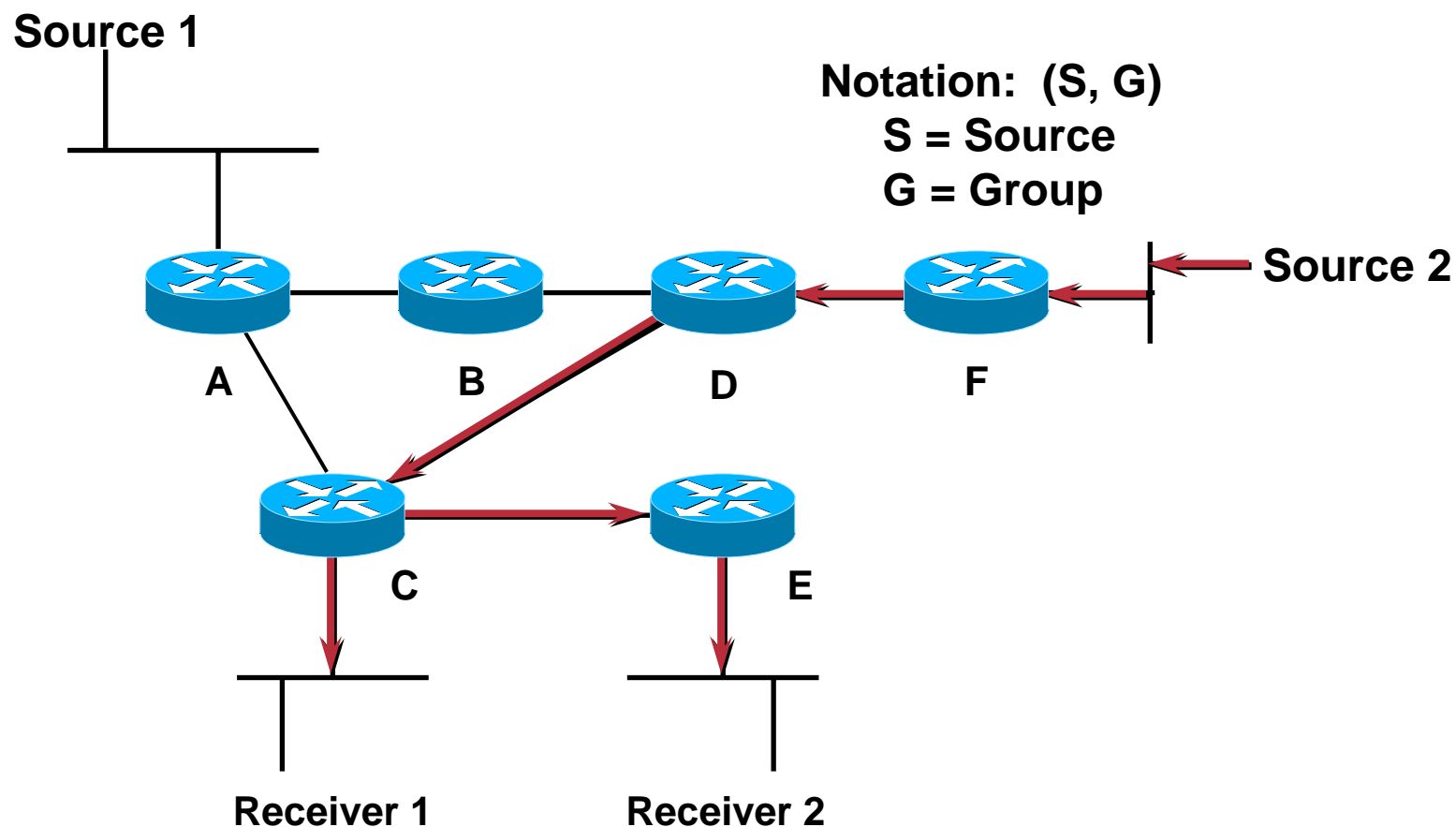   - **Use highest Next-Hop IP address.**

SRC 10.1.1.1

A

B   C

D   E

1.1.1.1   **Join**   1.1.2.1

E0   E1

F

E2

R1

Unicast Route Table		
Network	Intfc	Nxt Hop
10.1.0.0/24	E0	1.1.1.1
10.1.0.0/24	E1	1.1.2.1

Shortest Path or Source Distribution Tree

Source 1

Notation: (S, G)
S = Source
G = Group

Source 2

A   B   D   F

C   E

Receiver 1   **Receiver 2**

Multicast Distribution Trees

Shortest Path or Source Distribution Tree

Source 1

Notation: (S, G)
S = Source
G = Group

Source 2

A   **B**   **D**   **F**

C   **E**

Receiver 1   **Receiver 2**

Cisco Public   38

Shared Distribution Tree

Notation: (*, G)
* = All Sources
G = Group

A

B

D (RP)

F

C

E

(RP)   PIM Rendezvous Point

→   Shared Tree

Receiver 1

Receiver 2
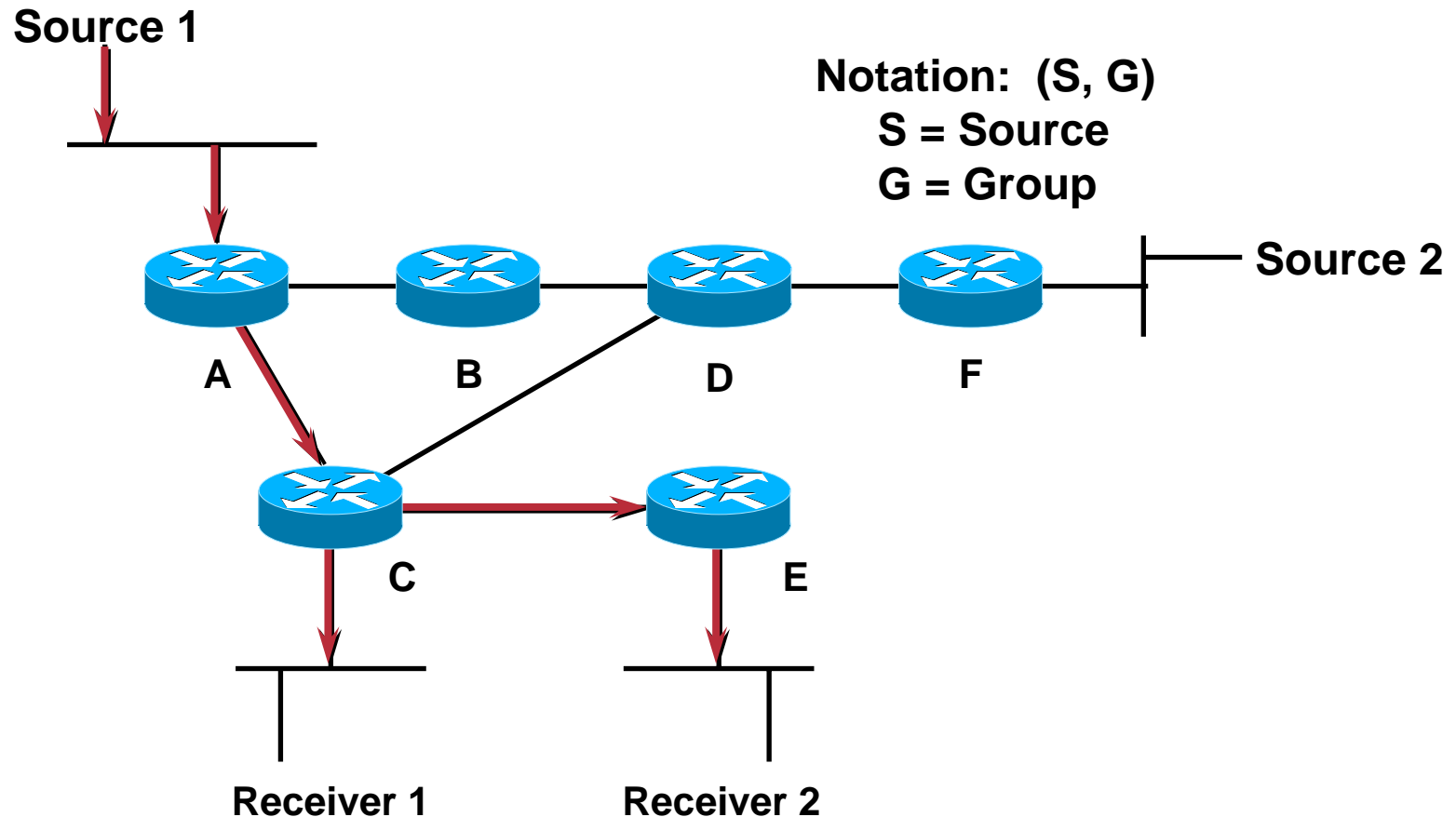
Multicast Distribution Trees

Shared Distribution Tree

Source 1

Notation: (*, G)
* = All Sources
G = Group

Source 2

A   B   D (RP)   F

C   E

(RP)   PIM Rendezvous Point

→   Shared Tree

→   Source Tree

Receiver 1   Receiver 2
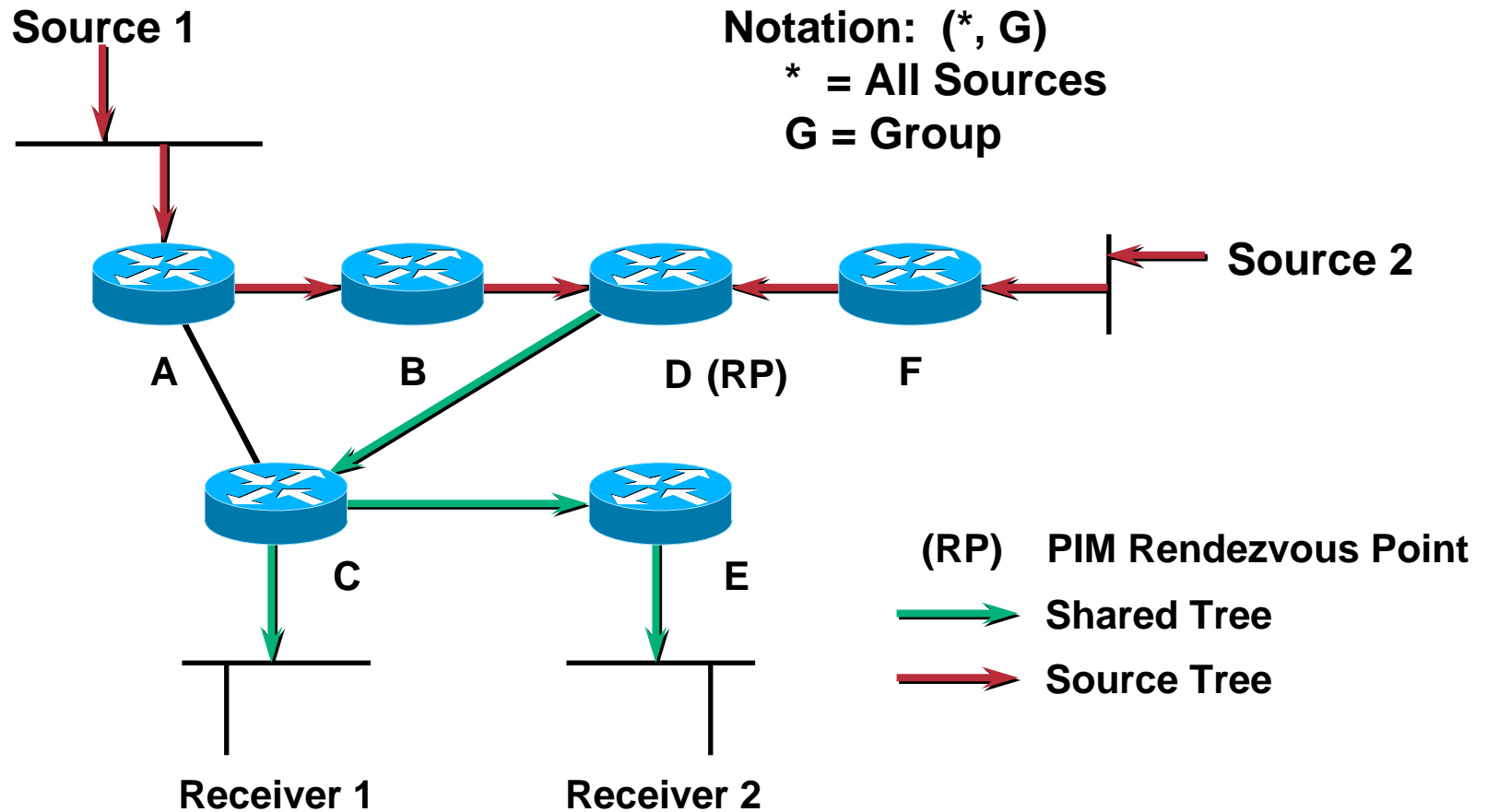
Multicast Distribution Trees
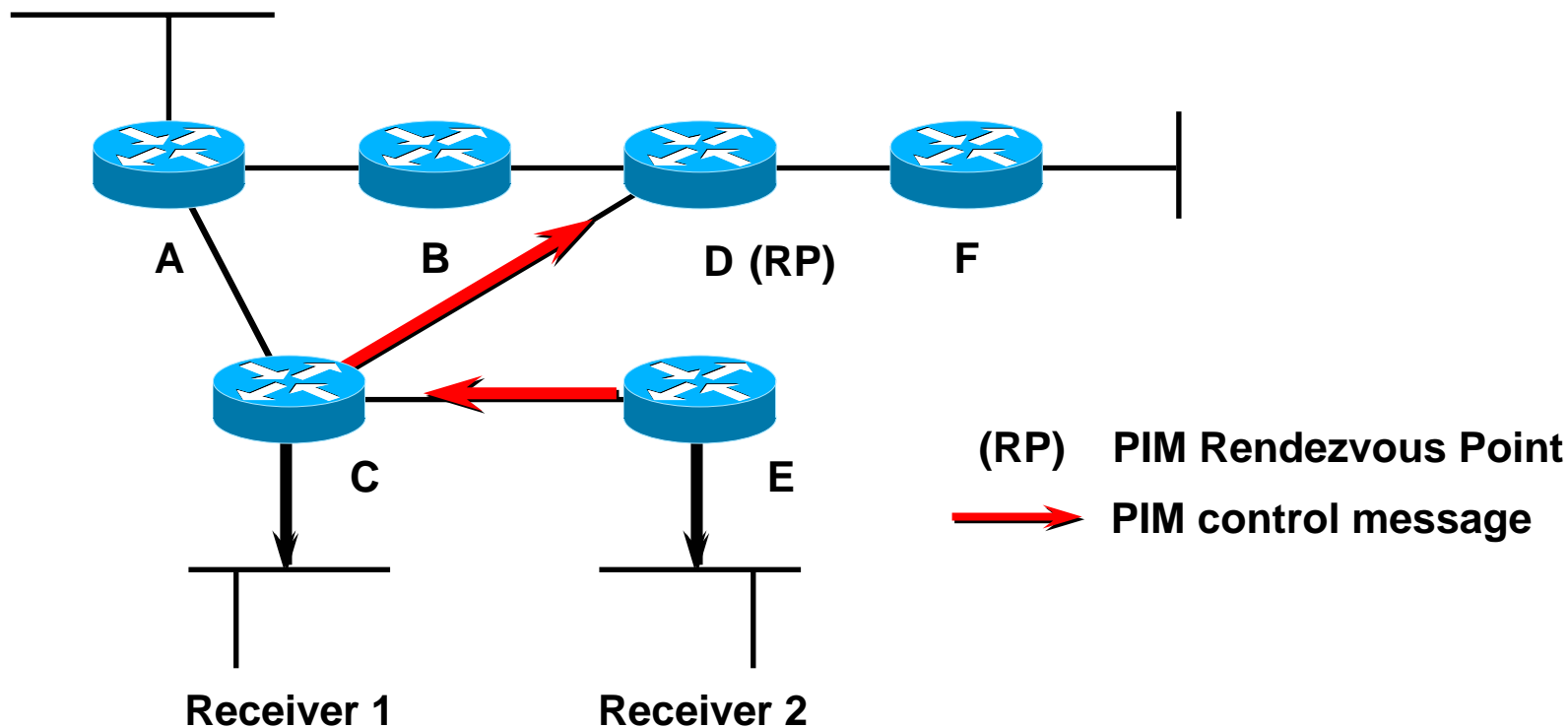
Characteristics of Distribution Trees

- **Source or Shortest Path trees**
  - **Uses more memory O(S x G) but you get optimal paths from source to all receivers; minimizes delay**

- **Shared trees**
  - **Uses less memory O(G) but you may get sub-optimal paths from source to all receivers; may introduce extra delay**

Multicast Tree creation

- **PIM Join/Prune Control Messages**
  - Used to create/remove Distribution Trees
- **Shortest Path trees**
  - PIM control messages are sent toward the Source
- **Shared trees**
  - PIM control messages are sent toward RP

Cisco Public

Multicast Distribution Tree creation

Shared Tree Example

A

B

D (RP)

F

C

E

Receiver 1

Receiver 2

(RP)   PIM Rendezvous Point

PIM control message

Cisco Public   43

PIM Protocol Variants

Cisco Public 44
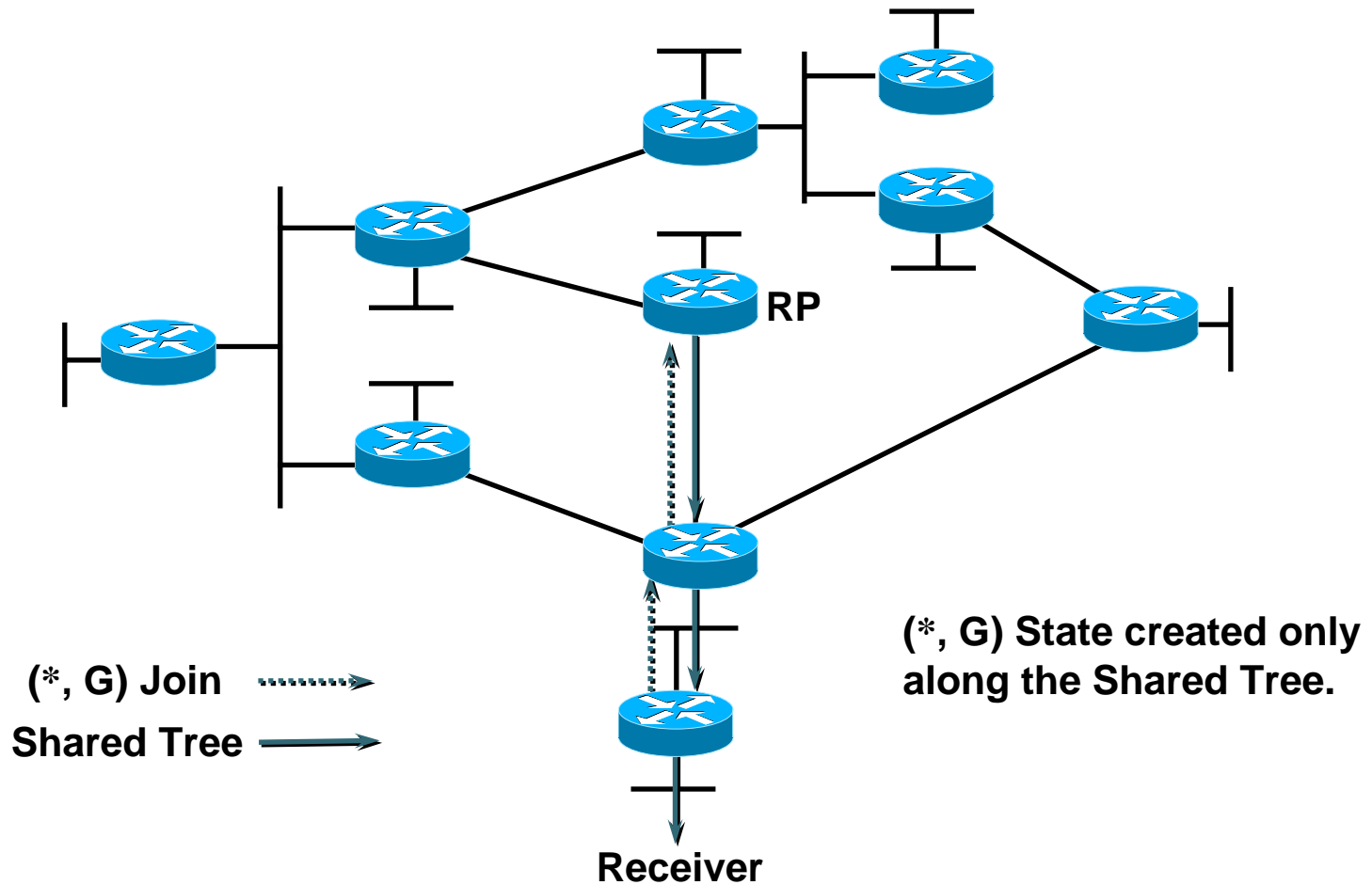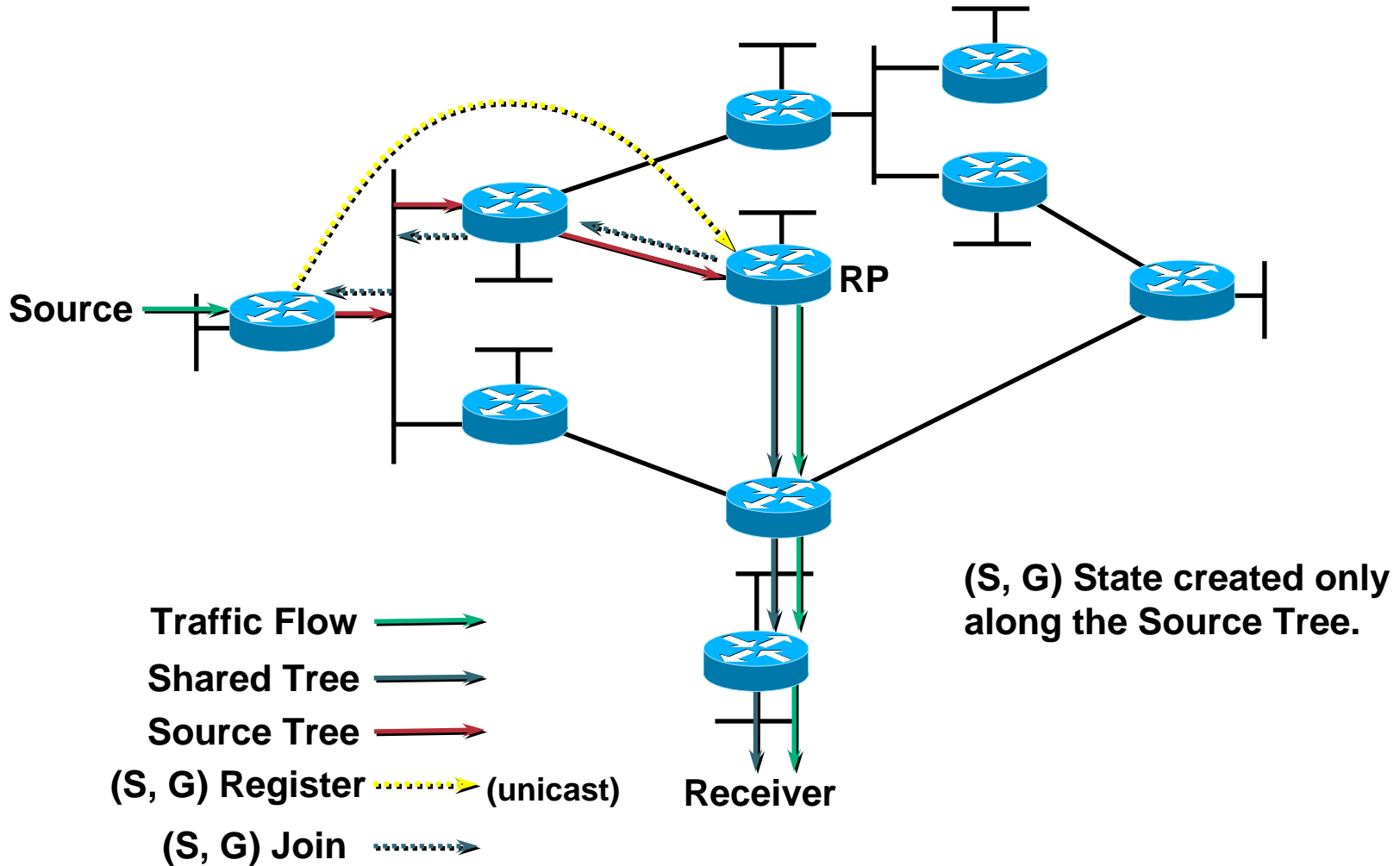
Major deployed PIM variants

- **PIM ASM**
  - **Sparse mode / RP / SPT / Shared Tree**
- **PIM SSM**
  - **Source Specific Multicast, no RP, SPT only**
- **PIM BiDir**
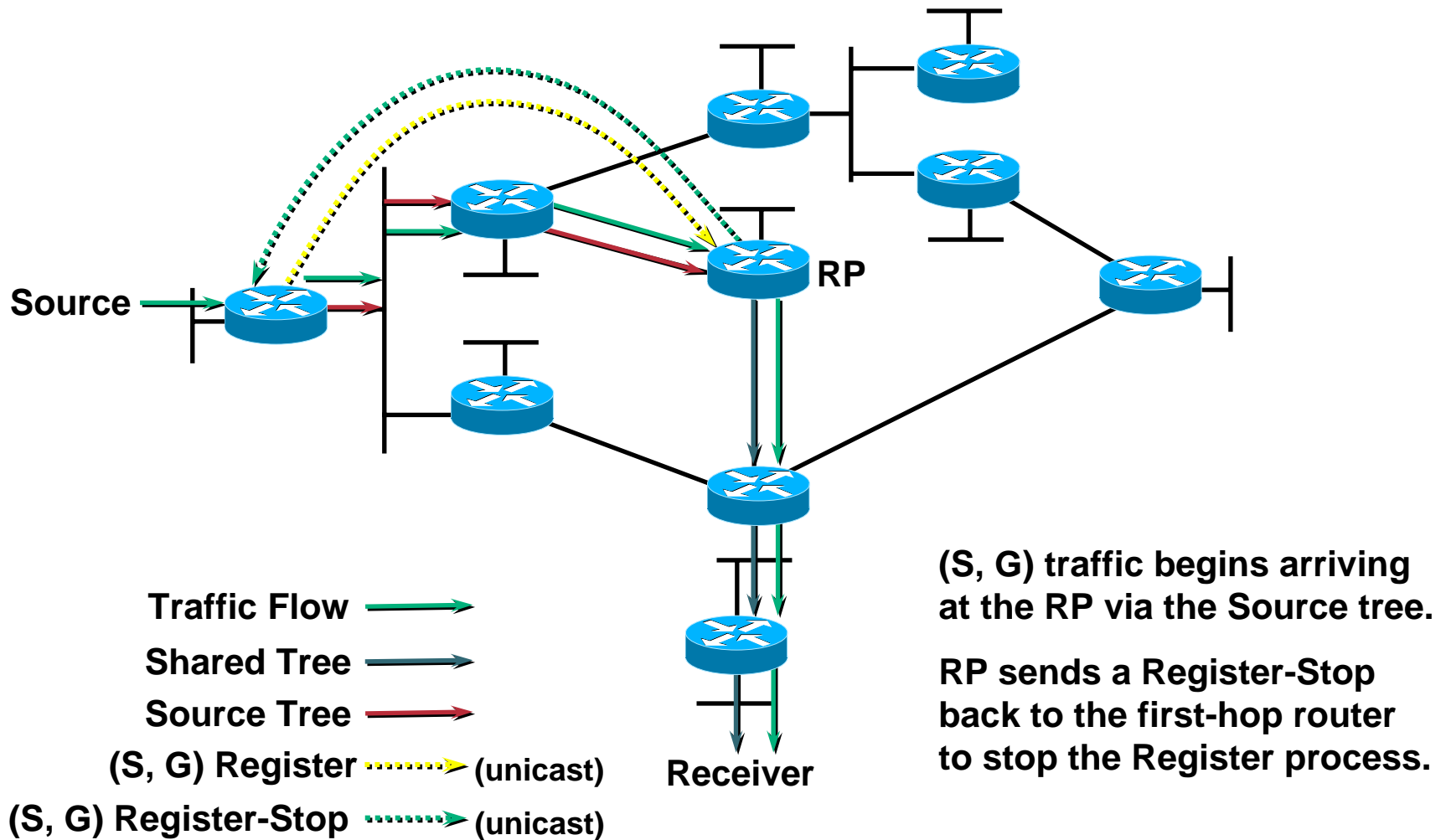 - **BiDirectional PIM, no SPT, Shared tree only**

PIM-SM Shared Tree Join

(∗, G) Join ┄┄┄┄►

Shared Tree ────►

RP

(∗, G) State created only along the Shared Tree.

Receiver

Cisco Public   46

PIM-SM Sender Registration

Source

RP

(S, G) State created only along the Source Tree.

Traffic Flow	→
Shared Tree	→
Source Tree	→
(S, G) Register	┈┈> (unicast)
(S, G) Join	┈┈┈>

Receiver

Cisco Public 47

PIM-SM Sender Registration

Source

RP

Traffic Flow ——→
Shared Tree ——→
Source Tree ——→
(S, G) Register ·····→ (unicast)
(S, G) Register-Stop ·····→ (unicast)

Receiver

(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.

Cisco Public 48

PIM-SM Sender Registration

Source

RP

Traffic Flow ⟶

Shared Tree ⟶

Source Tree ⟶

Source traffic flows natively along SPT to RP.

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

Receiver

Cisco Public 49

PIM-SM SPT Switchover

Source

Traffic Flow ⟶

Shared Tree ⟶

Source Tree ⟶

(S, G) Join ┅┅┅⟶

RP

Receiver

Last-hop router joins the Source Tree.

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

Cisco Public

PIM-SM SPT Switchover

Source

RP

Traffic Flow

Shared Tree

Source Tree

(S, G)RP-bit Prune

Receiver

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.

Cisco Public

PIM-SM SPT Switchover

Source

RP

Traffic Flow ⟶
Shared Tree ⟶
Source Tree ⟶

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

Receiver

Cisco Public 52

PIM-SM SPT Switchover

Source

RP

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

Traffic Flow ⟶
Shared Tree ⟶
Source Tree ⟶
(S, G) Prune ┈┈┈▶

Receiver

Cisco Public 53

PIM-SM SPT Switchover

Source

Traffic Flow ⟶
Shared Tree ⟶
Source Tree ⟶

RP

(S, G) Traffic flow is now only flowing to the Receiver via a single branch of the Source Tree.

Receiver

PIM-SM FFF

" 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."
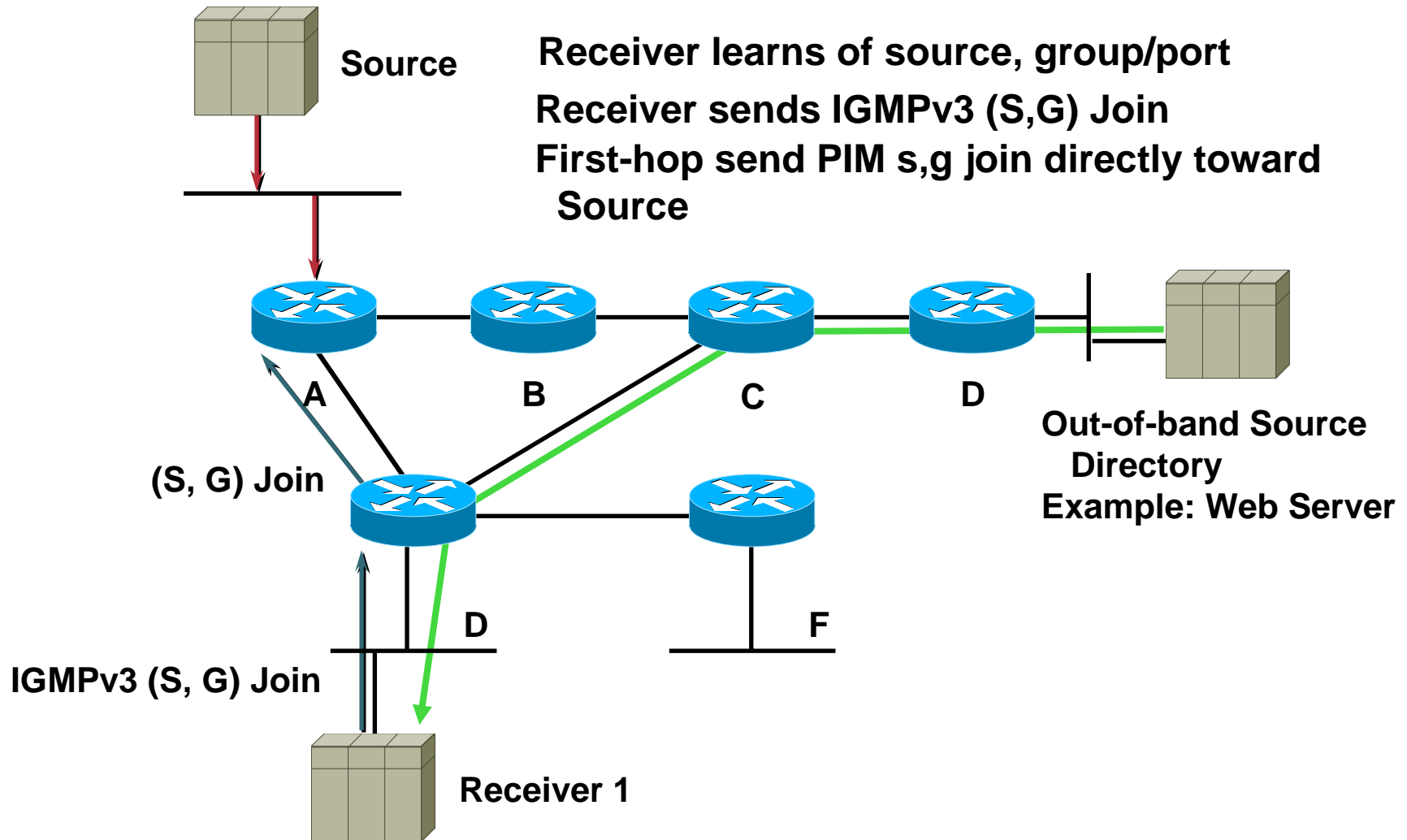
PIM-SM Frequently Forgotten Fact

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 inter-domain multicast routing**
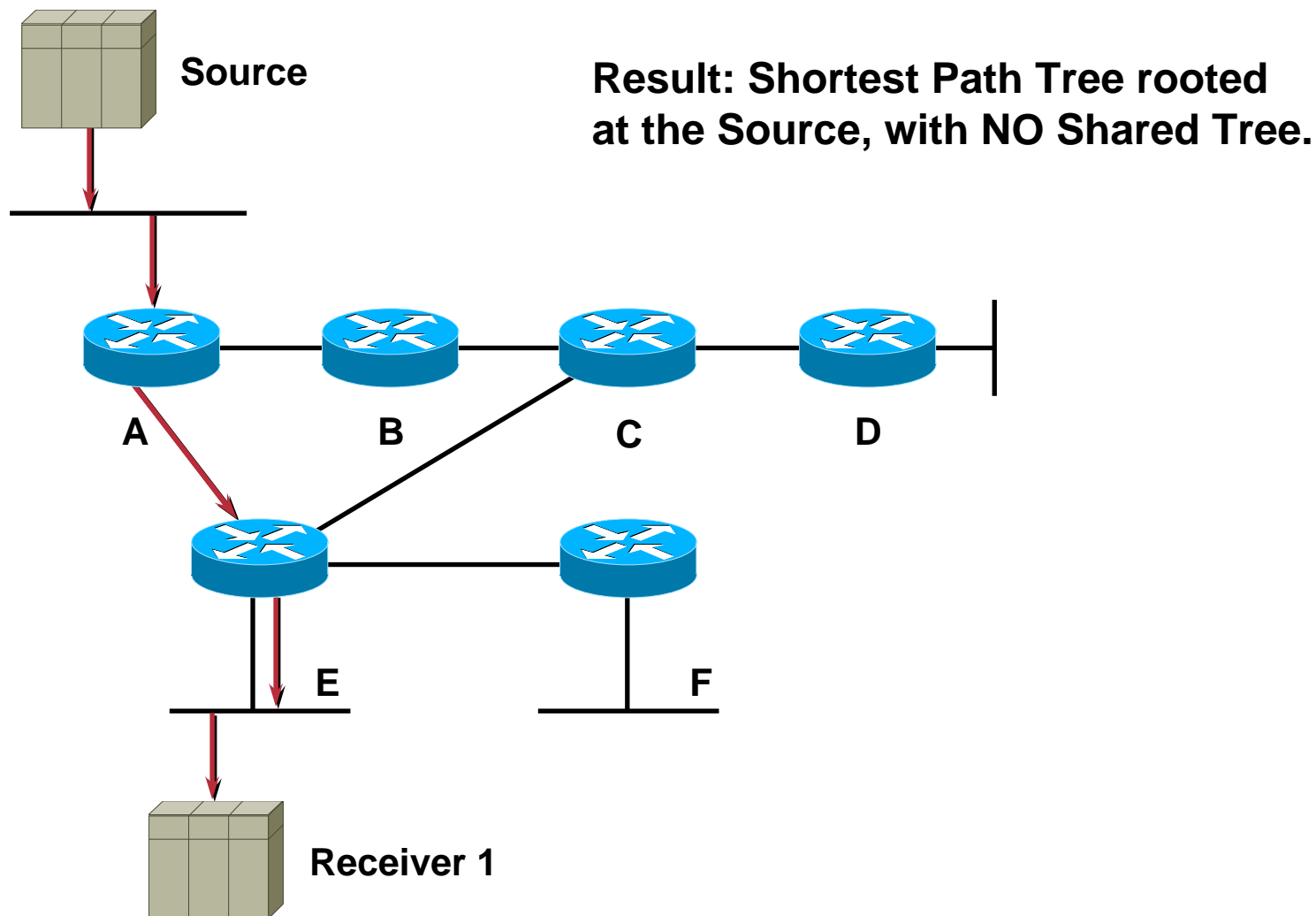   - **When used with MBGP, MSDP and/or SSM**

Source Specific Multicast

- **Assume a One-to-Many Multicast Model.**
  - **Example: Video/Audio broadcasts, Stock Market data**
- **Why does PIM-SM need a Shared Tree?**
  - **So that hosts and 1st hop routers can learn who the active source is for the group.**
- **What if this was already known?**
  - **Hosts could use IGMPv3 to signal exactly which (S,G) SPT to join.**
  - **The Shared Tree & RP wouldn't be necessary.**
  - **Different sources could share the same Group address and not interfere with each other.**
- **Result: Source Specific Multicast (SSM)**
- **RFC 3569 An Overview of Source-Specific Multicast (SSM)**
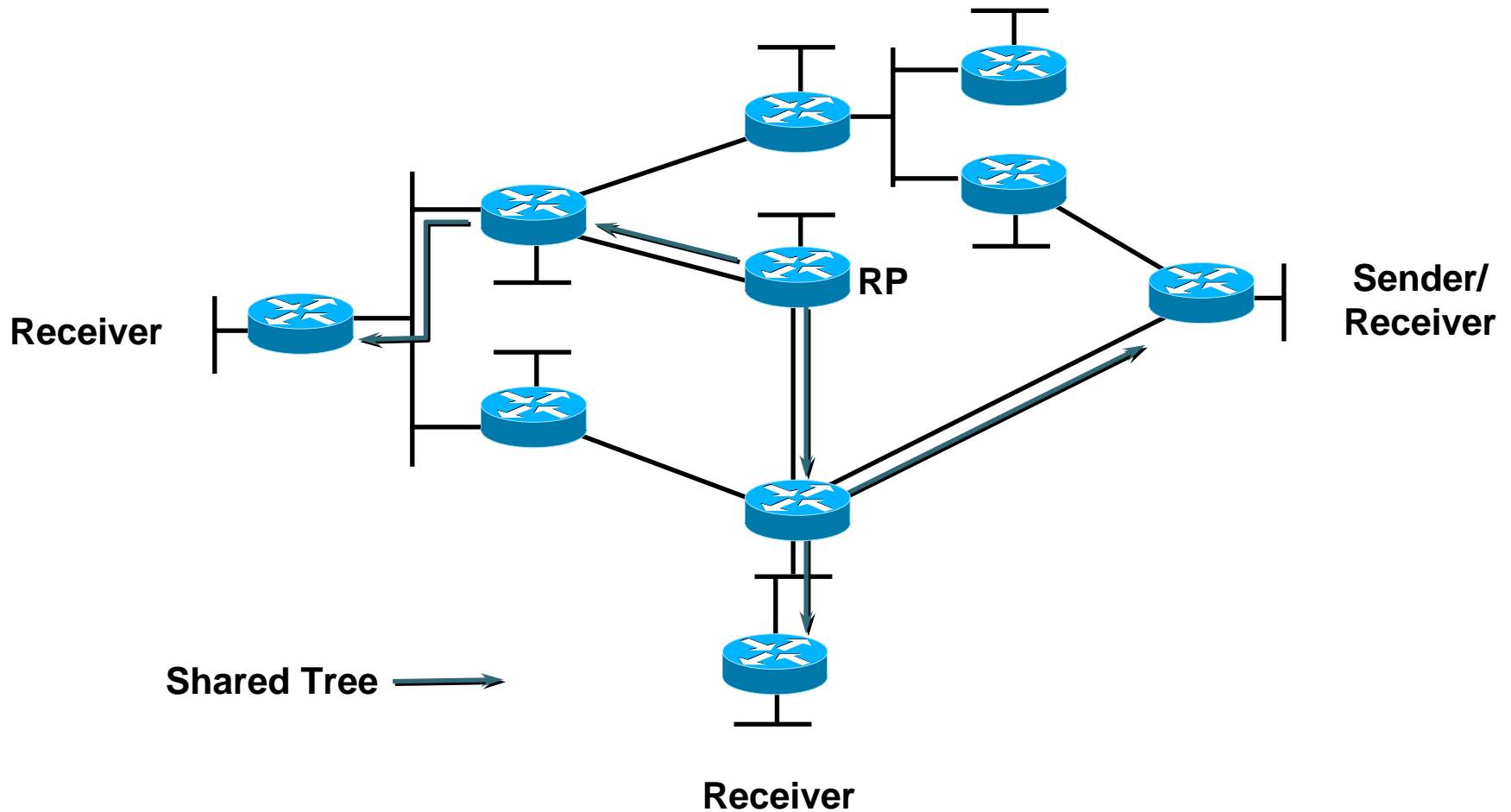
Cisco Public

PIM Source Specific Mode

Source

Receiver learns of source, group/port

Receiver sends IGMPv3 (S,G) Join

First-hop send PIM s,g join directly toward Source

A   B   C   D

(S, G) Join

Out-of-band Source Directory
Example: Web Server

D   **F**

IGMPv3 (S, G) Join

Receiver 1

PIM Source Specific Mode

Source

Result: Shortest Path Tree rooted at the Source, with NO Shared Tree.

A   B   C   D

E   F

Receiver 1

PIM-SSM - Evaluation

- **Ideal for applications with one source sending to many receivers**

- **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.**
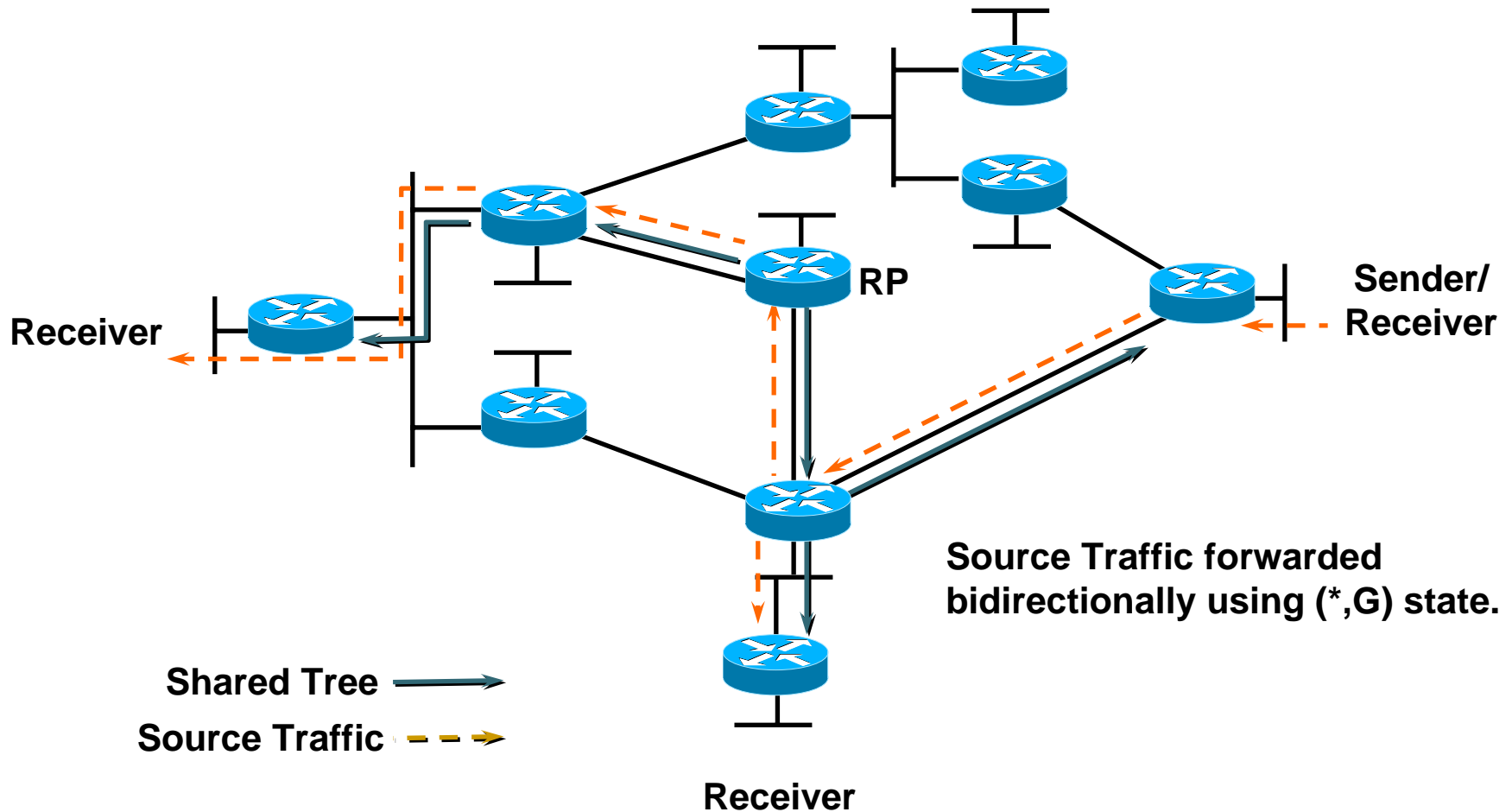   - **Not received by host application.**

Many-to-Any State Problem

- **Creates huge amounts of (S,G) state**
  - **State maintenance workloads skyrocket**
    - **High OIL fanouts make 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**

Bidirectional PIM—Overview

RP

Sender/
Receiver

Receiver

Shared Tree ——————➤

Receiver

Cisco Public 62

Bidirectional PIM—Overview

RP

Sender/
Receiver

Receiver

Source Traffic forwarded
bidirectionally using (*,G) state.

Shared Tree ⟶

Source Traffic - - - ▶

Receiver

Cisco Public 63

Bidir PIM–Evaluation

- **Ideal for Many to Many applications**

- **Drastically reduces network mroute state.**
    - **Eliminates ALL (S,G) state in the network.**
        - **SPT's 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.**

RP choices

Cisco Public

How does the network know about the RP ?

- **Static configuration**

- **AutoRP**

- **BSR**

- **MSDP Anycast**

Cisco Public 66

Static RP's

- **Hard-coded RP address**
  - **When used, must be configured on every router**
  - **All routers must have the same RP address**
  - **RP fail-over not possible**
    - **Exception: If Anycast RPs are used.**

- **Command**
  - `ip pim rp-address <address> [group-list <acl>] [override]`
  - **Optional group list specifies group range**
    - **Default: Range = 224.0.0.0/4 (Includes Auto-RP Groups!!!!)**
  - **Override keyword "overrides" Auto-RP information**
    - **Default: Auto-RP learned info takes precedence**

Auto-RP—From 10,000 Feet

MA
A

MA
B

Announce   C   Announce

C-RP
1.1.1.1

Announce   D   Announce

C-RP
2.2.2.2
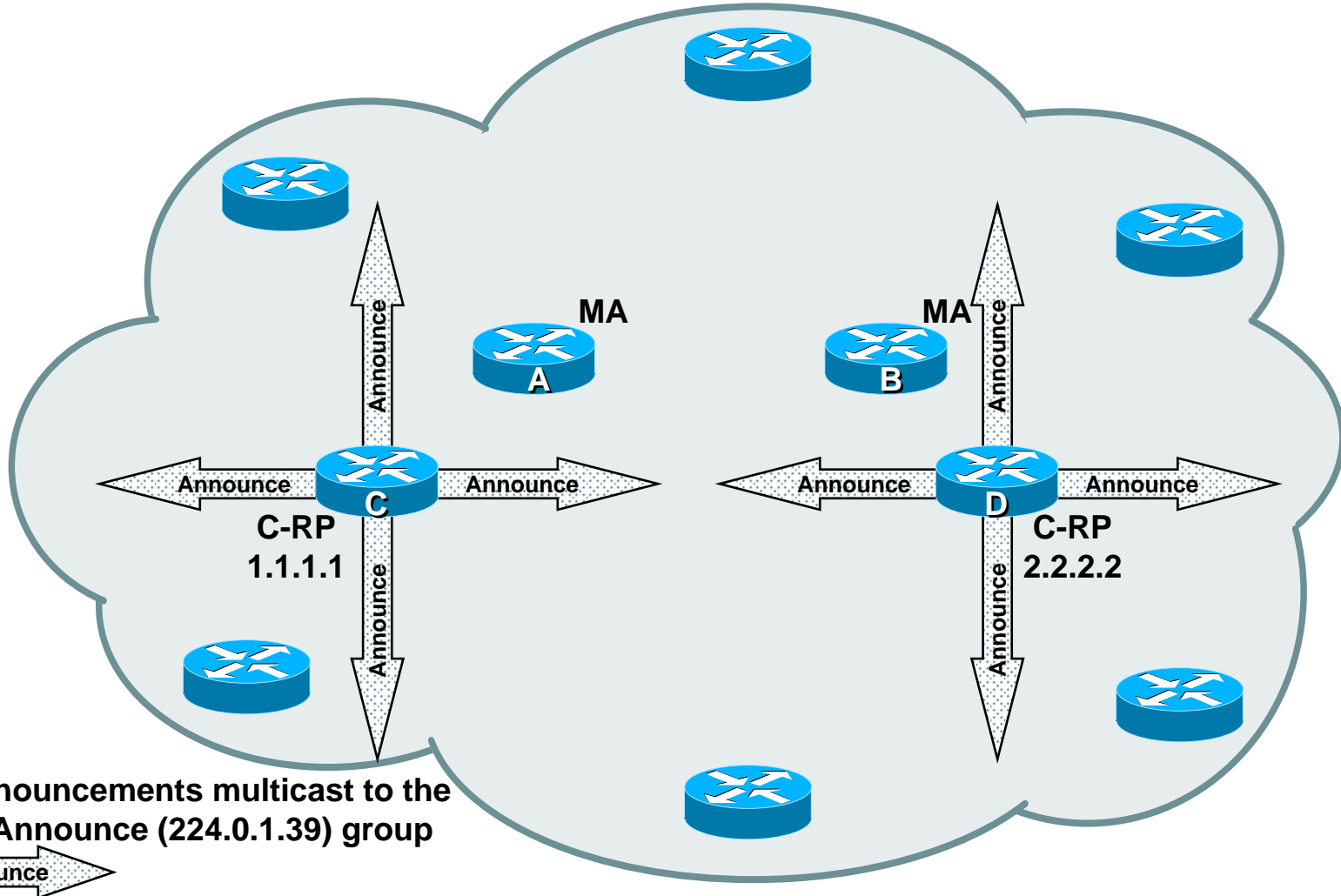
RP-Announcements multicast to the
Cisco Announce (224.0.1.39) group

Announce

Cisco Public   68

Auto-RP—From 10,000 Feet

MA

MA

A

B

C

D

C-RP
1.1.1.1

C-RP
2.2.2.2

Discovery

RP-Discoveries multicast to the Cisco Discovery (224.0.1.40) group

Discovery

Cisco Public
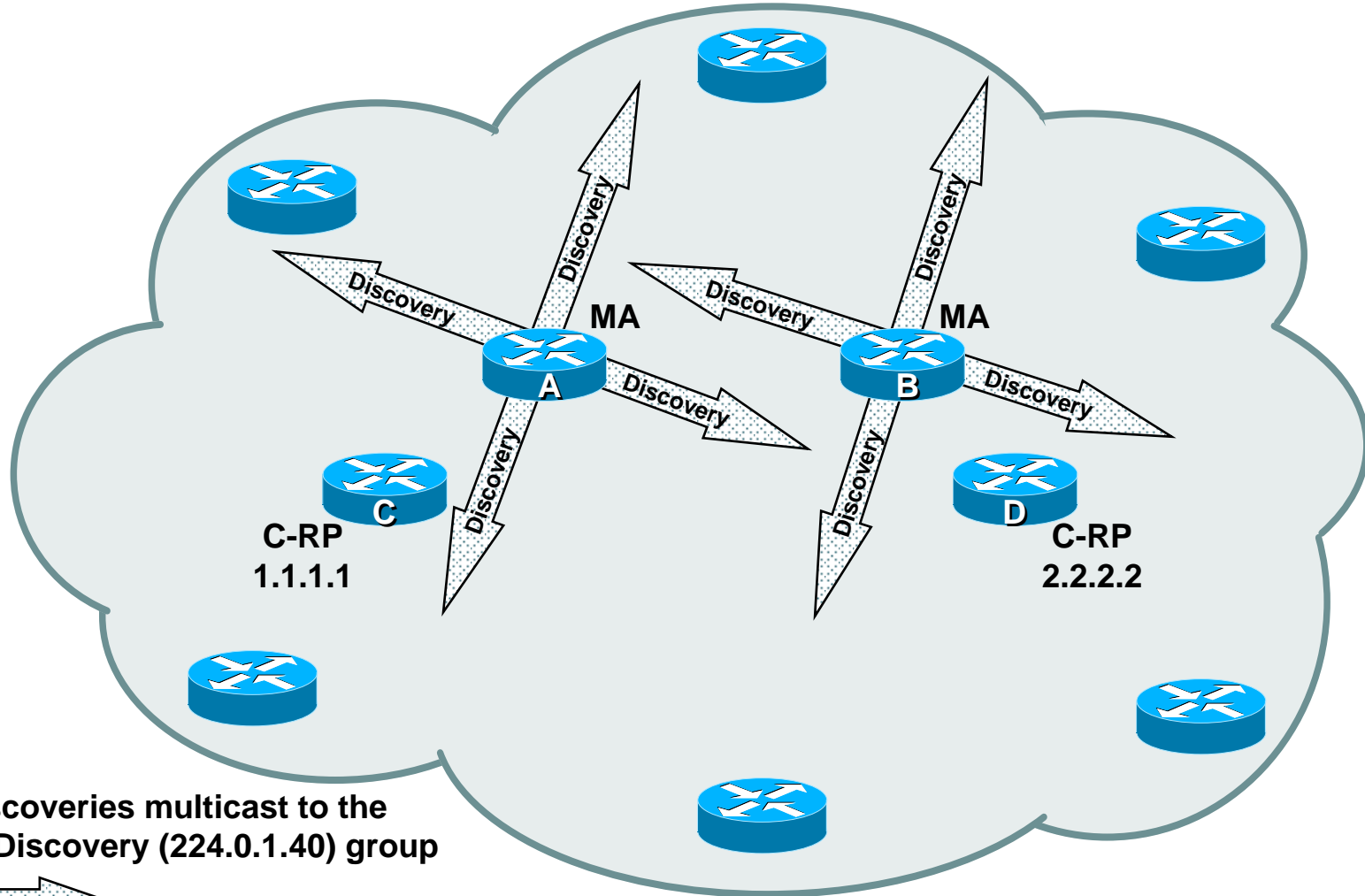
BSR Overview

- **A single Bootstrap Router (BSR) is elected**
  - **Multiple Candidate BSR's (C-BSR) can be configured**
    - **Provides backup in case currently elected BSR fails**
  - **C-RP's send C-RP announcements to the BSR**
    - **C-RP announcements are sent via unicast**
    - **BSR stores ALL C-RP announcements in the "RP-set"**
  - **BSR periodically sends BSR messages to all routers**
    - **BSR Messages contain entire RP-set and IP address of BSR**
    - **Messages are flooded hop-by-hop throughout the network away from the BSR**
  - **All routers select the RP from the RP-set**
    - **All routers use the same selection algorithm; select same RP**
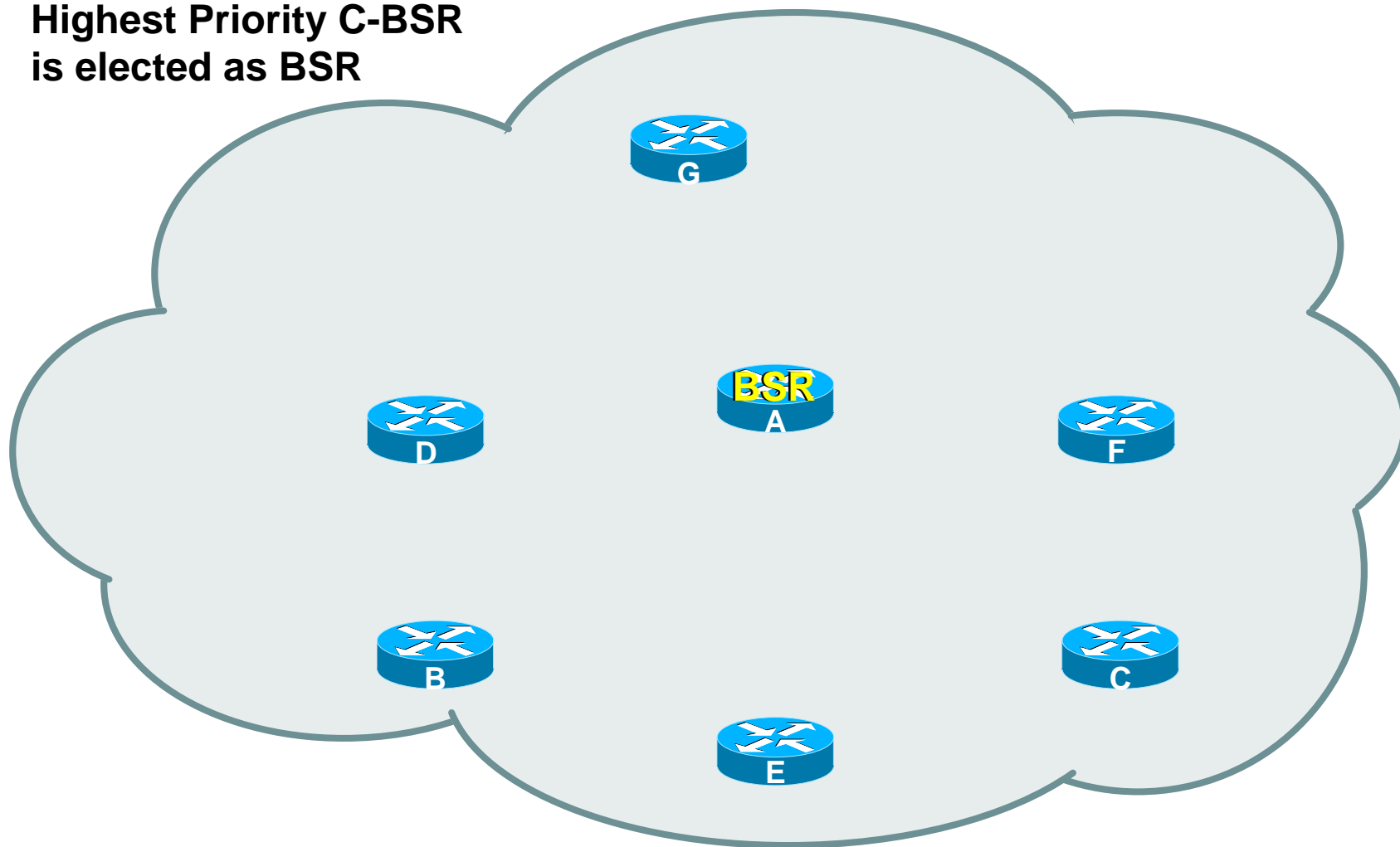- **BSR cannot be used with Admin-Scoping**
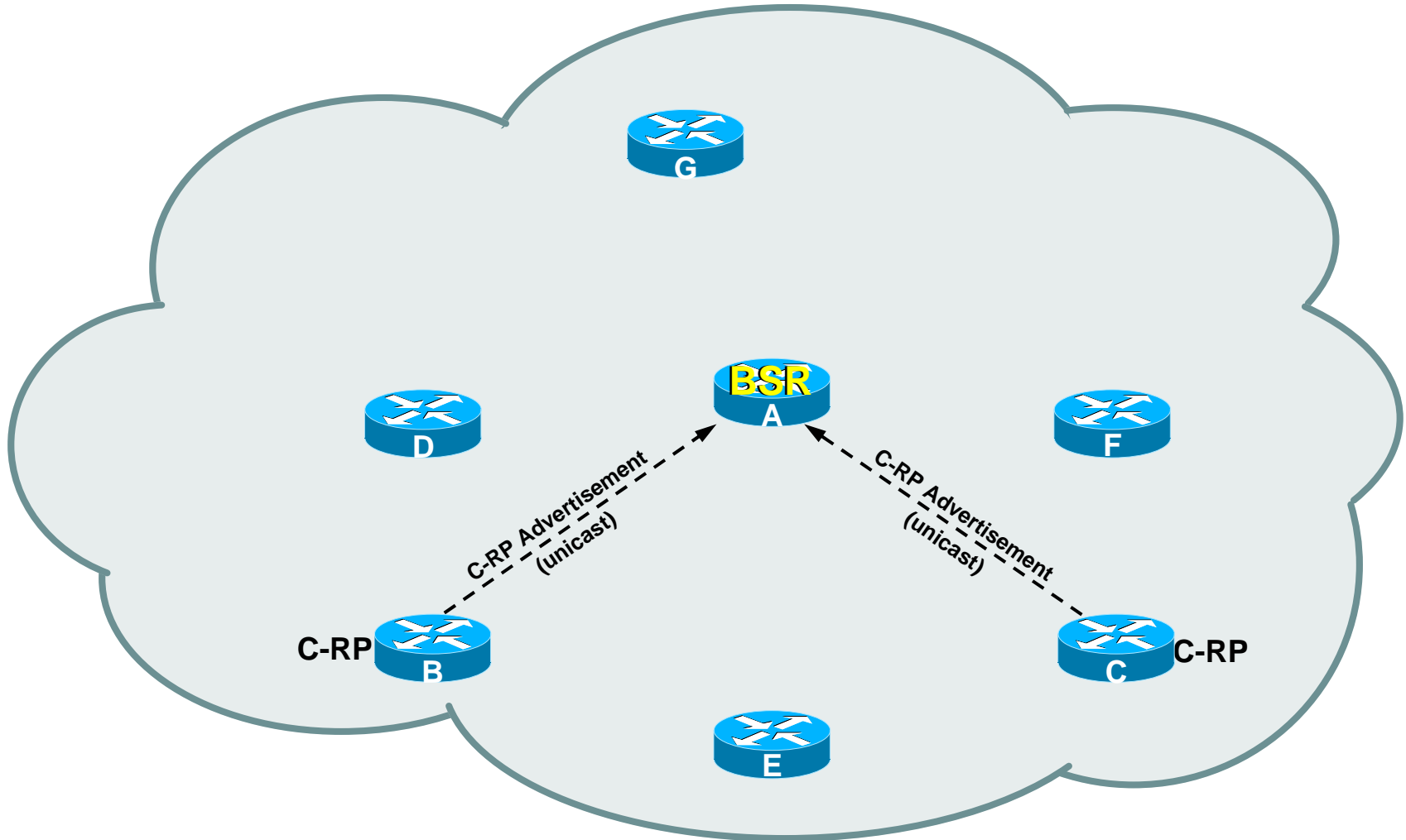
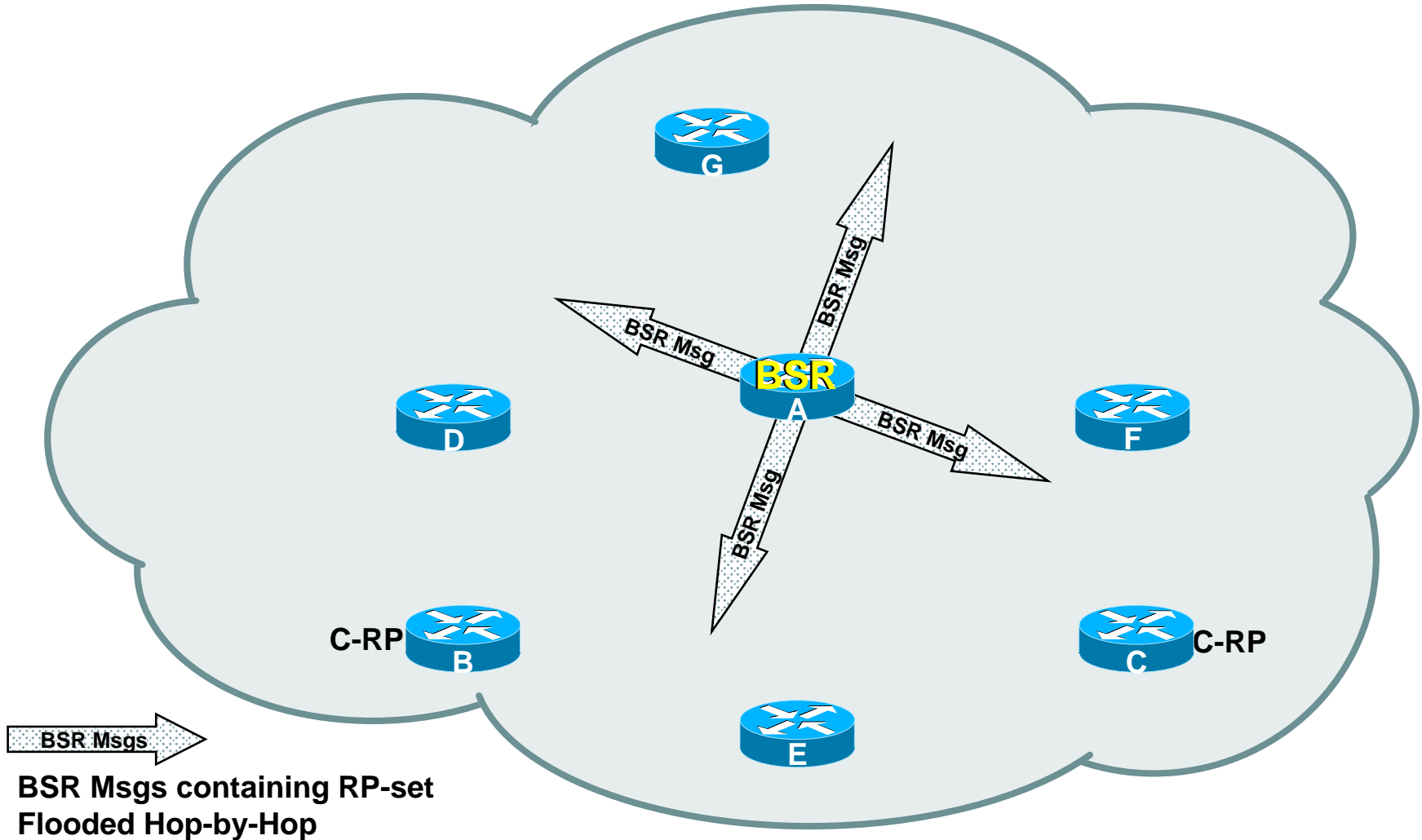# BSR – From 10,000 Feet

BSR Election Process

BSR Msgs

BSR Msgs Flooded Hop-by-Hop

Cisco Public   71
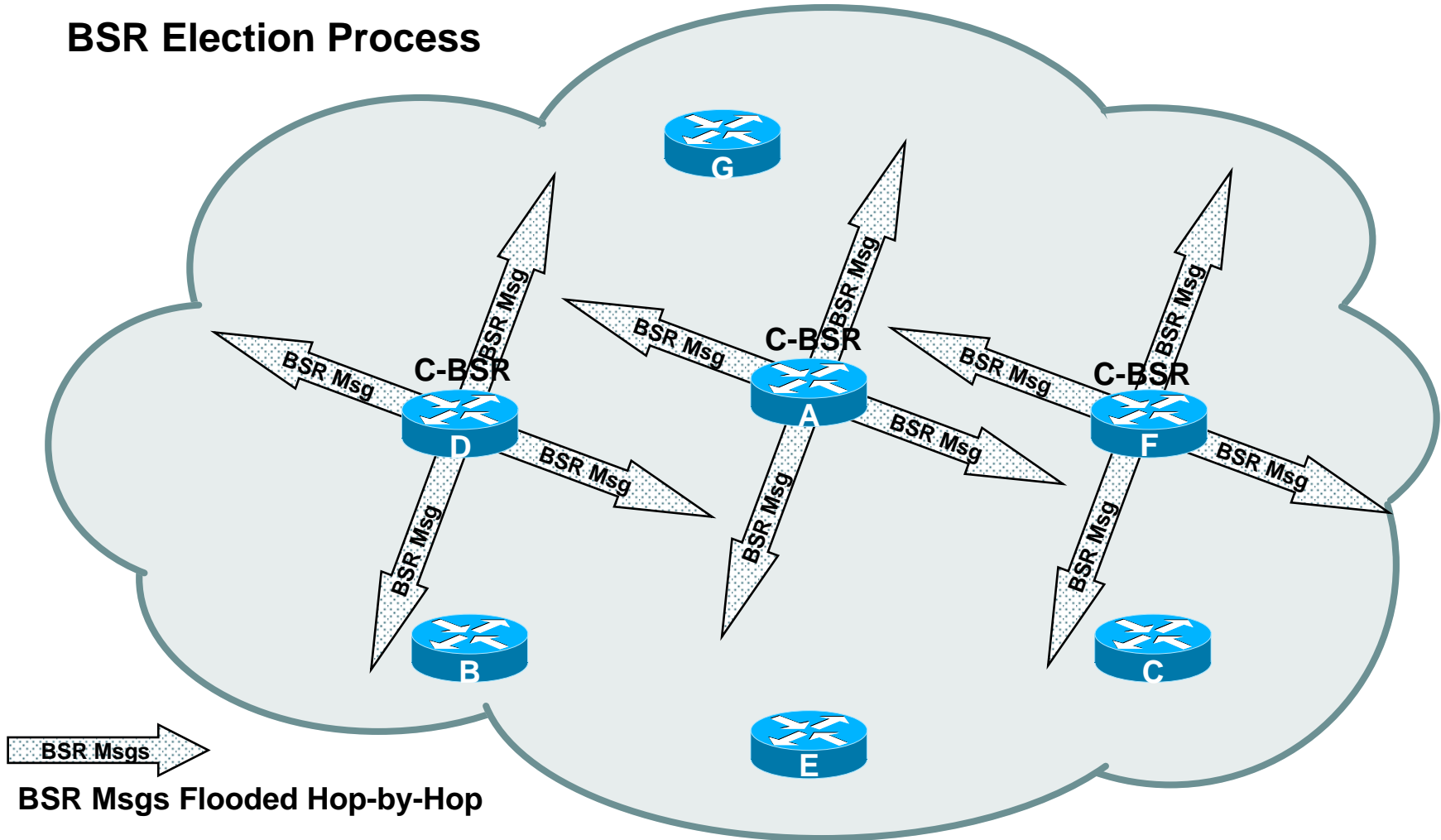
BSR – From 10,000 Feet

**Highest Priority C-BSR
is elected as BSR**

Cisco Public 72

BSR – From 10,000 Feet

Cisco Public 73

BSR – From 10,000 Feet

BSR Msgs

**BSR Msgs containing RP-set
Flooded Hop-by-Hop**

Cisco Public 74

Multicast at Layer 2

Cisco Public

L2 Multicast Frame Switching

## Problem: Layer 2 Flooding of Multicast Frames

- **Typical L2 switches treat multicast traffic as unknown or broadcast and must "flood" the frame to every port**

- **Static entries can sometimes be set to specify which ports should receive which group(s) of multicast traffic**

- **Dynamic configuration of these entries would cut down on user administration**

PIM

Multicast M

Cisco Public 76
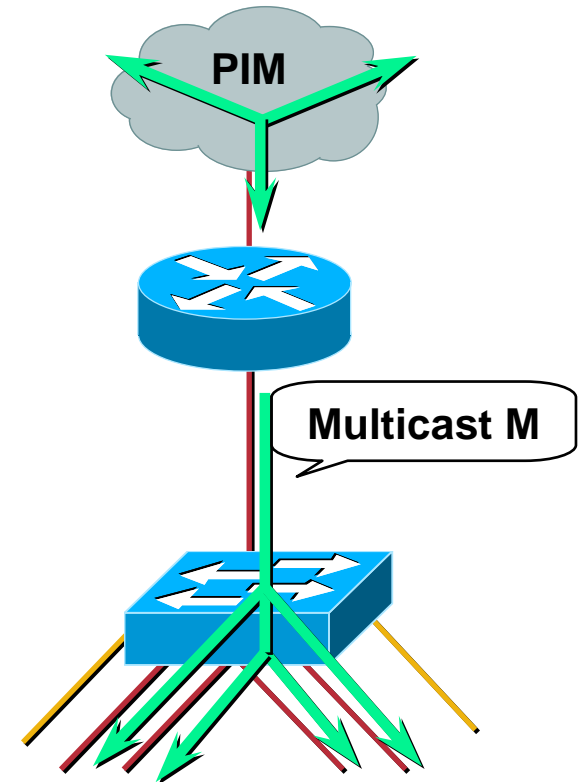
IGMPv1-v2 Snooping

- **Switches become "IGMP" aware**
- **IGMP packets intercepted by the NMP or by special hardware ASICs**
  - **Requires special hardware to maintain throughput**
- **Switch must examine contents of IGMP messages to determine which ports want what traffic**
  - **IGMP membership reports**
  - **IGMP leave messages**
- **Impact on low-end Layer-2 switches:**
  - **Must process ALL Layer 2 multicast packets**
  - **Admin. load increases with multicast traffic load**
  - **Generally results in switch *Meltdown* !!!**

PIM
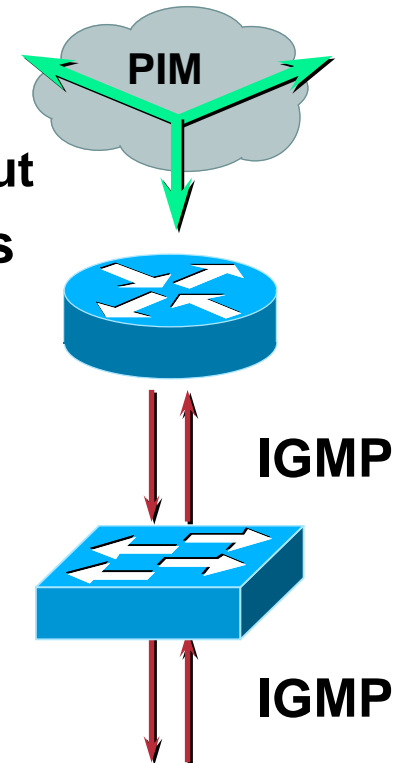
IGMP

IGMP

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
    - Permits (S,G) state to be maintained by switch
      - » Currently not implemented by any switches.
    - May be necessary for full IGMPv3 functionality

Summary—Frame Switches

- ## 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*!**

            » **Shouldn't be a problem when IGMPv3 is implemented**

Interdomain IP Multicast

Cisco Public

MBGP Overview

- **MBGP: Multiprotocol BGP**
  - **Defined in RFC 2858 (extensions to BGP)**
  - **Can carry different types of routes**
    - **Unicast**
    - **Multicast**
  - **Both routes carried in same BGP session**
  - **Does not propagate multicast state info**
  - **Same path selection and validation rules**
    - **AS-Path, LocalPref, MED, …**

MBGP Overview

- **Separate BGP tables maintained**
  - **Unicast Routing Information Base (URIB)**
  - **Multicast Routing Information Base (MRIB)**
- **URIB**
  - **Contains unicast prefixes for unicast forwarding**
  - **Populated with BGP unicast NLRI**
    - **AFI = 1, Sub-AFI = 1**
- **MRIB**
  - **Contains *unicast* prefixes for RPF checking**
  - **Populated with BGP multicast NLRI**
    - **AFI = 1, Sub-AFI = 2**

MBGP Overview

- **MBGP allows divergent paths and policies**
  - **Same IP address holds dual significance**
    - **Unicast routing information**
    - **Multicast RPF information**
  - **For same IPv4 address two different NLRI with different next-hops**
  - **Can therefore support both congruent and incongruent topologies**
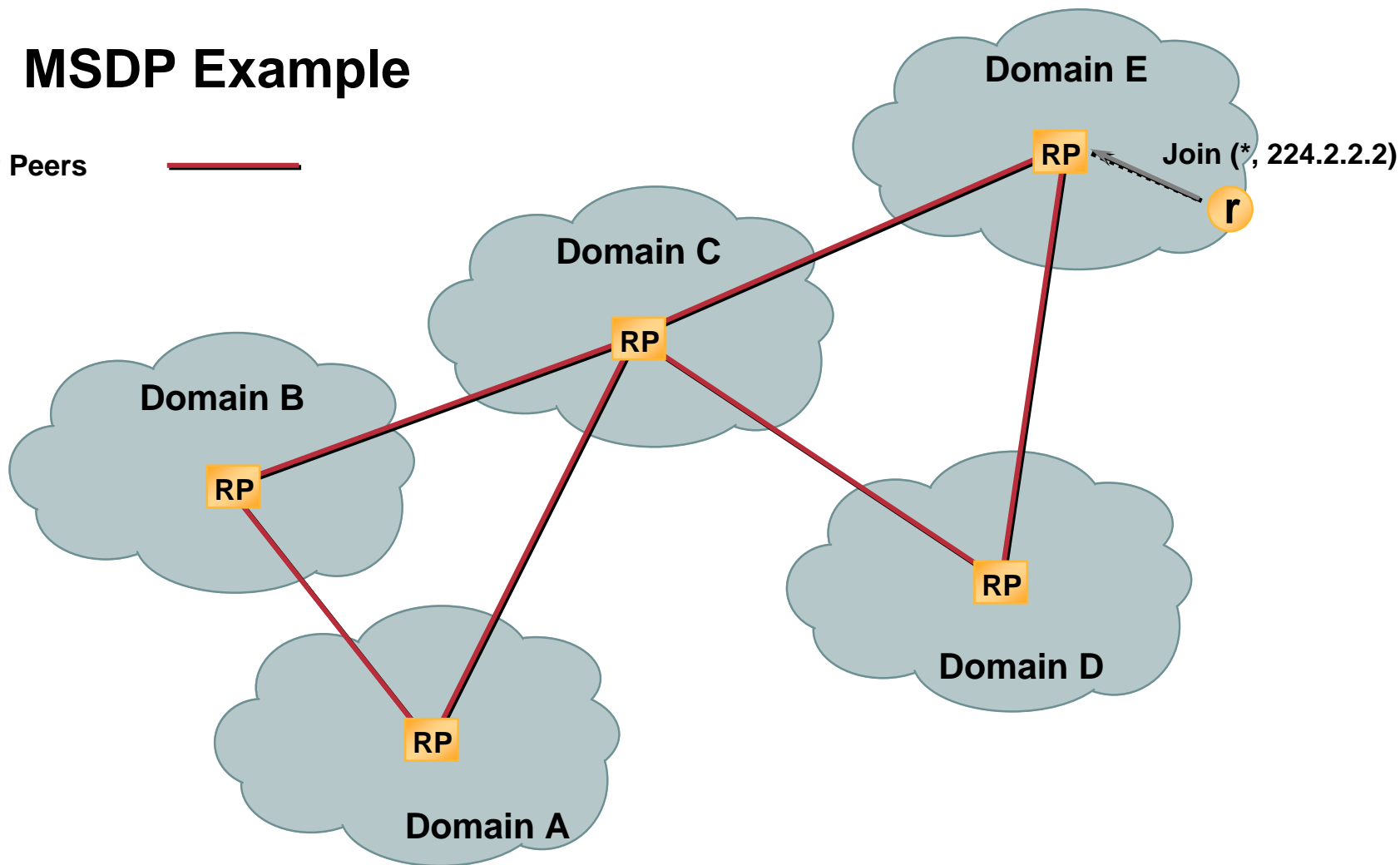
Cisco Public   83

MSDP

- **Based on RFC 3618**

- **Works with PIM-SM only**
  - **RP's knows about all sources in their domain**
    - **Sources cause a "PIM Register" to the RP**
    - **Tell RP's in other domains of it's sources**
      - » **Via MSDP SA (Source Active) messages**
  - **RP's know about receivers in a domain**
    - **Receivers cause a "(*, G) Join" to the RP**
    - **RP can join the source tree in the peer domain**
      - » **Via normal PIM (S, G) joins**

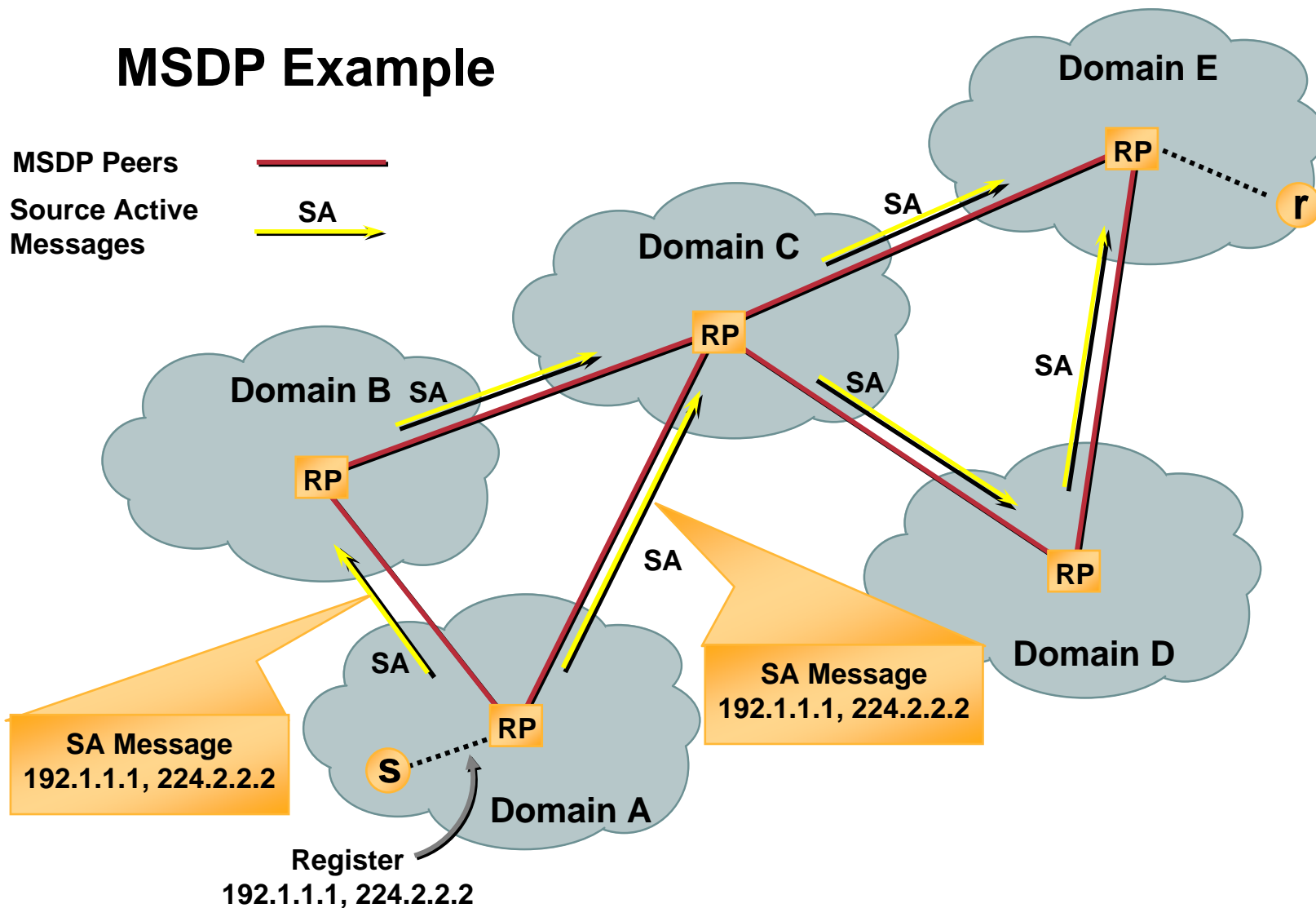# MSDP Overview

MSDP Example

MSDP Peers ———

Domain E

Domain C

Domain B

RP — Join (*, 224.2.2.2)

r

Domain D

Domain A

MSDP Overview

MSDP Example

Domain E

MSDP Peers

Source Active Messages SA
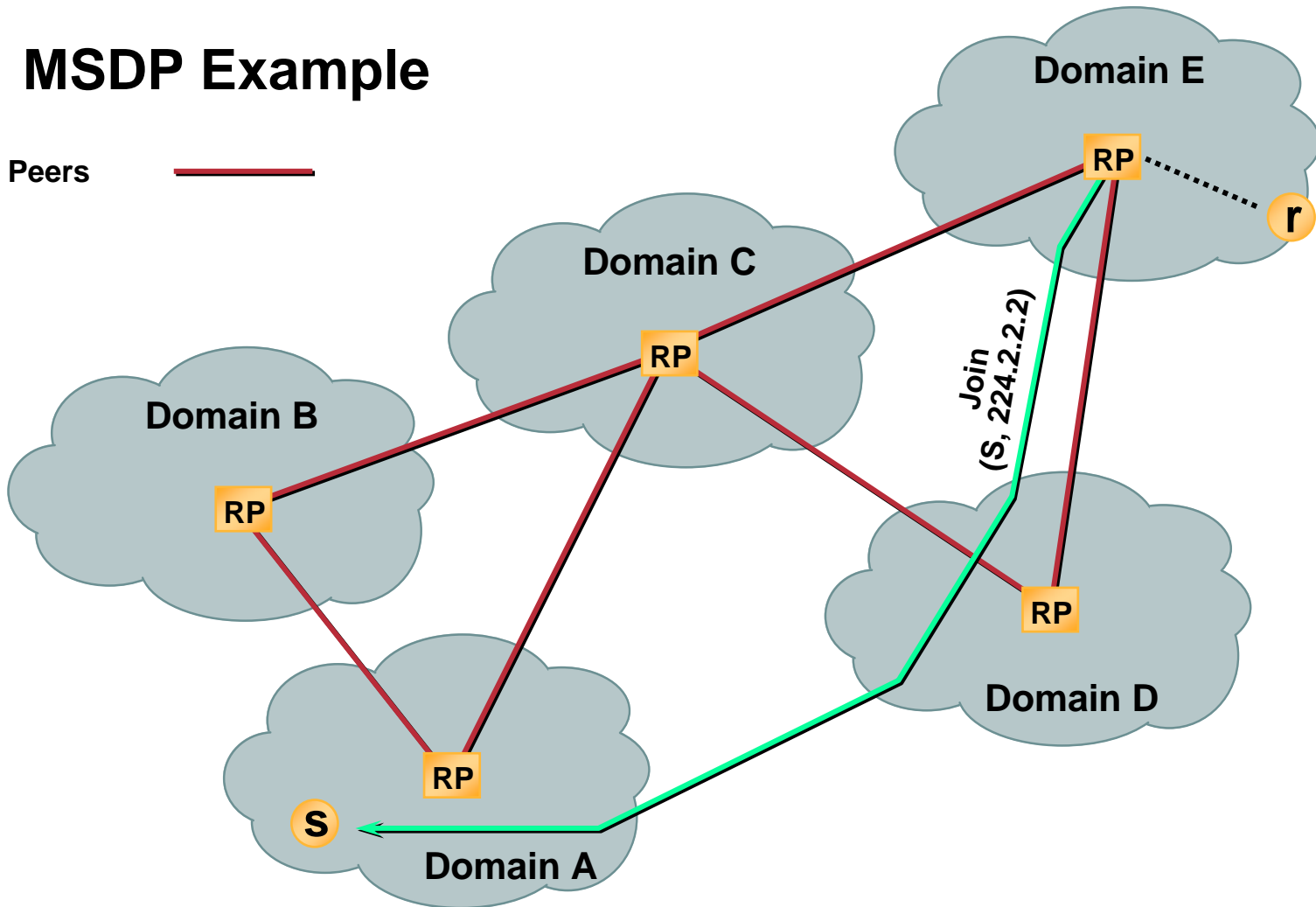
RP r

SA

Domain C

RP

SA

Domain B SA

RP SA

SA

RP

SA

**SA Message
192.1.1.1, 224.2.2.2**

Domain D

SA

**SA Message
192.1.1.1, 224.2.2.2**

**SA Message
192.1.1.1, 224.2.2.2**

RP

S

Domain A

**Register
192.1.1.1, 224.2.2.2**

MSDP Overview

MSDP Example

MSDP Peers ————

Domain E

Domain C

Domain B

RP

RP

RP

RP

RP

RP

r

Join
(S, 224.2.2.2)

S

Domain A

Domain D

MSDP Overview

MSDP Example

MSDP Peers ————————

Multicast Traffic ————————

Domain E

Domain C

Domain B

Domain D

Domain A

RP

RP

RP

RP

RP

S

r

MSDP Overview

MSDP Example

MSDP Peers ────────

Multicast Traffic ────────

Domain E

RP

r

Join
(S, 224.2.2.2)

Domain C

RP

Domain B

RP

RP

Domain D

RP

S

Domain A

MSDP Overview

MSDP Example

MSDP Peers ———————

Multicast Traffic ———————

Domain E

Domain C

Domain B

Domain D

Domain A

RP
RP
RP
RP
RP
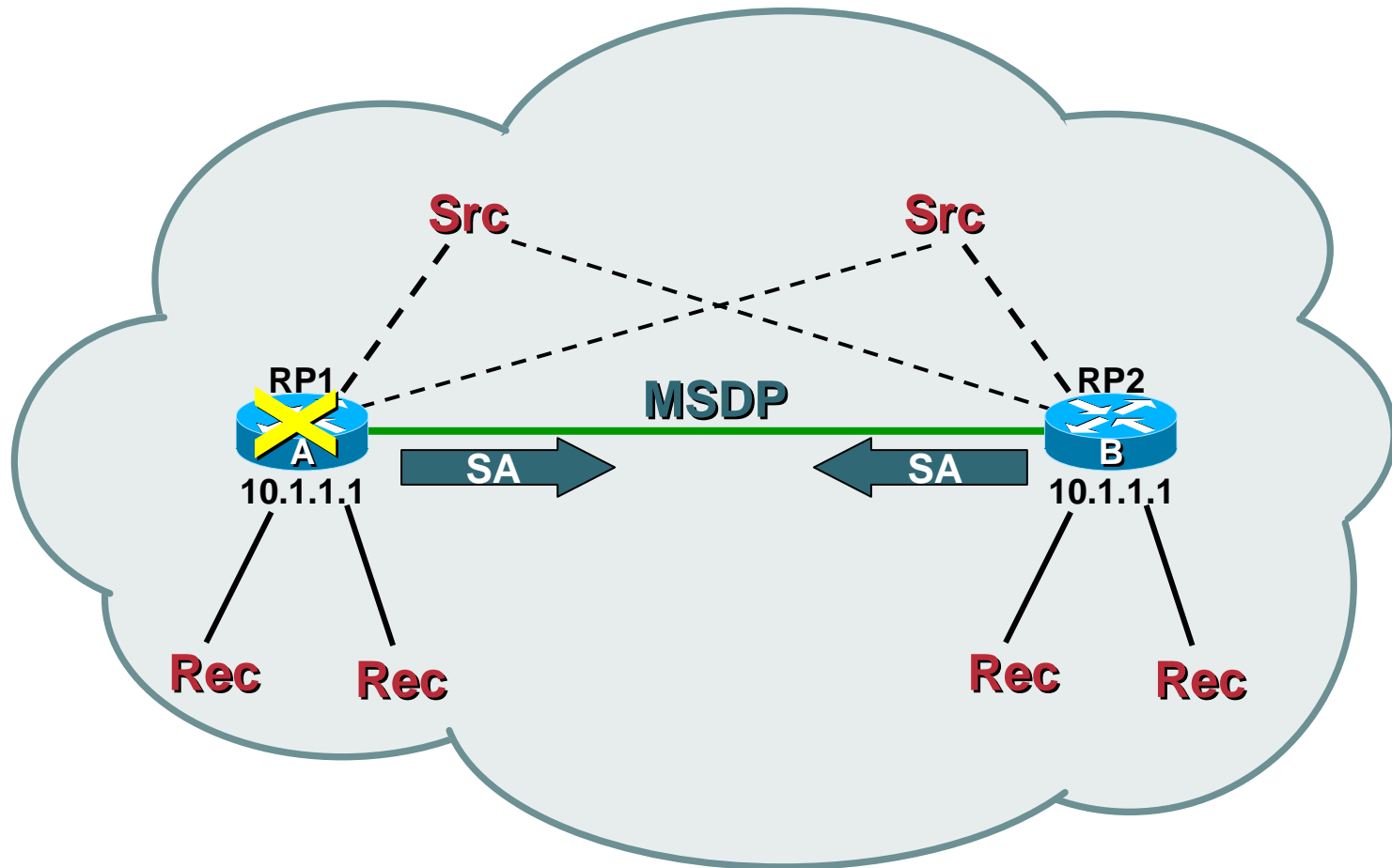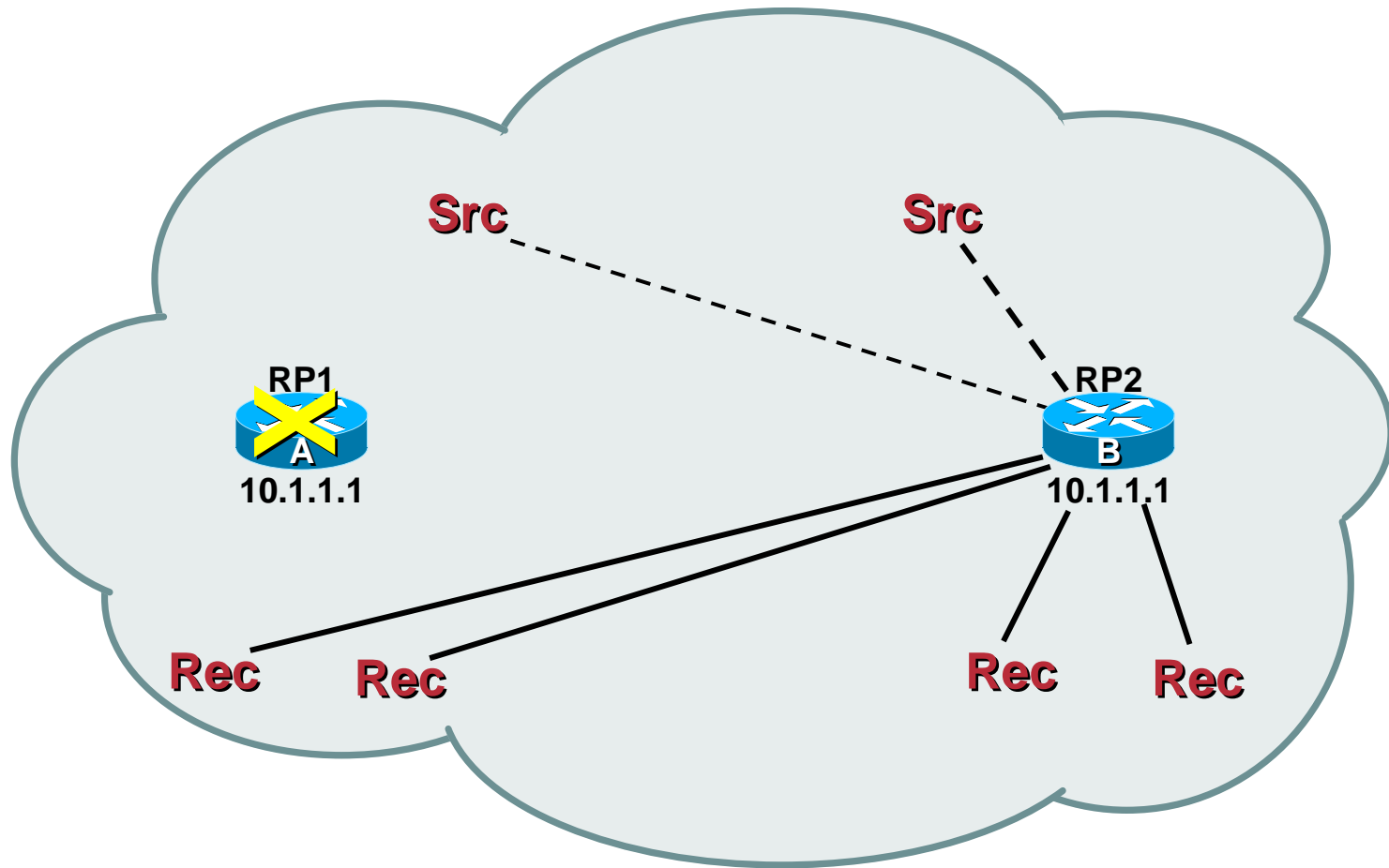
S

r
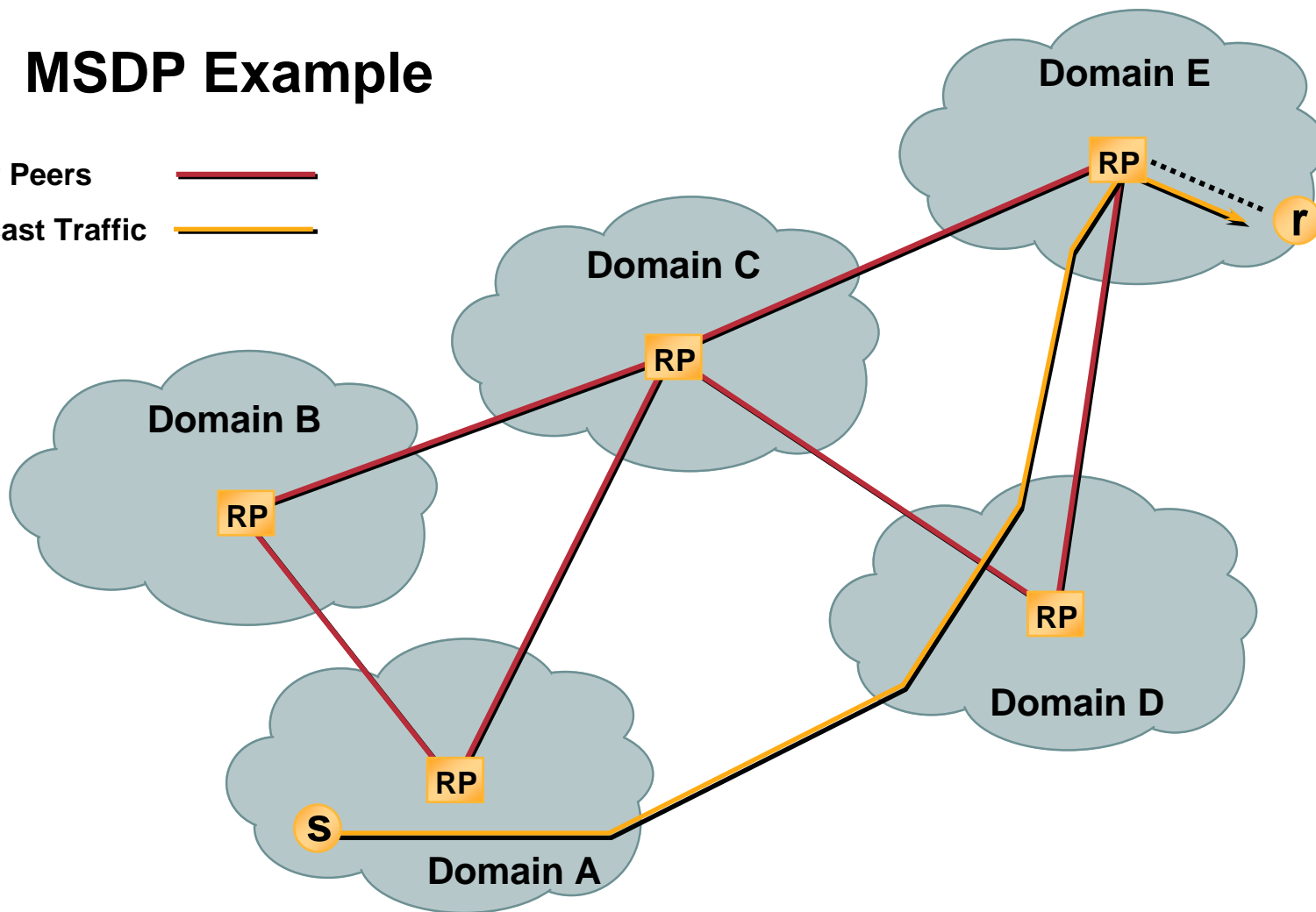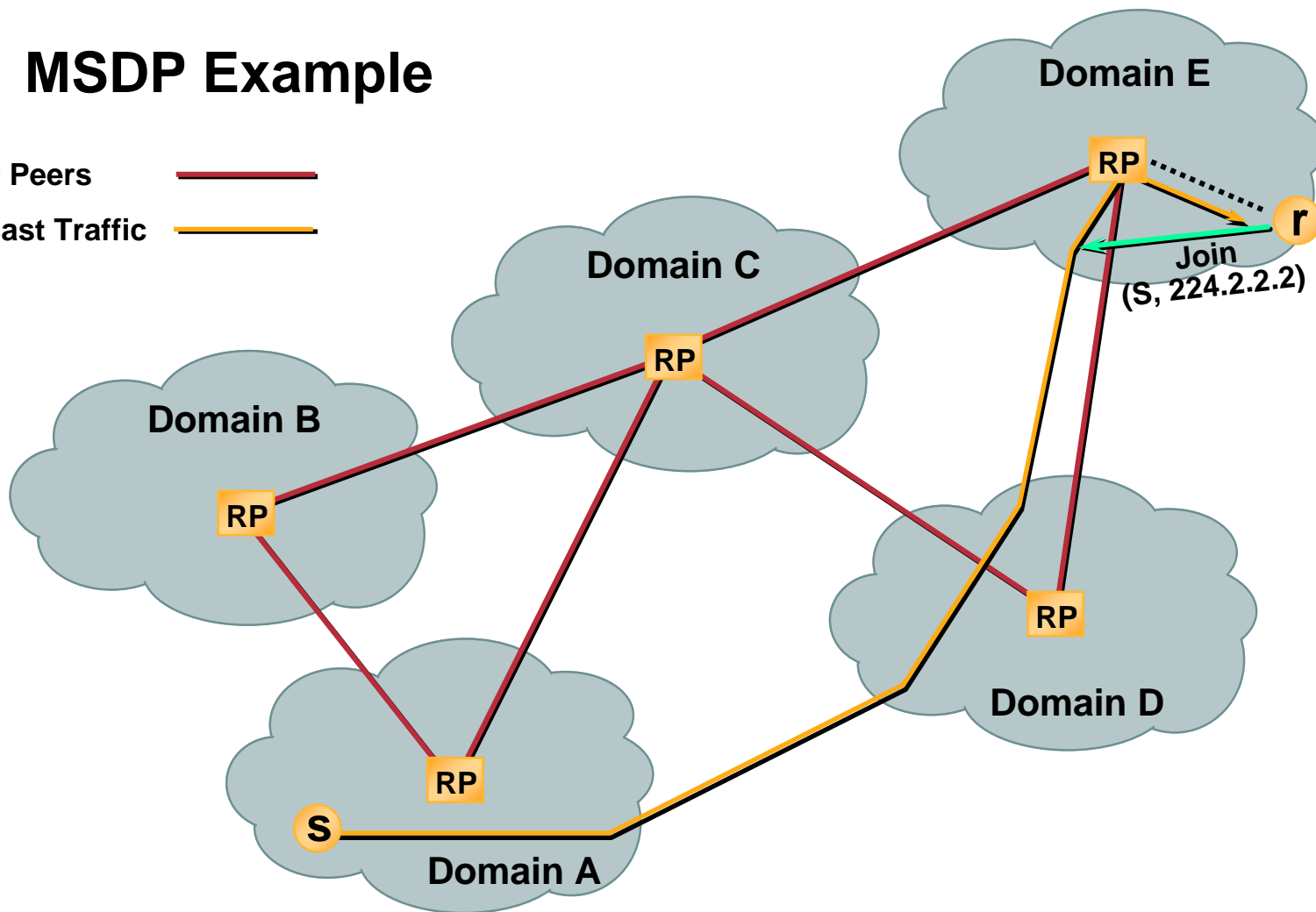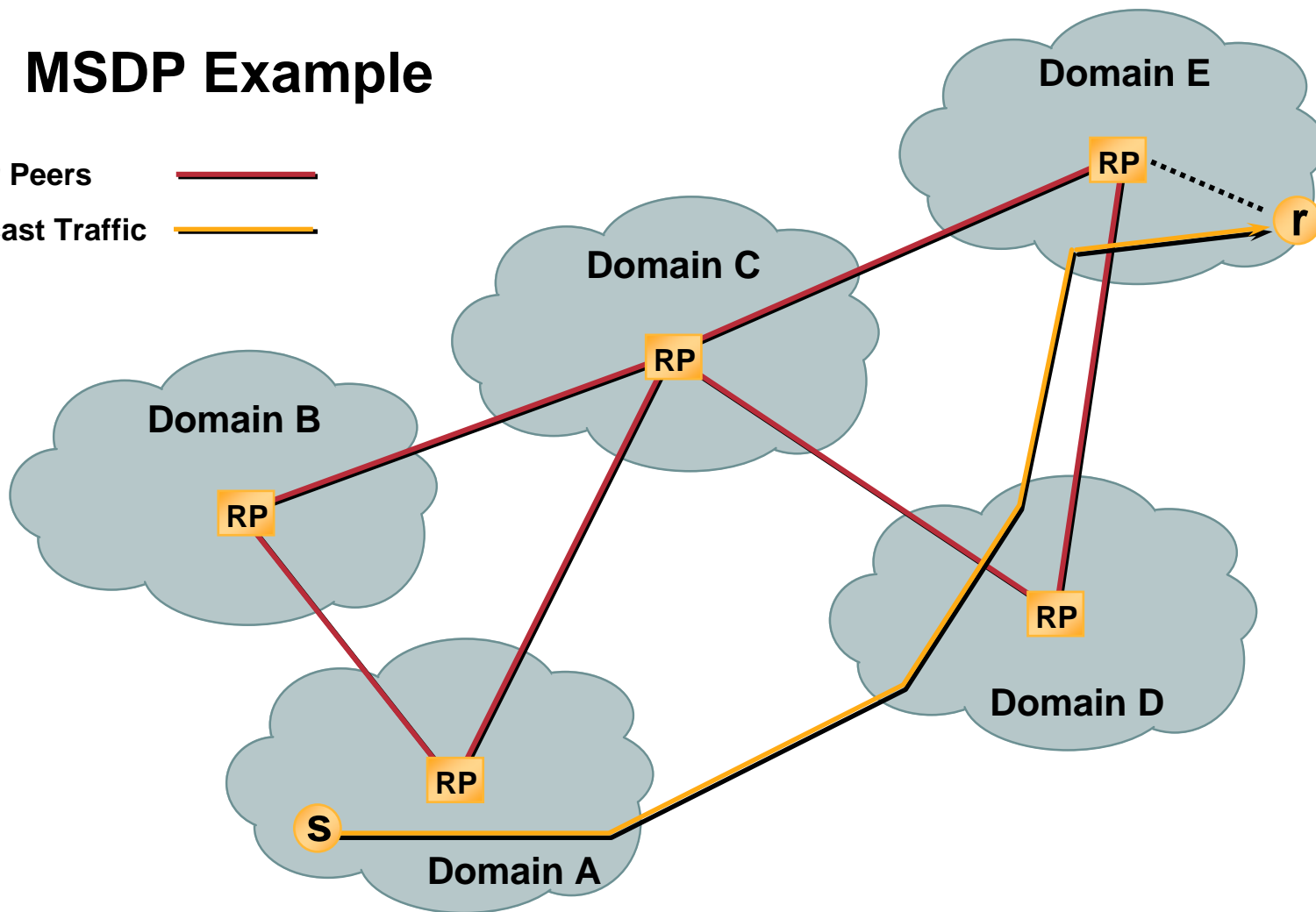
Anycast RP – Overview

- ## Uses single defined RP address
  - **Two or more routers have same RP address**
    - **RP address defined as a Loopback Interface.**
    - **Loopback address advertised as a Host route.**
  - **Senders & Receivers Join/Register with closest RP**
    - **Closest RP determined from the unicast routing table.**
  - **Can *never* fall back to Dense mode.**
    - **Because RP is statically defined.**

- ## MSDP session(s) run between all RPs
  - **Informs RPs of sources in other parts of network**
  - **RPs join SPT to active sources as necessary**

Anycast RP – Overview

RP1
A
10.1.1.1

RP2
B
10.1.1.1

MSDP

SA

SA

Src

Src

Rec

Rec

Rec

Rec

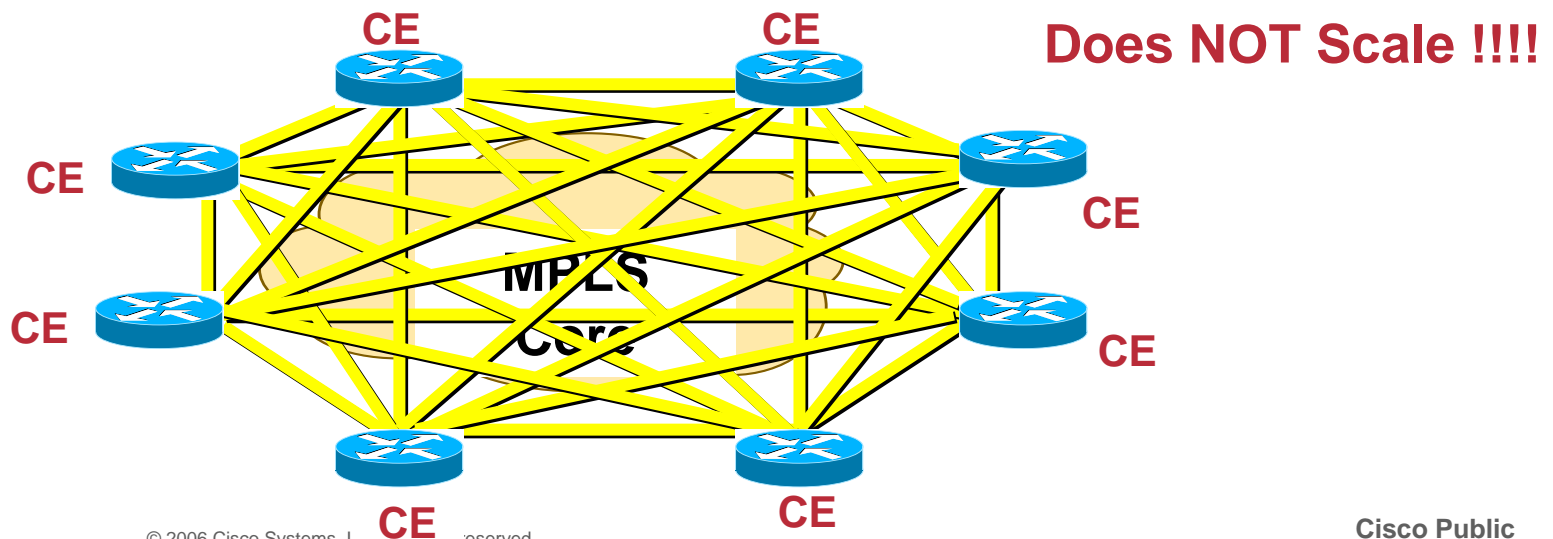# Anycast RP – Overview

Cisco Public

Latest Additions

Cisco Public

Multicast VPN – Customer Requirement

- **MPLS VPN customers want to run multicast within their VPNs**

- **Multicast deployment is expanding**

- **MPLS VPNs do not support multicast today**

- **Multicast options in MPLS VPNs today**
  - **GRE tunnels from CE to CE**

CE **CE** **Does NOT Scale !!!!**

CE

CE

MPLS Core

CE

CE

CE

CE

CE **CE**

Multicast VPN (MVPN)

- **Allows an ISP to provide its MPLS VPN customers the ability to transport their Multicast traffic across MPLS packet-based core networks**

- **Requires IPmc enabled in the core**

- **MPLS still used to support unicast**

- **A scalable architecture solution for MPLS networks based on native multicast deployment in the core**

- **Uses draft-rosen-vpn-mcast encapsulation and signaling to build MVPN Multicast VPN (MVPN)**

Multicast VPN – Terminology used

- **VPN: Virtual Private Network**
  - **Although different VPN models exist, the discussion here is for MPLS based VPNs**

- **MVPN: Multicast VPN**
  - **A VPN that supports multicast natively**

- **VRF: VPN Routing and Forwarding**
  - **per-site forwarding tables**

- **MVRF: Multicast VRF**
  - **A VRF that supports unicast and multicast tables**

Multicast VPN – Terminology used

- **MDT: Multicast Distribution Tree**
  - A multicast tree, built in the core network between PE and P routers that distributes multicast traffic between sites

- **Default-MDT:**
  - Default MDT group used for control traffic and flooding channel for dense mode and low bandwidth groups.

- **Data-MDT:**
  - MDT group created on demand for MVPN (S,G) pairs, usually high bandwidth traffic

Multicast VPN (MVPN) Concept & Fundamentals
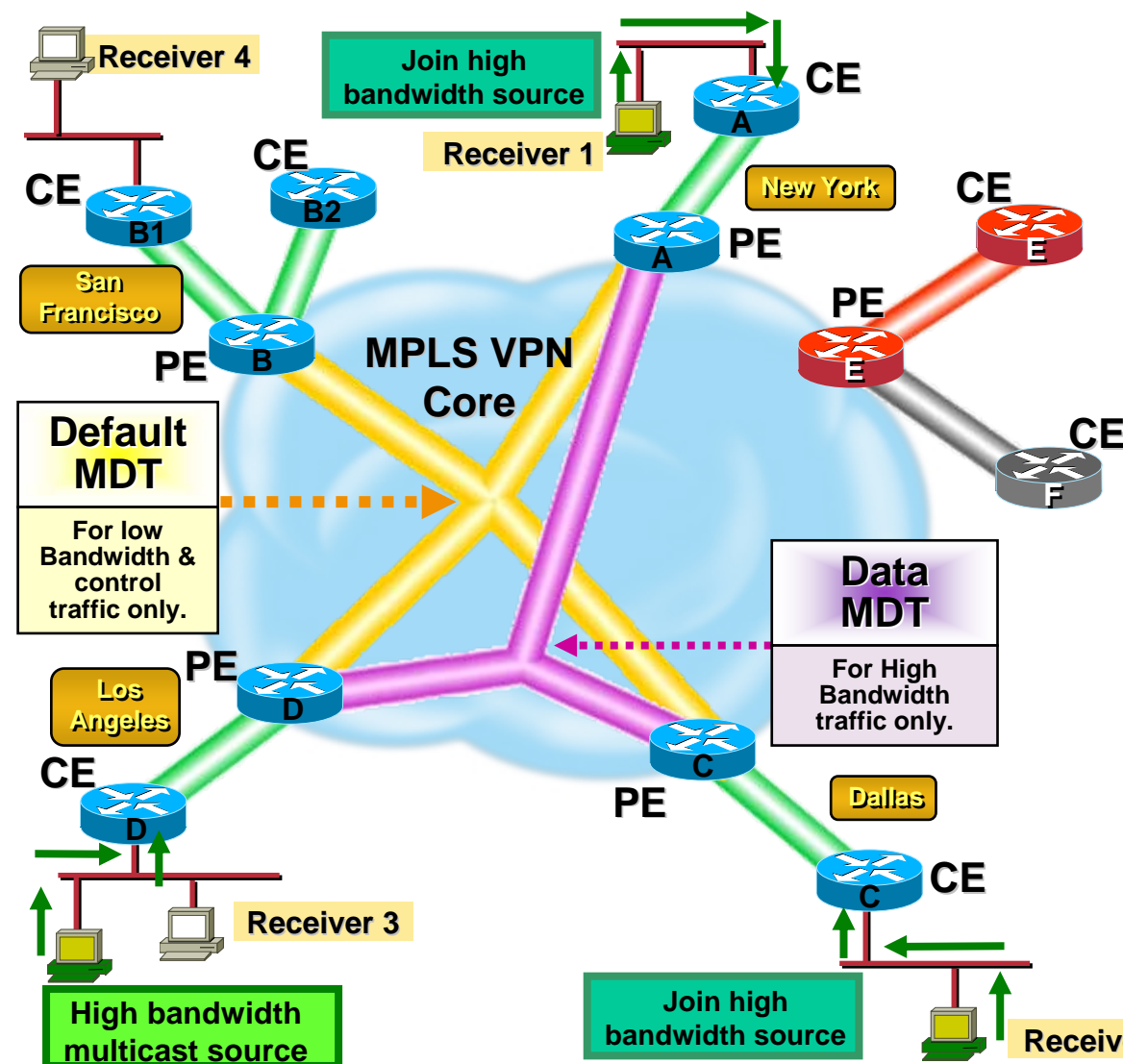
- **Customer CE devices joins the MPLS Core through provider's PE devices**

- **The MPLS Core forms a Default MDT for a given Customer**

- **A High-bandwidth source for that customer starts sending traffic**

- **Interested receivers 1 & 2 join that High Bandwidth source**

- **Data-MDT is formed for this High-Bandwidth source**

PGM: Specifications

- **RFC 3208 : Pragmatic Generic Multicast** (PGM) Reliable Transport Protocol Specification (experimental)

  – Defines an industry standard protocol for scalable, reliable multicast

  – PGM guarantees that a receiver in the group either receives all data packets from transmissions and repairs, **or is able to detect unrecoverable data packet loss** (just like TCP, but for IP multicast !)

- **draft-ietf-rmt-bb-pgmcc:** PGM congestion control (PGM-CC)

  – Provides TCP compatible flow control (eg: RED) based on slowest receiver (with minimum rate). Without CC, source app. controls rate.
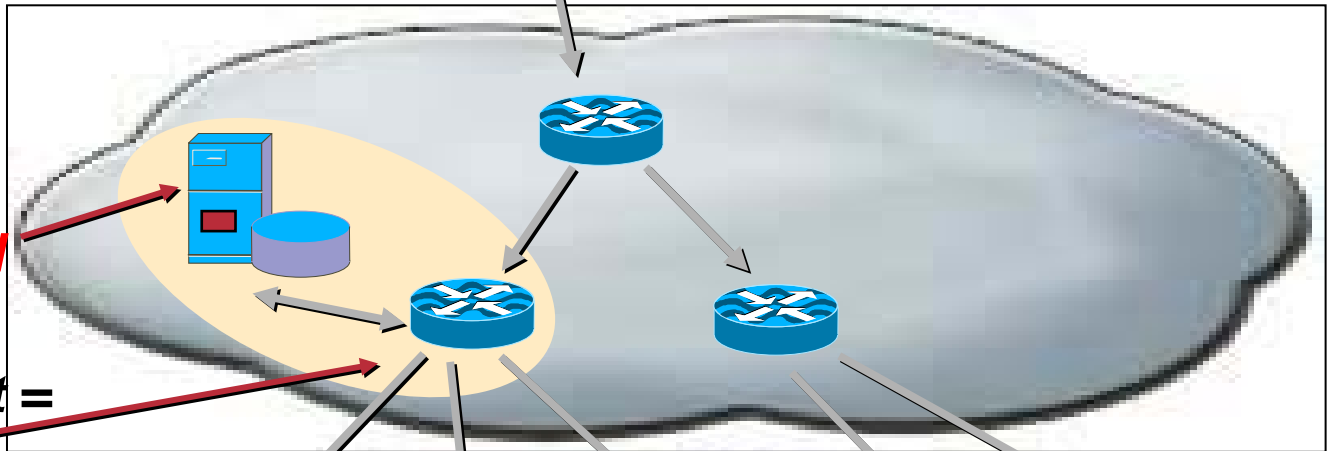
PGM: Components

Server

PGM
(Server/Source) Stack
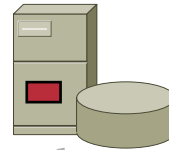
Network

Optional PGM
functions
*DLR: Designated
Local Repairer*

*Network Element =
Router Assist*

Host

PGM
(Host/Receiver) Stack

Cisco Public

PGM: Network Functions

- **PGM Network Element
  A PGM Network Element is an IP multicast router
  with PGM router assist.**

  - Can be deployed on specific routers in the path when needed.
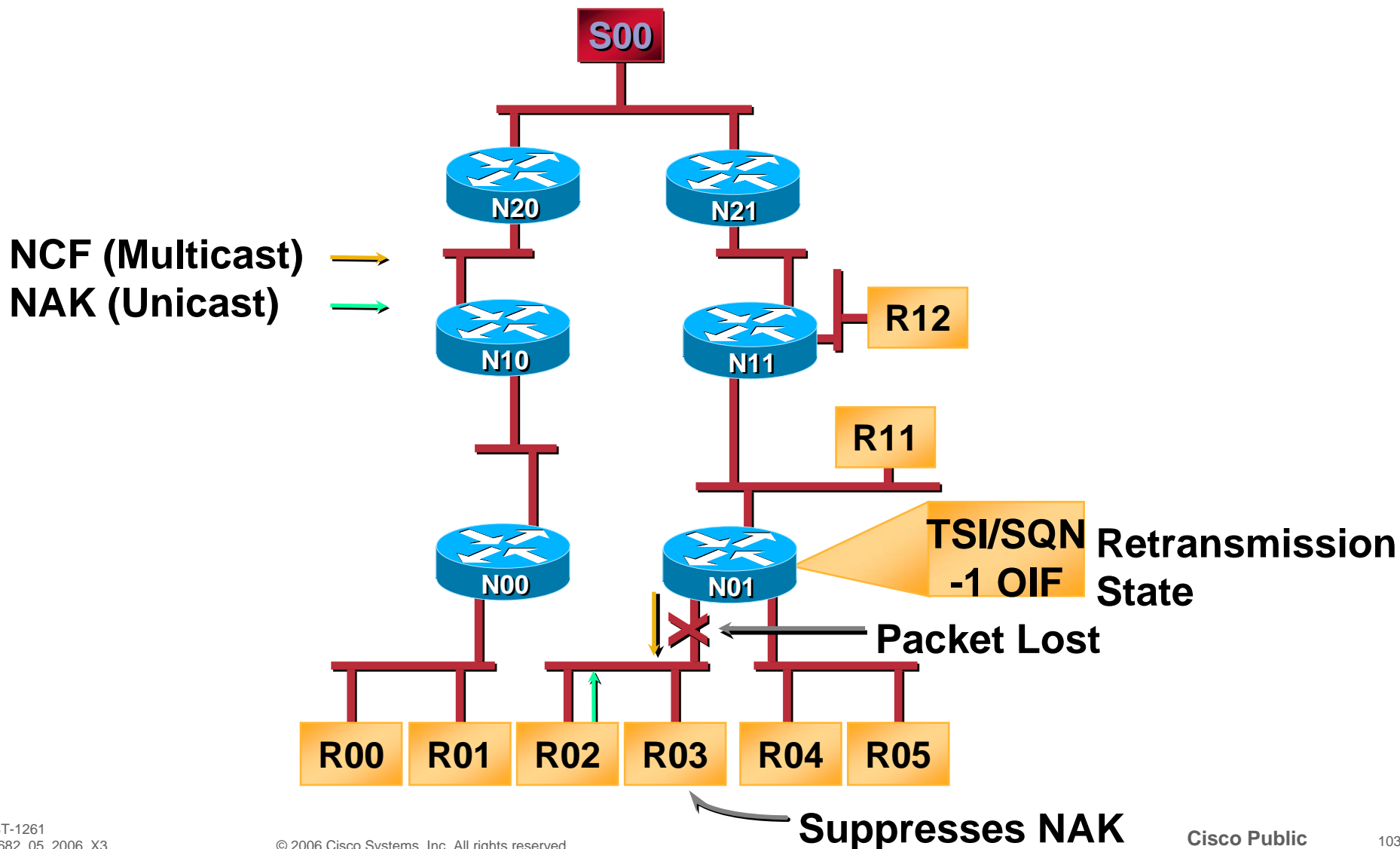
  - Provides NAK supression (inhibits NAK implosion)

  - Provides constrained retransmission
    (only those receivers that need retransmission will receive it)

- **DLR: Designated Local Repairer:**

  - Can be host or special network element

  - Buffers data from source and provides for retransmissions for all receiver hosts below it in the distribution tree.

PGM Router Assist
NAK elimination and constrained retransmission (1)
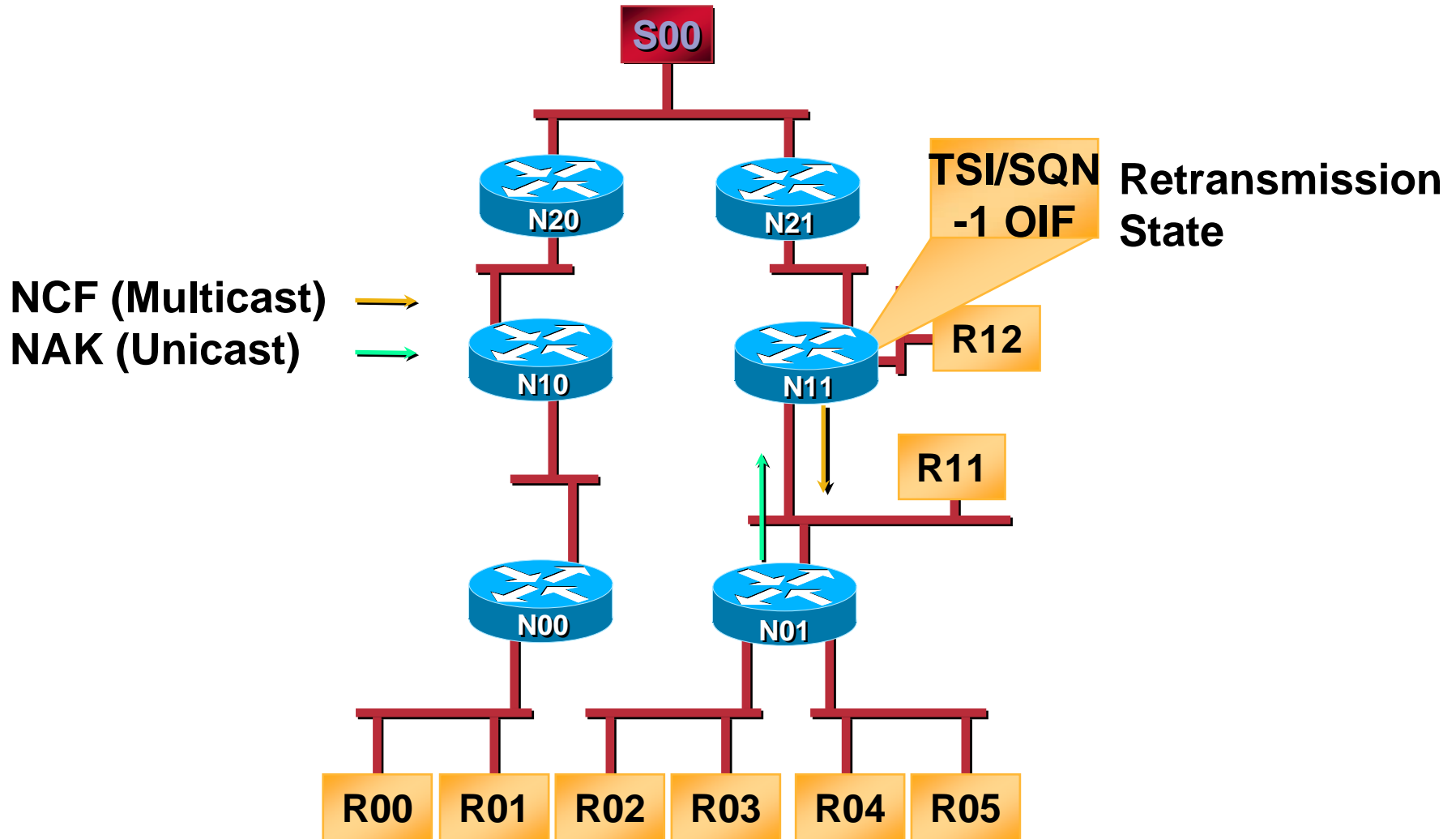
S00

N20   N21

NCF (Multicast) →
NAK (Unicast) →

N10   N11   R12

R11

N00   N01   TSI/SQN -1 OIF   Retransmission State

Packet Lost

R00 R01 R02 R03 R04 R05

Suppresses NAK

PGM Router Assist
NAK elimination and constrained retransmission (2)

S00

N20

N21

TSI/SQN -1 OIF **Retransmission State**

NCF (Multicast) →

NAK (Unicast) →

N10

N11

R12

R11

N00

N01

R00 **R01** **R02** **R03** **R04** **R05**

PGM Router Assist
NAK elimination and constrained retransmission (3)

S00

TSI/SQN -1 OIF

Retransmission State

N20

N21

NCF (Multicast)

NAK (Unicast)

N10

N11

R12

R11

N00

N01

R00 R01 R02 R03 R04 R05

Cisco Public
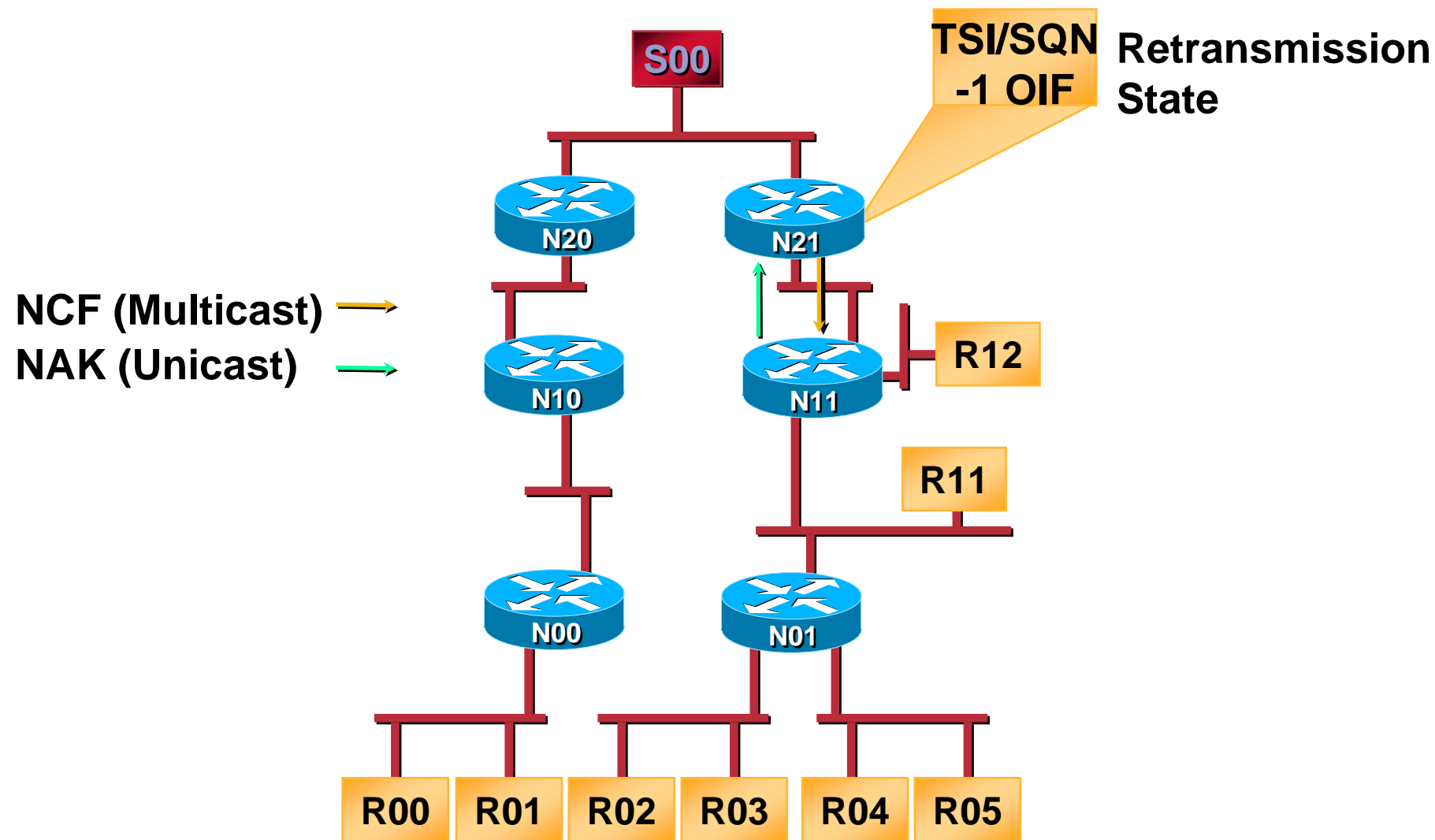
IPv4 versus IPv6 Multicast

IP Service	IPv4 Solution	IPv6 Solution
Address Range	32-bit, class D	128-bit (112-bit Group)
Routing	Protocol Independent All IGPs,and BGP4+	Protocol Independent All IGPs,and BGP4+ with v6 mcast SAFI
Forwarding	PIM-DM, PIM-SM, PIM-SSM, PIM-bidir	PIM-SM, PIM-SSM, PIM-bidir
Group Management	IGMPv1, v2, v3	MLDv1, v2
Domain Control	Boundary/Border	Scope Identifier
Inter-domain Solutions	MSDP across Independent PIM Domains	Single RP within Globally Shared Domains

IPv6 Multicast Addresses (RFC 3513)

128 bits

8	4	4		
FF	Flags	Scope	0	Interface-ID

1111 1111	Flags

F	F			P	T	scope

8 bits   8 bits

Flags = {
T or Lifetime, 0 if permanent, 1 if temporary
P proposed for unicast-based assignments
Others are undefined and must be zero
}

Scope = {
1 = interface-local
2 = link
4 = admin-local
5 = site
8 = organization
E = global
}
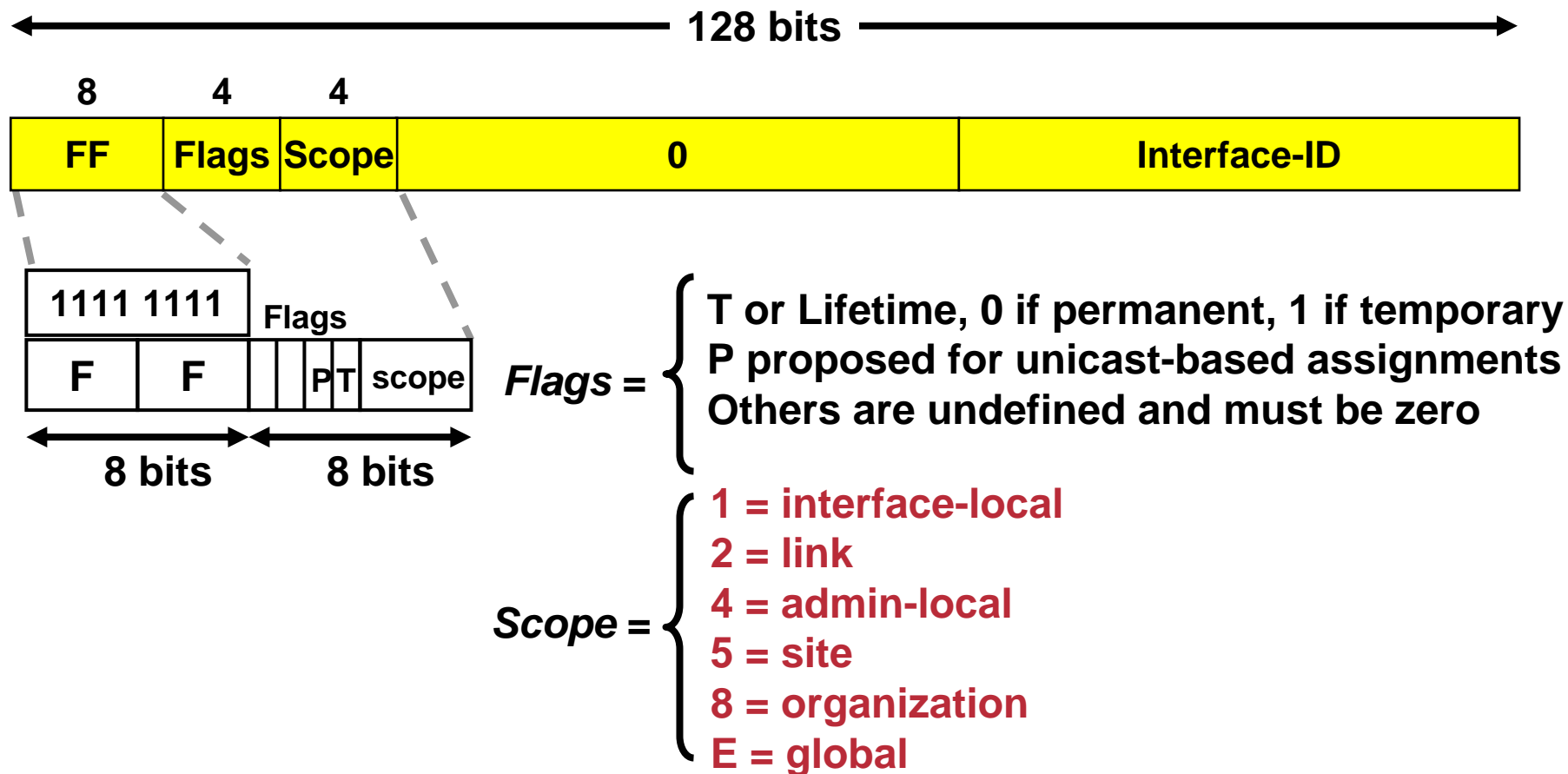
IPv6 Layer 2 Multicast Addressing Mapping

IPv6 Multicast Address

8	4	4	80	32
FF	Flags	Scope	High-Order	Low-Order

← 112 Bits →

80 Bits Lost

33-33-xx-xx-xx-xx

← 48 Bits →

Ethernet MAC Address
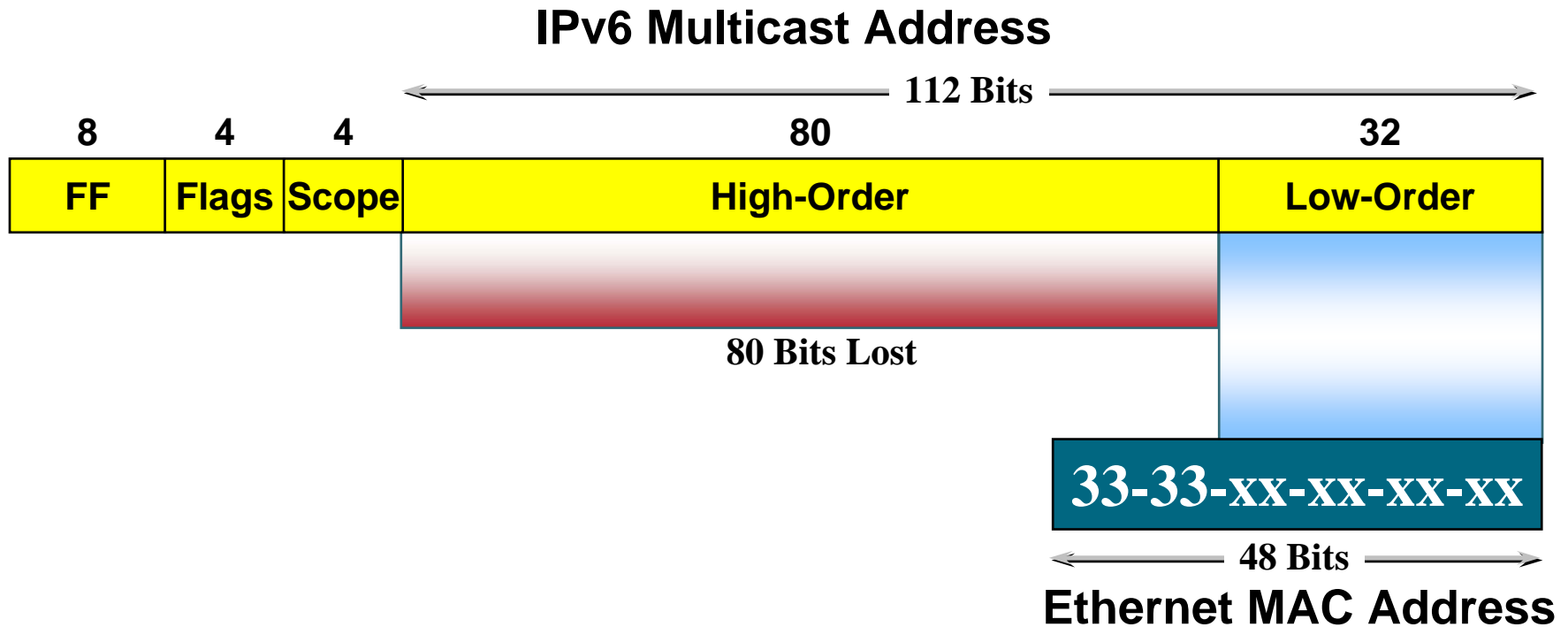
Unicast-based Multicast addresses

8	4	4	8	8	64	32
FF	Flags	Scope	Rsvd	Plen	Network-Prefix	Group-ID

- **RFC 3306 – Unicast-based Multicast Addresses**
  - **Similar to IPv4 GLOP Addressing**
  - **Solves IPv6 global address allocation problem.**
  - **Flags = 00PT**

    P = 1, T = 1   =>   Unicast-based Multicast address

- **Example:**
  - **Content Provider's Unicast Prefix**

    1234:5678:9abc::/64

  - **Multicast Address**

    FF36:0030:1234:5678:9abc::0001

Cisco Public

IP Routing for Multicast

- **RPF based on reachability to v6 source same as with v4 multicast**

- **RPF still protocol independent:**
  - **Static routes, mroutes**
  - **Unicast RIB: BGP, ISIS, OSPF, EIGRP, RIP, etc**
  - **Multi-protocol BGP (mBGP)**
  - **- support for v6 mcast sub-address family**
  - **- provide translate function for non-supporting peers**

IPv6 Multicast Forwarding

- **PIM-Sparse Mode (PIM-SM)**
  - **draft-ietf-pim-sm-v2-new-11.txt,**
- **PIM-Source Specific Mode (PIM-SSM)**
  - **RFC3569 SSM overview (v6 SSM needs MLDv2)**
  - **unicast prefix based multicast addresses ff30::/12**
  - **SSM range is ff3X::/32**
    - **Current allocation is from ff3X::/96**
- **PIM-bidirectional Mode (PIM-bidir)**
  - **draft-ietf-pim-bidir-07.txt**

RP mapping mechanisms for IPv6 PIM-SM

- **Static RP assignment**

- **BSR**

- **Auto-RP – no current plans**

- **Embedded RP**

Cisco Public

Embedded RP Addressing – RFC3956

8	4	4	4	4	8	64	32
FF	Flags	Scope	Rsvd	RPadr	Plen	Network-Prefix	Group-ID

- **Proposed new multicast address type**
  - Uses Unicast-Based Multicast addresses (RFC 3306)
- **RP Address is embedded in multicast address.**
- **Flag bits = 0RPT**

  R = 1, P = 1, T = 1   =>   Embedded RP Address
- **Network-Prefix::RPadr = RP address**
- **For each Unicast prefix you own, you now also own:**
  - 16 RPs for each of the 16 Multicast Scopes (256 total) with 2^32 multicast groups assigned to each RP (2^40 total)

Cisco Public
114

Embedded RP Addressing – Example

Multicast Address with Embedded RP address

8	4	4	4	4	8	64	32
FF	Flags	Scope	Rsvd	RPadr	Plen	Network-Prefix	Group-ID

FF76:0130:1234:5678:9abc::4321

1234:5678:9abc::1
Resulting RP address
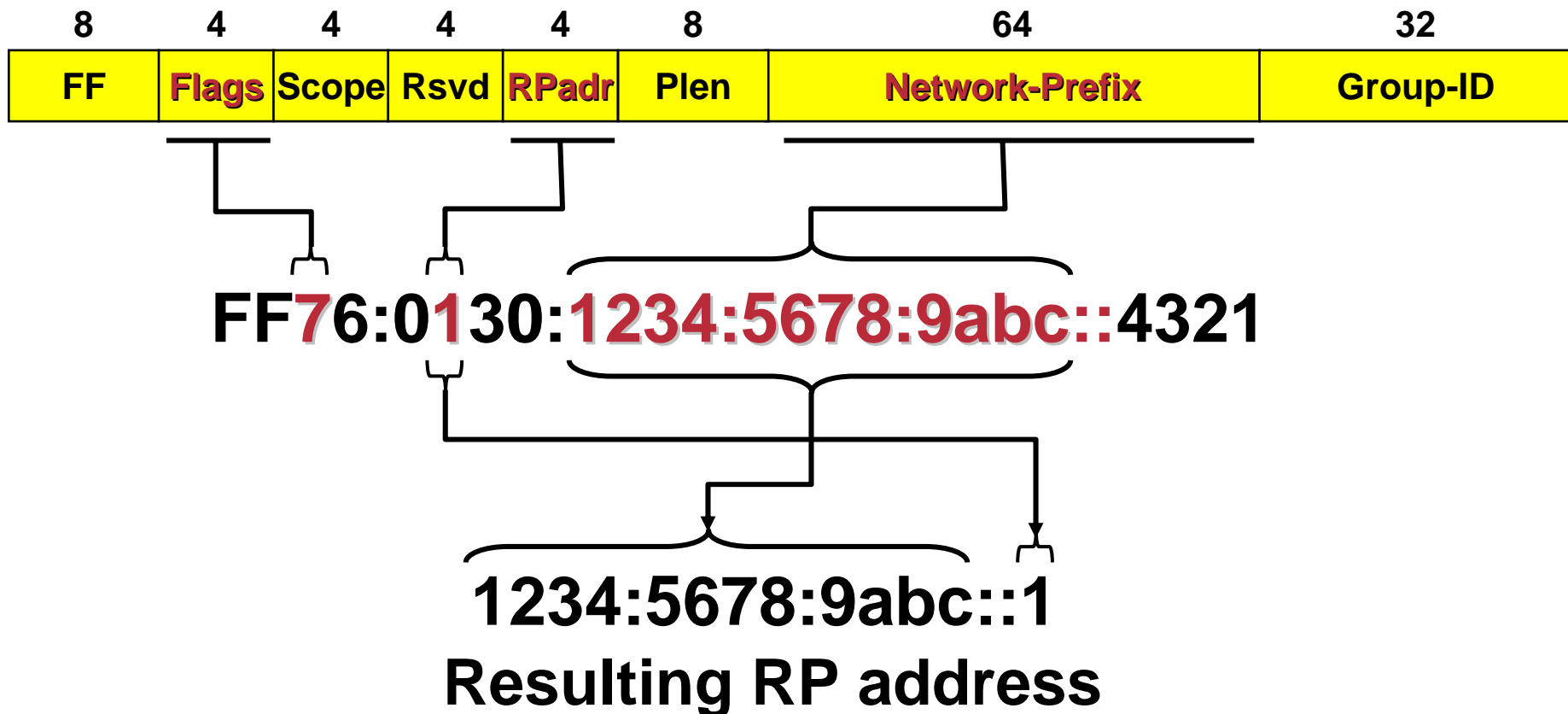
Multicast Listener Discover – MLD

- **MLD is equivalent to IGMP in IPv4**

- **MLD messages are transported over ICMPv6**

- **Version number confusion:**
  - **MLDv1 corresponds to IGMPv2**
    - **RFC 2710**
  - **MLDv2 corresponds to IGMPv3, needed for SSM**
    - **RFC 3810**

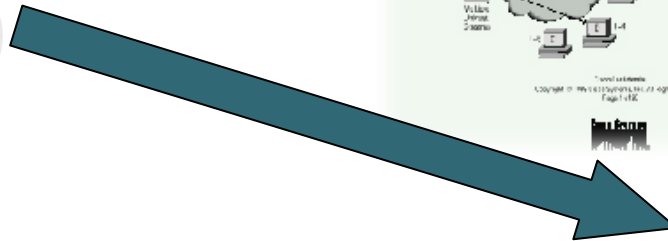- **MLD snooping**
  - **draft-ietf-magma-snoop-12.txt**

Now you know…

- **Why Multicast?**

- **Multicast Fundamentals**

- **PIM Protocols**

- **RP choices**

- **Multicast at Layer 2**

- **Interdomain IP Multicast**

- **Latest Additions IPv6, PGM and MVPN**

Cisco Public

More Information

- **White Papers**

- **Web and Mailers**

- **Cisco Press**

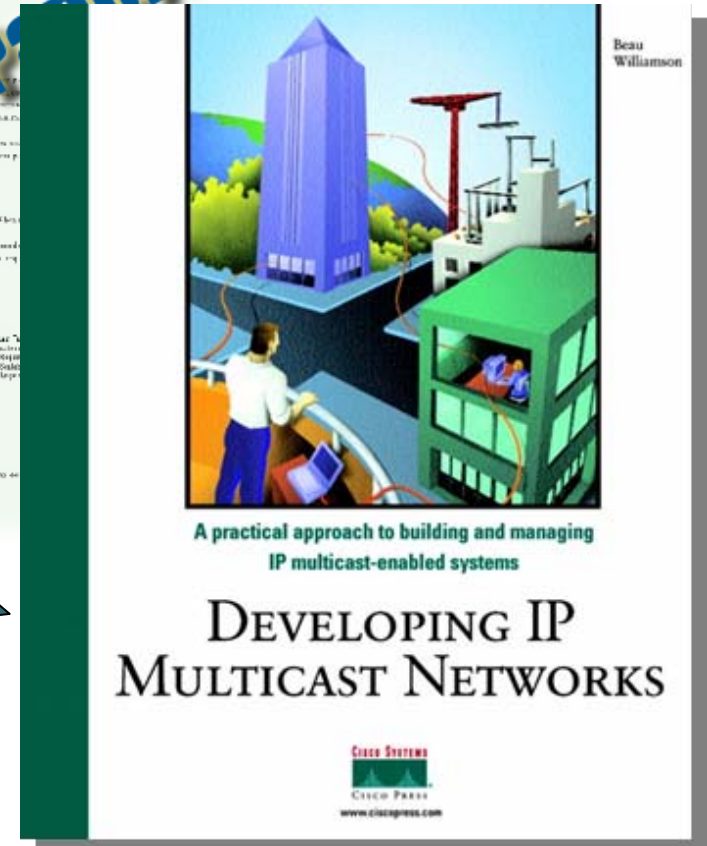**CCO Multicast page:**
   http://www.cisco.com/go/ipmulticast
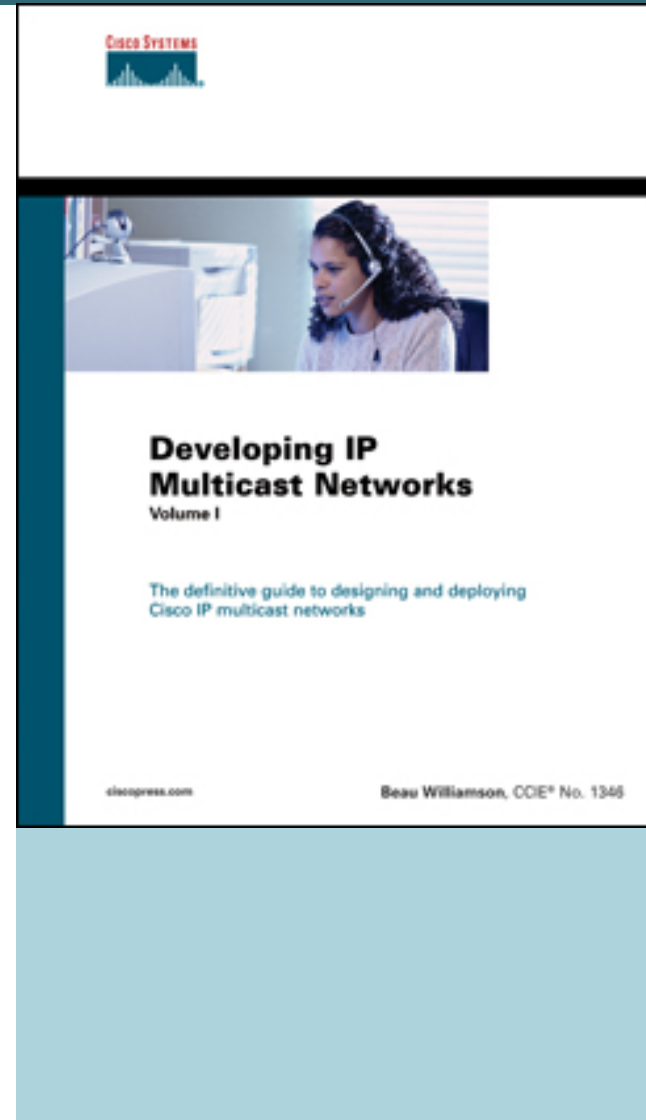**Customer Support Mailing List:**
   tac@cisco.com

RTFB = "Read the Fine Book"

Recommended Reading

- **Continue your Cisco Networkers learning experience with further reading from Cisco Press**

- **Check the Recommended Reading flyer for suggested books**

Available on site at the Cisco Company Store

CISCO SYSTEMS

Developing IP Multicast Networks
Volume I

The definitive guide to designing and deploying
Cisco IP multicast networks

ciscopress.com

Beau Williamson, CCIE® No. 1346

Cisco Public

Complete Your Online Session Evaluation

- **Win fabulous prizes; Give us your feedback**

- **Receive ten Passport Points for each session evaluation you complete**

- **Go to the Internet stations located throughout the Convention Center to complete your session evaluation**

- **Drawings will be held in the World of Solutions**
  - **Tuesday, June 20 at 12:15 p.m.**
  - **Wednesday, June 21 at 12:15 p.m.**
  - **Thursday, June 22 at 12:15 p.m. and 2:00 p.m.**

Cisco Public

Cisco Public 121