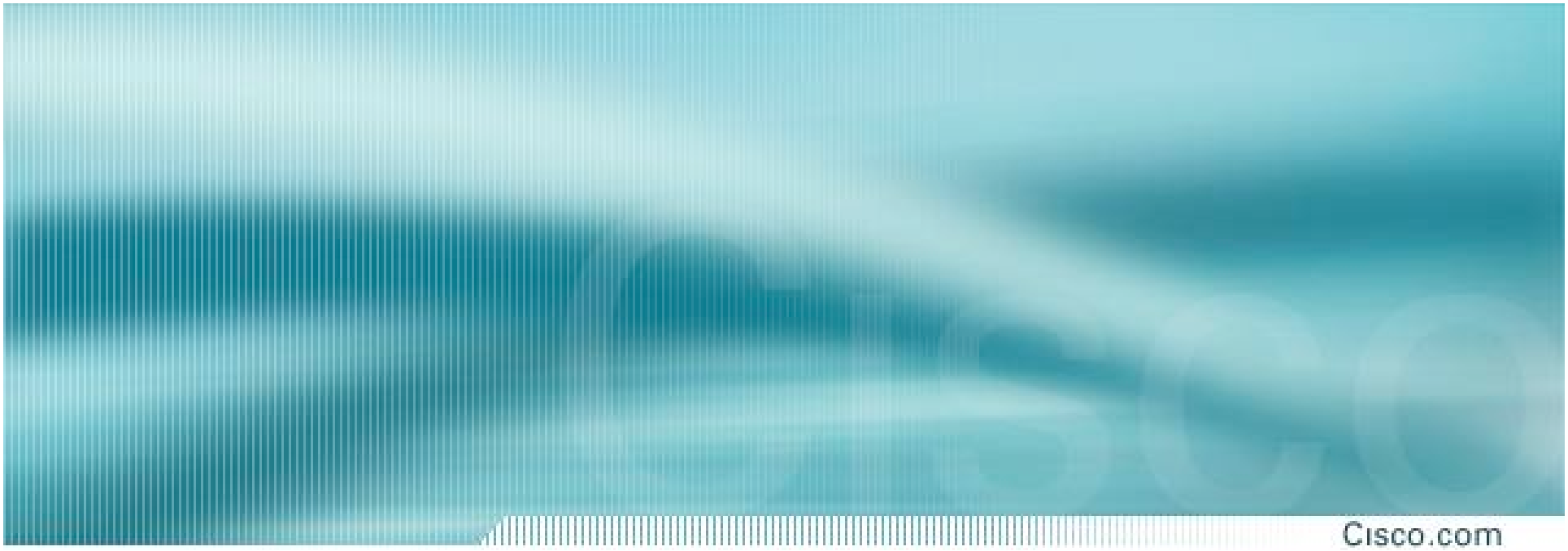


IP Multicast Workshop
Apricot Bali 2007

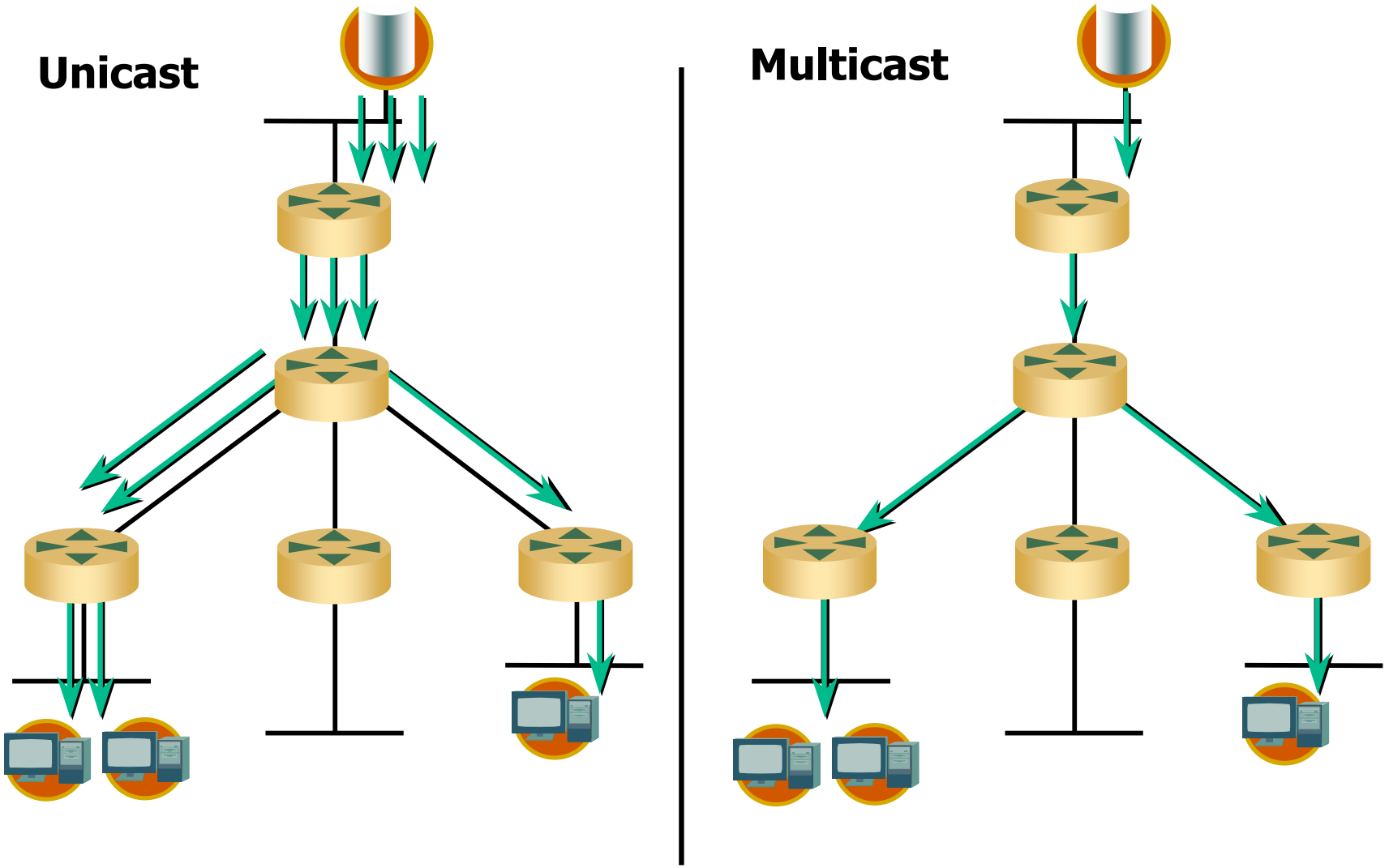
Agenda

- Introduction
- Multicast addressing
- Group Membership Protocol
- PIM-SM / SSM
- MSDP
- MBGP
- Summary

Agenda

- **Introduction**
- **Multicast addressing**
- **Group Membership Protocol**
- **PIM-SM / SSM**
- **MSDP**
- **MBGP**
- **Summary**

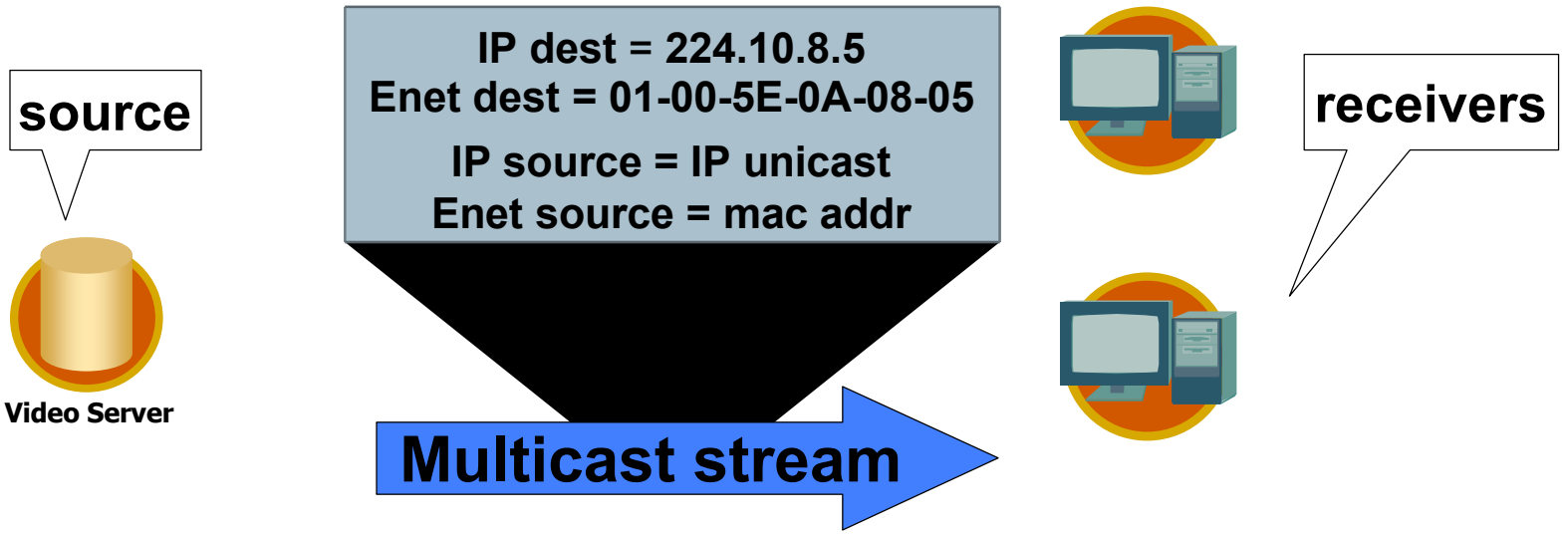
What is Multicasting?



Multicast Uses

- Any Applications with multiple receivers
  - 1-to-many or many-to-many
- Live Video distribution
- Collaborative groupware
- Periodic Data Delivery - "Push" technology
  - stock quotes, sports scores, magazines, newspapers
  - advertisements
- Server/Web-site replication
- Reducing Network/Resource Overhead
  - more efficient to establish multicast tree rather then multiple point-to-point links
- Resource Discovery
- Distributed Interactive Simulation (DIS)
  - wargames
  - virtual reality

Glossary of Terms: the basics



- Source = source of multicast stream
- Multicast stream = IP packet with multicast address as IP destination address. a.k.a. multicast group.
  - s,g (unicast source, group) reference
  - UDP packets (TTL > 1 for routed nets)
- Receiver = receiver (s) of multicast stream

IP Multicast building blocks

- The SENDERS send
  - Multicast Addressing - rfc1700
  - class D (224.0.0.0 - 239.255.255.255)
- The RECEIVERS inform the routers what they want to receive
  - Internet Group Management Protocol (IGMP) - rfc2236 -> version 2; rfc3376 -> version 3
- The ROUTERS make sure the packets make it to the correct subnets.
  - Multicast Routing Protocols (PIM-SM/SSM)
  - RPF (reverse path forwarding)

Multicast 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.
- Multicast Routing uses “Reverse Path Forwarding”: RPF

Multicast Forwarding

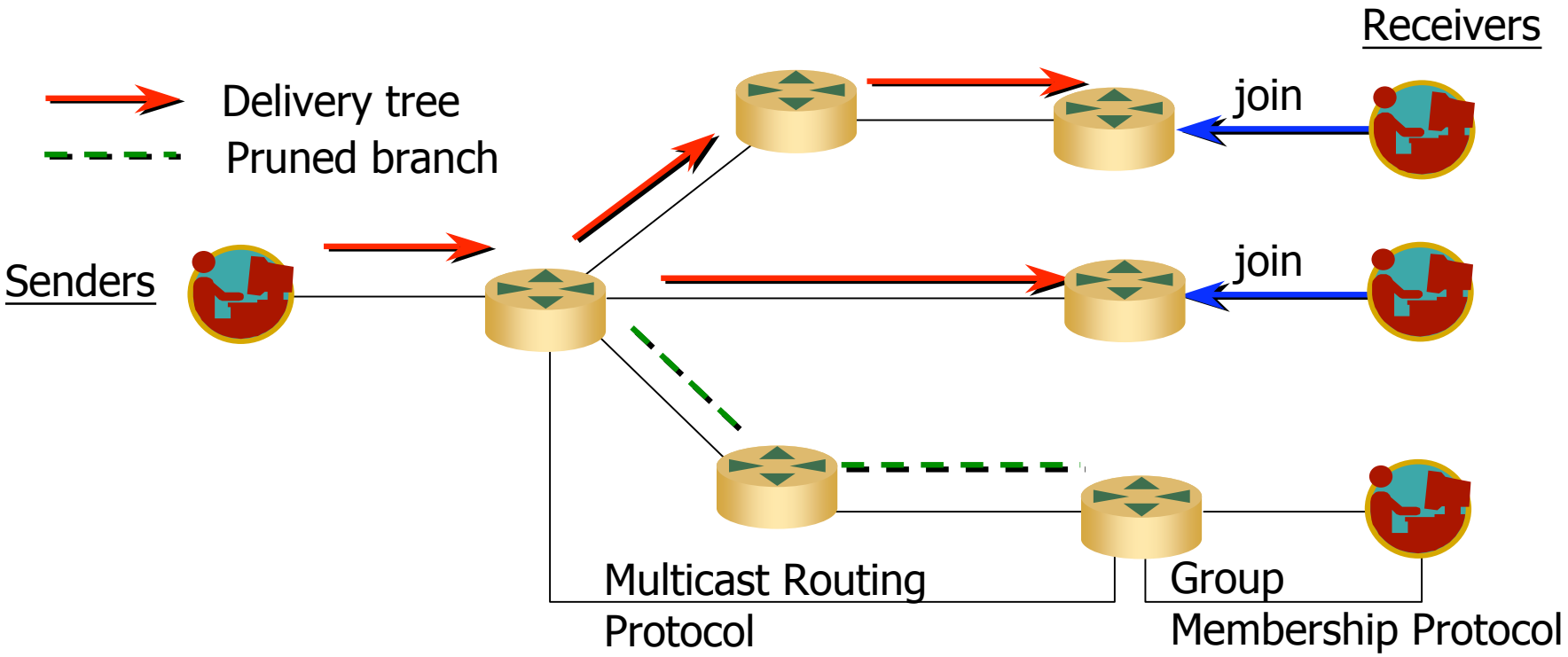
Reverse Path Forwarding (RPF)

- The source IP address of incoming multicast packets are checked against the unicast routing table.
- If the datagram arrived on the interface that would be used to send a unicast packet to the source; then the RPF check succeeds.
- RPF checks ensures packets won’t loop

IP Multicast Components

- **Group Membership Protocol - enables hosts to dynamically join/leave multicast groups. Membership info is communicated to nearest router**
  - **IGMPv1/v2**
  - **IGMPv3: specify the source also**
  - **Between receivers and "Last-hop" router**
- **Multicast Routing Protocol - enables routers to build a delivery tree between the sender(s) and receivers of a multicast group**
  - **PIM is the default protocol**
  - **Several "flavors" of PIM**
    - **Sparse**
    - **Dense: considered obsolete by many**
    - **Bidir**
    - **Source-Specific (SSM)**

IP Multicast Components

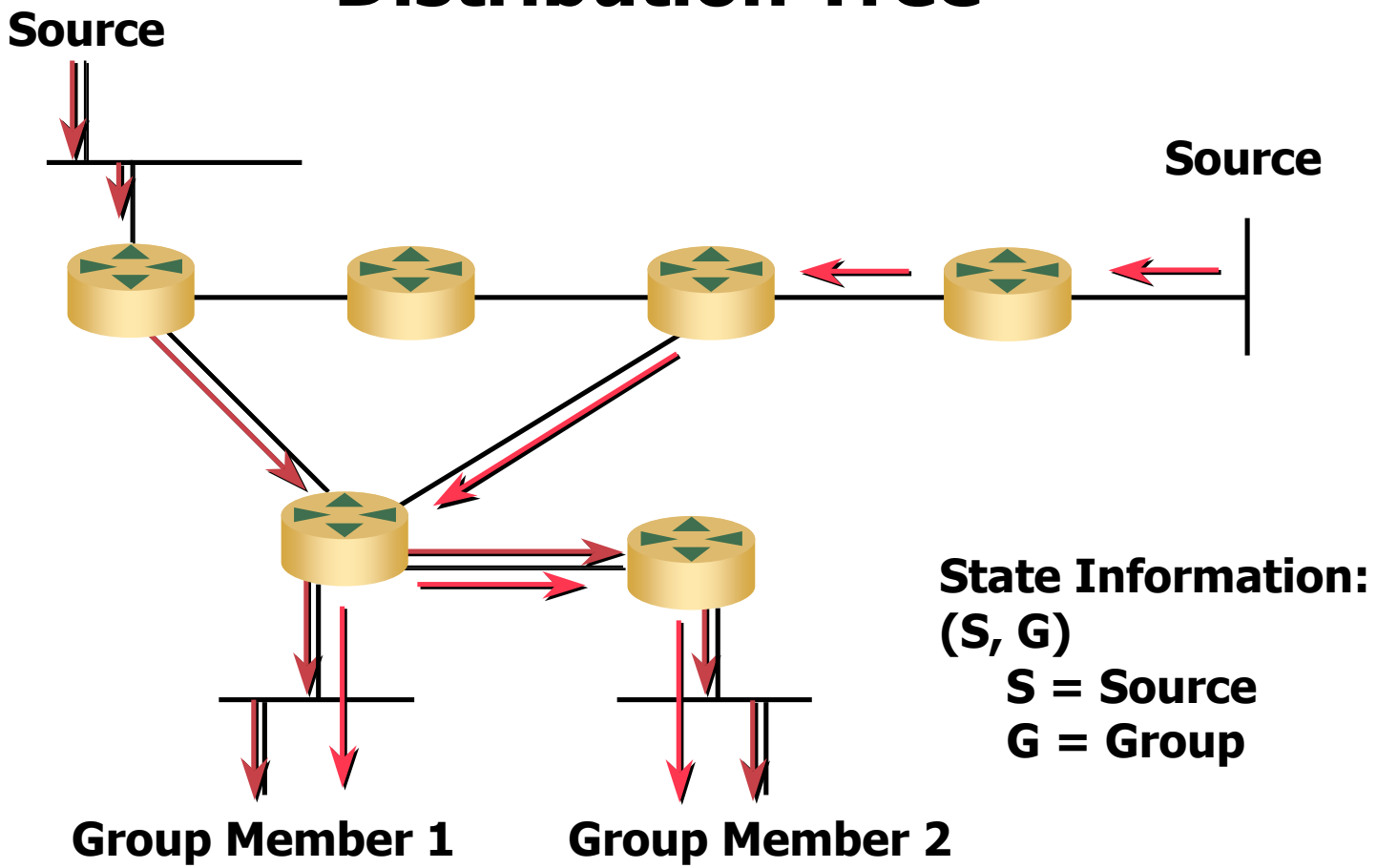


Multicast Distribution Trees

- **Shared path**
  - **All sources use the same path: (*,G)**
  - **The root of the tree is common for all sources.**
  - **Called RPT: Rendezvous Point Tree because of PIM**
  - **How the source packets arrive at the root of the shared tree is defined by the specific protocol.**
- **Shortest path**
  - **Each source has a unique path: (S,G)**
  - **Called SPT: Shortest Path tree**
  - **The root of the tree is the source itself**
- **Bi Directional**
  - **Allows for upstream forwarding on the shared path**

Multicast Distribution Trees

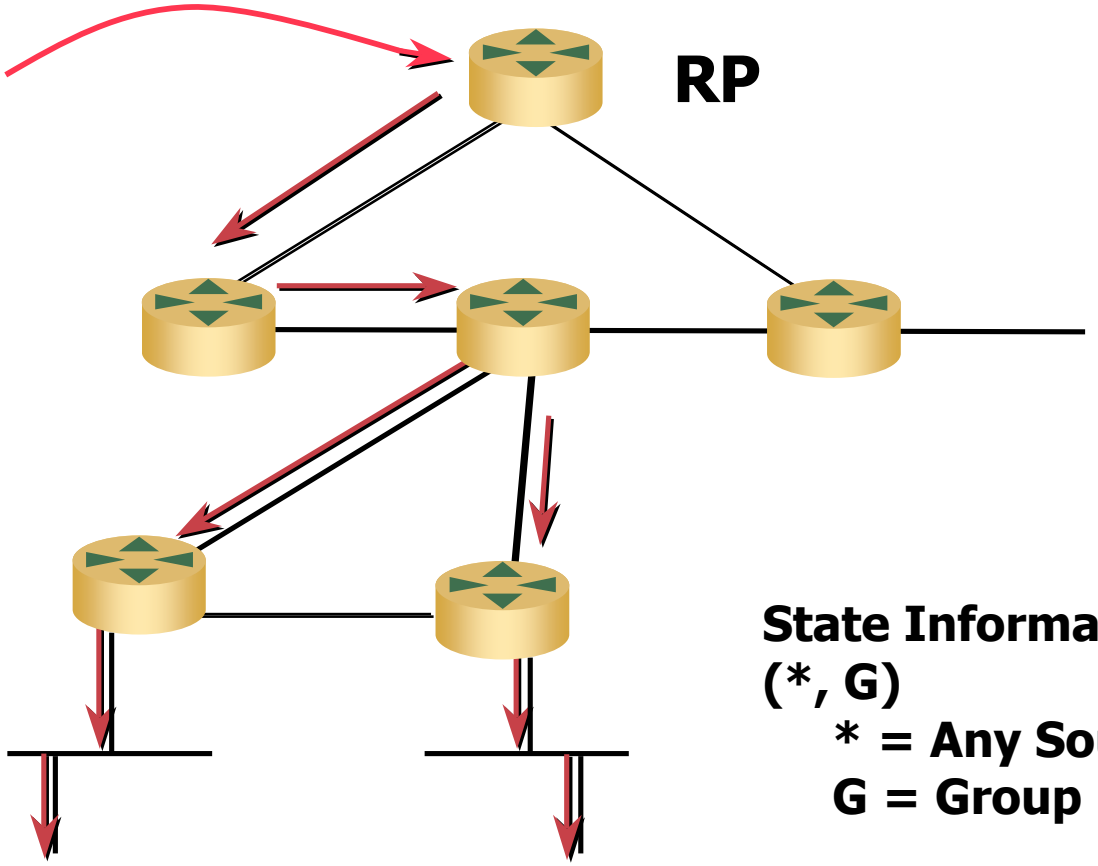
Shortest Path or Source Based Distribution Tree



Multicast Distribution Trees

Shared or RP Distribution Tree

Sources deliver packets to the root of the tree as defined by the protocol used



State Information:
(*, G)
* = Any Source
G = Group

Multicast Distribution Trees

- **Source or Shortest Path trees**
  - **More resource intensive; requires more state→ n(S x G)**
  - **You get optimal paths from source to all receivers, minimizes delay**
  - **Best for one-to-many distribution**
- **Shared or trees**
  - **Uses fewer resources; less memory →n(G)**
  - **You may get sub optimal paths from source to all receivers, depending on topology**
  - **The RP (core) itself and its location *may* affect performance**

Multicast Modes (theories)

- **Dense**
  - **Assumes a receiver on almost every LAN on the network**
  - **Defaults to broadcast**
- **Sparse**
  - **Assumes receivers are widely separated**
  - **Practice has shows that sparse-mode multicast distribution accomplishes everything that dense-mode does.**

PIM varieties

- Dense
  - Considered obsolete
- Sparse
  - Widely deployed
  - Complicated
  - Would like to see obsolete
- Source Specific Multicast (SSM)
  - No problems building the tree to anywhere in the Multicast-enabled Internet
- Bidir
  - Saves state

Agenda

- Introduction
- **Multicast addressing**
- Group Membership Protocol
- PIM-SM / SSM
- MSDP
- MBGP
- Summary

Multicast Addressing

- IP Multicast Group Addresses
  - 224.0.0.0–239.255.255.255
  - Class “D” Address Space
    - High order bits of 1st Octet = “1110”
- Use of the address space is being more tightly defined.
- RFC 3171:IANA Guidelines for IPv4 Multicast Address Assignments
  - Link-local
  - Internetwork control block
  - SSM
  - GLOP

Multicast Addressing

- http://www.iana.org/assignments/multicast-addresses
- Examples of Reserved & Link-local Addresses
  - 224.0.0.0 - 224.0.0.255 reserved & not forwarded
  - 239.0.0.0 - 239.255.255.255 Administrative Scoping
  - 224.0.0.1 - All local hosts
  - 224.0.0.2 - All local routers
  - 224.0.0.4 - DVMRP
  - 224.0.0.5 - OSPF
  - 224.0.0.6 - Designated Router OSPF
  - 224.0.0.9 - RIP2
  - 224.0.0.13 - PIM
  - 224.0.0.15 - CBT
  - 224.0.0.18 - VRRP

Internet Network Control Block

- 224.0.1.x
- Addresses in the Internetwork Control block are used for protocol control that must be forwarded through the Internet.
  - Examples include 224.0.1.1 (NTP [RFC2030]) and 224.0.1.68 (mdhcpdiscover [RFC2730]).

Dynamic Address Allocation

- SDR The defacto
  - 224.2.0.0 – 224.2.255.255 (224.2/16) SDP/SAP Block
- Still used, but not required
- Will not scale well
  - Limited address space
  - Single directory application for ALL content?!?!?
- Web links should prevail

Multicast Addressing

- Administratively Scoped Addresses – rfc2365
  - 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: 239.253.0.0/16
    - Organization-local scope: 239.192.0.0/14

Multicast Addressing

- GLOP addresses
  - Provides globally available private Class D space
  - 233.x.x/24 per AS number
  - RFC3180 (obsoletes 2770)

How?

- AS number = 16 bits

  Insert the 16 ASN into the middle two octets of 233/8

Online Glop Calculator:

http://www.shepfarm.com/multicast/glop.html

Multicast Addressing

- SSM - RFC 4607: Source-Specific Multicast for IP
  - 232/8 – IANA assigned
  - One-to-many ONLY (no shared trees)
  - Guaranties ONE source on any delivery tree
    - Content security (no 'Captain Midnight')
  - Reduced protocol dependence – more later..
  - Solves address allocation issues for interdomain one-to many
    - ~tree address is 64 bits – S,G
  - Host must learn of source address out-of-band (web page)
  - Requires host-to-router source AND group request
    - IGMPv3 include-source list
  - Hard-coded behavior in 232/8 in most router implementations
    - RFC 4608: Source-Specific Protocol Independent Multicast in 232/8
    - Configurable to expand range

Multicast Addresses - Layer 2

- **RFC 1700 - ethernet**

```
                                    224.        10.         8.          5        <-- Class D IP address
0000 0001   0000 0000   0101 1110   0xxx xxxx   xxxx xxxx   xxxx xxxx     <-- IANA's reserved block 01-00-5E
        |                                |
  Multicast Bit                  0 = Internet Multicast
                                 1 = IANA reserved
0000 0001   0000 0000   0101 1110   0000 1010   0000 1000   0000 0101     <-- MAC address 01-00-5E-0A-08-05
```

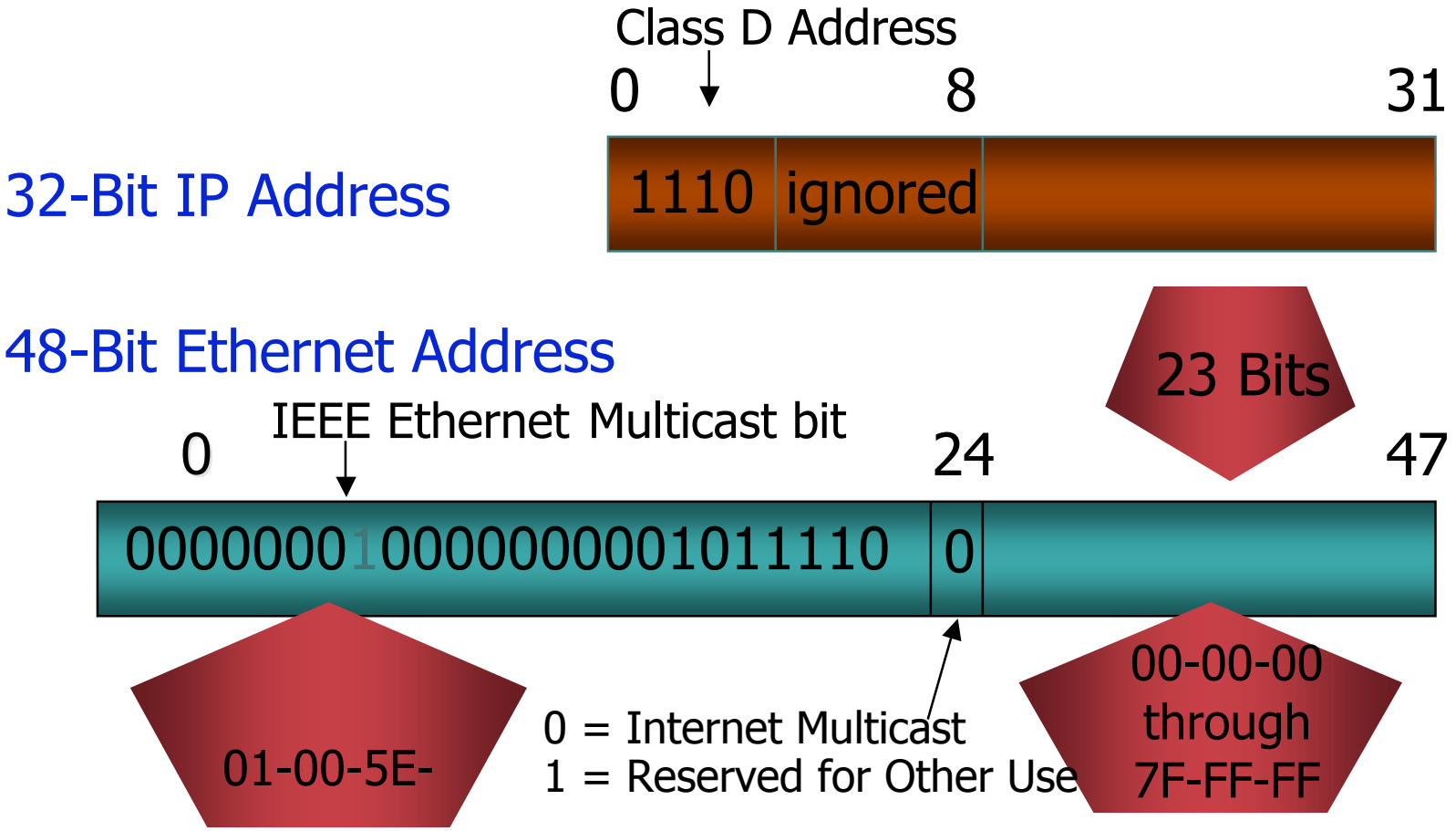
224.10.8.5 multicast stream maps to 01-00-5E-0A-08-05 ethernet layer 2 address.

- **rfc 1469 TR**
- **rfc 1390 FDDI**
- **rfc 2226 & 2022 - ATM**
- **rfc 1209 SMDS (broadcast)**

Ethernet Multicast Addressing

IANA Owns 01-00-5E Vendor Address Block

Half of It Assigned for IP Multicast



Agenda

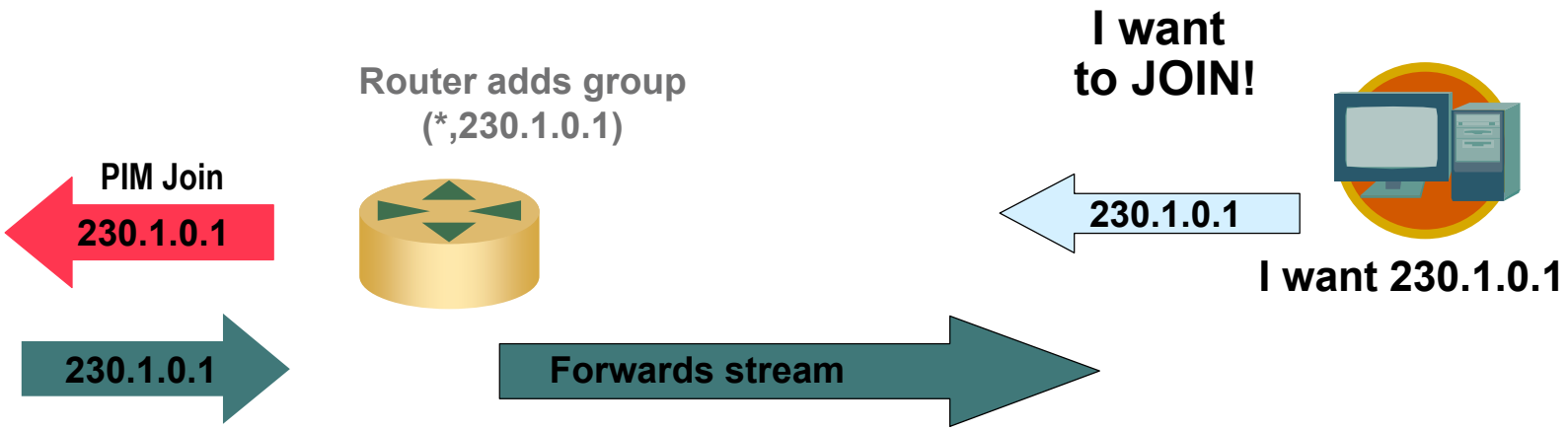
Cisco.com

- Introduction
- Multicast addressing
- **Group Membership Protocol**
- PIM-SM / SSM
- MSDP
- MBGP
- Summary

Internet Group Membership(management) Protocol

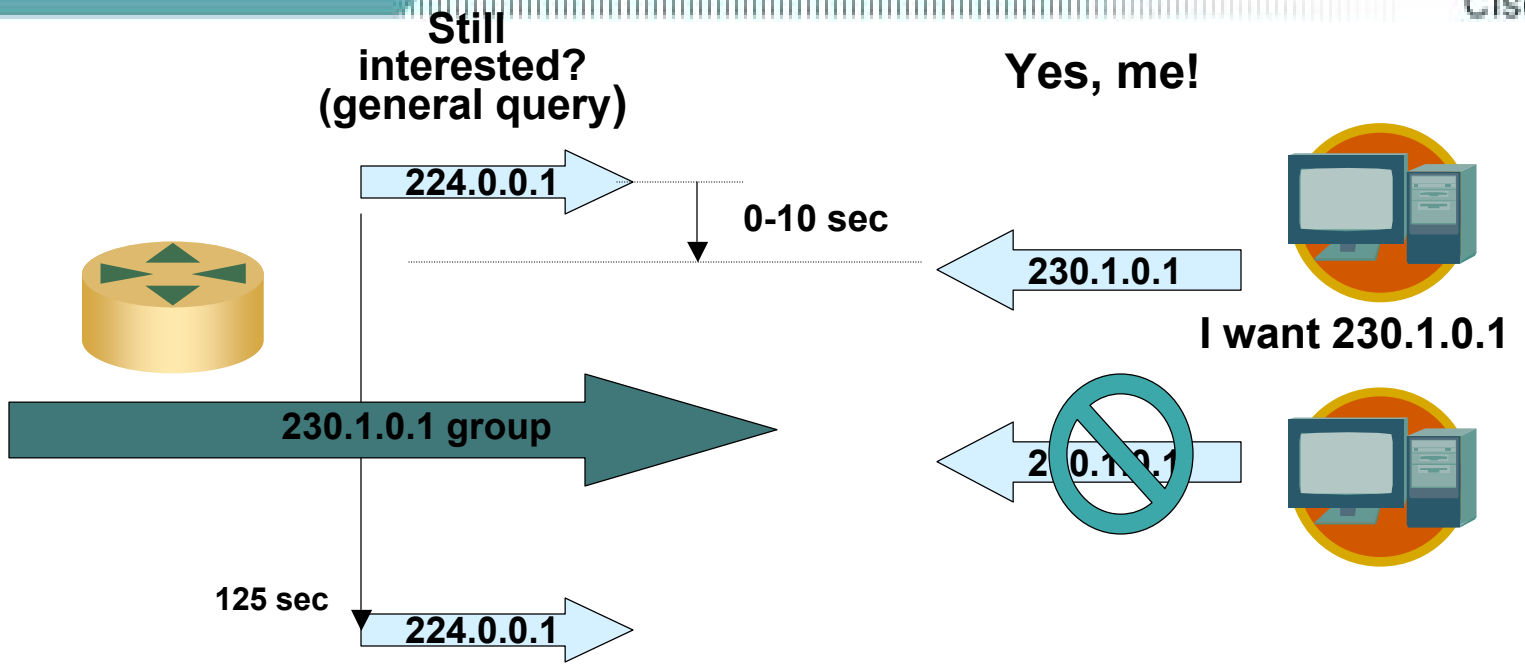
- Routers solicit group membership from directly connected hosts
- RFC 2236 specifies version 2 of IGMP
  - Widely supported
- RFC 3376 specifies version 3 IGMP
  - provides source include-list capabilities (SSM!)
  - Host support:
    Linux 2.6(16.1), Window XP
    FreeBSD patch
    NOT IN MacOSX!! Send Bug reports to Apple.

IGMP Protocol Flow - Join a Group



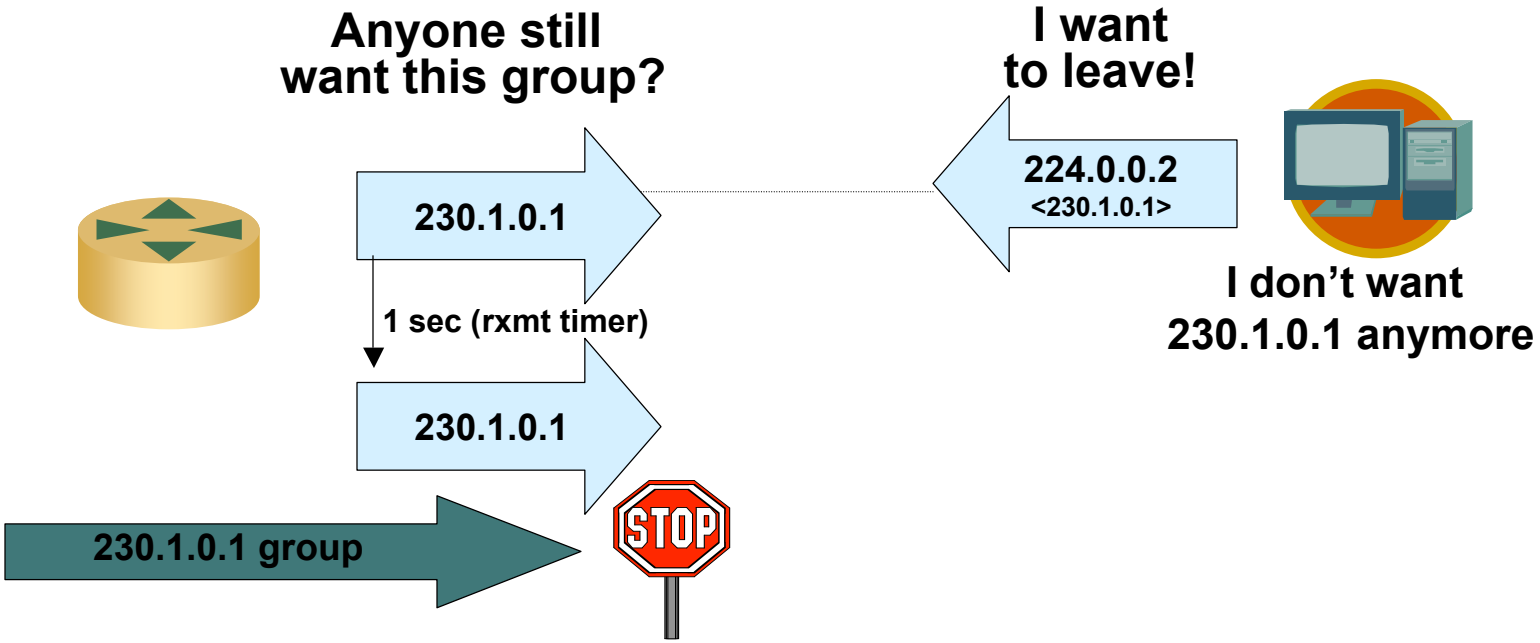
- Hosts can send unsolicited *join* membership messages – called reports in the RFC (usually more than 1)
- Router triggers group membership request to PIM.

IGMP Protocol Flow - Querier



- Hosts respond to *query* to indicate (new or continued) interest in group(s)
- After 260 sec with no response, router times out group

IGMP Protocol Flow - Leave a Group

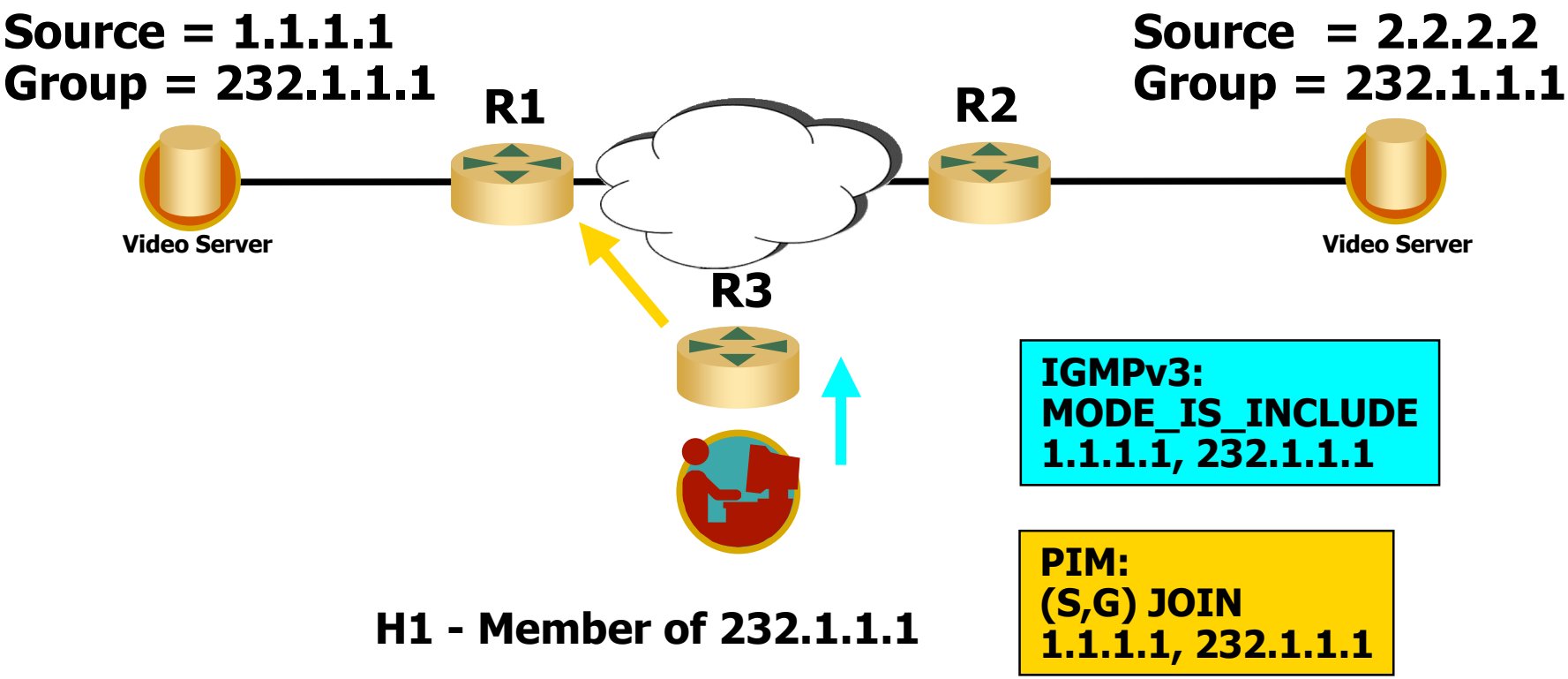


- Hosts that support IGMP v2 send *leave* messages to the "all-routers" group.
- Router follows up with 2 *group-specific queries*
- IGMP v1 hosts leave by not responding to *queries* (260 sec timeout)

IGMPv3

RFC3376

Hosts to listen only to a specified subset of sources



IGMP Details

- **Router:**
  - **Sends Membership Query messages to All Hosts (224.0.0.1)**
    query-interval = 125 secs default
  - **Router with lowest IP address is Querier**
  - **PIM DR listens for reports and adds group to membership list for that interface (Implementation specific)**
  - **Timeout (Group member interval) default:**
    (robust-count x query-interval) + (1 x query-response-interval) = **260 sec**
  - **Robust-count - provides fine-tuning to allow for expected packet loss on a subnet. Default = 2**

IGMPv2 Details

- **Host:**
  - **Sends Membership Report,**
    waits 0-10 sec (default) after receiving a query.
    Hosts listen to other host reports
    Only 1 host responds
  - **Membership reports are sent to the group address (e.g. 224.10.8.5)**
  - **Leave messages to All Routers (224.0.0.2)**

IGMP Enhancements

- **IGMP Version 2**
  - **Multicast router with lowest IP address is elected querier IGMPv1 was mcast protocol specific and potentially conflicted.**
  - **Group-Specific Query message is defined. Enables router to transmit query to specific multicast address rather than to the "all-hosts" address of 224.0.0.1**
  - **Leave Group message is defined. Last host in group wishes to leave, it sends Leave Group message to the "all-routers" address of 224.0.0.2. Router then transmits Group-Specific query and if no reports come in, then the router removes that group from the list of group memberships for that interface**

IGMPv3 Enhancements

- Include: list the specific sources in a group to receive
  - Include {NULL} is equivalent to an IGMPv2 group leave
  - Include (S,G): I want this source and group
- Exclude: list the specific sources in a group to NOT receive
  - Exclude {NULL) is equivalent to an IGMPv2 group join
  - Exclude (S,G): I want everything -except- this source and group.
- Reports are sent to 224.0.0.22, NOT to the group address
  - Works better with L2 switches. No snooping required.
- No report suppression: explicit tracking

Agenda

- Introduction
- Multicast addressing
- Group Membership Protocol
- **PIM-SM / SSM**
- MSDP
- MBGP
- Summary

PIM

- Protocol Independent Multicast
  - http://ietf.org/html.charters/pim-charter.html
  - RFC 3973: PIM-Dense Mode
  - RFC 4601: PIM-Sparse Mode
  - RFC 4602: Proposed Standard Requirements Analysis
  - RFC 4610: PIM Anycast-RP
  - Draft-ietf-pim-sm-bsr-09: bootstrap router, RP distribution
  - draft-ietf-pim-bidir-08.txt: bi-directional PIM
- Depends on Unicast Routing table for forwarding decision

PIM-DM

- Data is flooded to the boundaries of the PIM domain and pruned back
- Current RFC has a “state-refresh” mechanism that avoids periodic flood-and-prune requirements of original design
- Implementations are either non-existent or based on old RFC
- Considered obsolete by most, but appears very useful in some networks. (Talk to the RFC authors)
- Can not be used interdomain without a lot of trouble

PIM-SM

- Receiver initiated
- Explicit join: data sent only to locations where receivers exist
- Rendezvous Point (RP) is used for source discovery
- All routers in a PIM domain must have RP mapping
- Last-hop router initiates join to the SPT.
- May remain on the shared-tree
  - Useful for some topologies
  - Being replaced by bi-dir
- Source-tree state is refreshed when data is forwarded and with Join/Prune control messages
- Can be used interdomain with MSDP

PIM-SSM (source specific multicast)

- Uses only the SPT
- Source discovery is not part of the protocol
- Requires IGMPv3 on the receivers
  - Some work-arounds
- Very simple to deploy and understand
- Works interdomain without any other special configurations
- Can be used with Automatic Multicast Tunnels (AMT)

PIM-Bidir (bi-directional)

- Uses only the shared-tree
- Reduces state on the routers by at least an order of magnitude
  - Amazon.com went from thousands of (S,G) entries to less than 200 (*,G) routes
- Currently only useful within a contiguous domain
- The root of the tree is virtual (does not require a physical RP) so can be moved easily and rapidly within an enterprise
- With a well-implemented BSR, backup and resource allocation is very easy and reliable.

Which PIM do you want to use?

- SSM
  - Easiest to understand
  - Easiest to deploy
  - Most reliable
  - Interdomain
- Bidir
  - Within a domain with many, many sources
- Sparse-mode
  - If receivers can not do IGMPv3
- Never use dense

When to use which PIM

- SSM
  - One-to-many applications with well-known sources
  - a good source discovery protocol (TBD)
- Bidir
  - Many-to-many applications with transient sources
  - When supporting 100,000+ sources
  - To keep routing simple
- Sparse-mode
  - One-to-many when receivers are not IGMPv3 capable
  - Few-to-few and no source discovery protocol works
  - Try to avoid

PIM Sparse-mode operation

- We cover sparse-mode because it has all the basic concepts required to understand the other modes
- It is the most widely deployed
- Even though we may want it to go away, it won't be any time soon

PIM-SM Operation
Designated Router (DR)

- **Neighboring PIM-SM routers multicast periodic "Hello" messages to each other - default 30 secs.**
  - **Hello-interval tunable for faster convergence**
- **On receipt of a Hello message**
  - **a router stores the IP address and priority for that neighbor**
- **Router with highest Priority is selected as the DR**
  - **If the priorities are the same, highest IP address**
- **DR is only important on first-hop and last hop**
  - **On last-hop sends "Join/Prune" messages toward RP/source**
  - **On first-hop: send "Register" messages to the RP**

PIM Sparse-Mode:RP

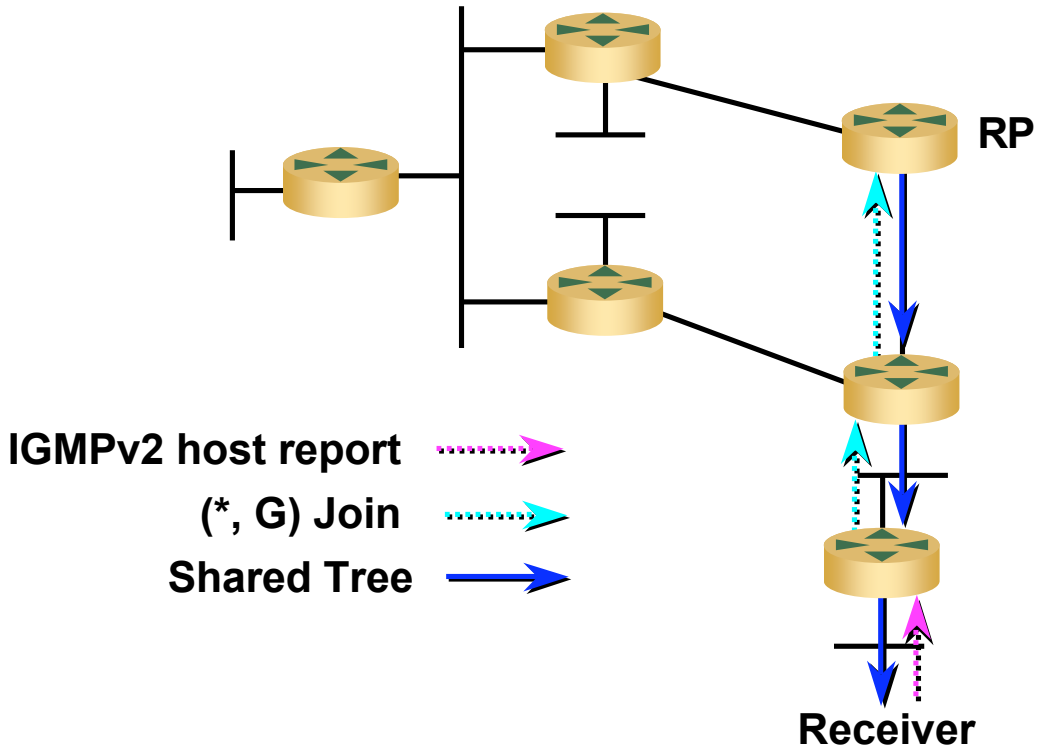
- Allows Source Trees or Shared Trees
- Rendezvous Point (RP)
  - Provides network source discovery
  - Root of shared tree
- Typically use shared tree to bootstrap source tree
- RP's can be learned via:
  - Static configuration
  - Auto-RP
  - Bootstrap Router

    DCOS has an excellent BSR implementation.

Setting up PIM forwarding

The following slides provide a basic overview of how PIM-SM works and only provides a general overview

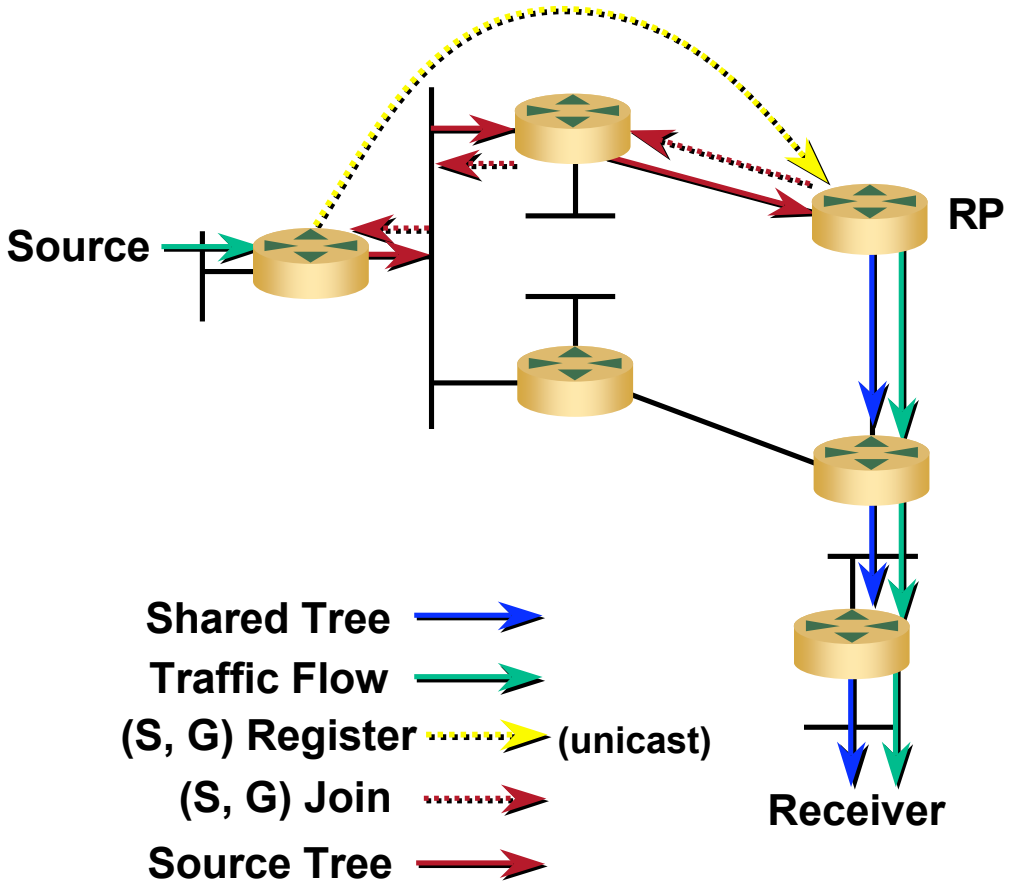
PIM-SM Shared Tree Join



Receiver announces desire to join group G with igmpv2 host report – (*,G).

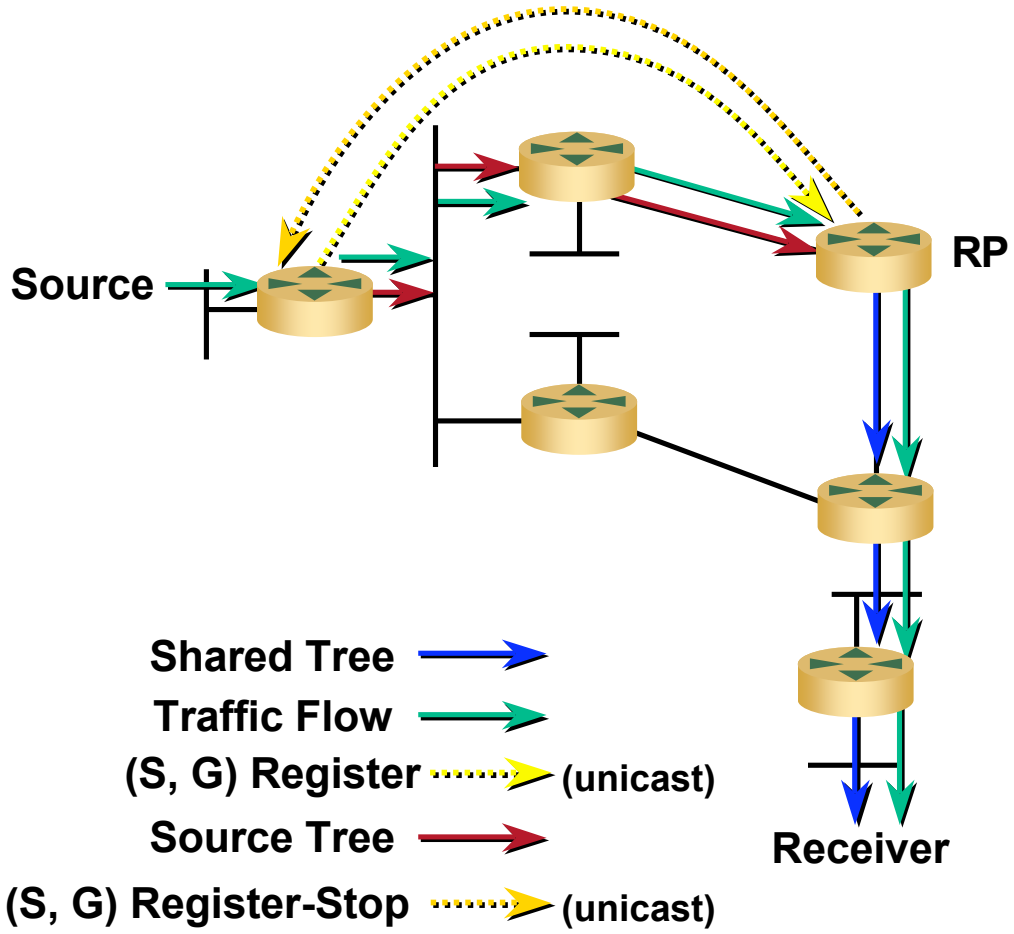
(*, G) State created from the RP to the receiver.

PIM-SM Sender Registration



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

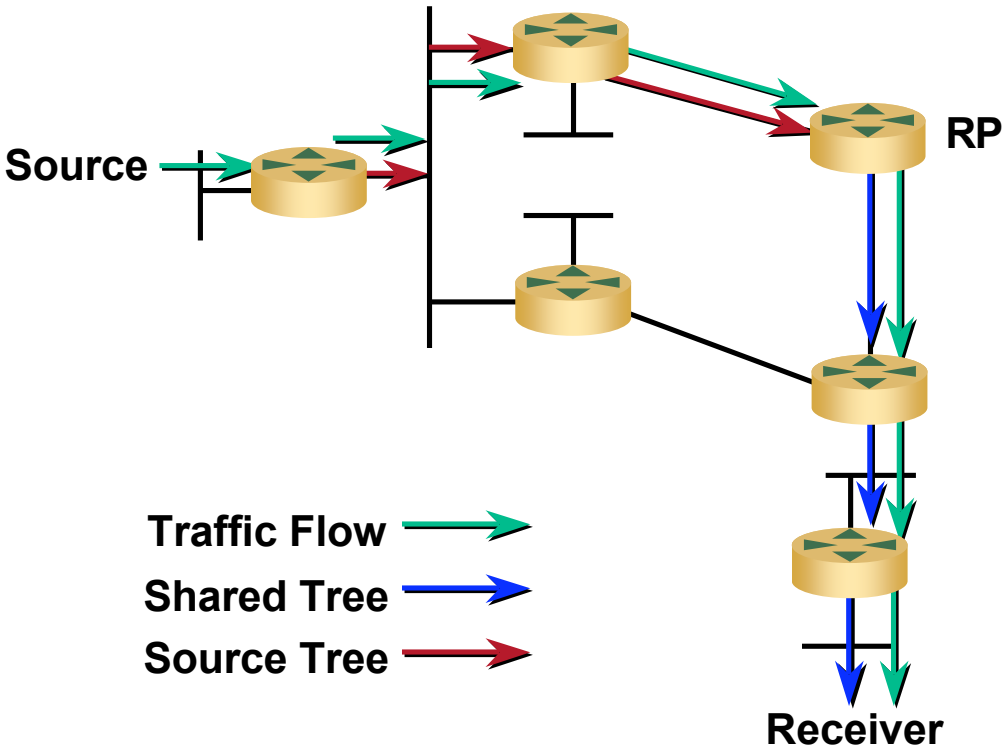
PIM-SM Sender Registration



(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.

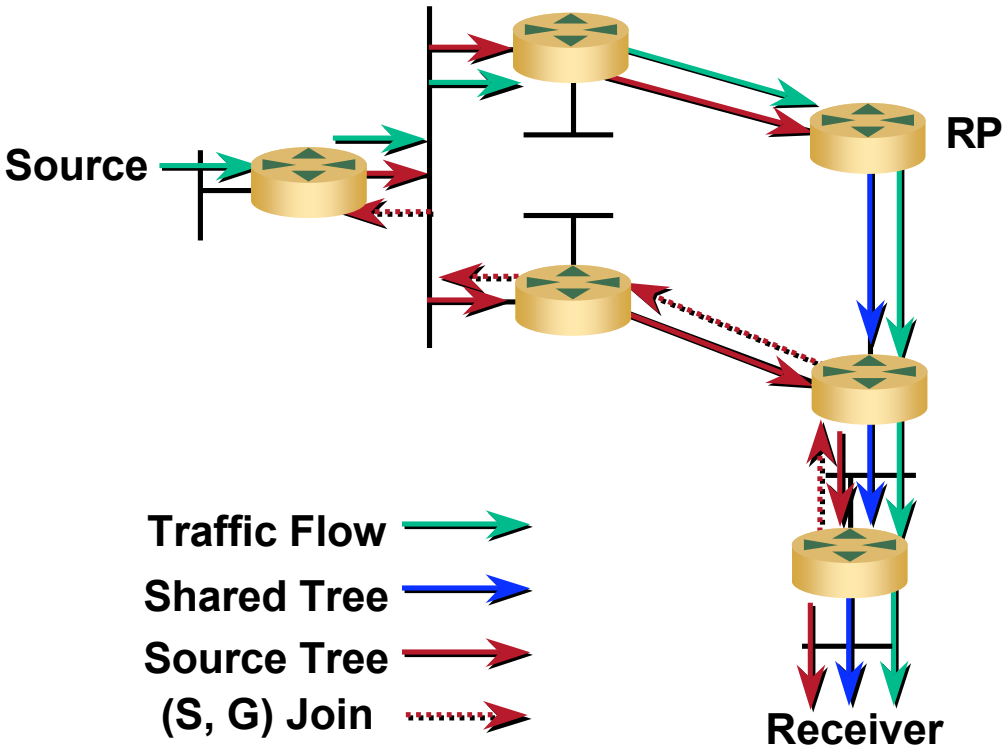
PIM-SM Sender Registration



Source traffic flows natively along SPT to RP.

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

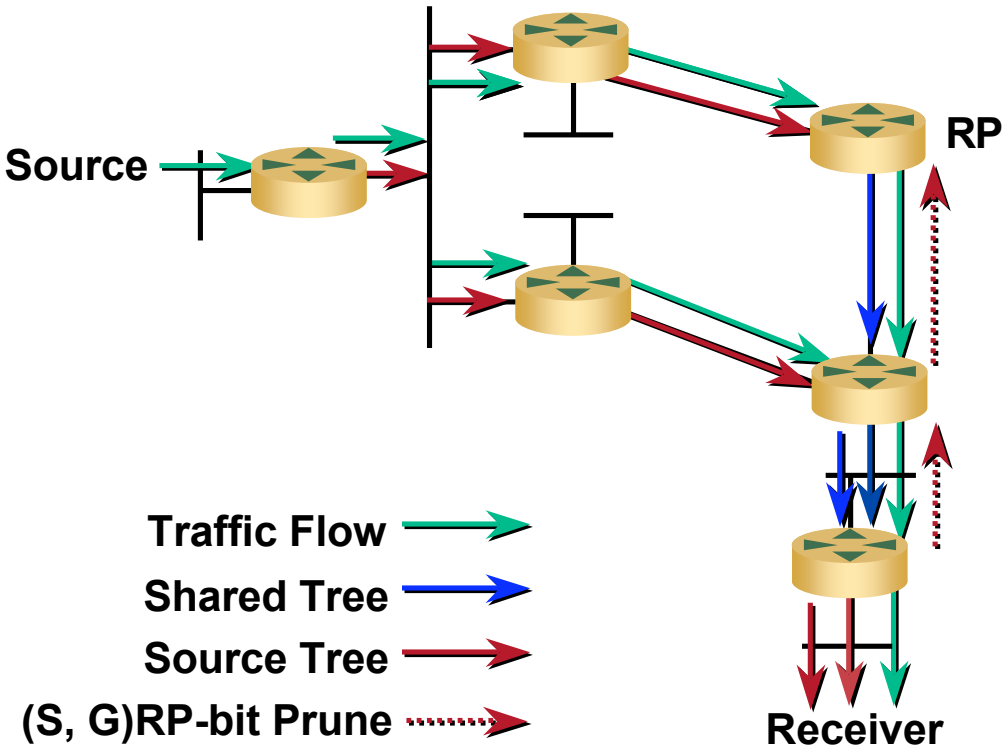
PIM-SM SPT Cutover



Last-hop router joins the Source Tree.

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

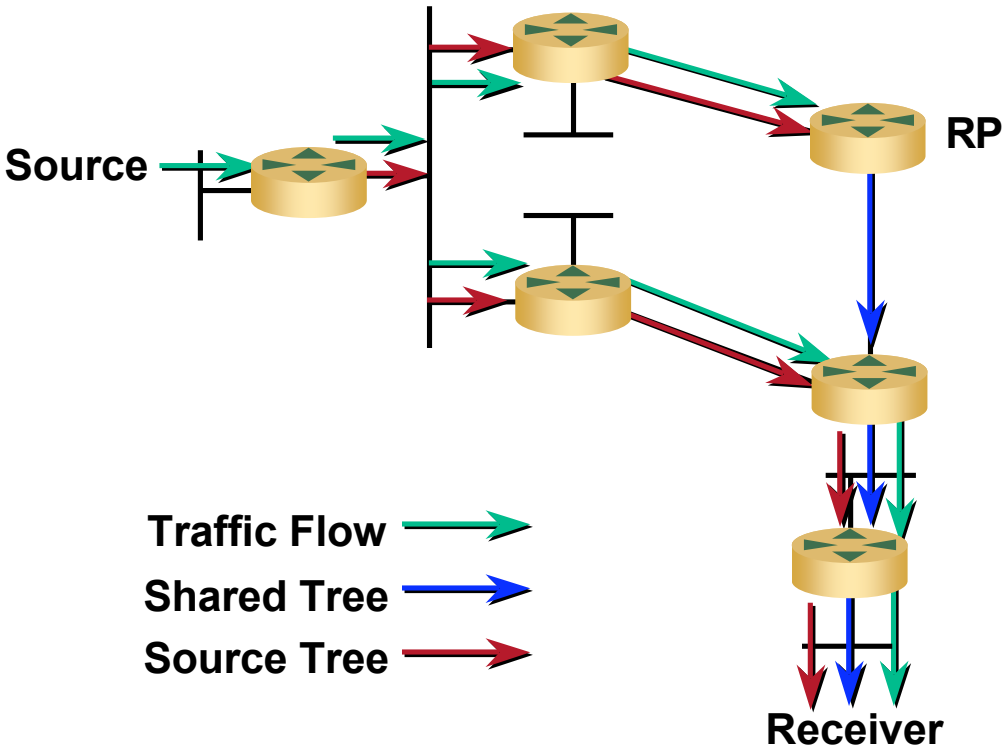
PIM-SM SPT Cutover



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

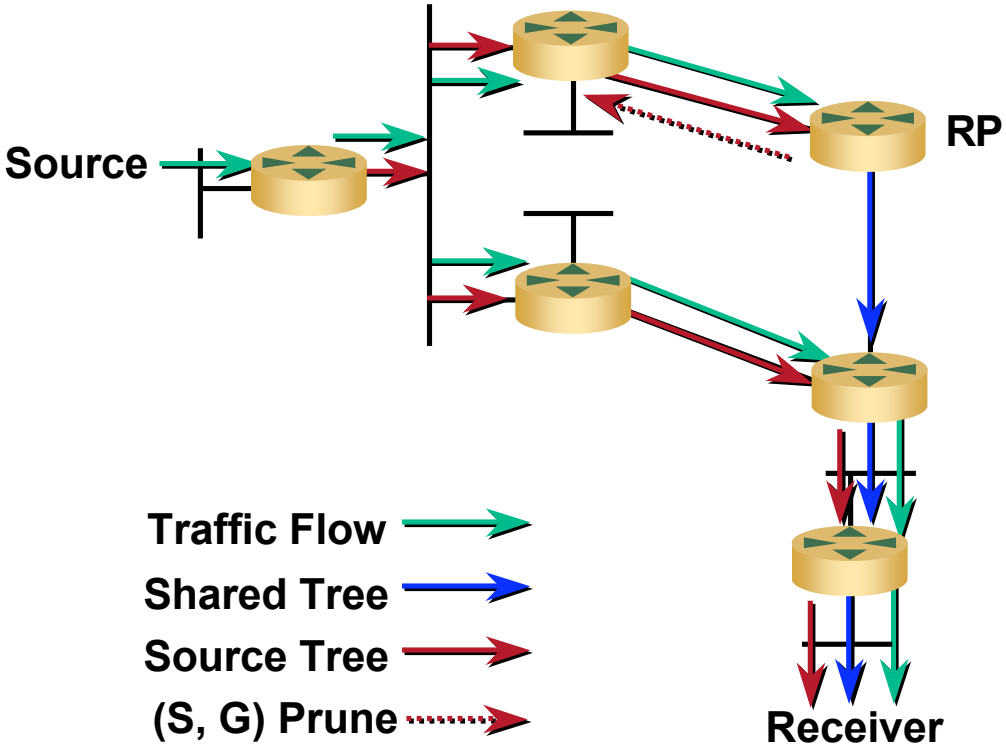
Additional (S, G) State is created along the Shared Tree to prune off (S, G) traffic.

PIM-SM SPT Cutover



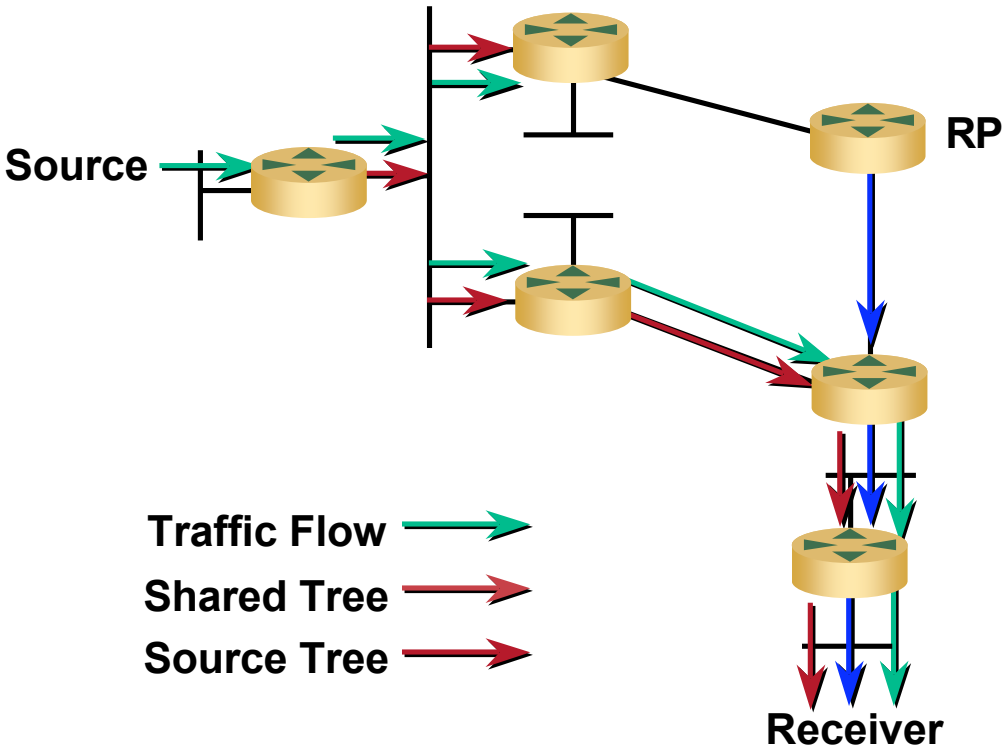
(S, G) Traffic flow is now pruned off of the Shared Tree and is flowing to the Receiver via the Source Tree.

PIM-SM SPT Cutover



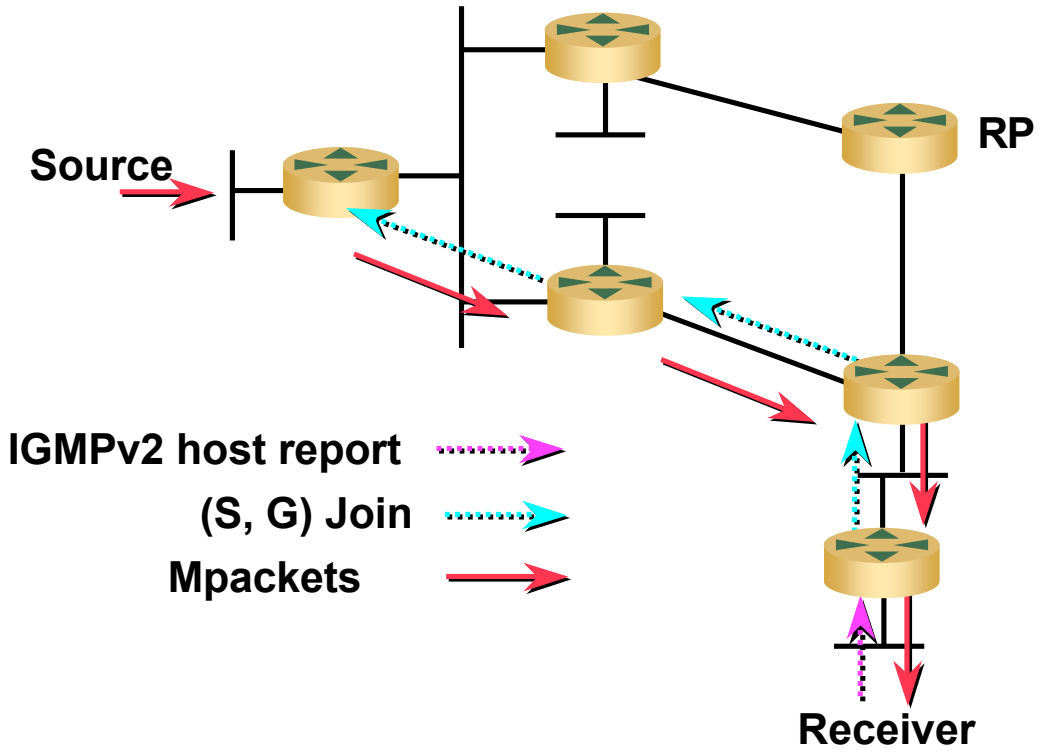
(S, G) traffic flow is no longer needed by the RP so it Prunes the flow of (S, G) traffic.

PIM-SM SPT Cutover



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

PIM-SSM Source Tree Join



Source is sending

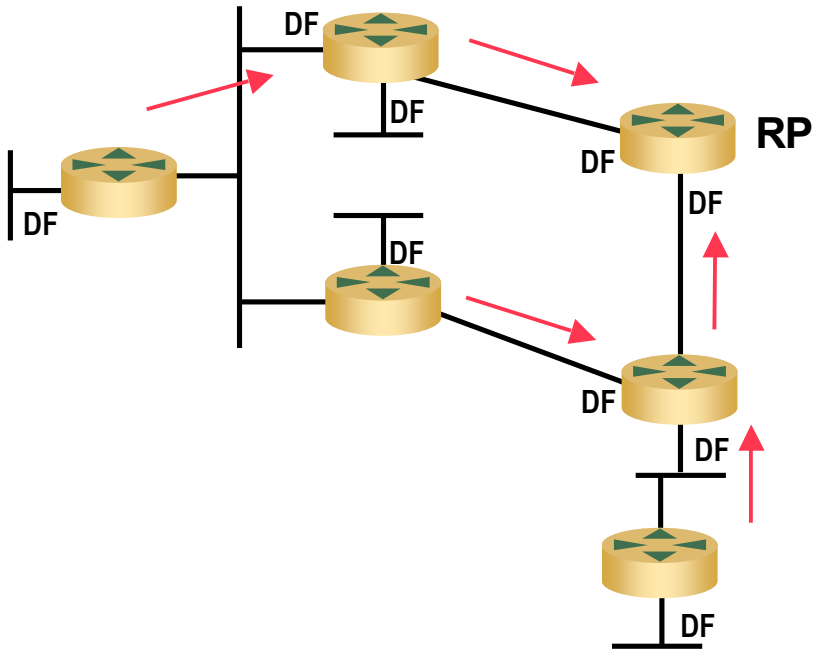
Receiver announces desire to join group G with igmpv3 host report – (S,G).

Pim Joins sent toward the source create (S, G) State.

Mpackets flow

Exactly like joining the SPT in PIM-SM

PIM-BiDir



- RP is announced
- DF elected per LAN
- Upstream forwarding is set

- Downstream forwarding is exactly like PIM-SM -except- (*,G) joins are sent to the DF

PIM Configuration

All Modes of PIM

```
interface <interface>
 ip pim sparse-mode
```

Sparse-mode and Bidir require an RP

```
ip pim rp-address 198.58.3.254 [bidir]
```

SSM: 232/8 is the default

```
ip pim ssm-range <range>
```

PIM RP Configuration

- **Options**
  - **Static: often recommended, especially for ISPs**
  - **Auto-RP: IOS only, becoming obsolete**
  - **Bootstrap router: RFC standard, works well in enterprise**

- **Static RP Configuration on ALL routers in PIM domain**

```
ip pim rp-address 198.58.3.254
```

- **Bootstrap Router (BSR): on just the BSR**

```
ip pim rp-candidate loopback 0
ip pim bsr-candidate loopback 0
```

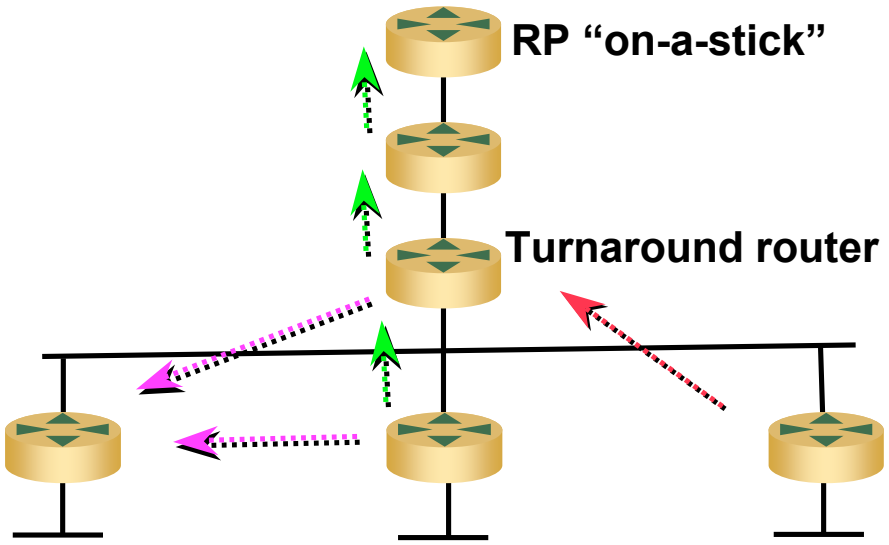
PIM Routing Triangle

- **PIM Sparse-mode routing is a triangle**
  - **Shortest-path from the RP to the receiver, Shared-tree or RP-tree (RPT)**
  - **Shortest-path from the Source to the receiver (SPT)**
  - **Shortest-path from the Source to the RP (X-line)**
    X on IOS mroutes shows the router is a “turnaround” router
- **Know which forwarding path you are on or you *will* be confused.**

The ‘turnaround’ router

- Also known as “router on a stick”
- Occurs when the shared-path and the shortest-path intersect on a LAN
- Most noticeable when one receiver stays on the shared-tree while another joins to the shortest-path tree
  - (*,G) joins are forwarded up to the RP removing (S,G) RP-bit
  - RP joins to the SPT creating (S,G) state: RPF toward source
  - A second receiver causes (S,G,RP-bit) prunes changing state again
- This repeats over and over with mpackets periodically being received at the RP and pruned back again.

The ‘turnaround’ router



In IOS the “Turnaround router” sets the X-flag and sends (S,G) joins toward the source, blocking (*,G) only joins from being sent to the RP.

(S,G) Join

(*,G) Join

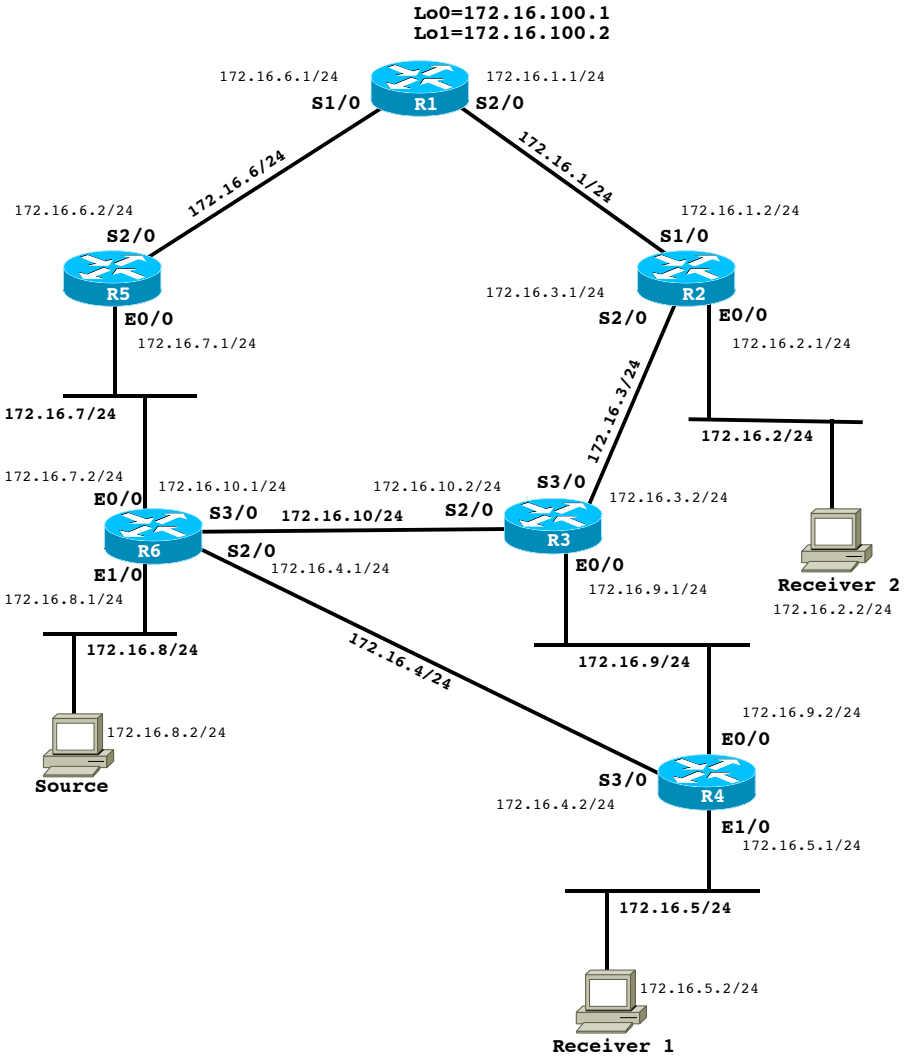
(*,G) Join with RP-bit prune

LAB #1
PIM-SM Mechanics - SSM / ASM / BiDir

Cisco.com

- Get your username and password from the instructor
- Once your are logged in, DO NOT start the lab until instructed
- Lab templates or cfgs: PIM-Mechanics
- Refer to your lab handout

LAB #1
PIM-SM Mechanics - SSM / ASM / BiDir



Agenda

Cisco.com

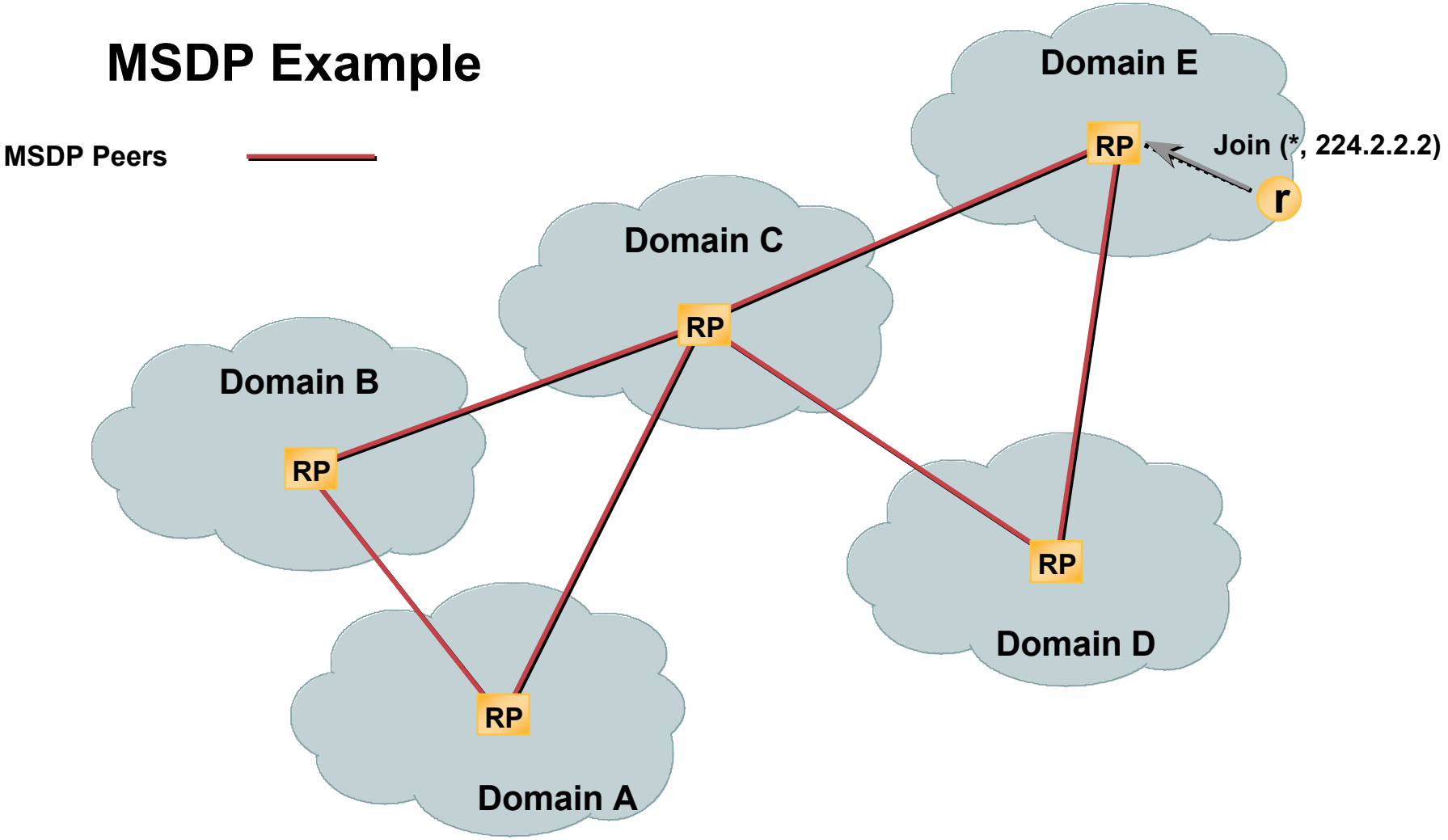
- Introduction
- Multicast addressing
- Group Membership Protocol
- PIM-SM / SSM
- **MSDP**
- MBGP
- Summary

MSDP Overview

- Uses inter-domain source trees only.
  - RP's know about all sources in their domain
    - Sources cause a "PIM Register" to the RP
    - Can tell RP's in other domains of its sources
      - Via MSDP SA (Source Active) messages
  - RP's know about receivers in their 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
      - Only necessary if there are receivers for the group
    - Last-hop routers then join source tree directly.

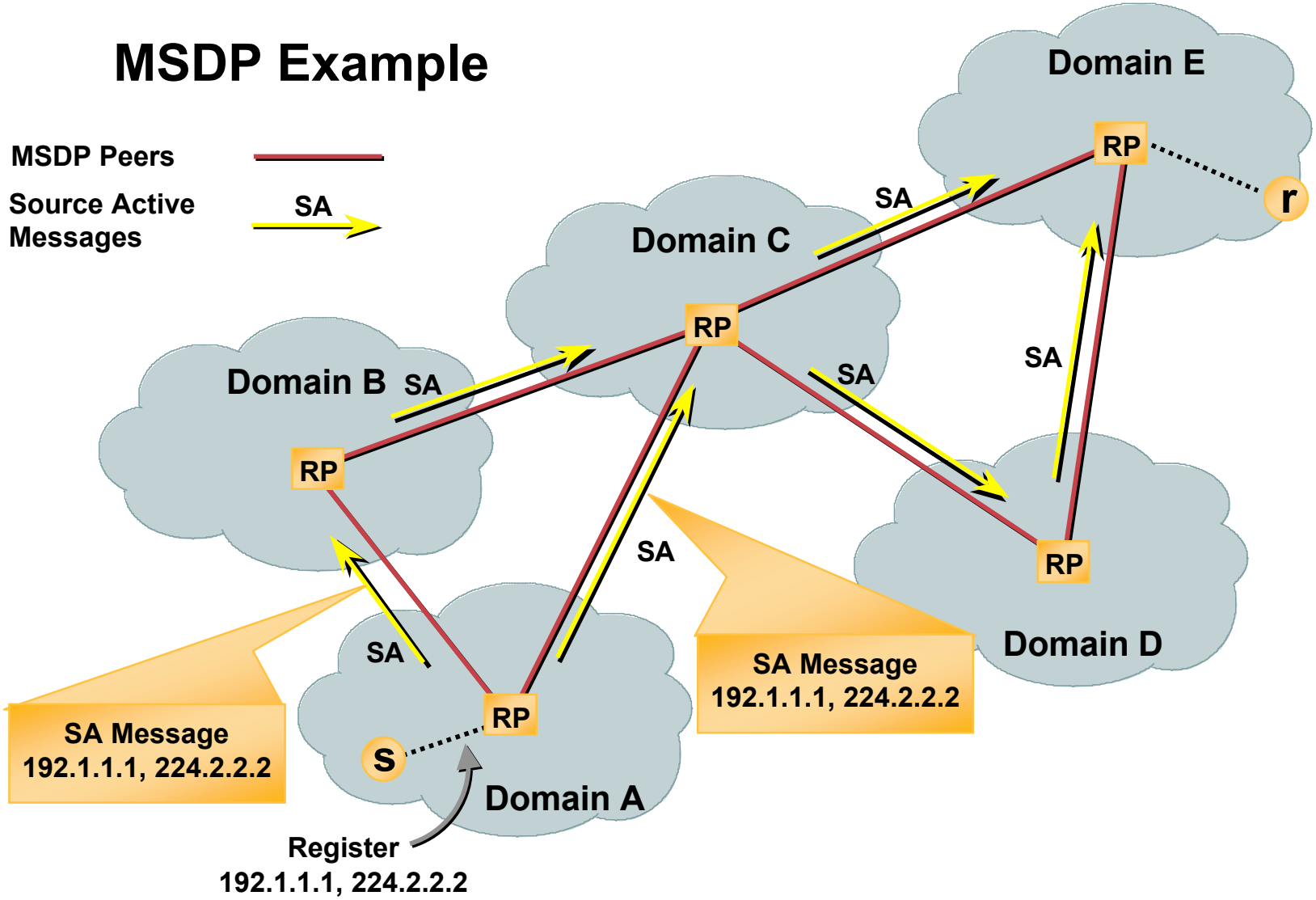
MSDP Overview

MSDP Example



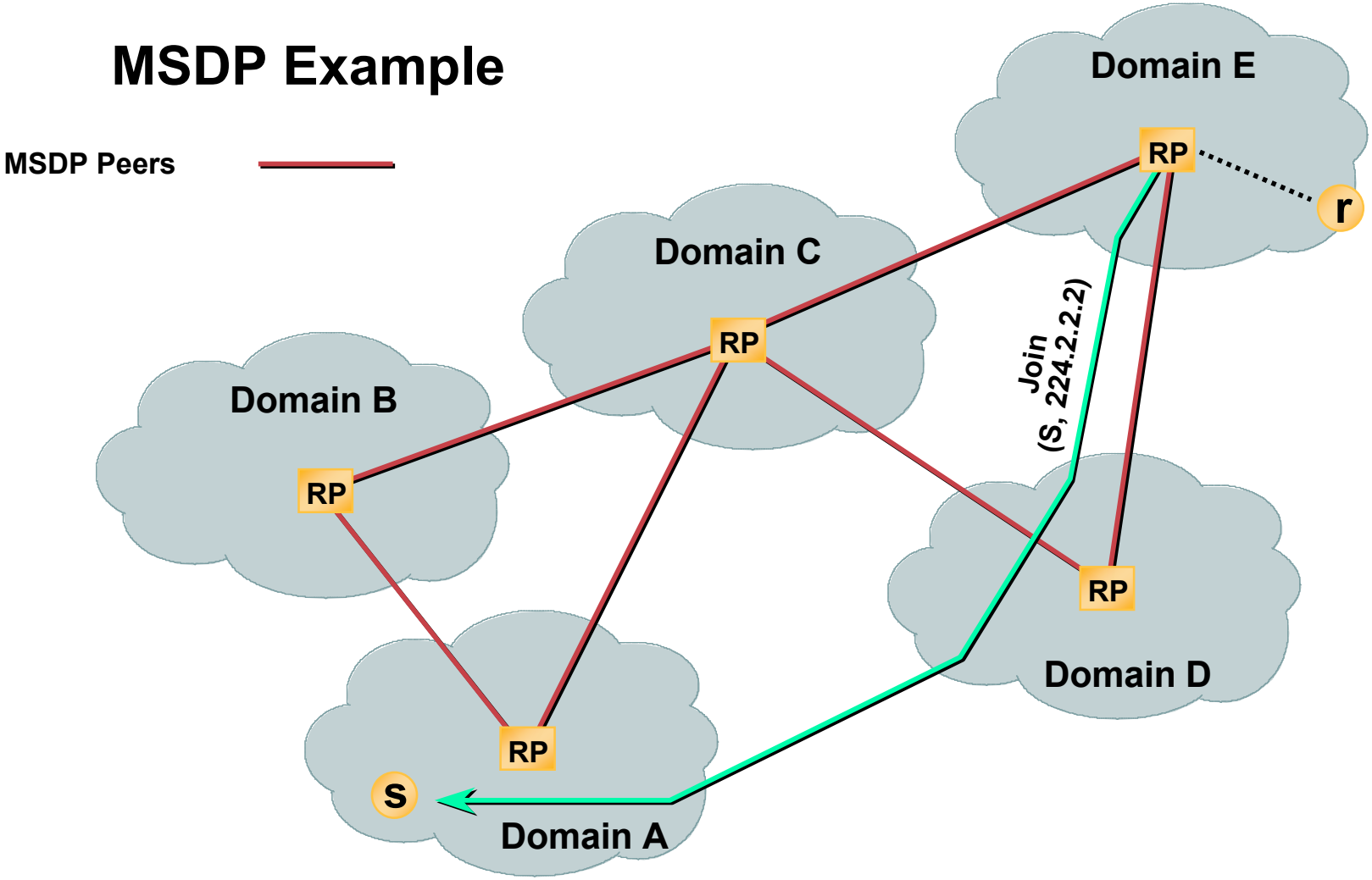
MSDP Overview

MSDP Example



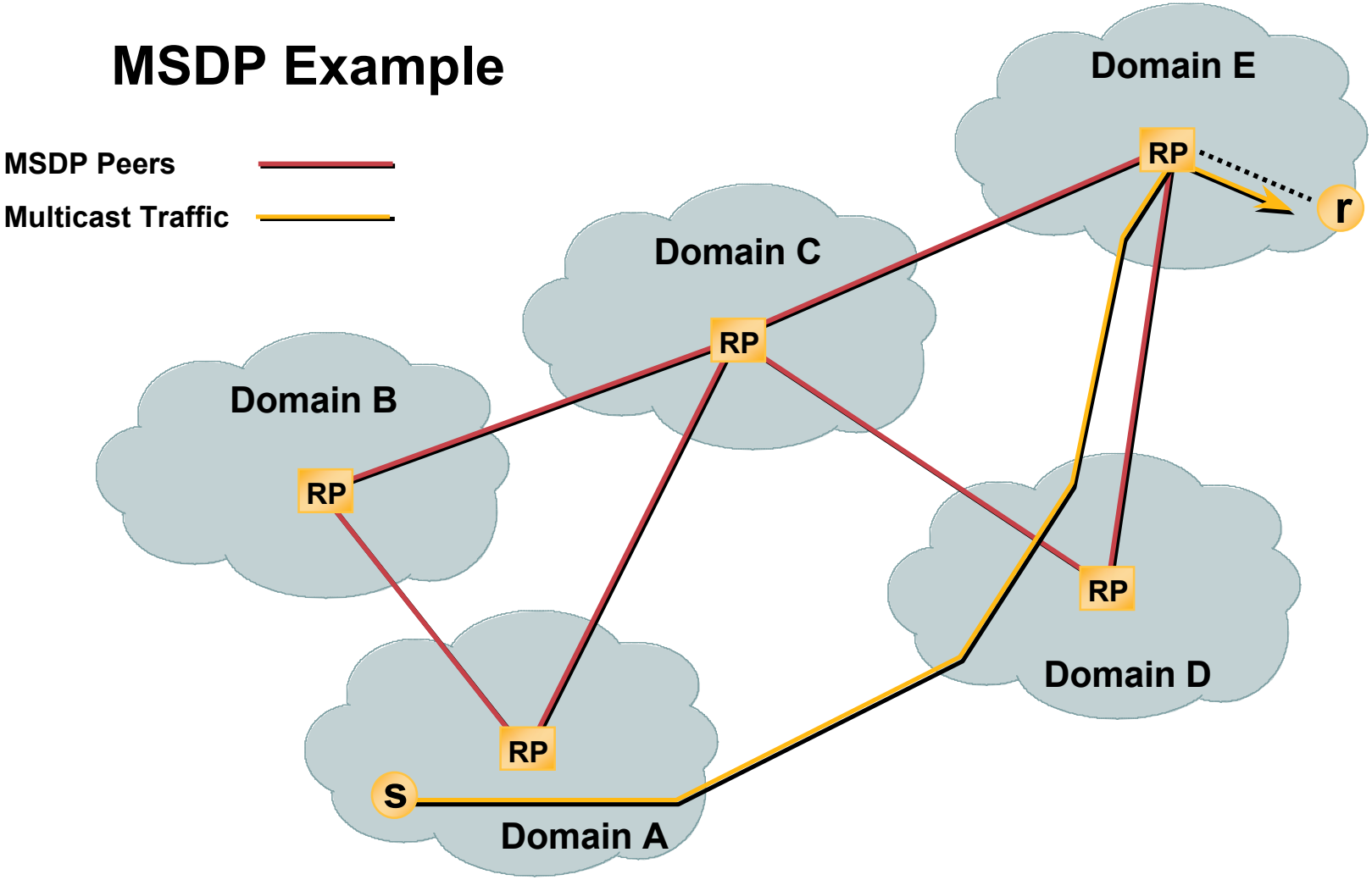
MSDP Overview

MSDP Example



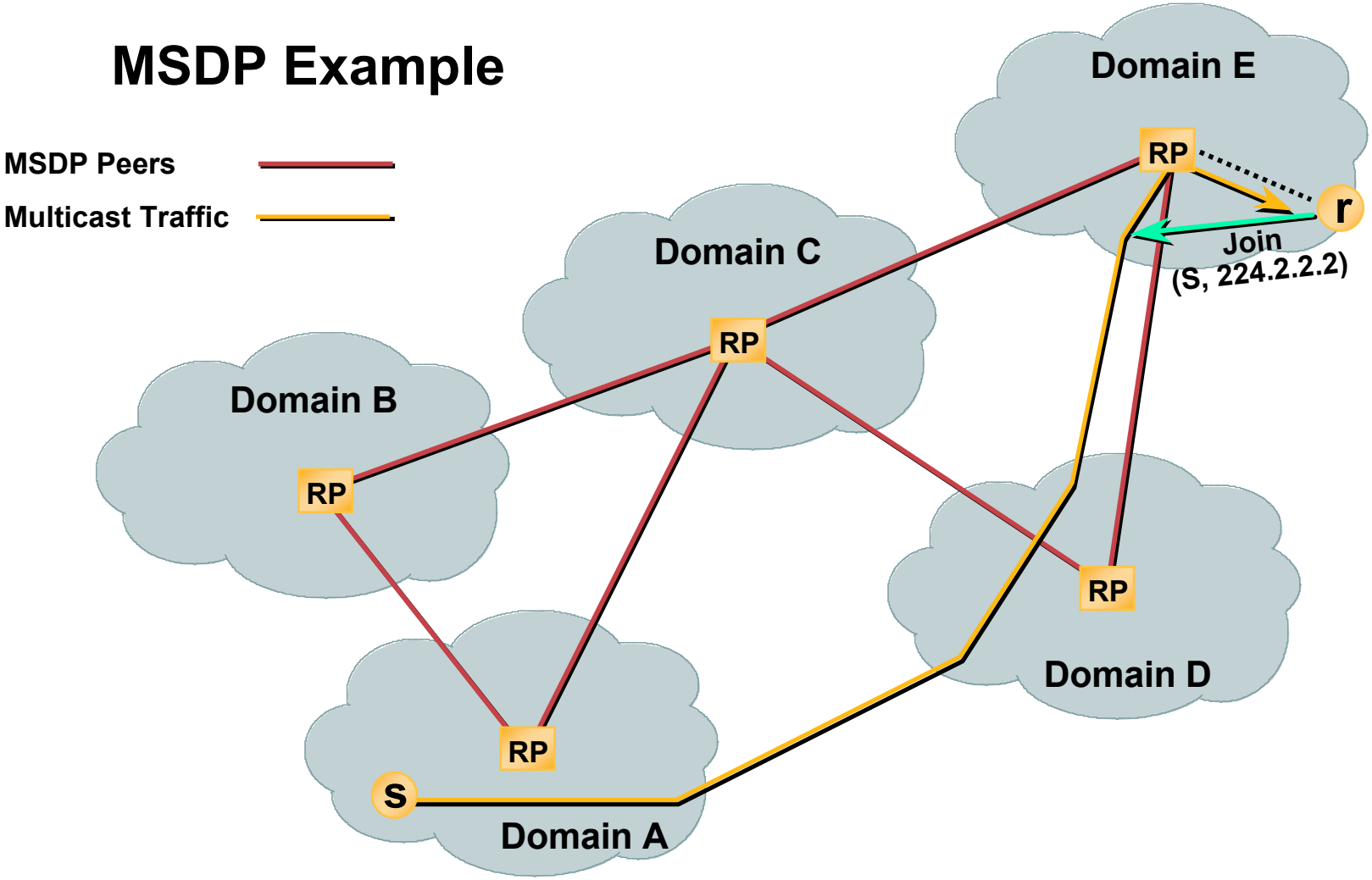
MSDP Overview

MSDP Example



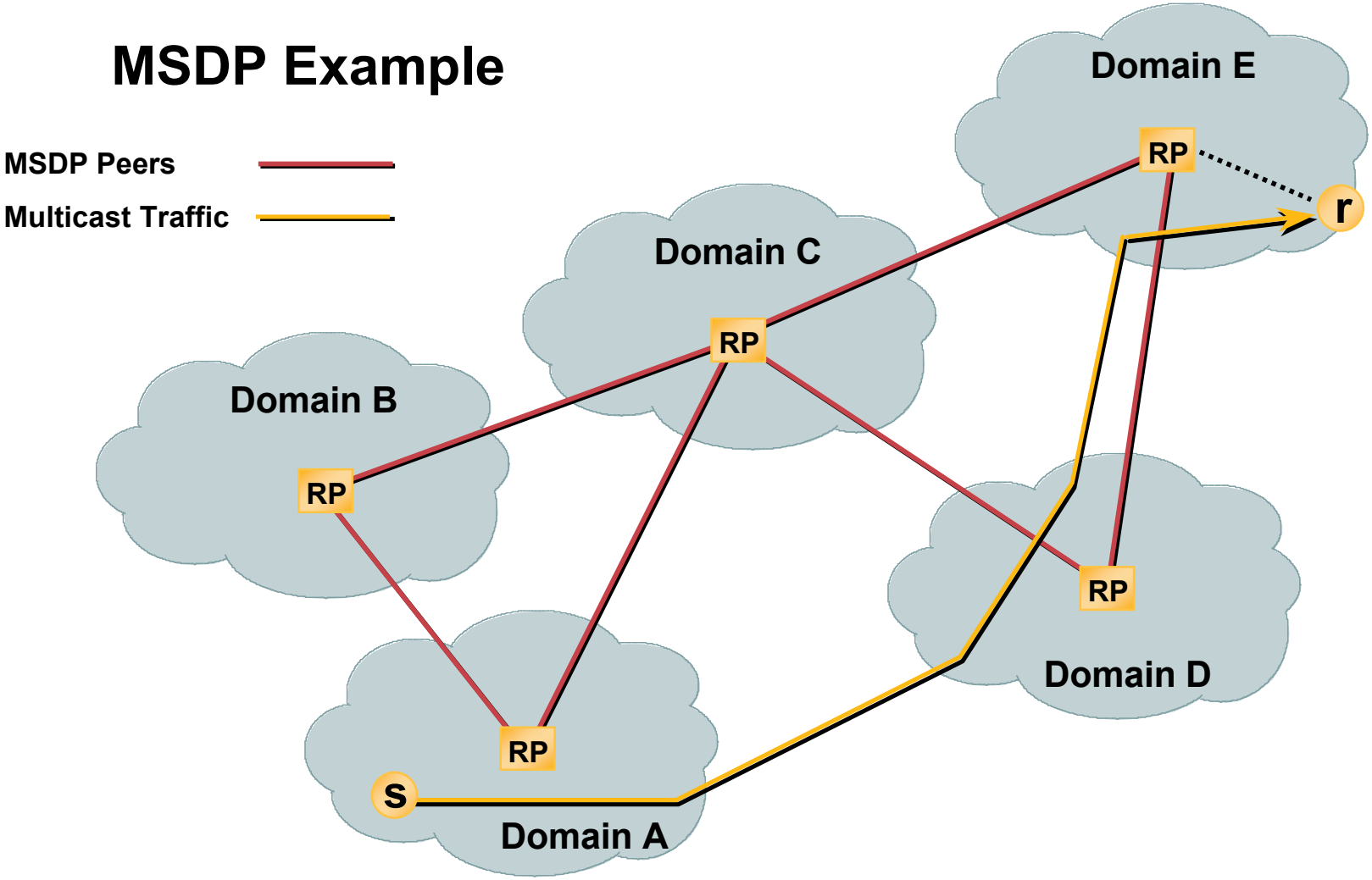
MSDP Overview

MSDP Example



MSDP Overview

MSDP Example



MSDP Peers

- MSDP Peers configured similar to BGP
- Peers connect using TCP port 639
  - Lower address peer initiates connection
  - Higher address peer waits in LISTEN state
- Peers send keepalives every 60 secs.
- Connection reset after 75 seconds
  - If no MSDP packets or keepalives are received

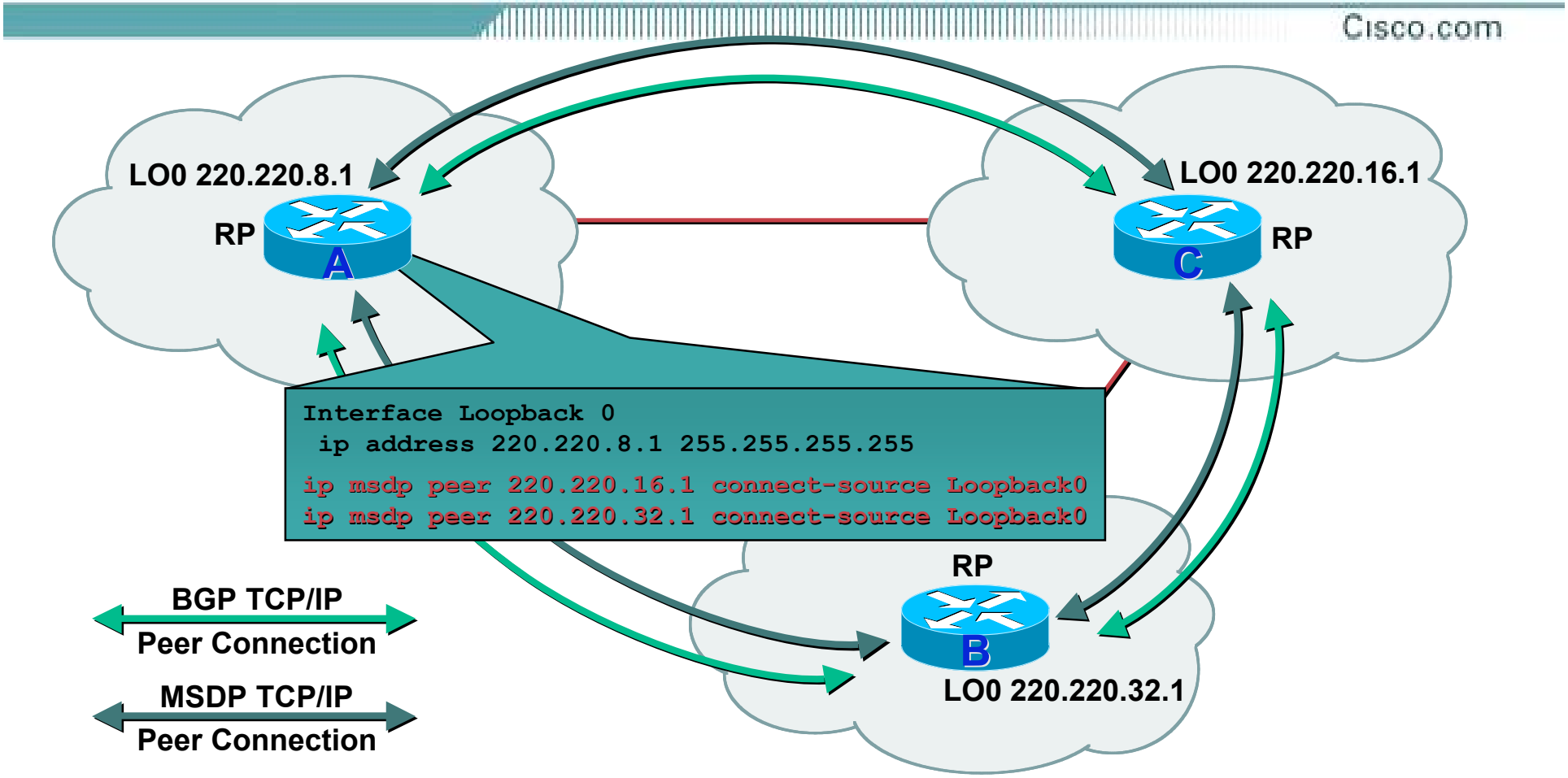
MSDP Peers

- **MSDP peers normally *must* run BGP!**
  - **BGP NLRI is used to RPF check SA messages.**

    May use NLRI from M-Table, U-Table or both.
  - **RPF check prevents SA's from looping.**

    (More on that later.)
- **Exceptions:**
  - **When peering with only a single MSDP peer.**
  - **When using an MSDP Mesh-Group.**

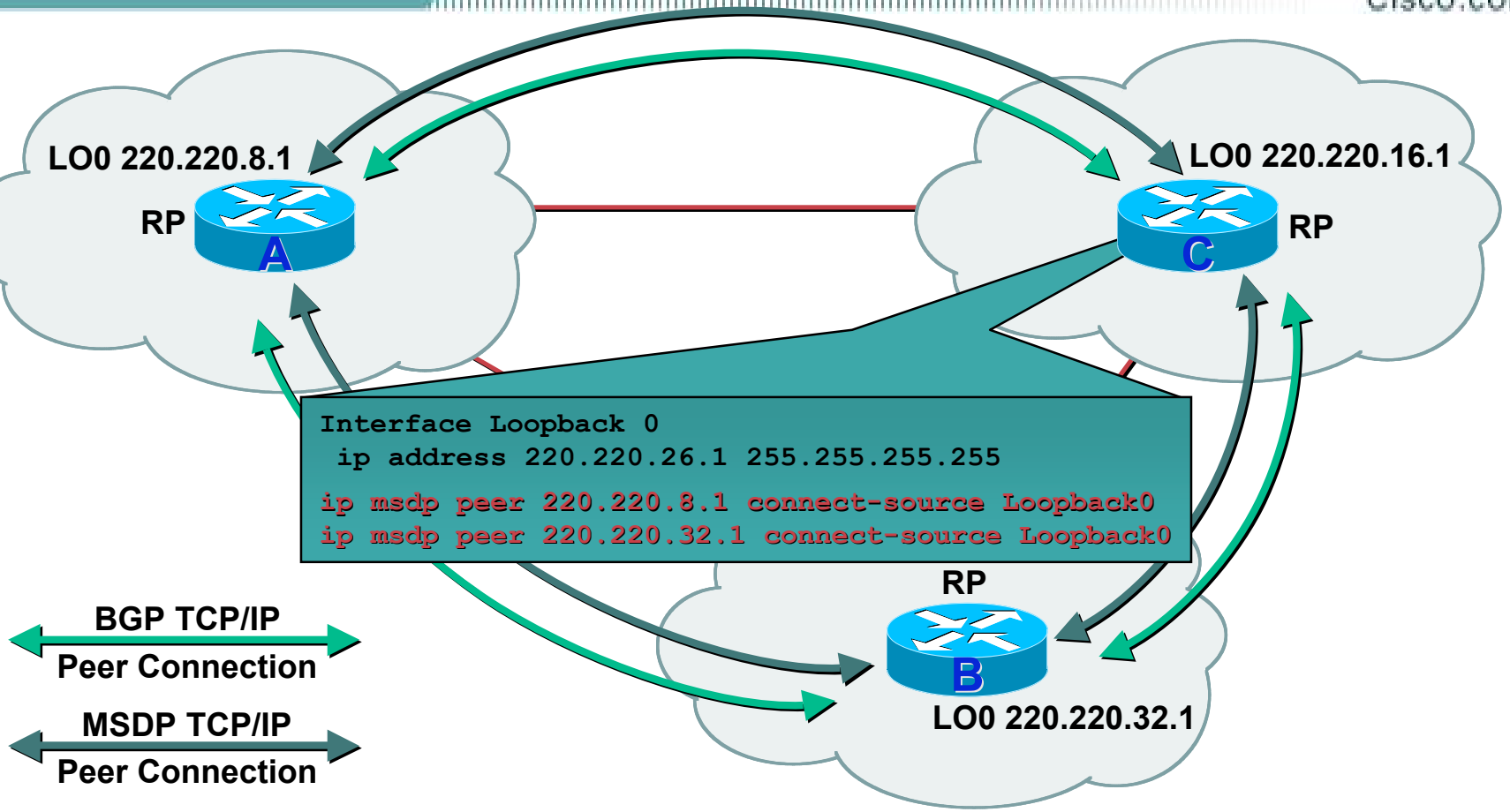
MSDP Peers



- **MSDP peer connections are established using the MSDP "peer" configuration command**

```
ip msdp peer <ip-address> [connect-source <intfc>]
```

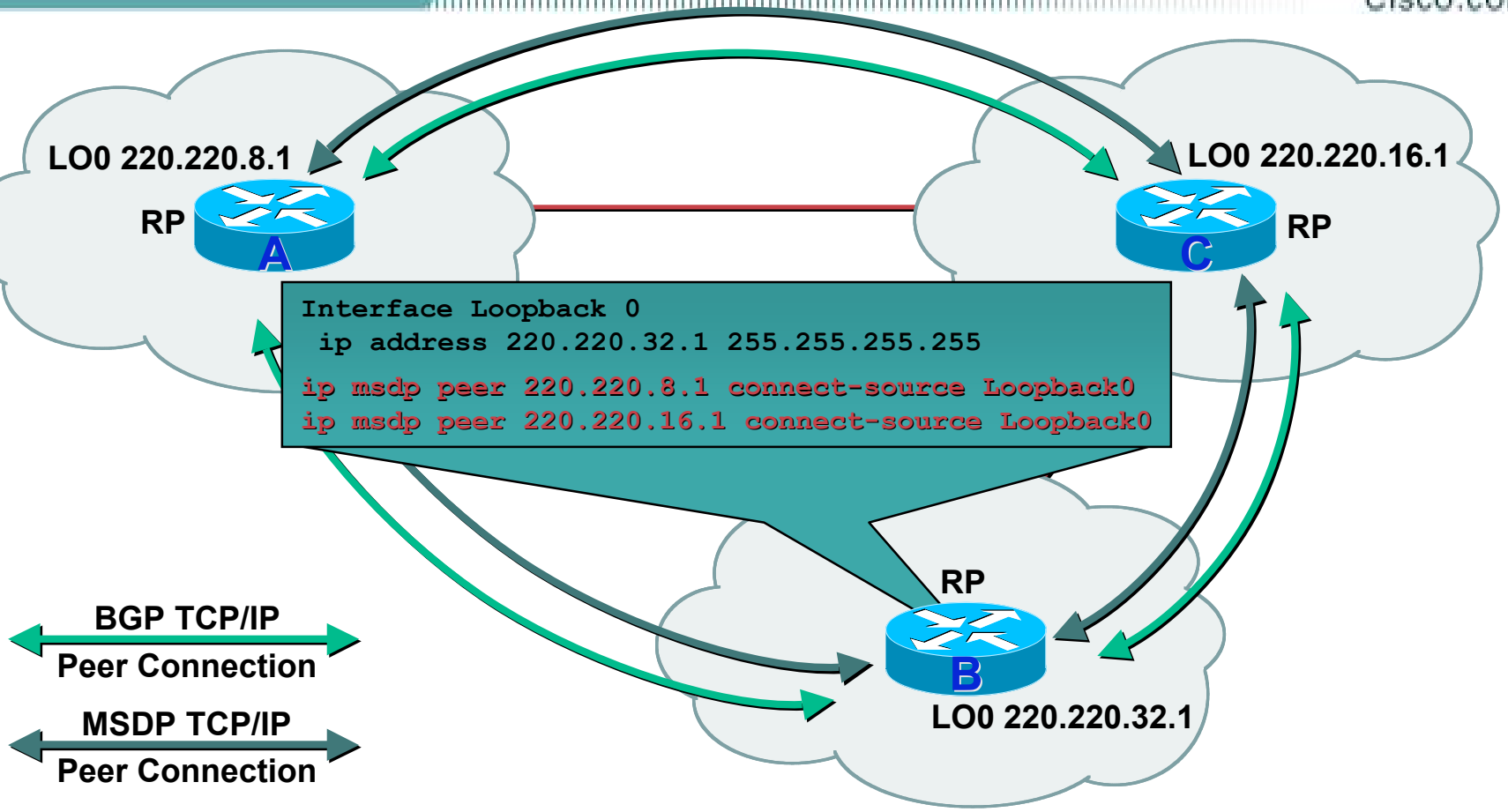
MSDP Peers



- **MSDP peer connections are established using the MSDP "peer" configuration command**

```
ip msdp peer <ip-address> [connect-source <intfc>]
```

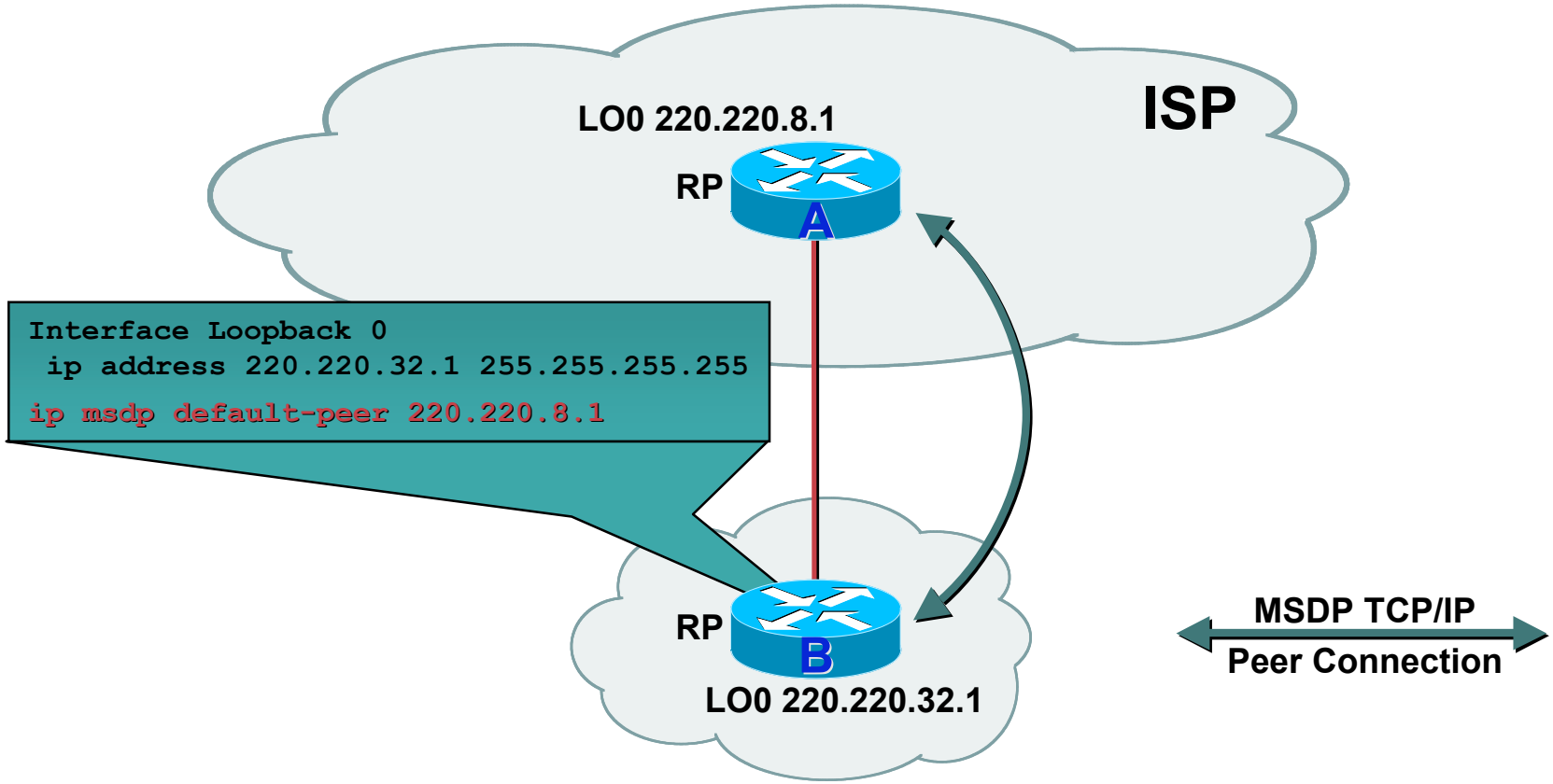
MSDP Peers



- **MSDP peer connections are established using the MSDP "peer" configuration command**

```
ip msdp peer <ip-address> [connect-source <intfc>]
```

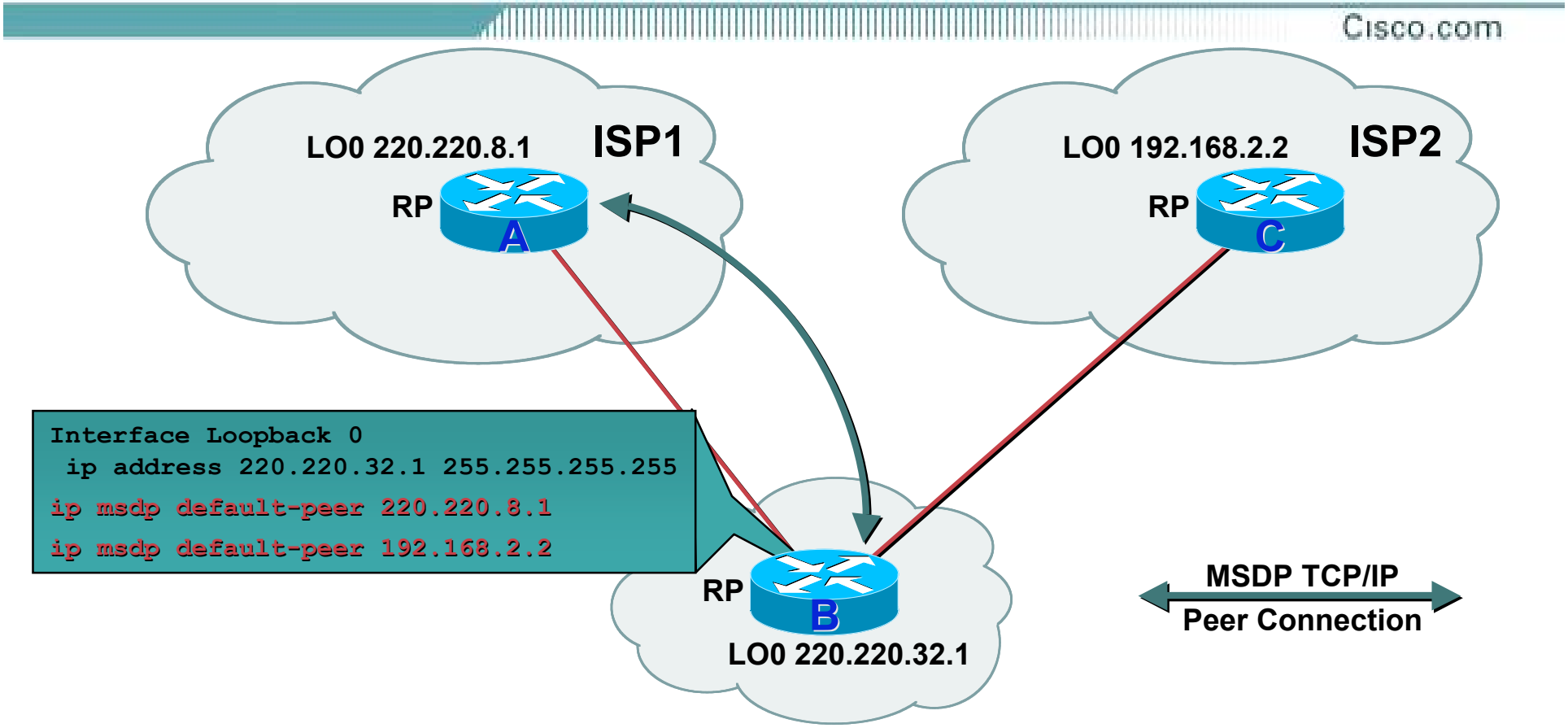
MSDP Peers



- **Stub-networks may use “default” peering without being a BGP peer by using the MSDP “default-peer” configuration command.**

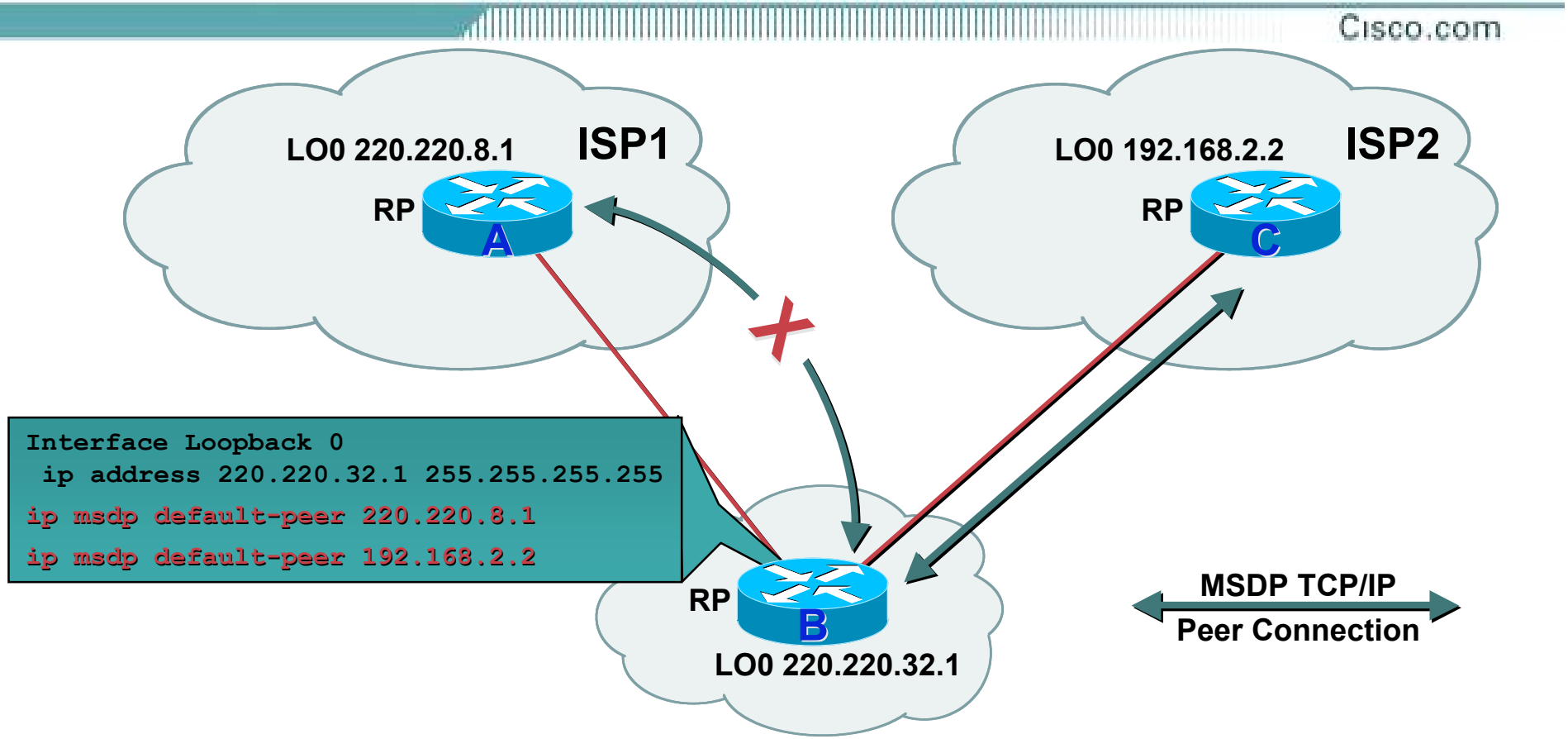
```
ip msdp default-peer <ip-address>
```

MSDP Peers



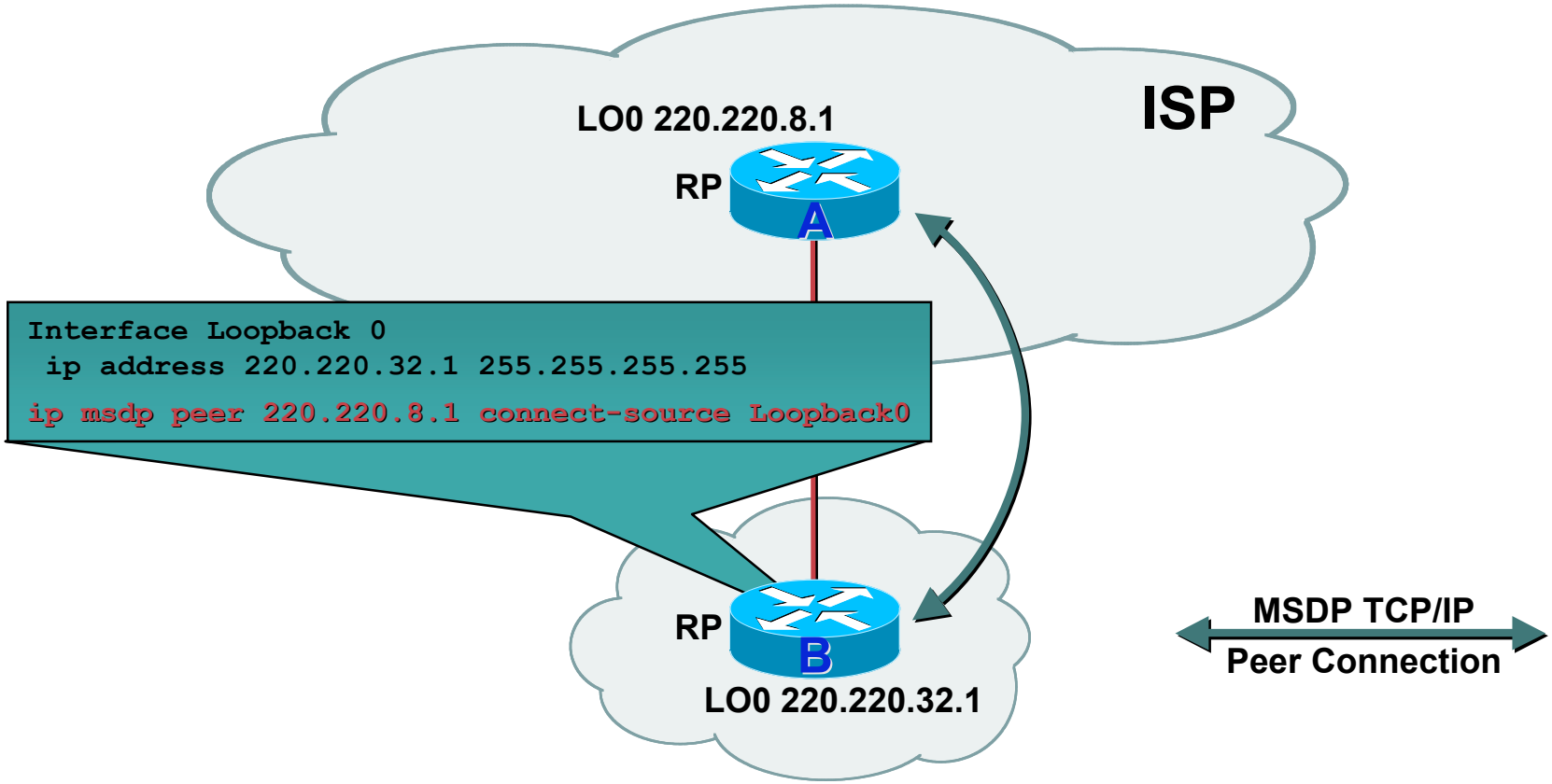
- **Multiple "default-peers" may be configured in case connection to first default-peer goes down.**

MSDP Peers



- When connection to first ‘default-peer’ is lost, the next one in the list is tried.

MSDP Peers



- **Stub-networks configured with only a single MSDP peer are treated in the same manner as when a single "default-peer" is configured. (i.e. BGP is not required.)**

SA Message Contents

- **MSDP Source Active (SA) Messages**
  - Used to advertise active Sources in a domain
  - Can also carry 1st multicast packet from source
    - Hack for Bursty Sources (a'la SDR)
  - SA Message Contents:
    - IP Address of Originating RP
    - Number of (S, G)'s pairs being advertised
    - List of active (S, G)'s in the domain
    - Encapsulated Multicast packet [optional]

Originating SA Messages

- Local Sources
  - RP's only originate SA's for local sources
    - Denoted by the "A" flag on an (S,G) entry on RP
  - A source is local if:
    - The RP received a "Register" for (S, G), or
    - The source is directly connected to RP

Originating SA Messages

- Use ‘msdp redistribute’ to control what SA’s are originated.
  - Think of this as ‘msdp sa-originate-filter’ function

```
ip msdp redistribute [list <acl>]
                     [asn <aspath-acl>]
                     [route-map <map>]
```

    Filter by (S,G) pair using ‘list <acl>’

    Filter by AS-PATH using ‘asn <aspath-acl>’

    Filter based on route-map ‘<map>’

  - Omitting all acl’s stops all SA origination

```
Example: ip msdp redistribute
```

  - Default: Originate SA’s for all local sources

    If ‘msdp redistribute’ command is not configured

Originating SA Messages

- SA messages are triggered when any new source in the local domain goes active.
  - Initial multicast packet is encapsulated in an SA message.
    - This is an attempt at solving the bursty-source problem

Originating SA Messages

- **Encapsulating Initial Multicast Packets**
  - **Can bypass TTL-Thresholds**

    Original TTL is inside of data portion of SA message

    SA messages sent via Unicast with TTL = 255
- **Requires special command to control**

  `ip msdp ttl-threshold <peer-address> <ttl>`
  - **Encapsulated multicast packets with a TTL lower than <ttl> for the specific MSDP peer are not forwarded or originated.**

Originating SA Messages

- Once a minute
  - Router scans mroute table
  - If group = sparse AND router = RP for group

    For each (S,G) entry for the group:

    If the '`msdp redistribute`' filters permits

    AND if the source is a local source

    Then originate an SA message for (S,G)

Receiving SA Messages

- If SA message RPF checks OK
  - Store in SA Cache
  - If new SA cache entry
    - Immediately flood SA downstream
    - Set entry's SA-expire-timer to 6 minutes.
    - If RP for group and receivers exist
      - Create (S,G) entry and trigger (S,G) Join
  - If existing entry
    - Reset entry's SA-expire-timer to 6 minutes.
      - When timer = zero, entry has expired and is deleted.
- Else
  - Discard SA

SA Message Cache

- **Enabling SA Caching**

```
ip msdp cache-sa-state [list <acl>]
```

  - **Caching is now on by default.**
    - **Beginning with IOS versions 12.1(7), 12.0(14)S1.**
      - **Cannot be turned off.**
  - **Router caches all SA messages.**
    - **Cached (S, G) entries timeout after 6 minutes.**
      - **If not refreshed by another (S,G) SA message.**
  - **Once per minute, router scans SA cache.**
    - **Sends SA downstream for each entry in cache.**

SA Message Caching

- **Listing the contents of the SA Cache**

```
show ip msdp sa-cache [<group-or-source>] [<asn>]
```

```
sj-mbone# show ip msdp sa-cache
MSDP Source-Active Cache - 1997 entries
(193.92.8.77, 224.2.232.0), RP 194.177.210.41, MBGP/AS 5408, 00:01:51/00:04:09
(128.119.167.221, 224.77.0.0), RP 128.119.3.241, MBGP/AS 1249, 06:40:59/00:05:12
(147.228.44.30, 233.0.0.1), RP 195.178.64.113, MBGP/AS 2852, 00:04:48/00:01:11
(128.117.16.142, 233.0.0.1), RP 204.147.128.141, MBGP/AS 145, 00:00:41/00:05:18
(132.250.95.60, 224.253.0.1), RP 138.18.100.1, MBGP/AS 668, 01:15:07/00:05:55
(128.119.40.229, 224.2.0.1), RP 128.119.3.241, MBGP/AS 1249, 06:40:59/00:05:12
(130.225.245.71, 227.37.32.1), RP 130.225.245.71, MBGP/AS 1835, 1d00h/00:05:29
(194.177.210.41, 227.37.32.1), RP 194.177.210.41, MBGP/AS 5408, 00:02:53/00:03:07
(206.190.42.106, 236.195.60.2), RP 206.190.40.61, MBGP/AS 5779, 00:07:27/00:04:04
                                .
                                .
                                .
```

- **Clearing the contents of the SA Cache**

```
clear ip msdp sa-cache [<group-address> | group-name]
```

Filtering Incoming/Outgoing SA Messages

- **SA Filter Command:**

```
ip msdp sa-filter {in|out} <peer-address> [list <acl>]
                                          [route-map <map>]
```

  - Filters (S,G) pairs to / from peer based on specified ACL.
  - Can filter based on AS-Path by using optional route-map clause with a path-list acl.
  - You can filter flooded and originated SA's based on a specific peer, incoming and outgoing.
- **Caution: Filtering SA messages can break the Flood and Join mechanism!**

Recommended MSDP SA Filter


```
! domain-local applications
access-list 111 deny   ip any host 224.0.2.2      !
access-list 111 deny   ip any host 224.0.1.3      ! Rwhod
access-list 111 deny   ip any host 224.0.1.24     ! Microsoft-ds
access-list 111 deny   ip any host 224.0.1.22     ! SVRLOC
access-list 111 deny   ip any host 224.0.1.2      ! SGI-Dogfight
access-list 111 deny   ip any host 224.0.1.35     ! SVRLOC-DA
access-list 111 deny   ip any host 224.0.1.60     ! hp-device-disc
!-- auto-rp groups
access-list 111 deny   ip any host 224.0.1.39
access-list 111 deny   ip any host 224.0.1.40
!-- scoped groups
access-list 111 deny   ip any 239.0.0.0 0.255.255.255
!-- loopback, private addresses (RFC 1918)
access-list 111 deny   ip 10.0.0.0 0.255.255.255 any
access-list 111 deny   ip 127.0.0.0 0.255.255.255 any
access-list 111 deny   ip 172.16.0.0 0.15.255.255 any
access-list 111 deny   ip 192.168.0.0 0.0.255.255 any
access-list 111 permit ip any any
!-- Default SSM-range. Do not do MSDP in this range.
access-list 111 deny   ip any 232.0.0.0 0.255.255.255
access-list 111 permit ip any any
```

See “ftp://ftp-eng.cisco.com/ipmulticast/msdp-sa-filter.txt” for the latest updates to this list.

SA Message RPF Checking

- **Purpose**
  - **Accept SA's via a single deterministic path**
    - **Ignore all other arriving SA's**
    - **Necessary to prevent SA's from looping endlessly**
- **Problem**
  - **Need to know MSDP topology of Internet**
    - **But, MSDP does not distribute topology data!**
- **Solution**
  - **Use BGP data to *infer* MSDP topology.**
    - **Impact:**
      - **The MSDP topology must follow BGP topology.**
      - **An MSDP peer must *generally* also be an BGP peer.**

SA Message RPF Checking

Cisco.com

- RPF Check Rules depend on peering
  - Rule 1: Sending MSDP peer = iBGP peer
  - Rule 2: Sending MSDP peer = eBGP peer
  - Rule 3: Sending MSDP peer != BGP peer
- Exceptions:
  - RPF check is skipped when:

    Sending MSDP peer = Originating RP

    Sending MSDP peer = Mesh-Group peer

    Sending MSDP peer = only MSDP peer

    (i.e. the ‘default-peer’ or the only ‘msdp-peer’ configured.)

SA Message RPF Checking

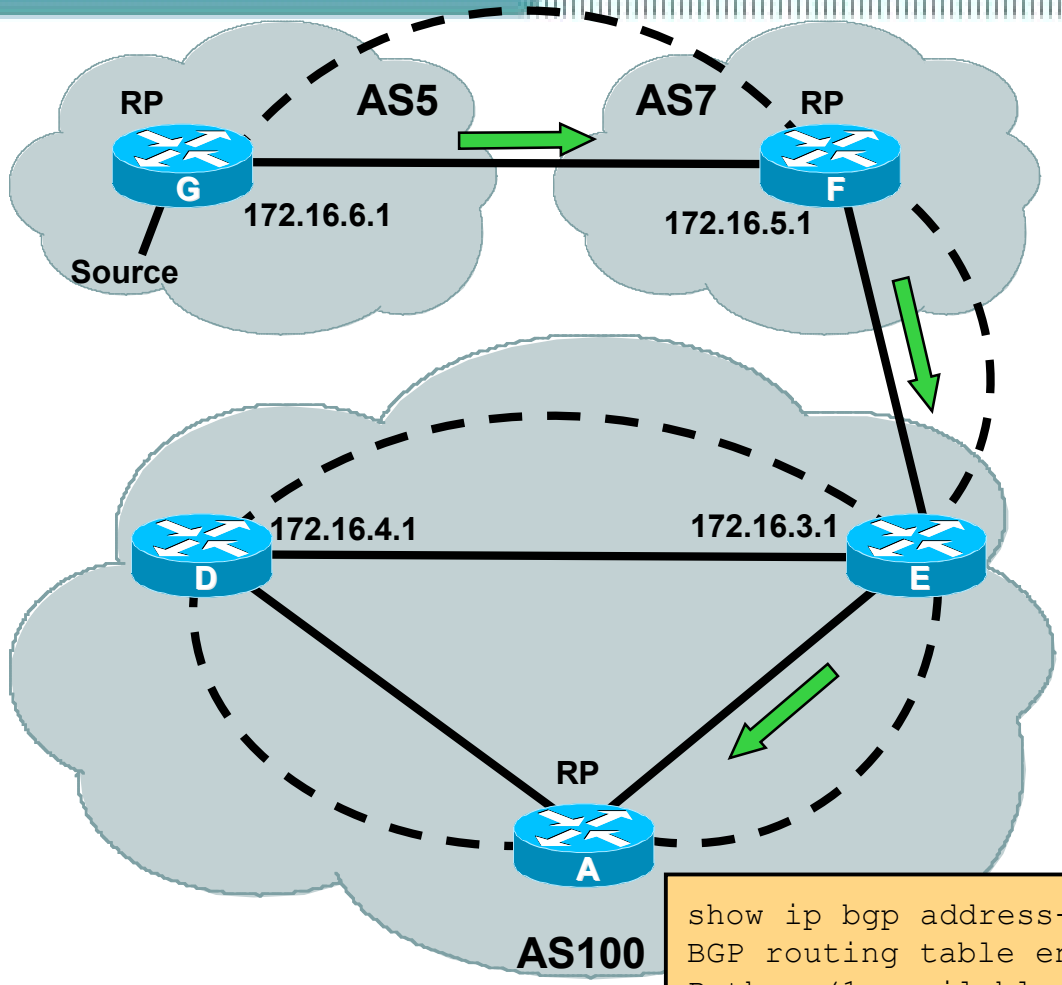
- **Determining Applicable RPF Rule**
  - **Use IP address of sending MSDP peer**

    Find BGP neighbor w/matching IP address
    IF (no match found)
    Apply Rule 3
    IF (matching neighbor = iBGP peer)
    Apply Rule 1
    ELSE {matching neighbor = eBGP peer}
    Apply Rule 2
- ***Implication***
  - ***The MSDP peer address must be configured using the same IP address as the BGP peer!***

RPF Check Rule 1

- When MSDP peer = iBGP peer
  - Find "Best Path" to RP in BGP Tables
    - Search M-Table first then U-Table.
    - If no path to Originating RP found, RPF Fails
  - Note "BGP Neighbor" that advertised path
    - (i.e IP Address of BGP peer that sent us this path)
    - *Warning:*
      - *This is not the same as the Next-hop of the path!!!*
      - *iBGP peers normally do not set Next-hop = Self.*
      - *This is also not necessarily the same as the Router-ID!*
  - Rule 1 Test Condition:
    - MSDP Peer address = BGP Neighbor address?
      - If Yes, RPF Succeeds

Rule1: MSDP peer = iBGP peer



BGP Peer

MSDP Peer

SA Message

iBGP peer address = 172.16.3.1
(advertising best-path to RP)

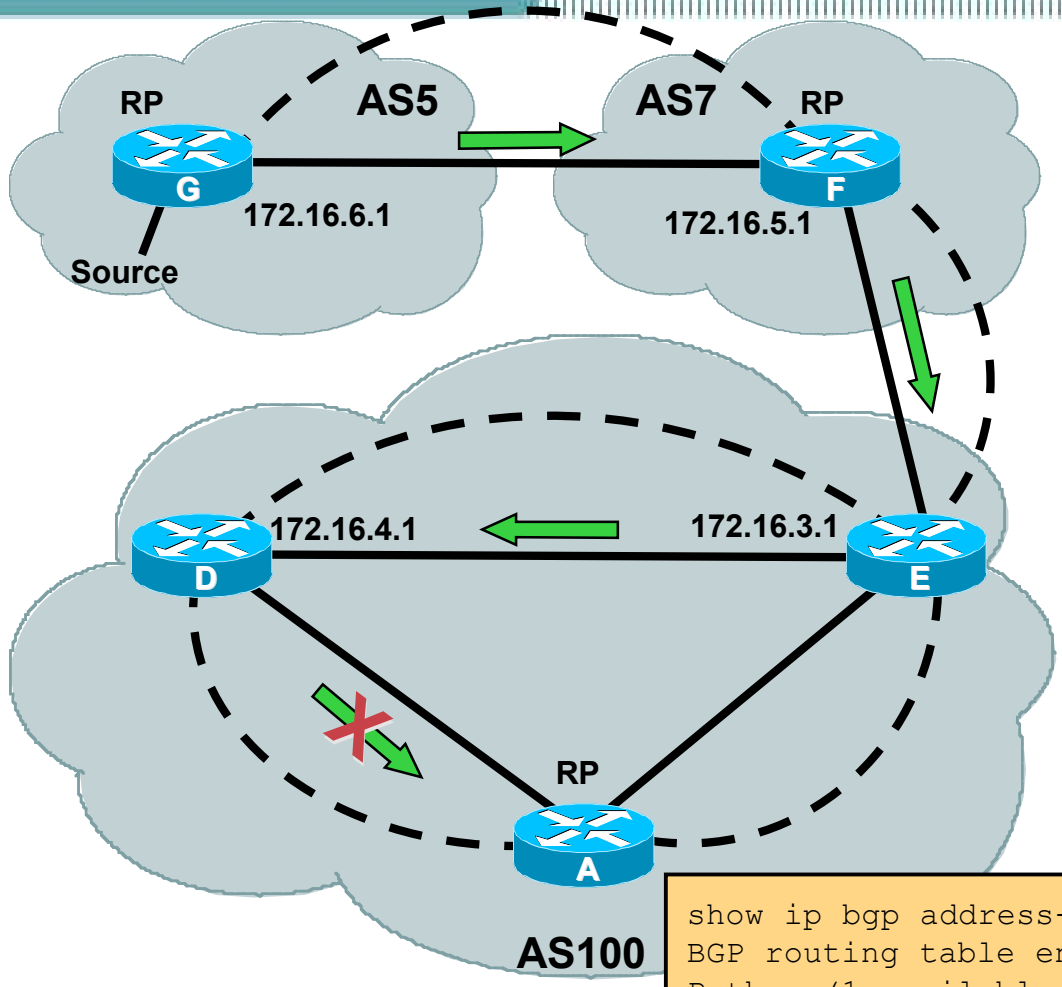
MSDP Peer address = 172.16.3.1

MSDP Peer address = iBGP Peer address

SA RPF Check Succeeds

```
show ip bgp address-family ipv4 multicast 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.3.1 (172.16.3.1)
```

Rule1: MSDP peer = iBGP peer



iBGP Peer address = 172.16.3.1
(advertising best-path to RP)

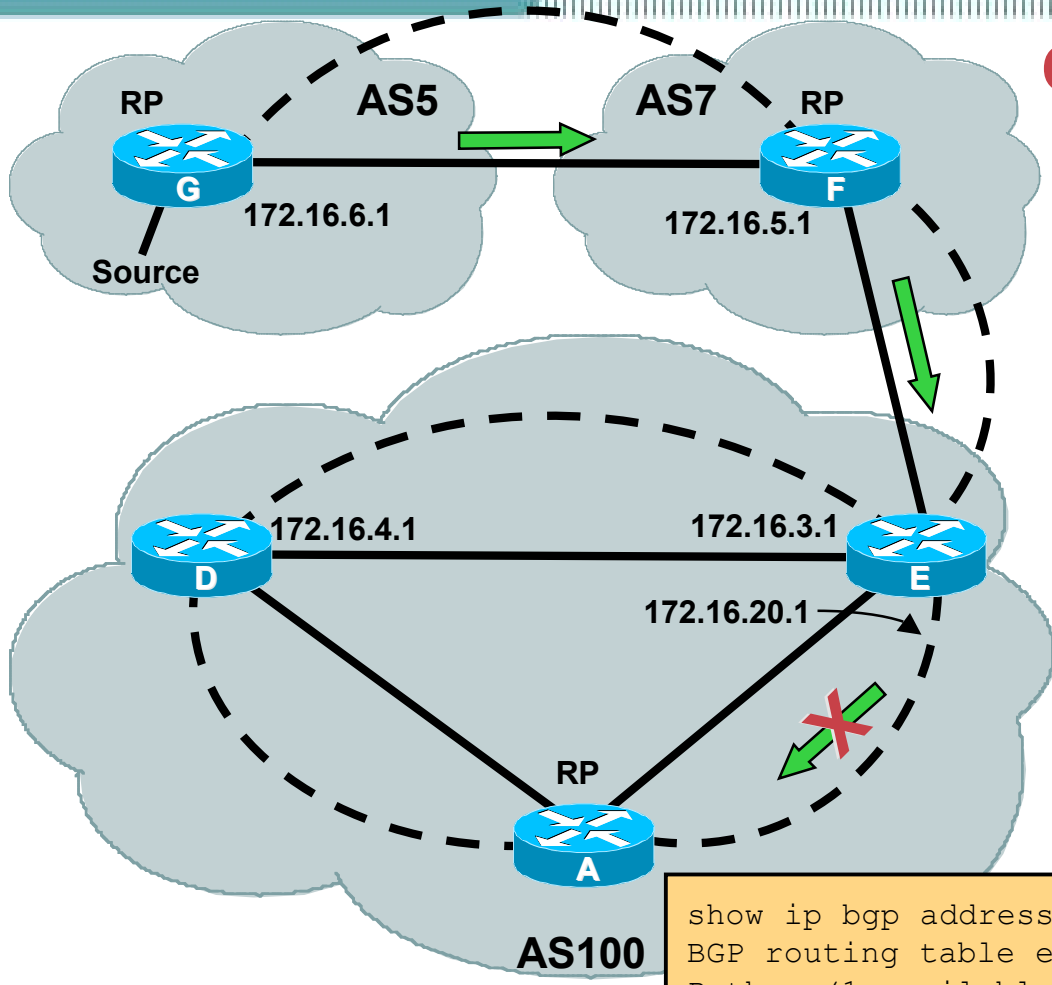
MSDP Peer address = 172.16.4.1

MSDP Peer address != iBGP Peer address

SA RPF Check Fails

```
show ip bgp address-family ipv4 multicast 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.3.1 (172.16.3.1)
```

Rule1: MSDP peer = iBGP peer



Common Mistake #1:

Failure to use same addresses for MSDP peers as iBGP peers!

iBGP Peer address = 172.16.3.1
(advertising best-path to RP)

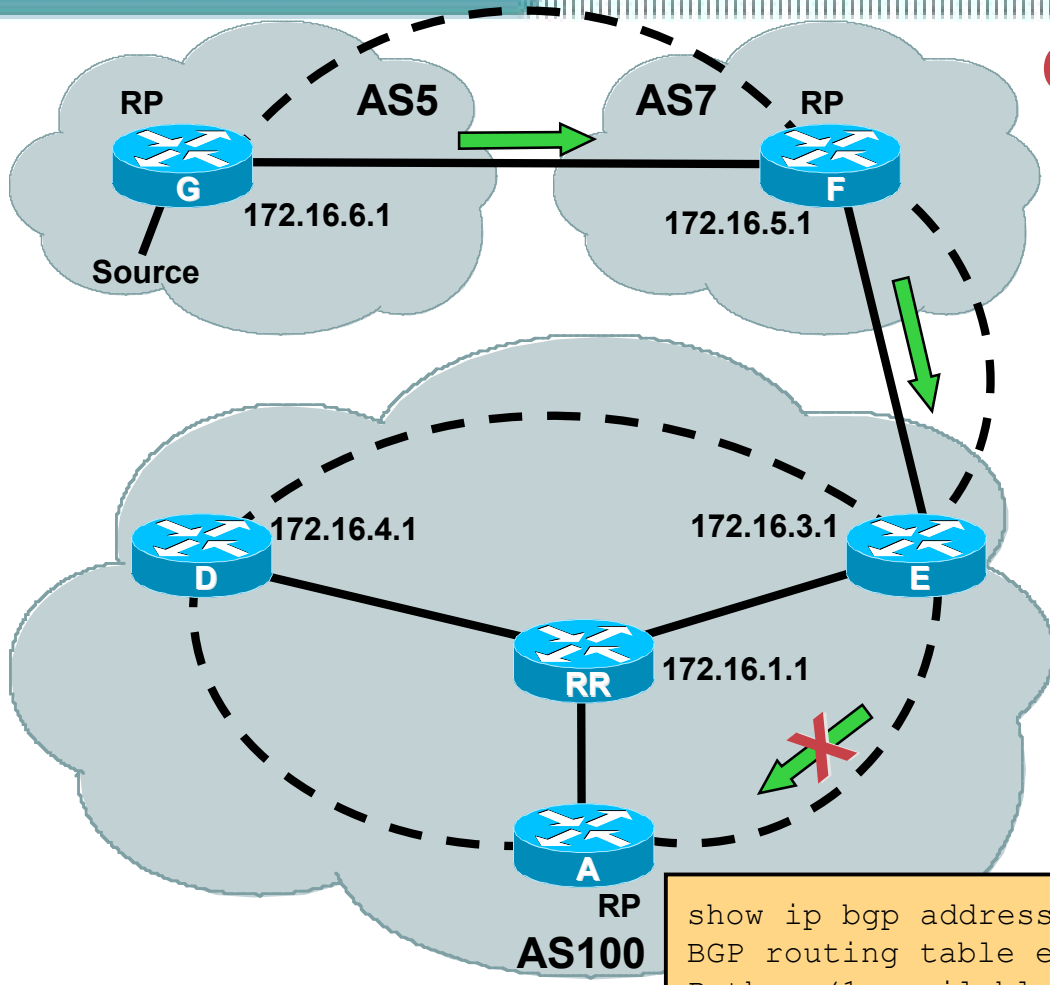
MSDP Peer address = 172.16.20.1

MSDP Peer address != iBGP Peer address

SA RPF Check Fails

```
show ip bgp address-family ipv4 multicast 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.3.1 (172.16.3.1)
```

Rule1: MSDP peer = iBGP peer



Common Mistake #2:

Failure to follow iBGP topology!
Can happen when RR's are used.

iBGP Peer address = 172.16.1.1
(advertising best-path to RP)

MSDP Peer address = 172.16.3.1

MSDP Peer address != iBGP Peer address

SA RPF Check Fails

```
show ip bgp address-family ipv4 multicast 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.1.1 (172.16.1.1)
```

RPF Check Rule 2

- When MSDP peer = eBGP peer
  - Find BGP “Best Path” to RP

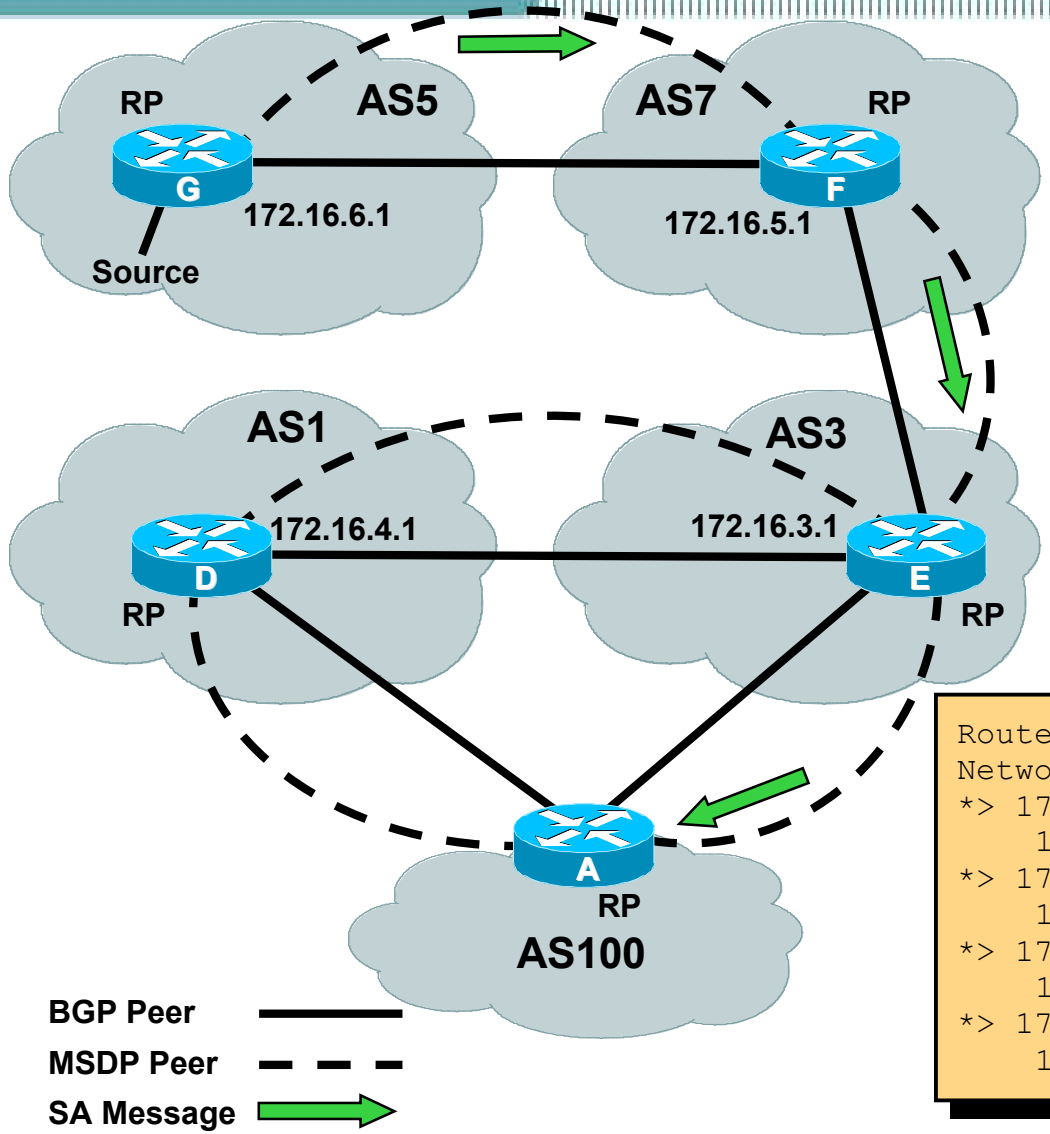
    Search M-Table first then U-Table.

    If no path to Originating RP found, RPF Fails
  - Rule 2 Test Condition:

    First AS in path to the RP = AS of eBGP peer?

    If Yes, RPF Succeeds

Rule2: MSDP peer = eBGP peer



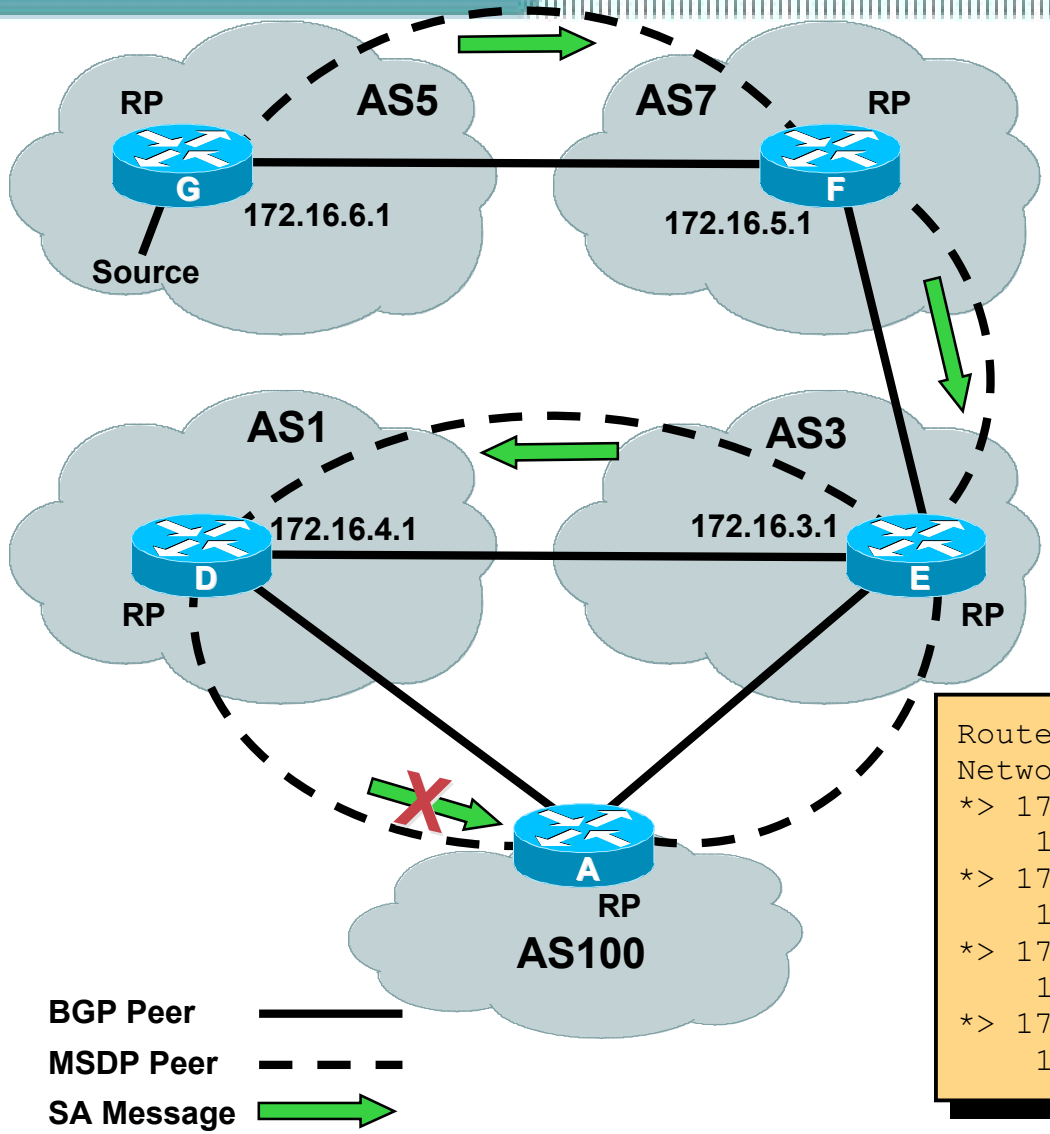
First-AS in best-path to RP = 3
AS of MSDP Peer = 3

First-AS in best-path to RP = AS of eBGP Peer

SA RPF Check Succeeds

```
Router A's ipv4 multicast BGP Table
Network              Next Hop       Path
*> 172.16.3.0/24     172.16.3.1     3 i
    172.16.3.0/24     172.16.4.1     1 3 i
*> 172.16.4.0/24     172.16.4.1     1 i
    172.16.4.0/24     172.16.3.1     3 1 i
*> 172.16.5.0/24     172.16.3.1     3 7 i
    172.16.5.0/24     172.16.4.1     1 3 7 i
*> 172.16.6.0/24     172.16.3.1     3 7 5 i
    172.16.6.0/24     172.16.4.1     1 3 7 5 i
```

Rule2: MSDP peer = eBGP peer



First-AS in best-path to RP = 3
AS of eBGP Peer = 1

First-AS in best-path to RP != AS of eBGP Peer

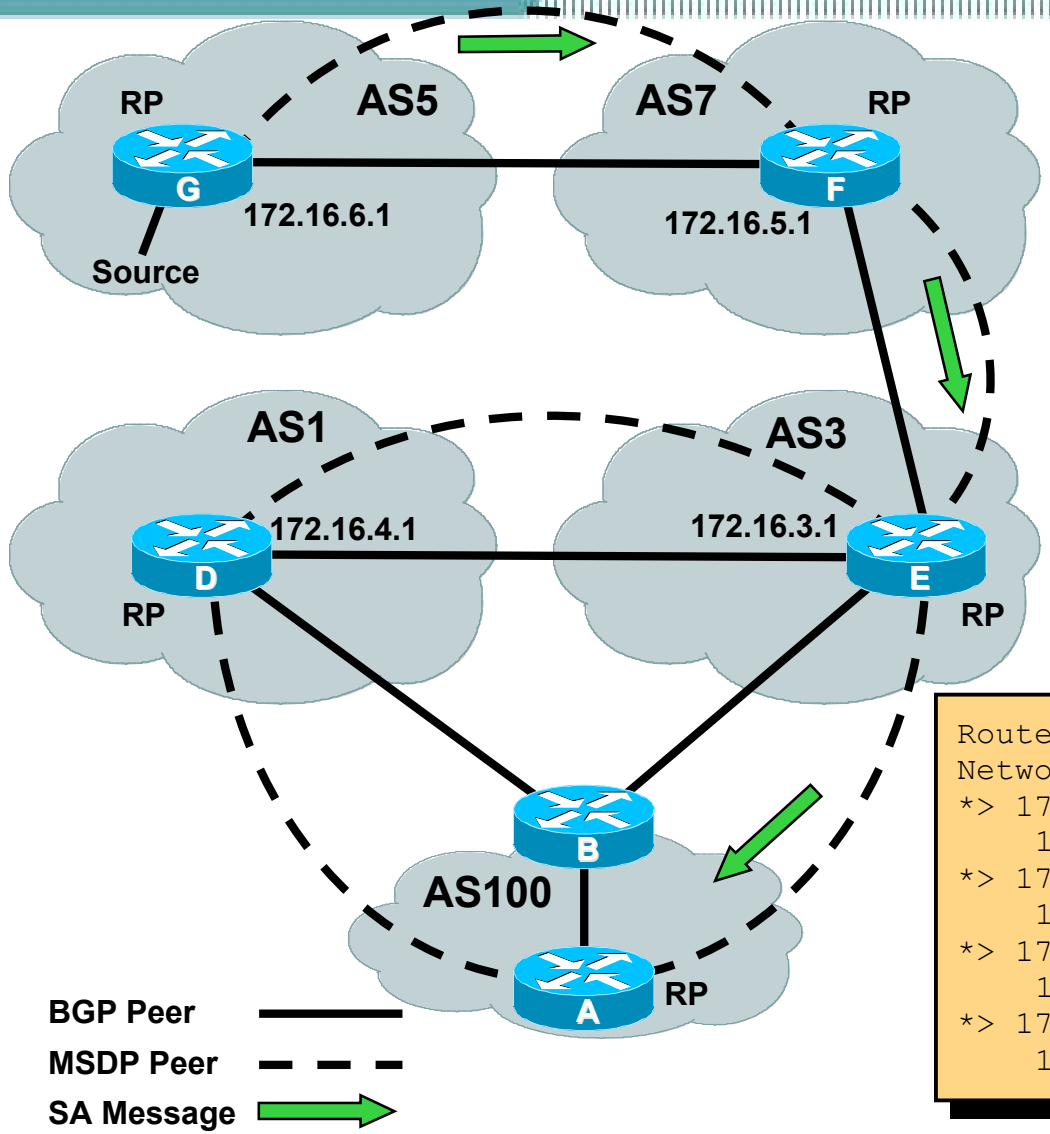
SA RPF Check Fails!

```
Router A's ipv4 multicast BGP Table
Network              Next Hop       Path
*> 172.16.3.0/24     172.16.3.1     3 i
    172.16.3.0/24     172.16.4.1     1 3 i
*> 172.16.4.0/24     172.16.4.1     1 i
    172.16.4.0/24     172.16.3.1     3 1 i
*> 172.16.5.0/24     172.16.3.1     3 7 i
    172.16.5.0/24     172.16.4.1     1 3 7 i
*> 172.16.6.0/24     172.16.3.1     3 7 5 i
    172.16.6.0/24     172.16.4.1     1 3 7 5 i
```

RPF Check Rule 3

- When MSDP peer != BGP peer
  - Find BGP "Best Path" to RP
    - Search M-Table first then U-Table.
      - If no path to Originating RP found, RPF Fails
  - Find BGP "Best Path" to MSDP peer
    - Search M-Table first then U-Table.
      - If no path to sending MSDP Peer found, RPF Fails
  - Note AS of sending MSDP Peer
    - Origin AS (last AS) in AS-PATH to MSDP Peer
  - Rule 3 Test Condition:
    - First AS in path to RP = Sending MSDP Peer AS ?
      - If Yes, RPF Succeeds

Rule3: MSDP peer != BGP peer



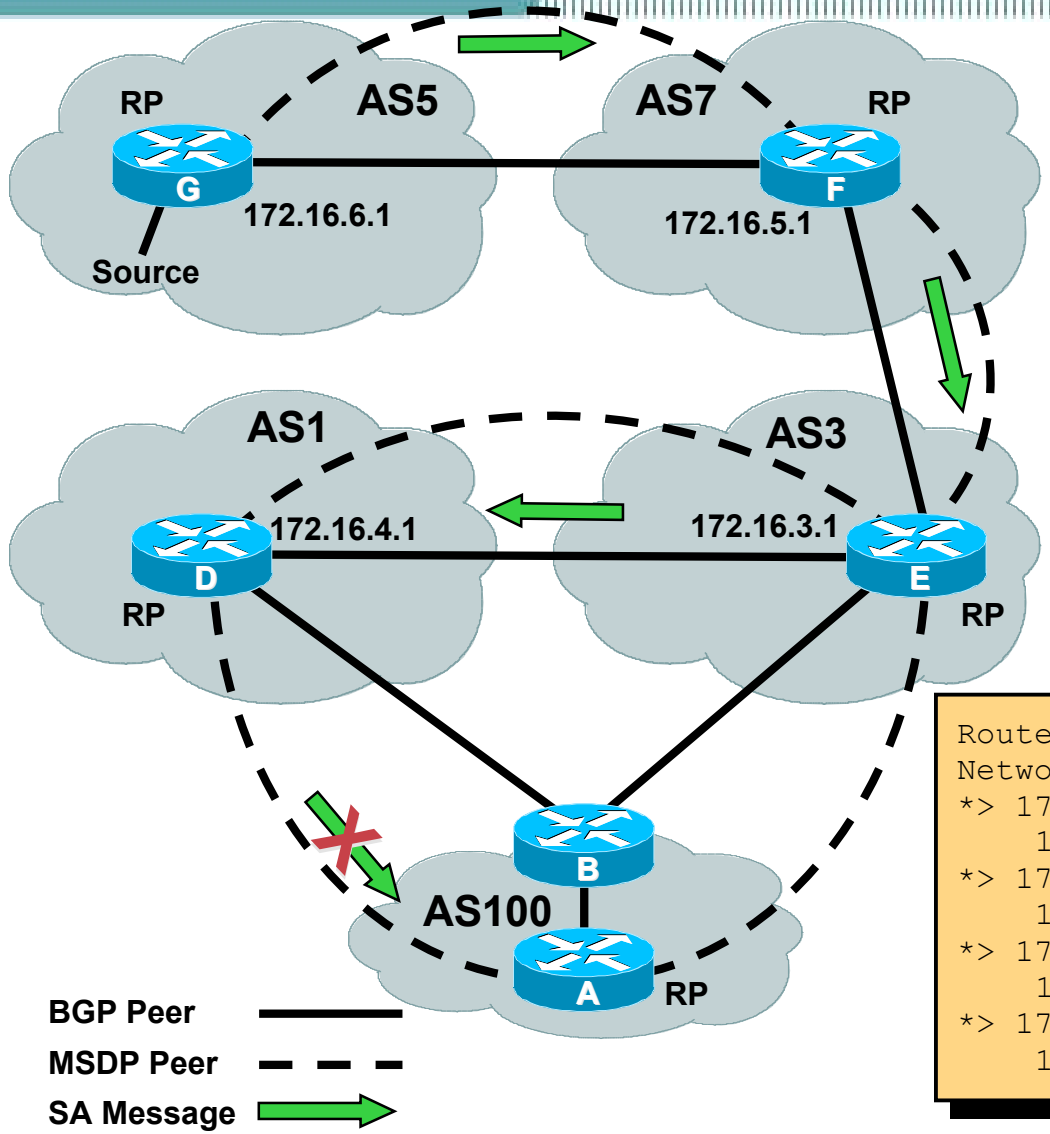
First-AS in best-path to RP = 3
AS of MSDP Peer = 3

First-AS in best-path to RP = AS of MSDP Peer

SA RPF Check Succeeds

```
Router A's ipv4 multicast BGP Table
Network              Next Hop        Path
*> 172.16.3.0/24     172.16.3.1      3 i
    172.16.3.0/24     172.16.4.1     1 3 i
*> 172.16.4.0/24     172.16.4.1      1 i
    172.16.4.0/24     172.16.3.1     3 1 i
*> 172.16.5.0/24     172.16.3.1      3 7 i
    172.16.5.0/24     172.16.4.1     1 3 7 i
*> 172.16.6.0/24     172.16.3.1      3 7 5 i
    172.16.6.0/24     172.16.4.1     1 3 7 5 i
```

Rule3: MSDP peer != BGP peer



First-AS in best-path to RP = 3
AS of MSDP Peer = 1

First-AS in best-path to RP != AS of MSDP Peer

SA RPF Check Fails

```
Router A's ipv4 multicast BGP Table
Network              Next Hop       Path
*> 172.16.3.0/24     172.16.3.1     3 i
    172.16.3.0/24     172.16.4.1     1 3 i
*> 172.16.4.0/24     172.16.4.1     1 i
    172.16.4.0/24     172.16.3.1     3 1 i
*> 172.16.5.0/24     172.16.3.1     3 7 i
    172.16.5.0/24     172.16.4.1     1 3 7 i
*> 172.16.6.0/24     172.16.3.1     3 7 5 i
    172.16.6.0/24     172.16.4.1     1 3 7 5 i
```

MSDP Mesh-Groups

- Optimises SA flooding.
  - Useful when 2 or more peers are in a group.
  - Requires full mesh of mesh group peers.
- Reduces amount of SA traffic in the net.
  - SA's not flooded to other mesh-group peers.
- Suspends RPF check of SA messages.
  - When received from a mesh-group peer.
  - SA's always accepted from mesh-group peers.
  - Eliminates need for BGP.

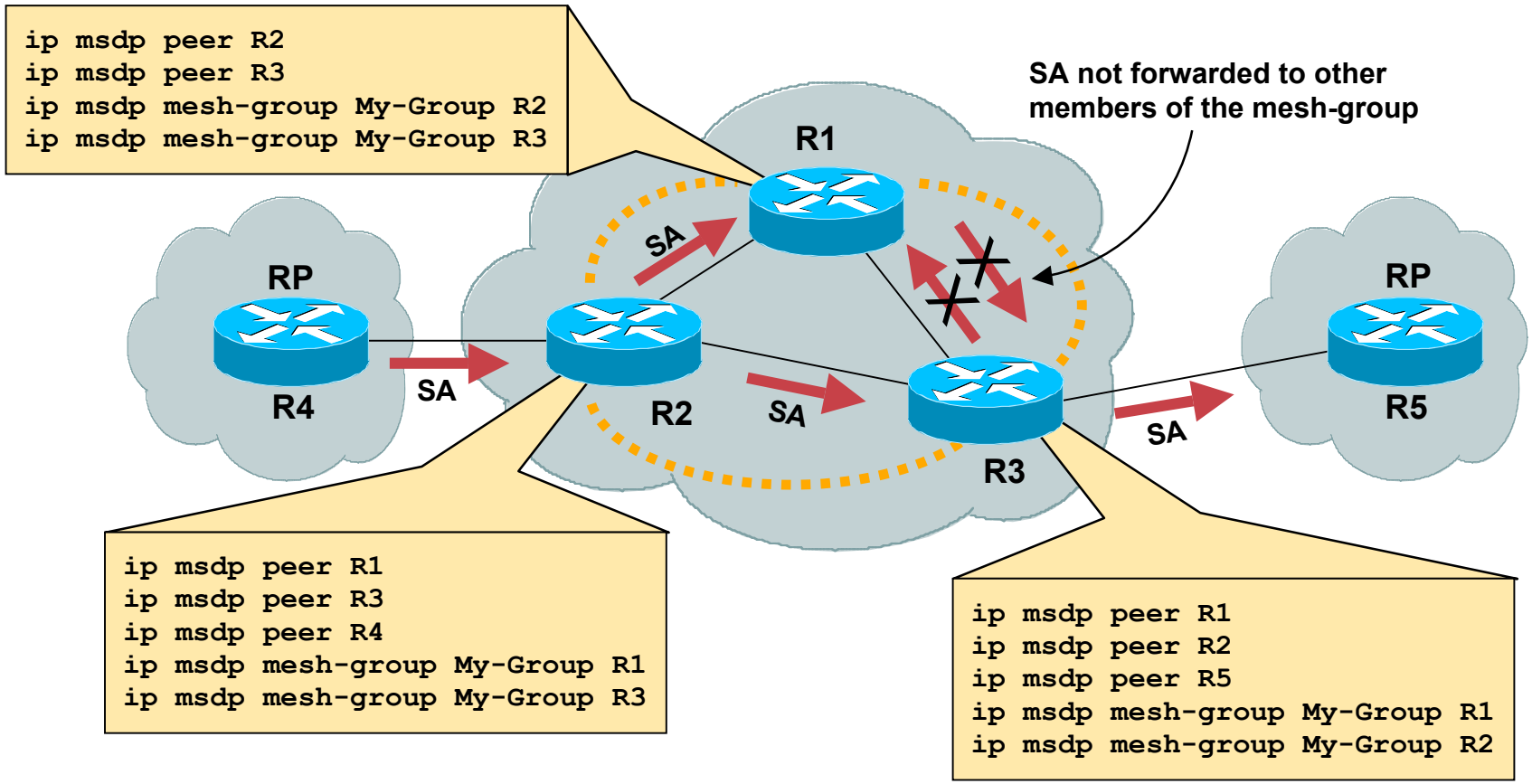
MSDP Mesh-Groups

- **Configured with:**

```
ip msdp mesh-group <name> <peer-address>
```

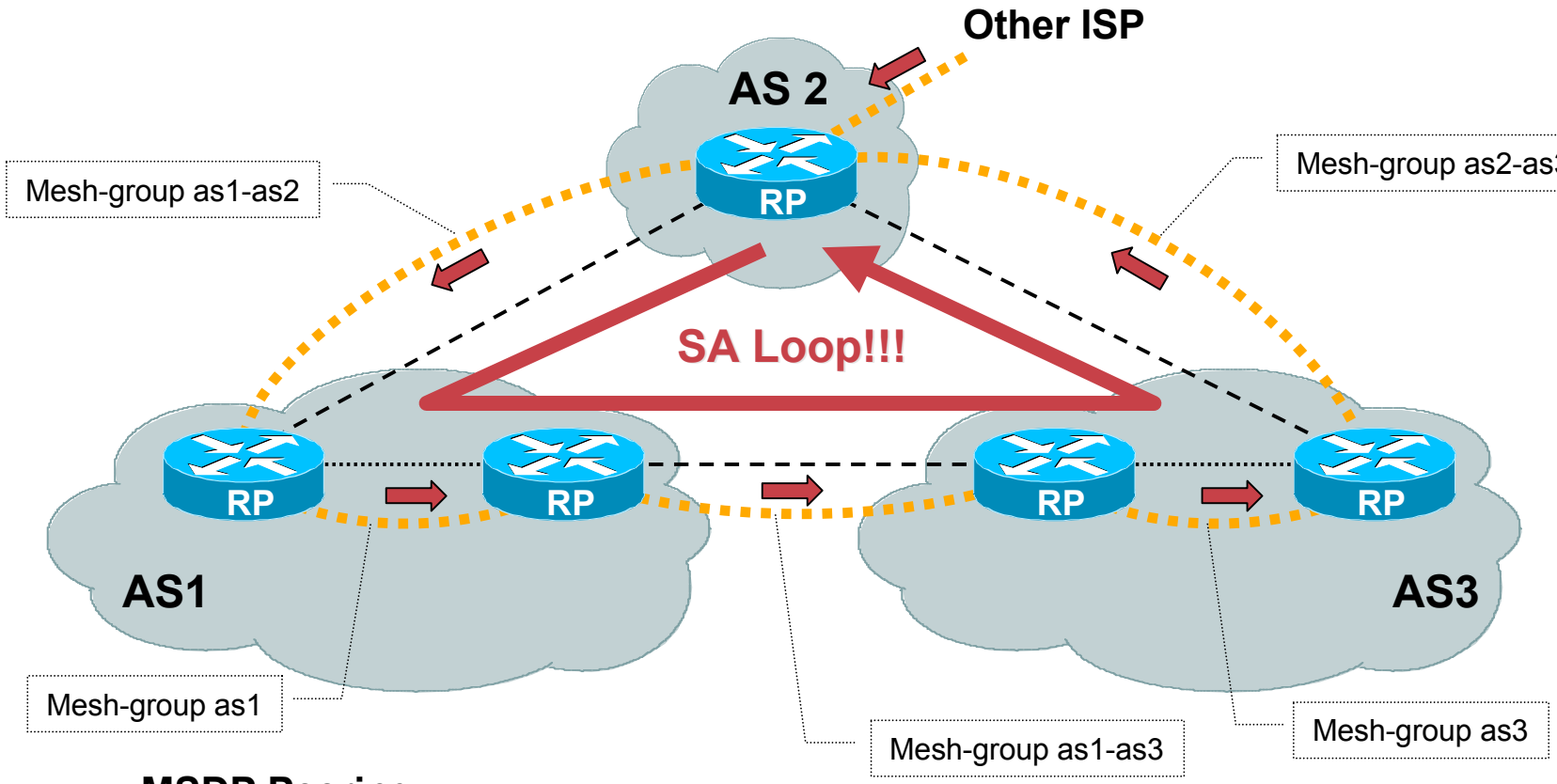
- **Peers in the mesh-group must be fully meshed.**
- **Multiple mesh-groups per router are supported.**

MSDP Mesh-Group Example



Avoid Mesh-Group Loops!!!

WARNING: There is no RPF check between Mesh-groups!!!



DON'T DO THIS!!!!

MSDP Mroute Flags

New ‘mroute’ Flags for MSDP

```
sj-mbone#show ip mroute summary
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, 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 - Advertised via MSDP
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.2.246.13), 5d17h/00:02:59, RP 171.69.10.13, flags: S
  (171.69.185.51, 224.2.246.13), 3d17h/00:03:29,   flags: TA
  (128.63.58.45, 224.2.246.13), 00:02:16/00:00:43, flags: M
  (128.63.58.54, 224.2.246.13), 00:01:16/00:01:43, flags: M
```

“M” flag indicates source was learned via MSDP

“A” flag indicates source is a *candidate* for advertisement by MSDP

MSDP Enhancements

- **New IOS command**

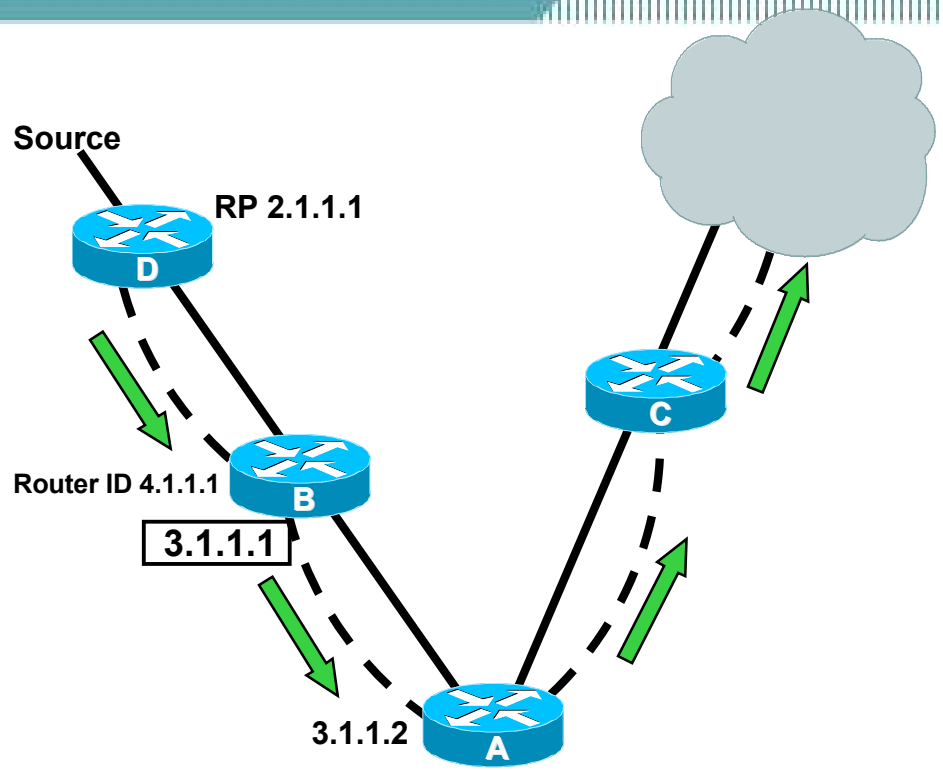
```
ip msdp new-rpf-rules
```

  - **MSDP SA RPF check using IGP**
  - **Accept SA's from BGP NEXT HOP**
  - **Accept SA's from closest peer along the best path to the originating RP**
  - **"show ip msdp rpf"**

MSDP RPF check using IGP

- **When MSDP peer = IGP peer (No BGP)**
  - **Find best IGP route to RP**
    - **Search URIB**
      - **If route to Originating RP found and:**
        - **If IGP next hop (or advertiser) address for RP is the MSDP peer and in UP state, then that is the RPF peer.**
        - **If route not found: Fall through to the next rule.**

IGP Rule: MSDP peer = IGP peer (Next hop)



MSDP Peer = 3.1.1.1

IGP next hop to originating RP = 3.1.1.1

IGP next hop to originating RP = MSDP peer

SA RPF Check Succeeds

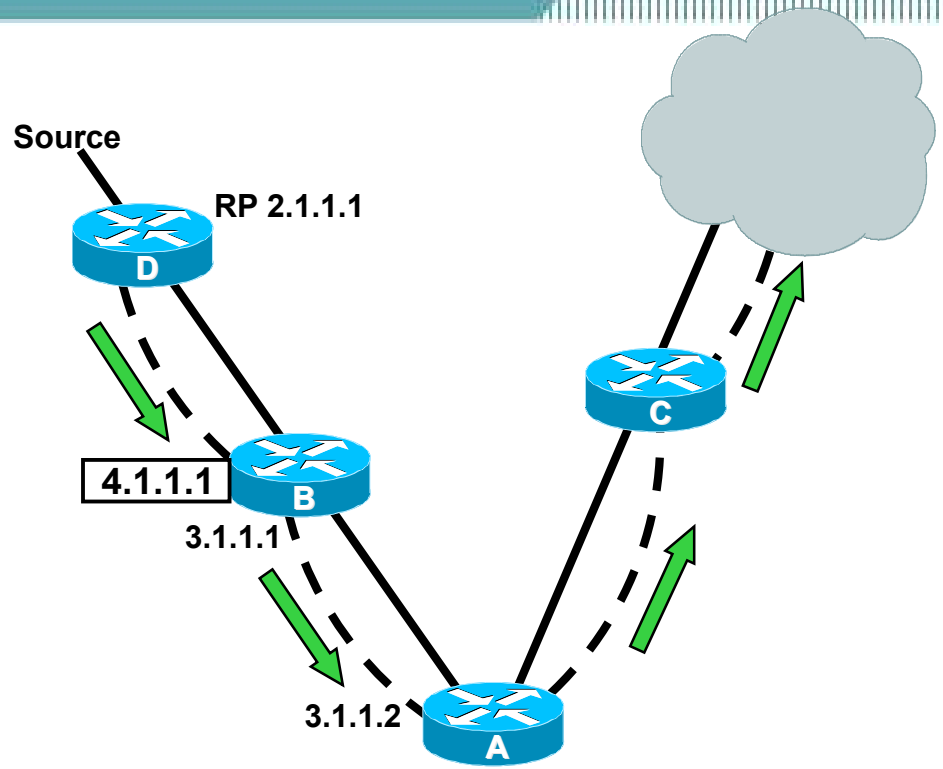
OSPF neighbor

MSDP Peer

SA Message

```
RouterA#show ip route 2.1.1.1
Routing entry for 2.1.1.0/24
  Known via "ospf 1", distance 110, metric 20, type intra area
  Last update from 3.1.1.1 on Ethernet2, 00:35:10 ago
  Routing Descriptor Blocks:
  * 3.1.1.1, from 4.1.1.1, 00:35:10 ago, via Ethernet2
      Route metric is 20, traffic share count is 1
```

IGP Rule: MSDP peer = IGP peer (Advertiser)



MSDP Peer = 4.1.1.1

IGP next hop to originating RP = ~~3.1.1.1~~

IGP advertiser to originating RP = 4.1.1.1

IGP advertiser to originating RP = MSDP peer

SA RPF Check Succeeds

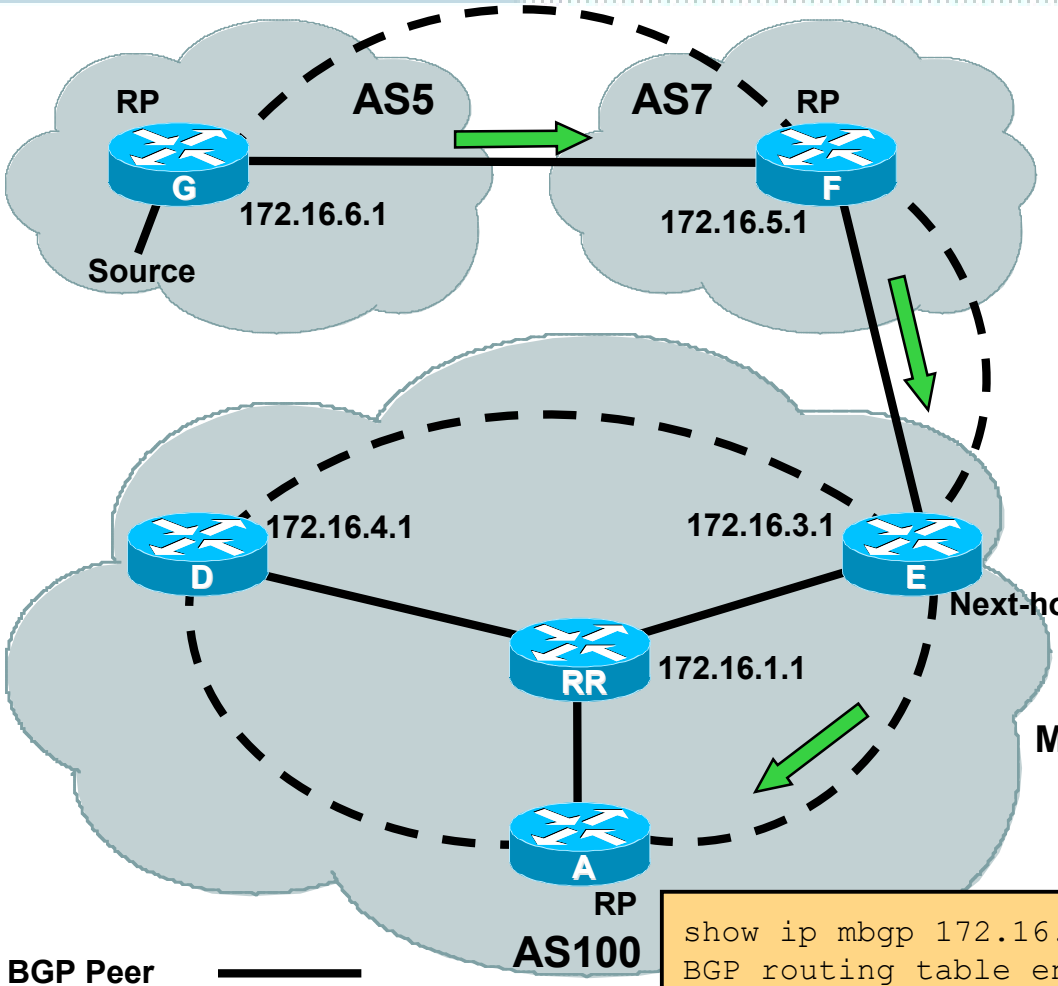
OSPF neighbor

MSDP Peer

SA Message

```
RouterA#show ip route 2.1.1.1
Routing entry for 2.1.1.0/24
  Known via "ospf 1", distance 110, metric 20, type intra area
  Last update from 3.1.1.1 on Ethernet2, 00:35:10 ago
  Routing Descriptor Blocks:
  * 3.1.1.1, from 4.1.1.1, 00:35:10 ago, via Ethernet2
      Route metric is 20, traffic share count is 1
```

SA's accepted from Next Hop



i(m)BGP Peer address = 172.16.1.1
(Advertiser of next hop)

MSDP Peer address = 172.16.3.1

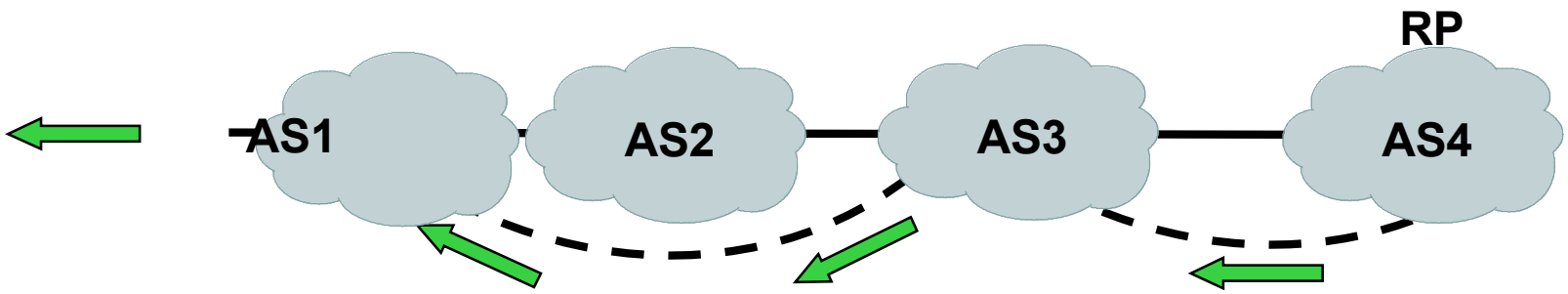
But, BGP next hop = 172.16.3.1

MSDP Peer address = BGP next hop address

SA RPF Check Succeeds

```
show ip mbgp 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.3.1 (metric 68096) from 172.16.1.1 (172.16.1.1)
```

Accept SA along RPF path



Existing Rule: If first AS in best path to the RP != MSDP peer
RPF Fails

New code: Choose peer in CLOSEST AS along best AS path to the RP.
Loosens rule a bit.

RPF Succeeds.

BGP Peer
MSDP Peer
SA Message

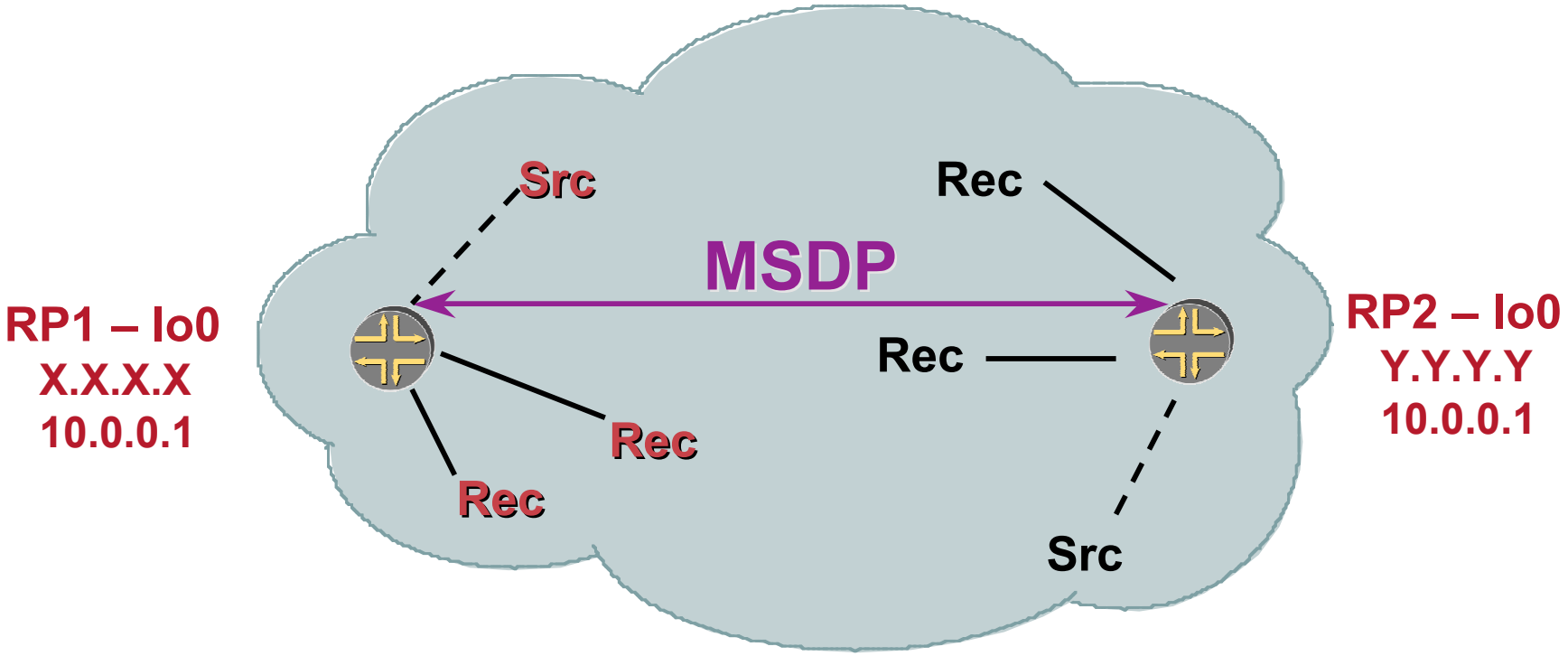
New MSDP RPF command


```
Router-A# show ip msdp rpf 2.1.1.1
RPF peer information for Router-B (2.1.1.1)
  RPF peer: Router-C (3.1.1.1)
  RPF route/mask: 2.1.1.0/24
  RPF rule: Peer is IGP next hop of best route
  RPF type: unicast (ospf 1)
```

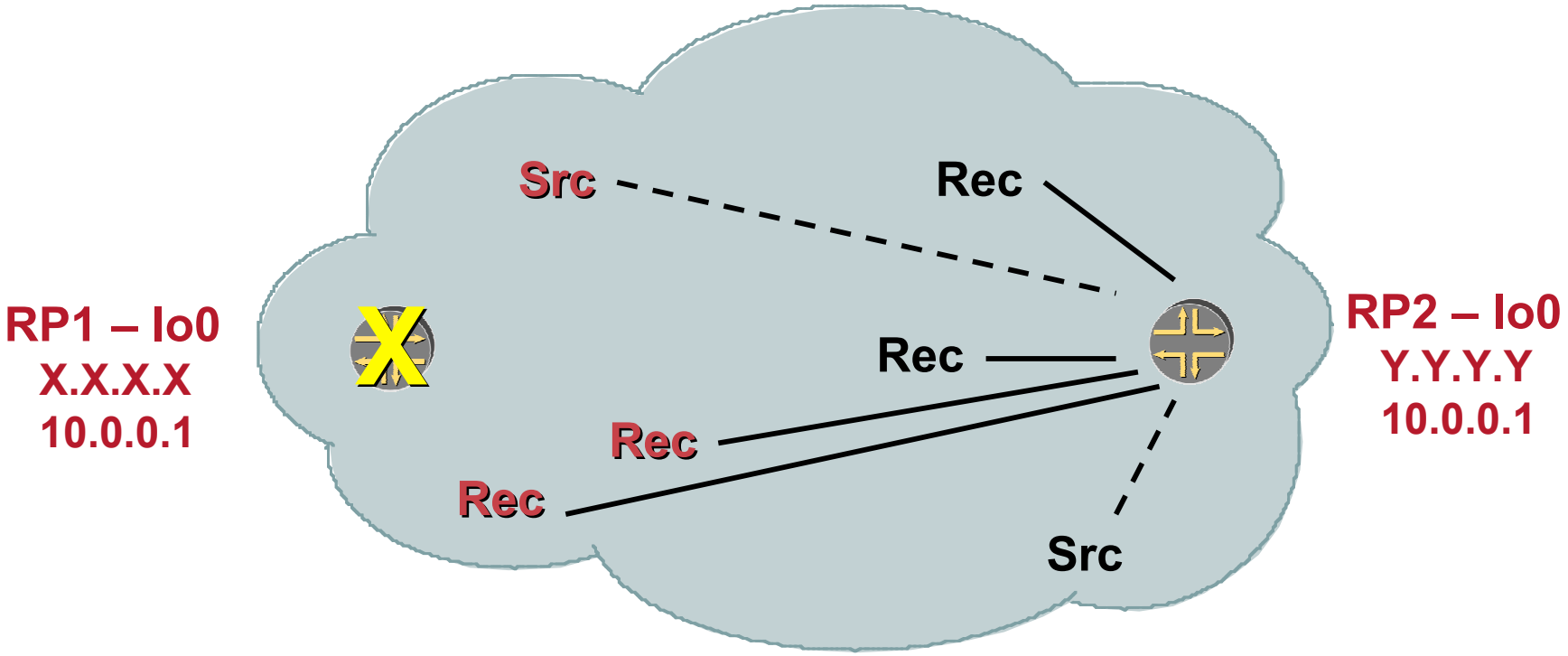
Anycast-RP

- draft-ietf-mboned-anycast-rp-08.txt
- Within a domain, deploy more than one RP for the same group range
- Sources from one RP are known to other RPs using MSDP
- Give each RP the same /32 IP address
- Sources and receivers use closest RP, as determined by the IGP
- Used intra-domain to provide redundancy and RP load sharing, when an RP goes down, sources and receivers are taken to new RP via unicast routing
  - Fast convergence!

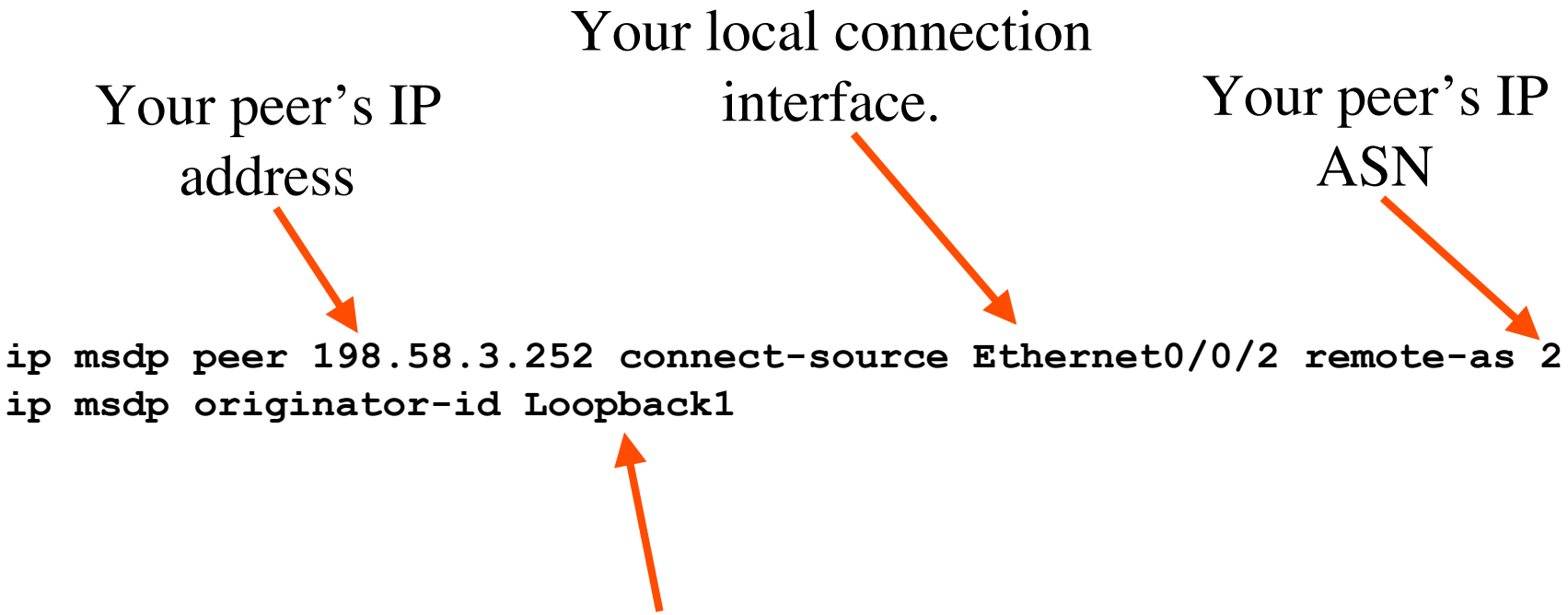
Anycast-RP



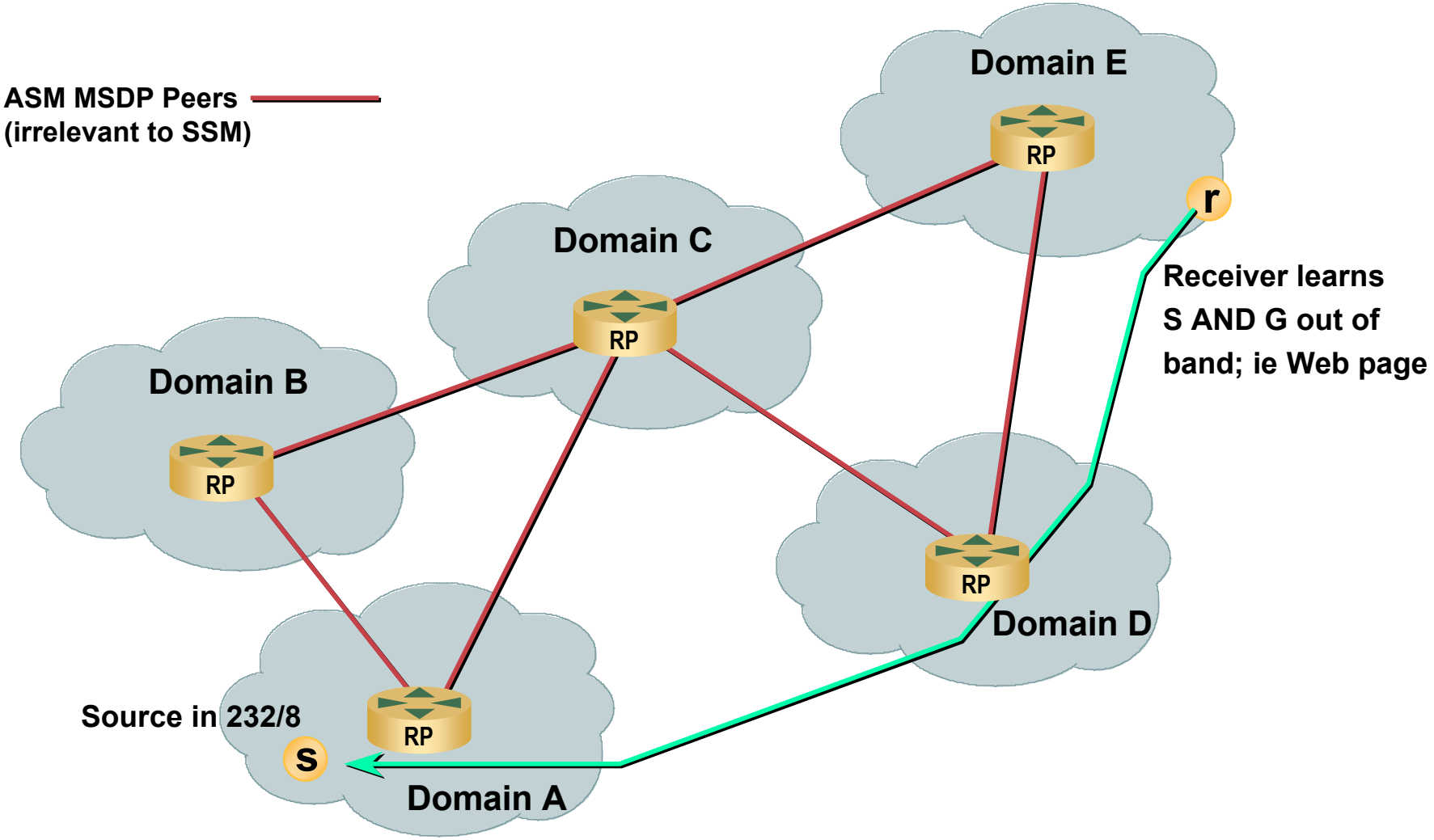
Anycast-RP



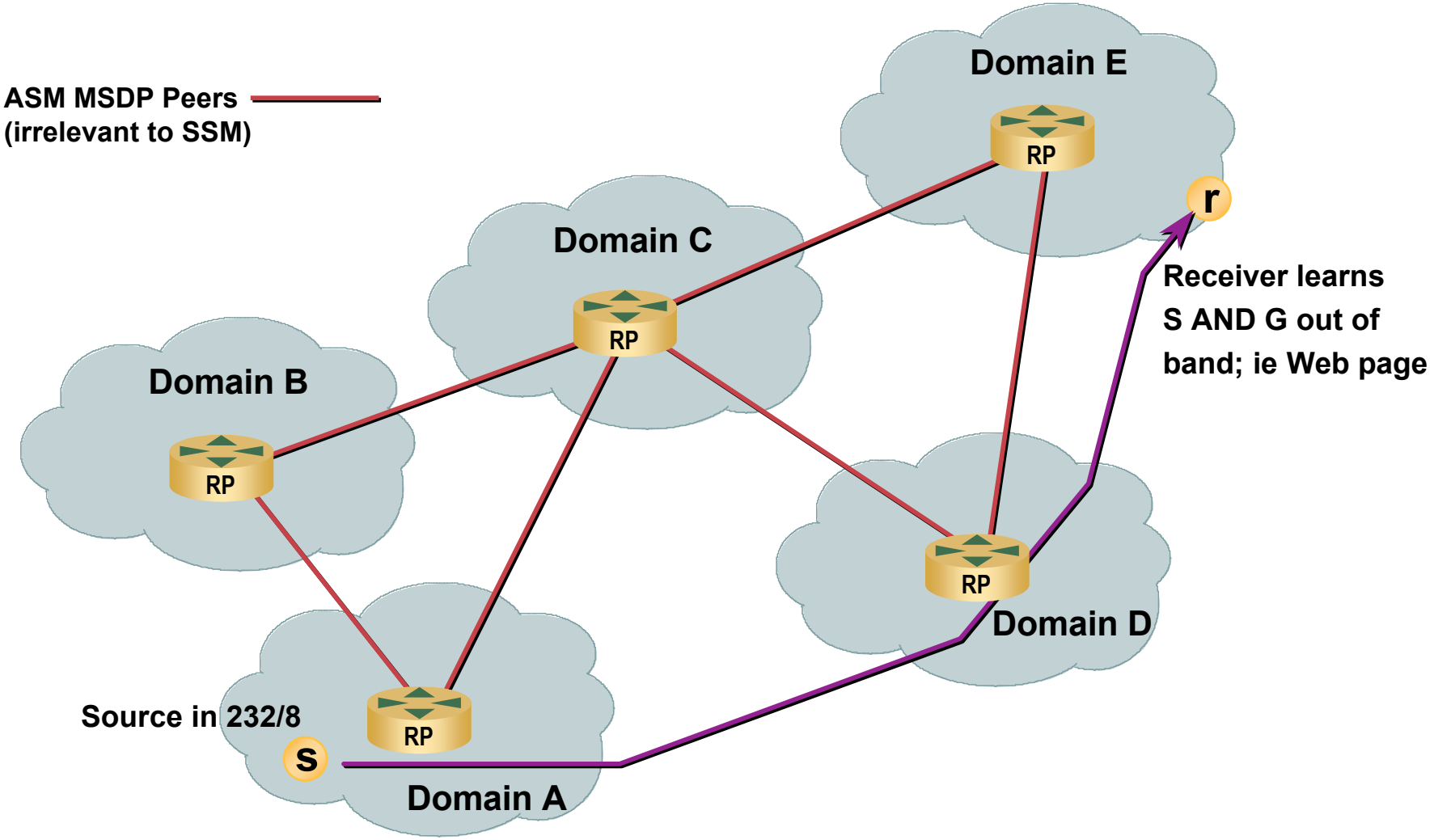
MSDP Configuration



MSDP wrt SSM – Unnecessary!



MSDP wrt SSM – Unnecessary!



Agenda

Cisco.com

- Introduction
- Multicast addressing
- Group Membership Protocol
- PIM-SM / SSM
- MSDP
- **MBGP**
- Summary

MBGP—Multiprotocol BGP

- MBGP overview
- MBGP capability negotiation
- MBGP NLRI exchange
- Configuration guidelines

MBGP

- Multiprotocol Extensions to BGP (RFC 2283).
- Tag unicast prefixes as multicast source prefixes for intra-domain mcast routing protocols to do RPF checks.
- WHY? Allows for interdomain RPF checking where unicast and multicast paths are non-congruent.
- DO I REALLY NEED IT?
  - YES, if:
    - ISP to ISP peering
    - Multiple-homed networks
  - NO, if:
    - You are single-homed

MBGP Overview

- **MBGP: Multiprotocol BGP (aka multicast BGP in multicast networks)**
  - **Defined in RFC 2283 (extensions to BGP)**
  - **Can carry different route types for different purposes**
    - **Unicast**
    - **Multicast**
  - **Both route types carried in same BGP session**
  - **Does not propagate multicast state information**
  - **Same path selection and validation rules**
    - **AS-Path, LocalPref, MED, …**

MBGP Overview

- New multiprotocol attributes
  - MP_REACH_NLRI
  - MP_UNREACH_NLRI
- MP_REACH_NLRI and MP_UNREACH_NLRI
  - Address Family Information (AFI) = 1 (IPv4)

    Sub-AFI = 1 (NLRI is used for unicast)

    Sub-AFI = 2 (NLRI is used for multicast RPF check)

    Sub-AFI = 3 (NLRI is used for both unicast and multicast RPF check)
- Allows for different policies between multicast and unicast

MBGP—Capability Negotiation

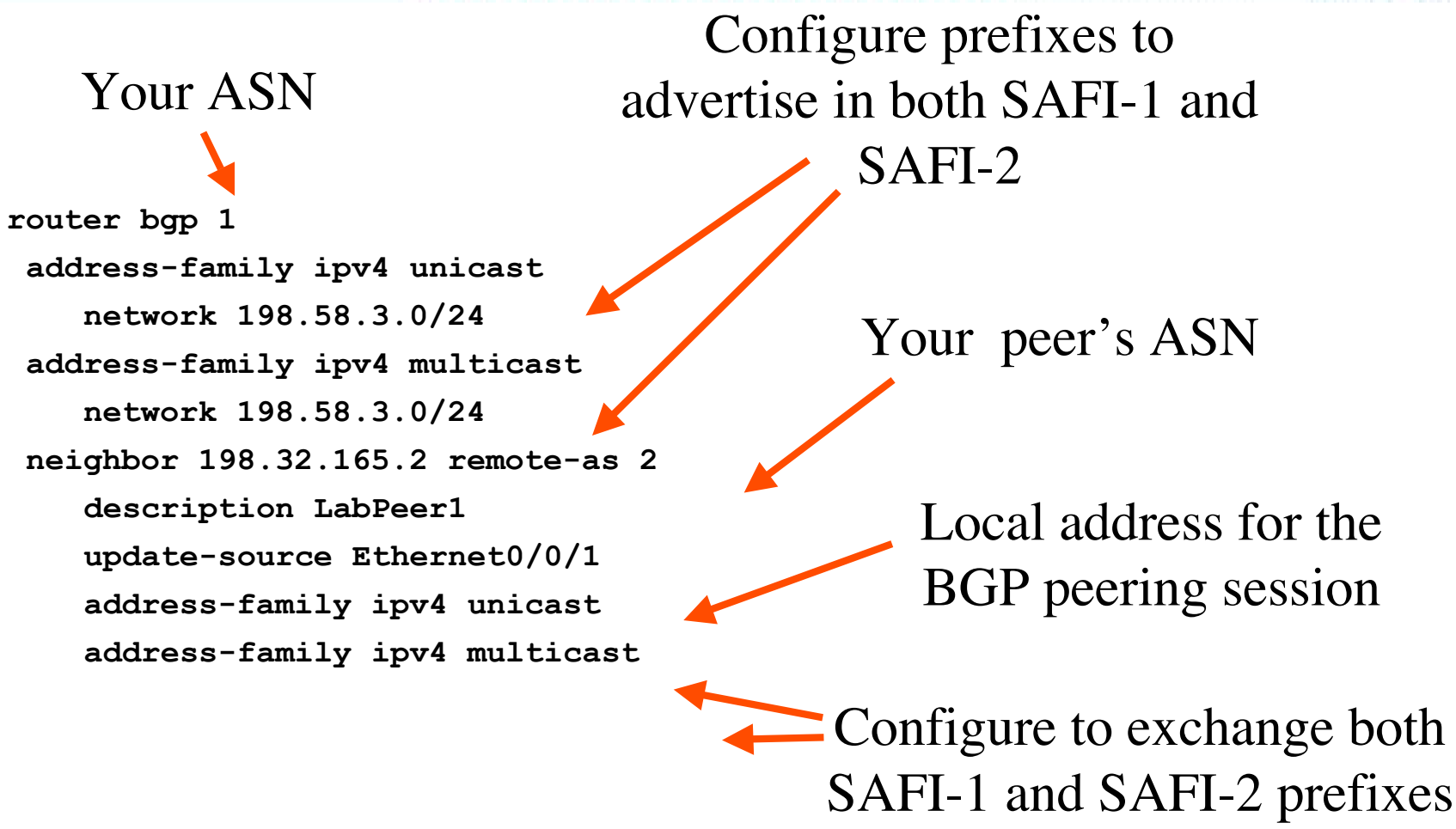
Cisco.com

- BGP routers establish BGP sessions through the OPEN message
- OPEN message contains optional parameters
- BGP session is terminated if OPEN parameters are not recognised
- New parameter: CAPABILITIES
  - Multiprotocol extension
  - Multiple routes for same destination
- Configures router to negotiate either or both NLRI
  - If neighbor configures both or subset, common NRLI is used in both directions
  - If there is no match, notification is sent and peering doesn’t come up
  - If neighbor doesn’t include the capability parameters in open, session backs off and reopens with no capability parameters
    - Peering comes up in unicast-only mode

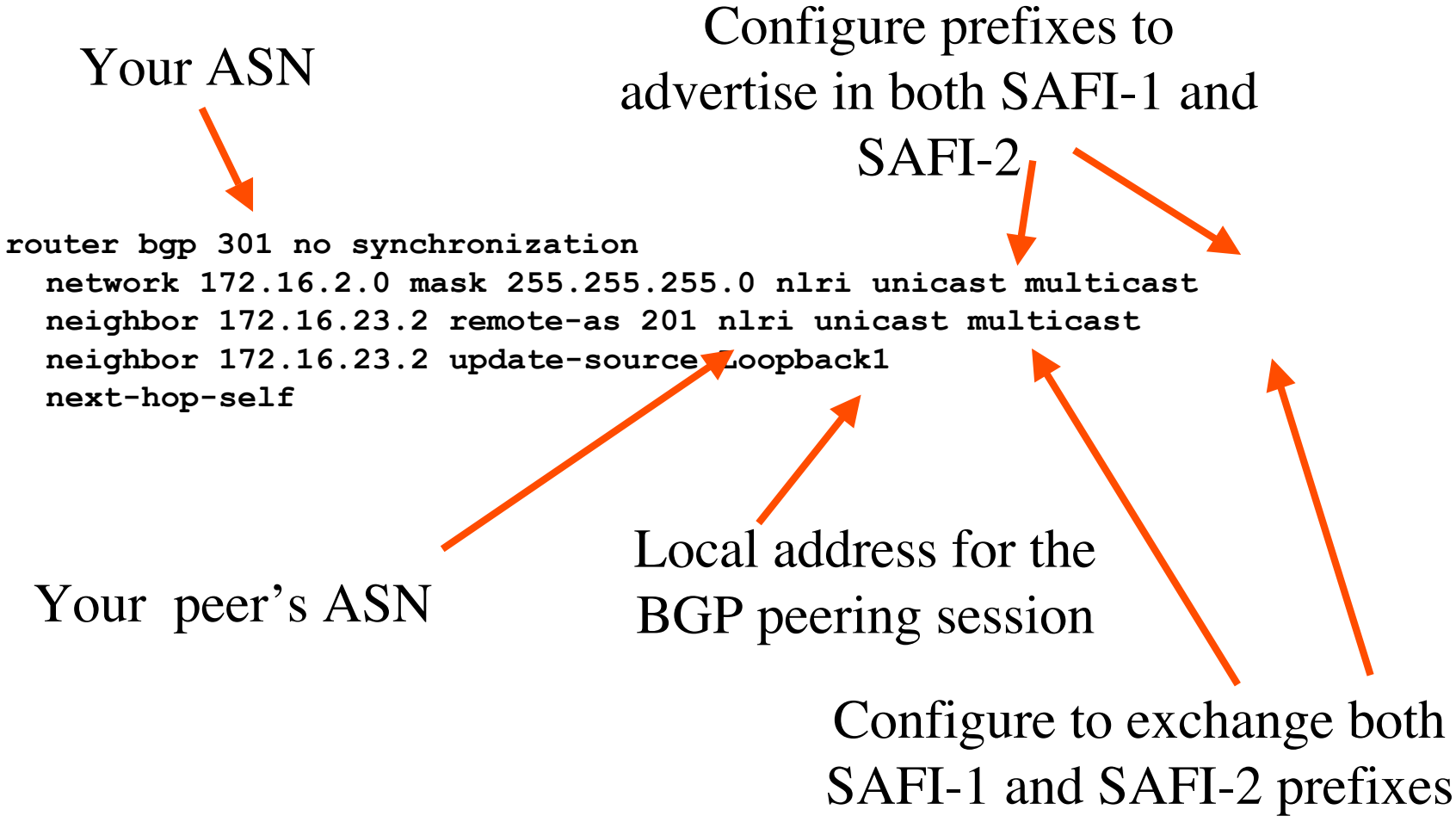
MBGP—Summary

- Solves part of inter-domain problem
  - Can exchange unicast prefixes for multicast RPF checks
  - Uses standard BGP configuration knobs
  - Permits separate unicast and multicast topologies if desired
- Still must use PIM to:
  - Build distribution trees
  - Actually forward multicast traffic
  - PIM-SM recommended

MBGP configuration (new)



MBGP configuration (original)

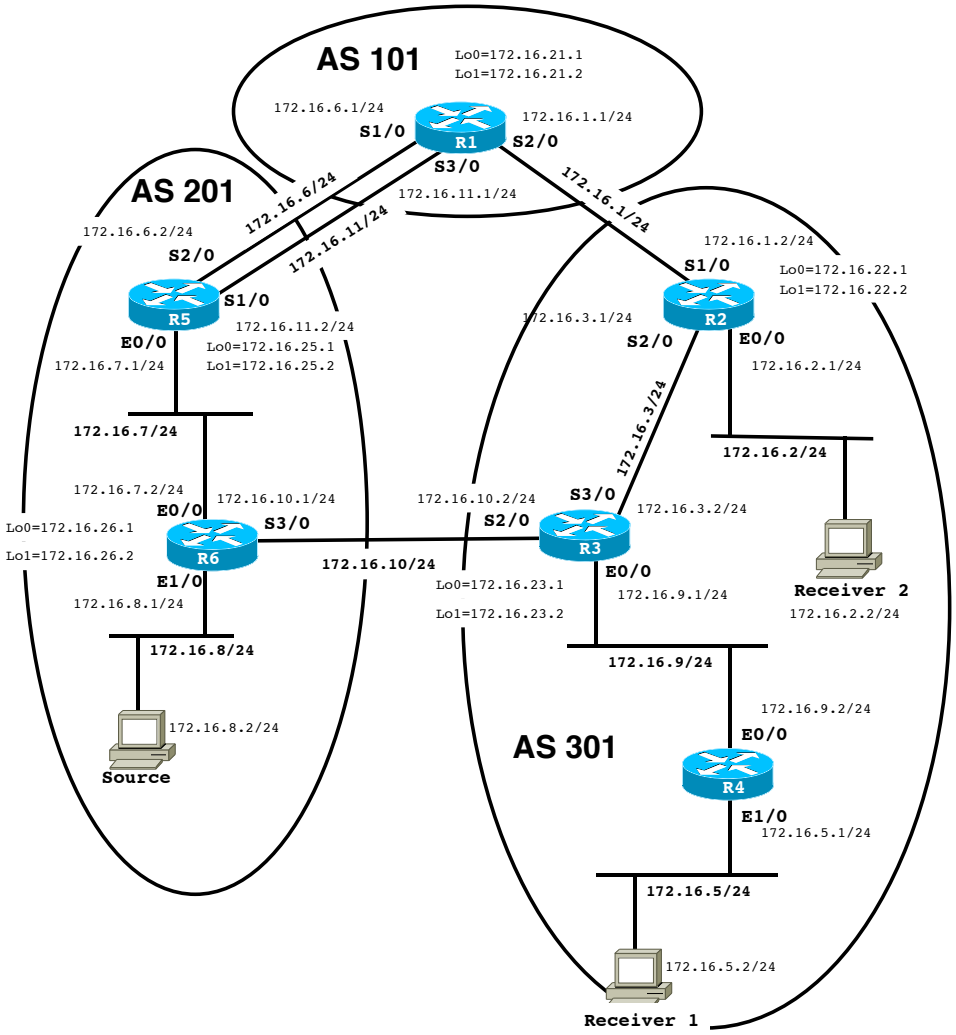


LAB #2
Interdomain Multicast

Cisco.com

- Do not launch lab until instructred to do so.
- Lab templates or cfgs: Interdomain-Multicast
- Refer to your lab handout

LAB #2
Interdomain Multicast



Agenda

Cisco.com

- Introduction
- Multicast addressing
- Group Membership Protocol
- PIM-SM / SSM
- MSDP
- MBGP
- **Summary**

The Soup

- **IGMP - Internet Group Management Protocol is used by hosts and routers to tell each other about group membership.**
- **PIM-SM - Protocol Independent Multicast-Sparse Mode is used to propagate forwarding state between routers.**
- **SSM - Source Specific Multicast utilizes a subset of PIM's functionality to guaranty source-only trees in the 232/8 range.**
- **MBGP - Multiprotocol Border Gateway Protocol is used to exchange routing information for interdomain RPF checking.**
- **MSDP - Multicast Source Discovery Protocol is used to exchange ASM active source information between RPs.**

Multicast Transit Design Objectives

- **PIM Border Constraints**
  - **Confine registers within domain**
  - **Confine local groups**
  - **Confine RP announcements**
  - **Control SA advertisements via MSDP**
- **Border RPF check**
  - **RPF check against unicast routes to multicast sources**
- **MSDP RPF check**
  - **RPF check toward RP in received SAs**

Internet IPMulticast

Cisco.com

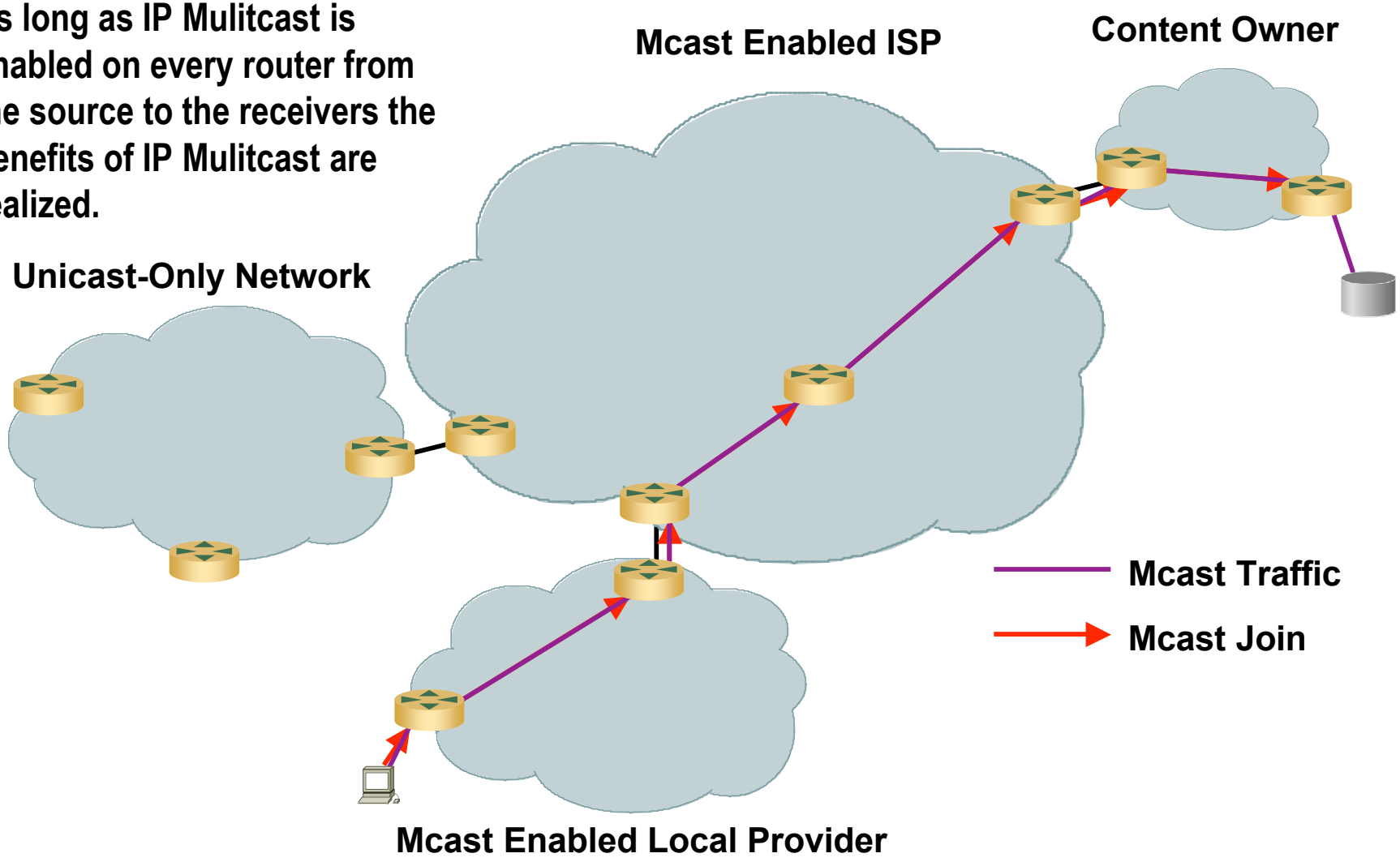
What worked?

What didn’t work?

What’s being done to fix it?

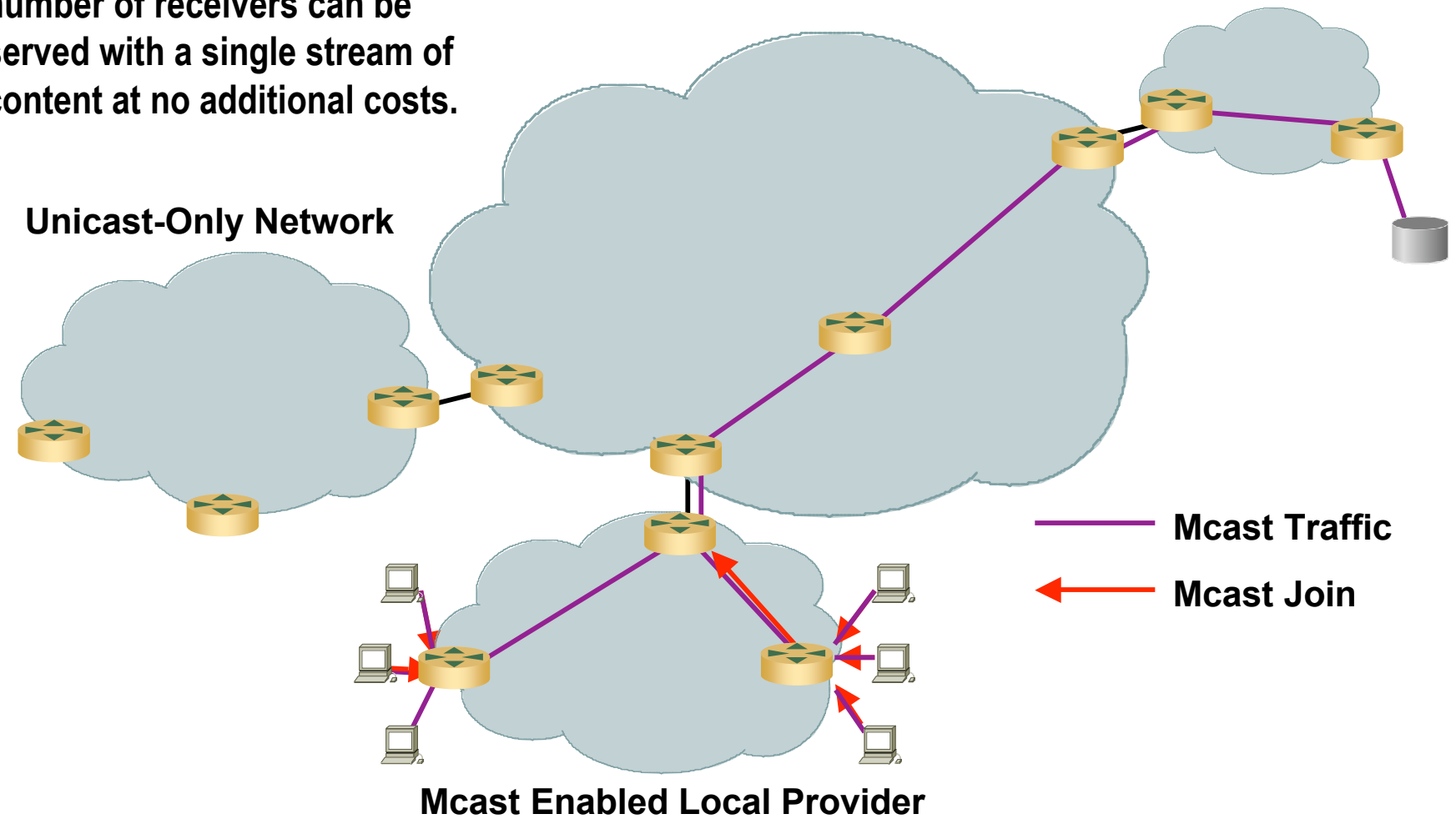
What Worked?

As long as IP Mulitcast is enabled on every router from the source to the receivers the benefits of IP Mulitcast are realized.



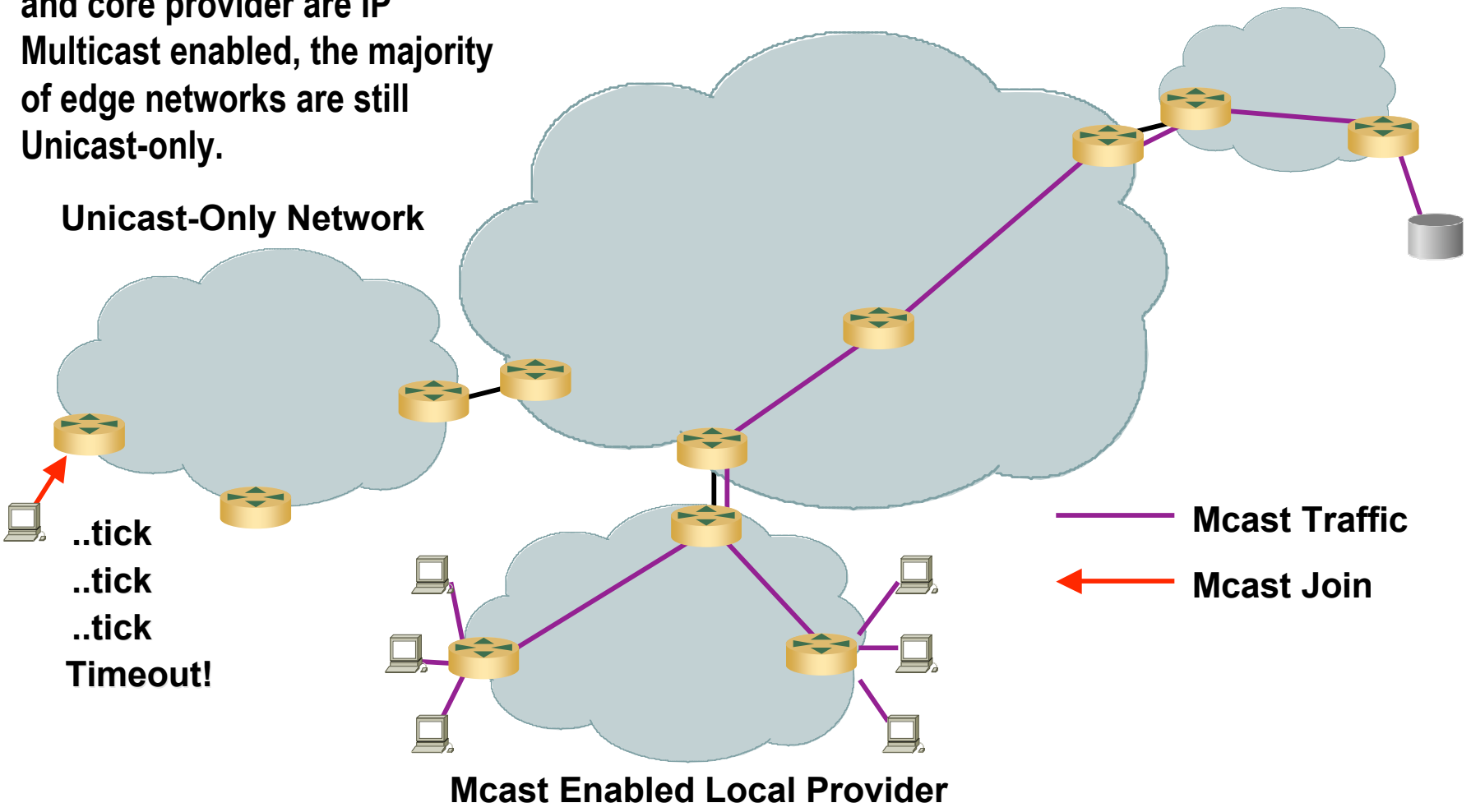
What Worked?

The benefits being an unlimited number of receivers can be served with a single stream of content at no additional costs.

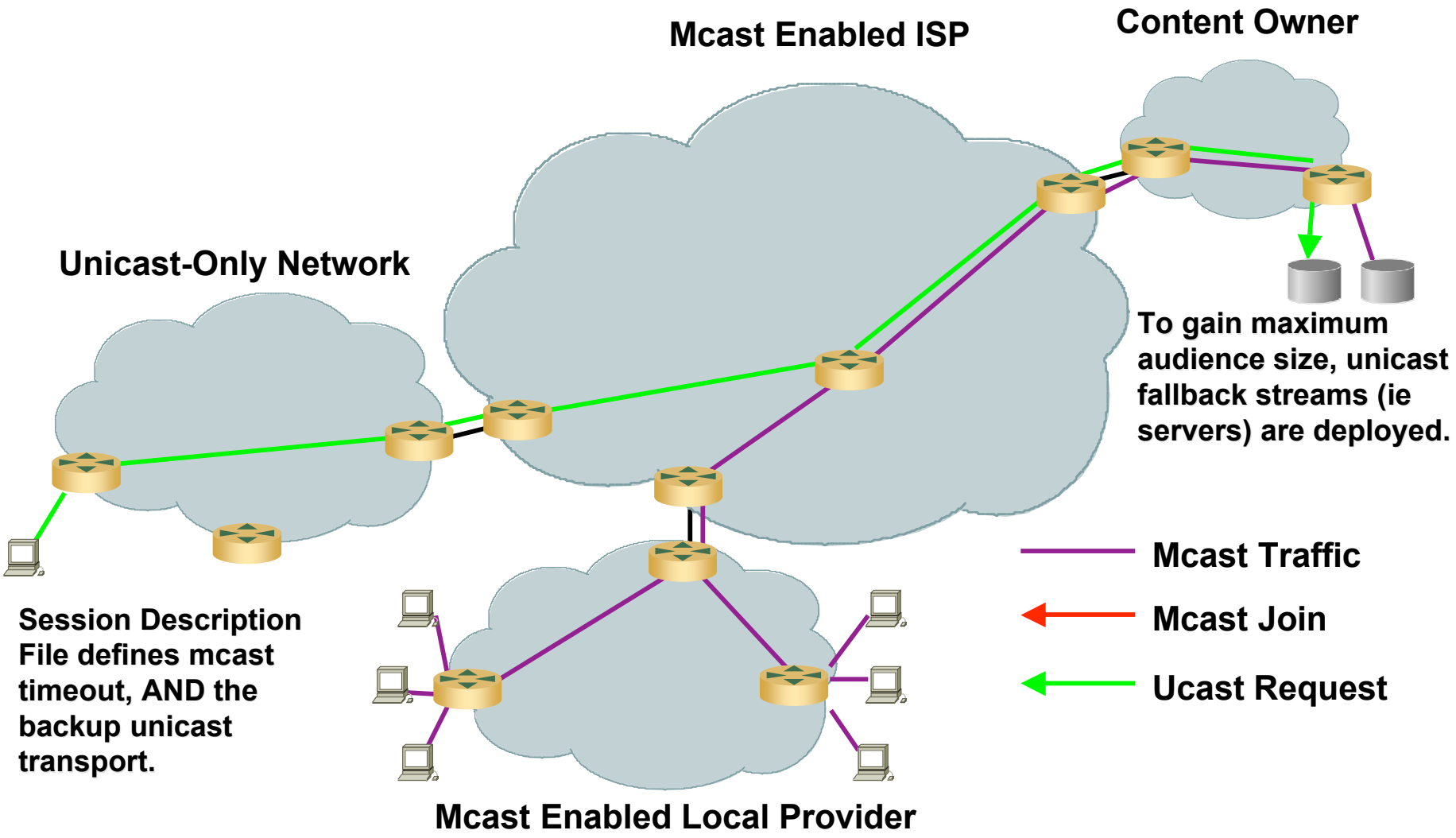


What Didn't?

Even though the content owner and core provider are IP Multicast enabled, the majority of edge networks are still Unicast-only.

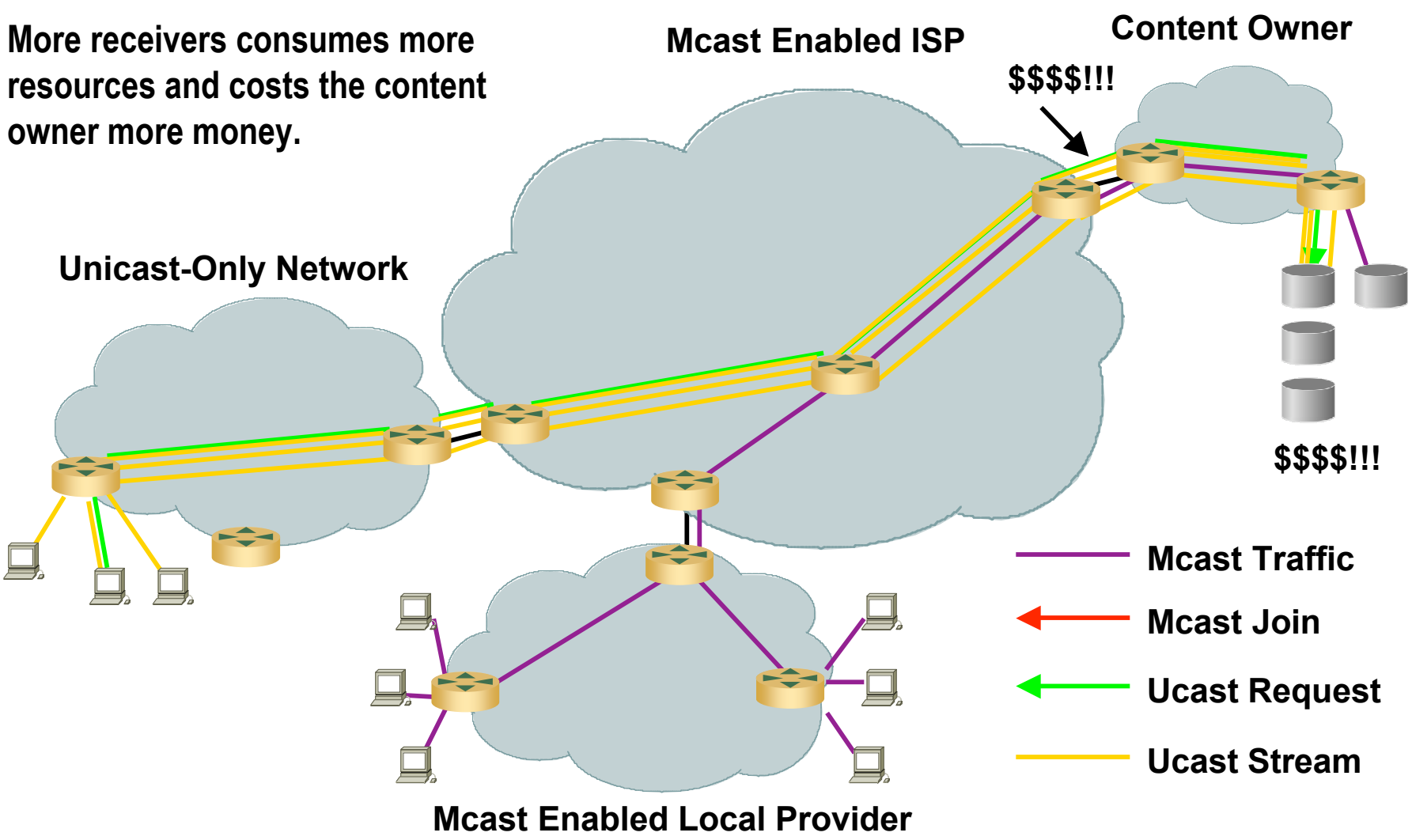


What Didn't?



What Didn't?

More receivers consumes more resources and costs the content owner more money.



What's Wrong?

- **Multicast in the Internet is an all-or-nothing solution**
  - **Each receiver must be on an IP Multicast enabled path.**
  - **Many core networks have IP Multicast enabled - but few edge networks do.**
- **Even Mcast-aware content owners are forced to provide unicast streams to gain audience size**
- **Unicast will never scale for streaming content**
  - **Splitters/Caches just distribute the problem**
    Still has a cost-per-user
  - **As receiver BW increases, problem gets worse.**
  - **Creates a non-functional business model**
  - **Will never bring rich content to IP.**

But multicast is being deployed, right?

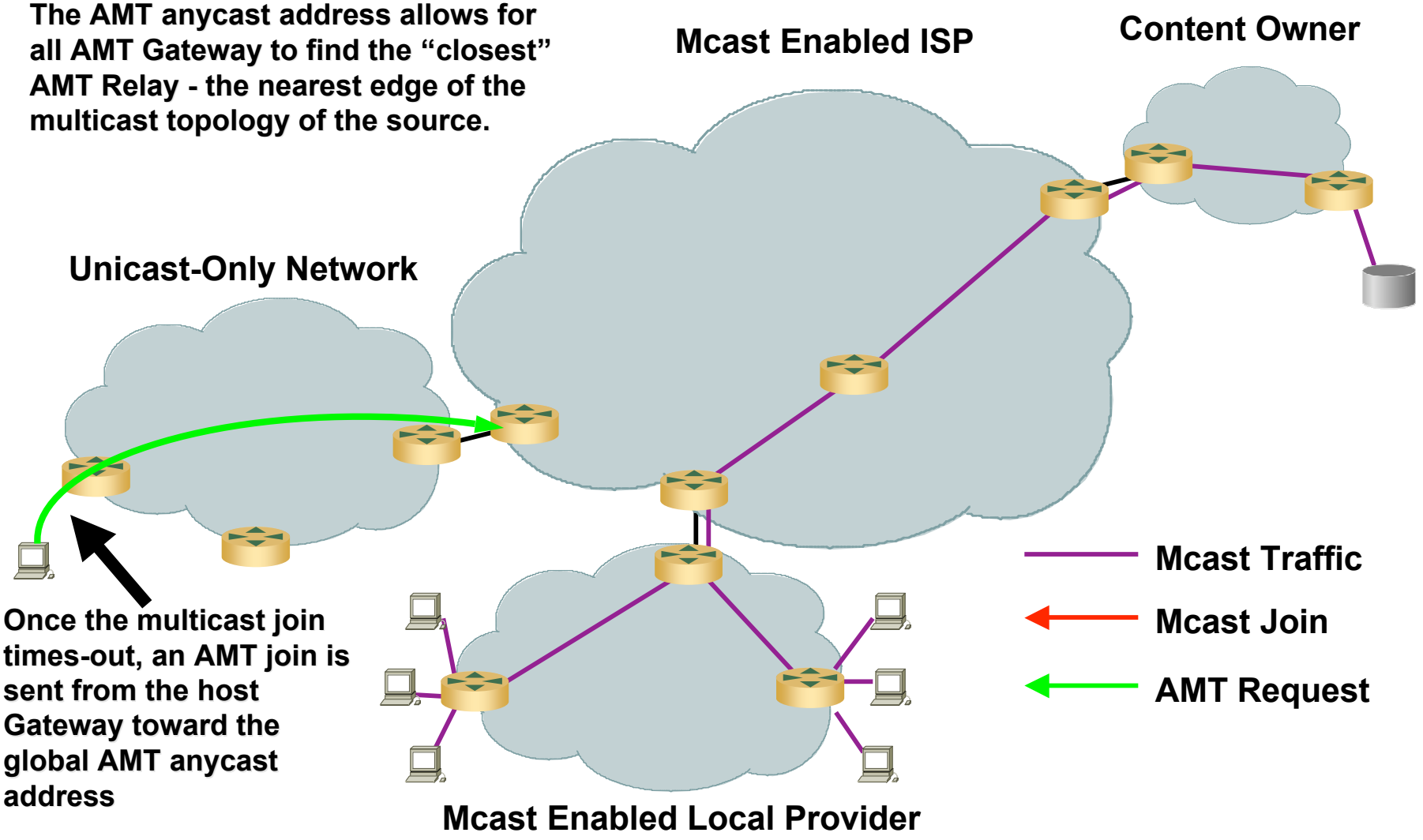
- **Edge (eyeball) Networks**
  - **Locally injected video content only**
  - **No external multicast peering**
  - **Mostly large established business models**
    - **Affiliate content monopoly**
    - **Cisco's primary efforts are focused at preserving these models**
- **Externally Sourced Video - Over The Top Video**
  - **Think Content Owner (not provider)**
  - **Established owners understand the benefits of mcast/bcast model**
  - **Newcomers need the benefits of mcast to compete**
  - **Mcast/ucast dynamic edge transition (AMT)**

AMT
Automatic Multicast Tunneling

- **Automatic IP Multicast without explicit Tunnels**
  - http://www.ietf.org/internet-drafts/draft-ietf-mboned-auto-multicast-05.txt
- **Allow multicast content distribution to extend to unicast-only connected receivers.**
  - **Bring the flat scaling properties of multicast to the Internet**
- **Provide the benefits of multicast wherever multicast is deployed.**
  - **Let the networks which have deployed multicast benefit from their deployment.**
- **Work seamlessly with existing applications**
  - **No OS kernel changes**

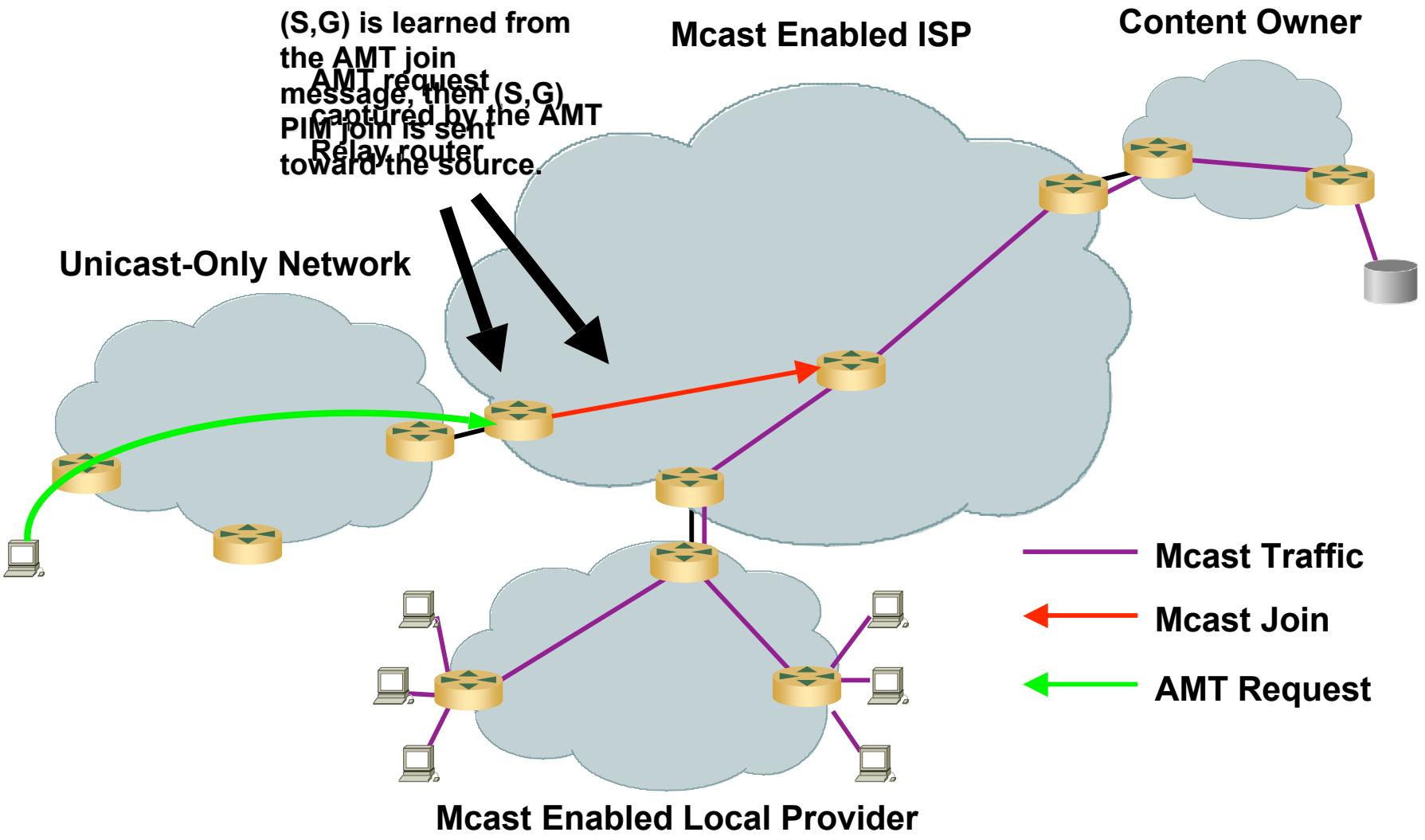
AMT
Automatic Multicast Tunneling

The AMT anycast address allows for all AMT Gateway to find the "closest" AMT Relay - the nearest edge of the multicast topology of the source.

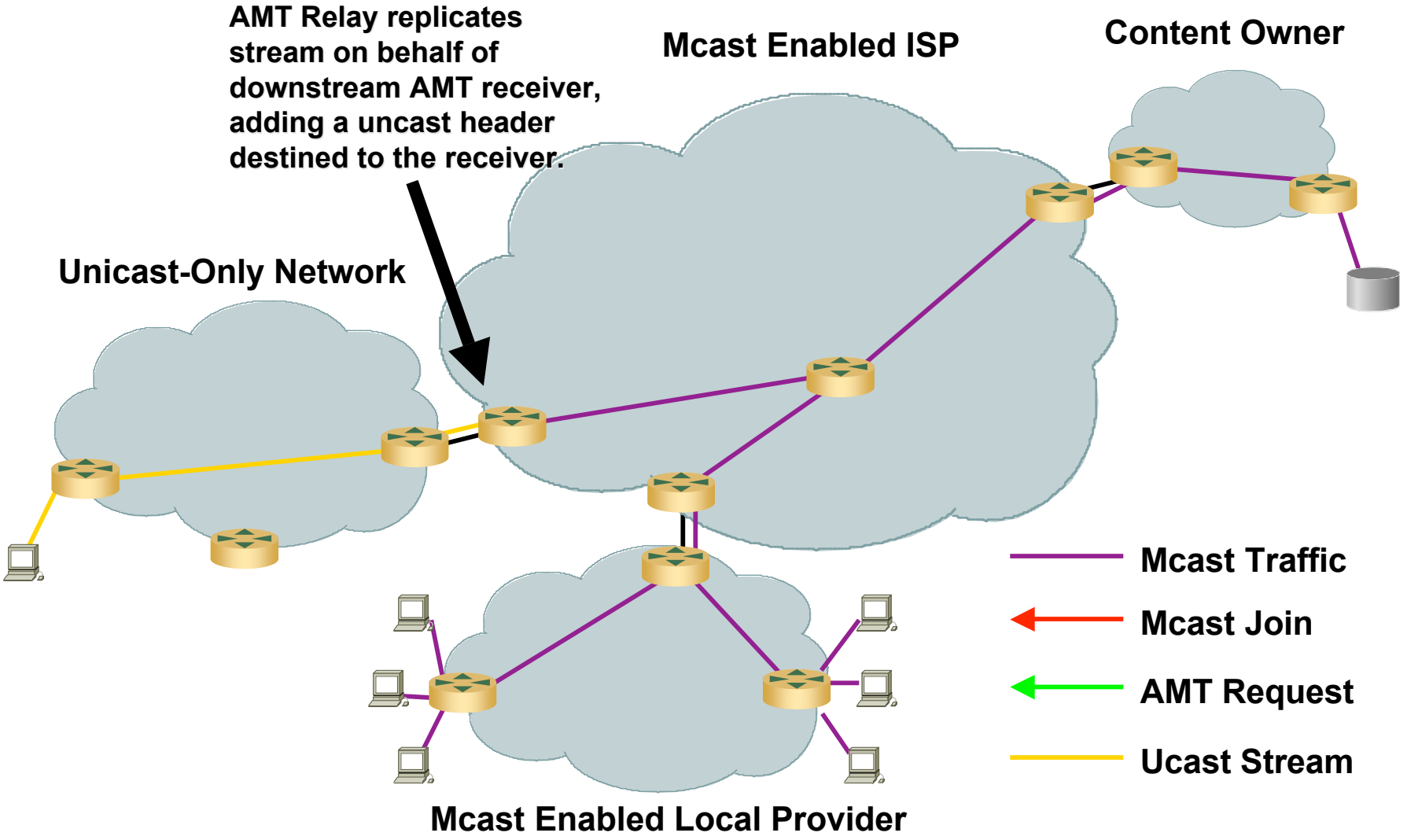


Once the multicast join times-out, an AMT join is sent from the host Gateway toward the global AMT anycast address

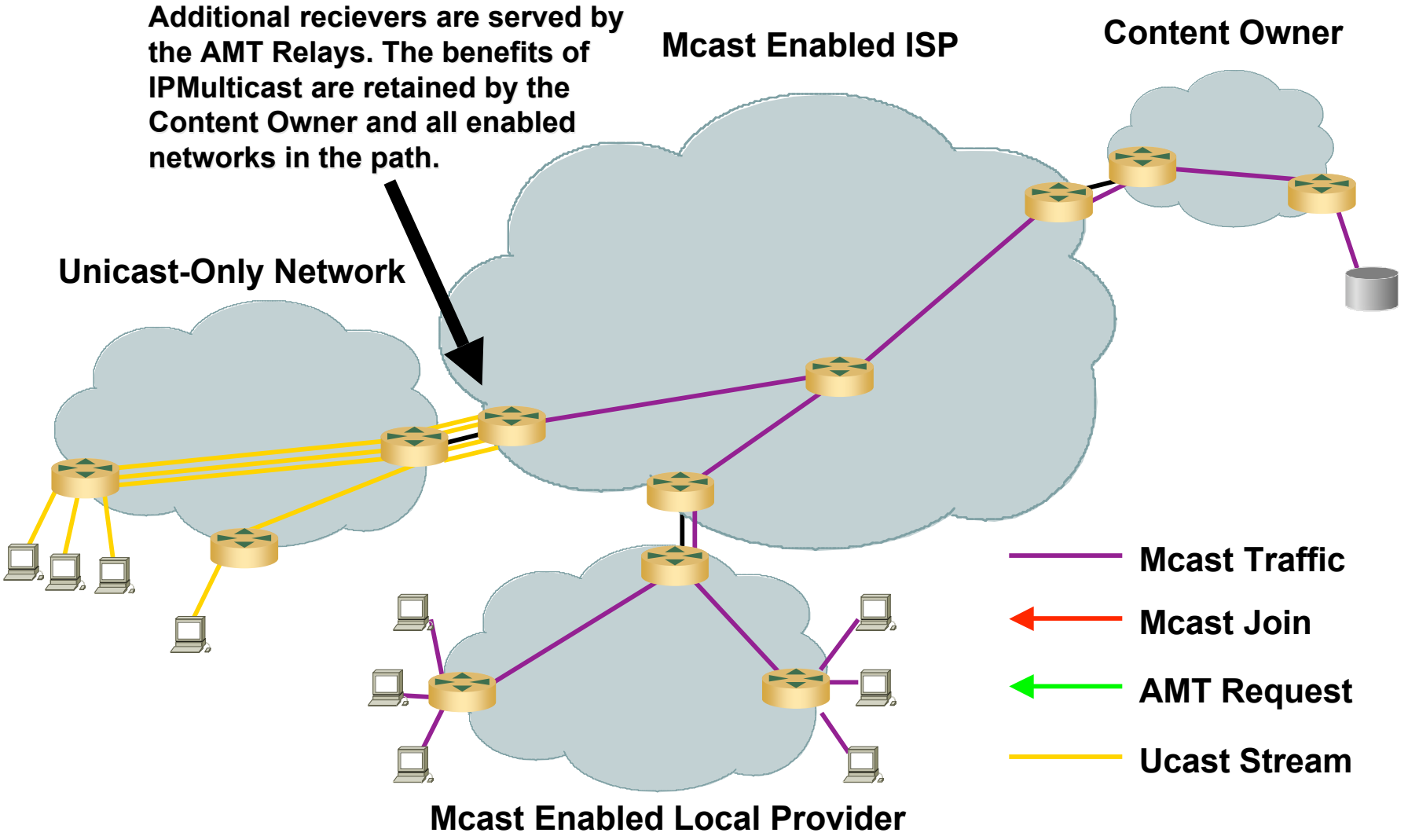
AMT
Automatic Multicast Tunneling



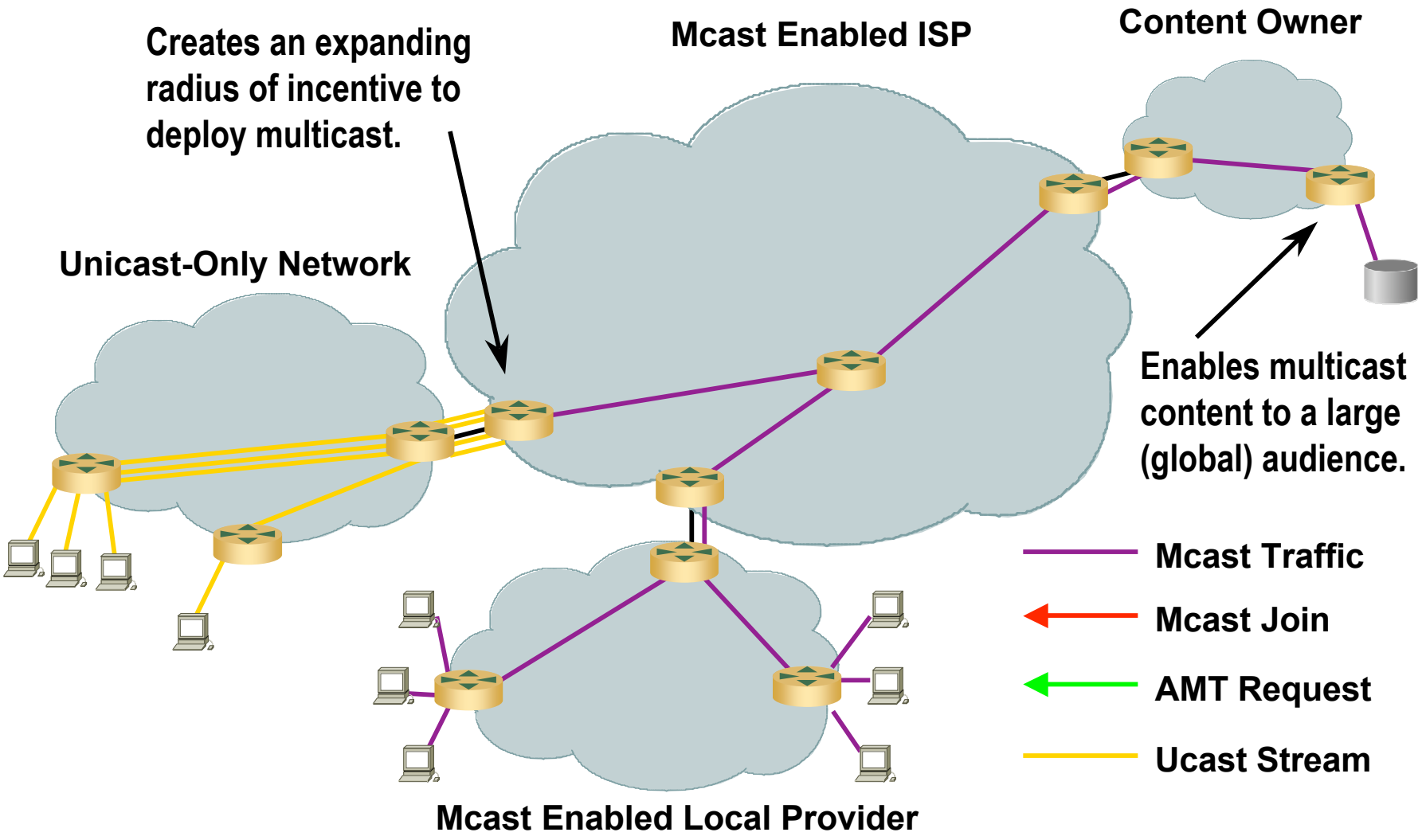
AMT
Automatic Multicast Tunneling



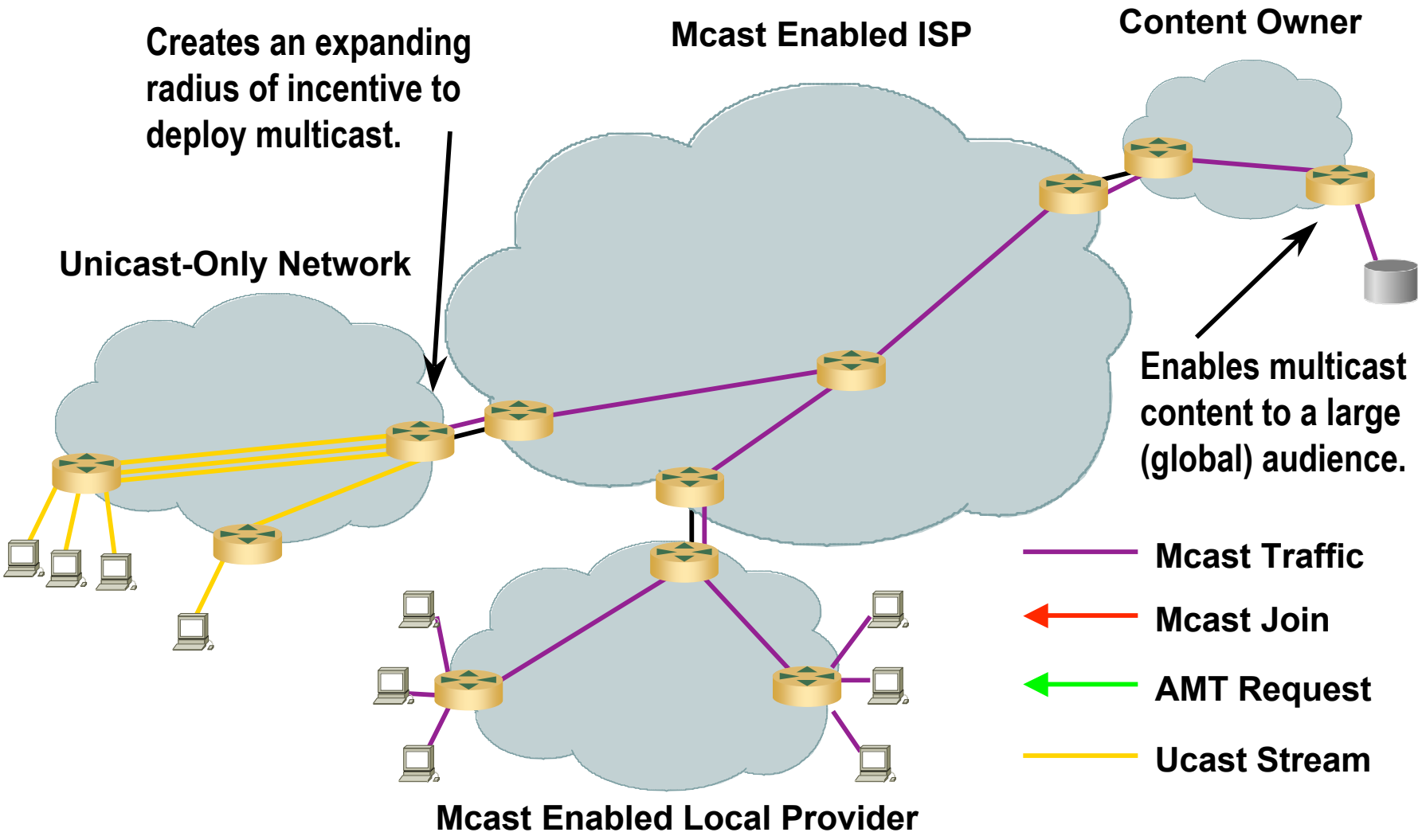
AMT
Automatic Multicast Tunneling



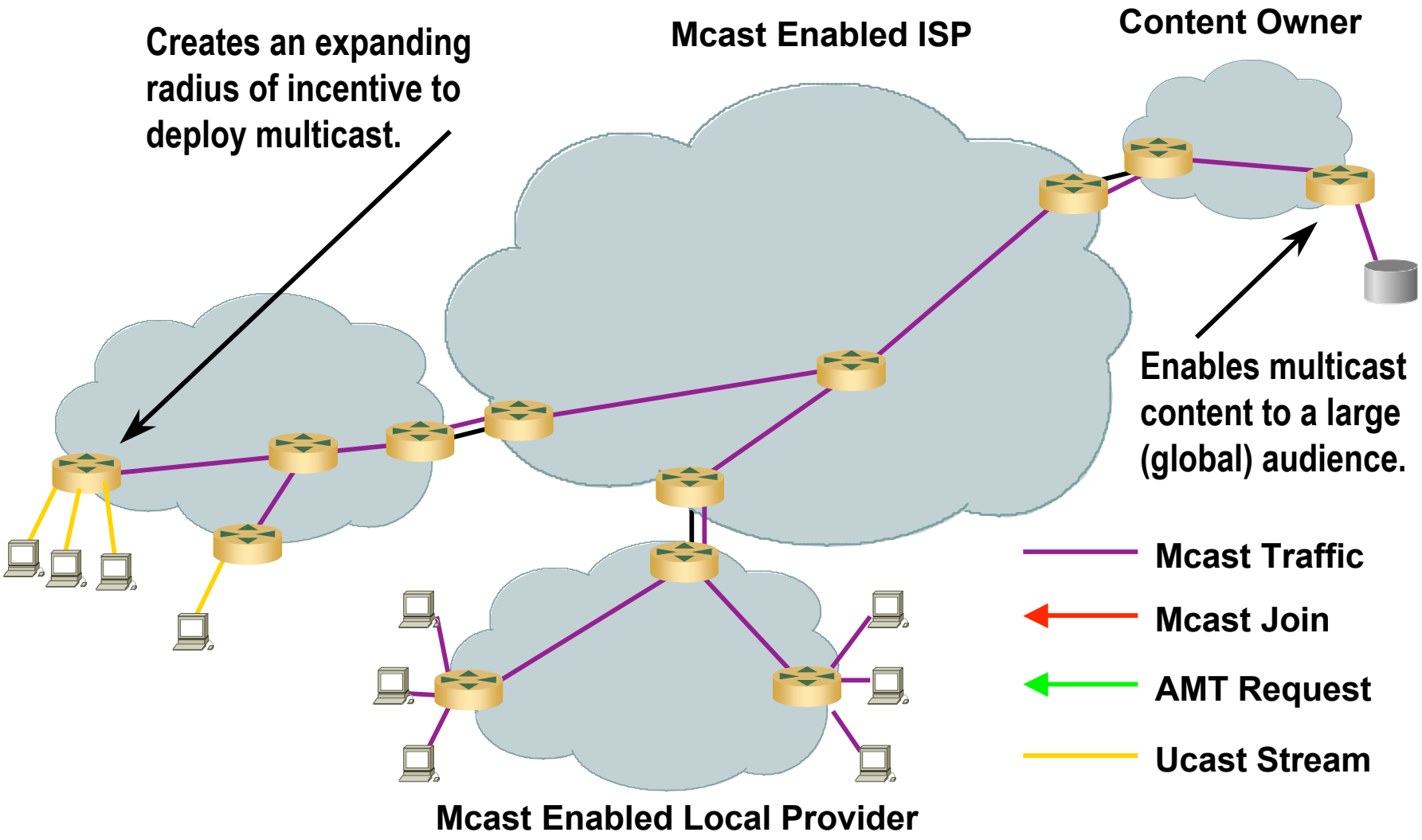
AMT
Automatic Multicast Tunneling



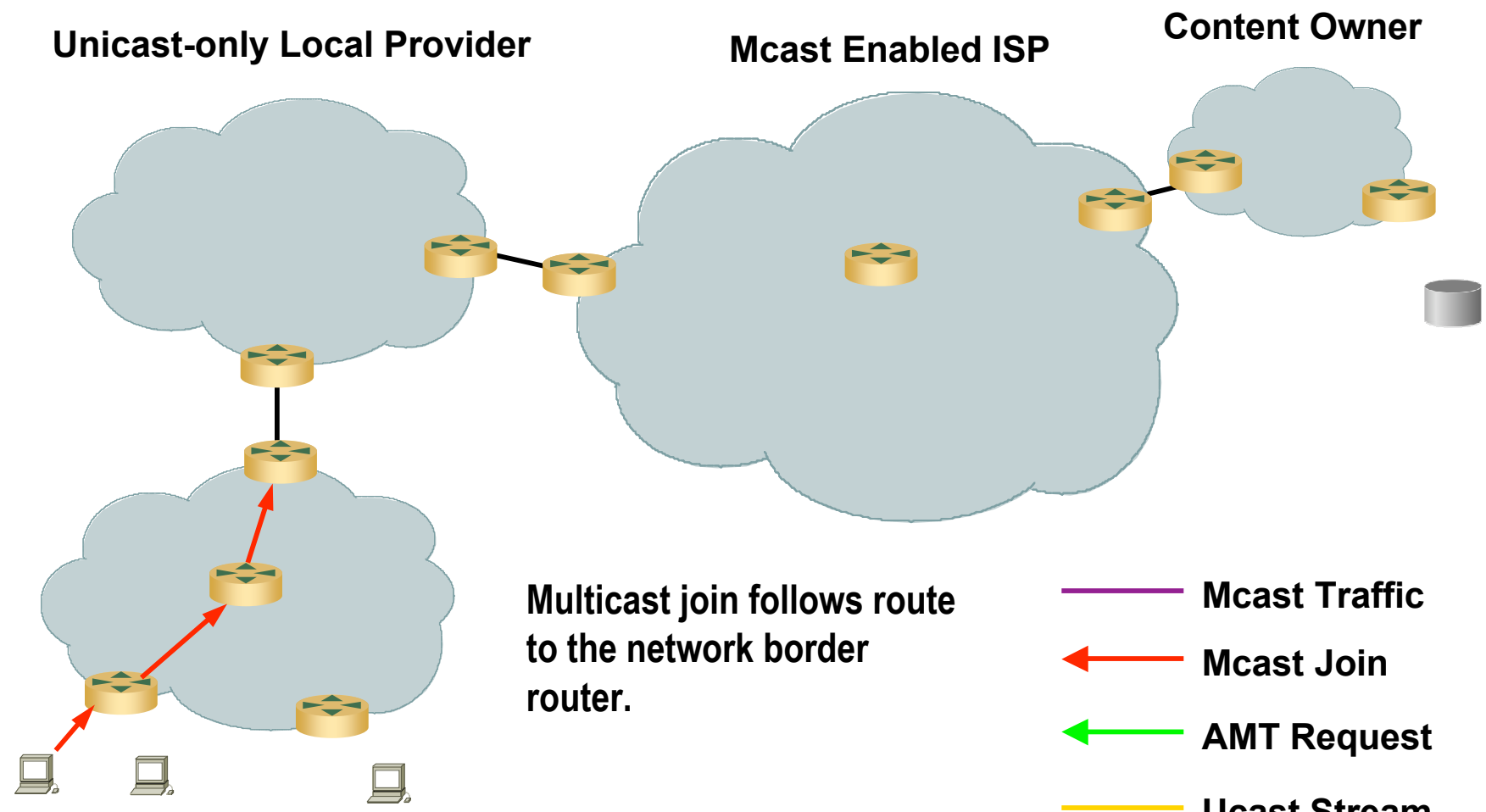
AMT
Automatic Multicast Tunneling



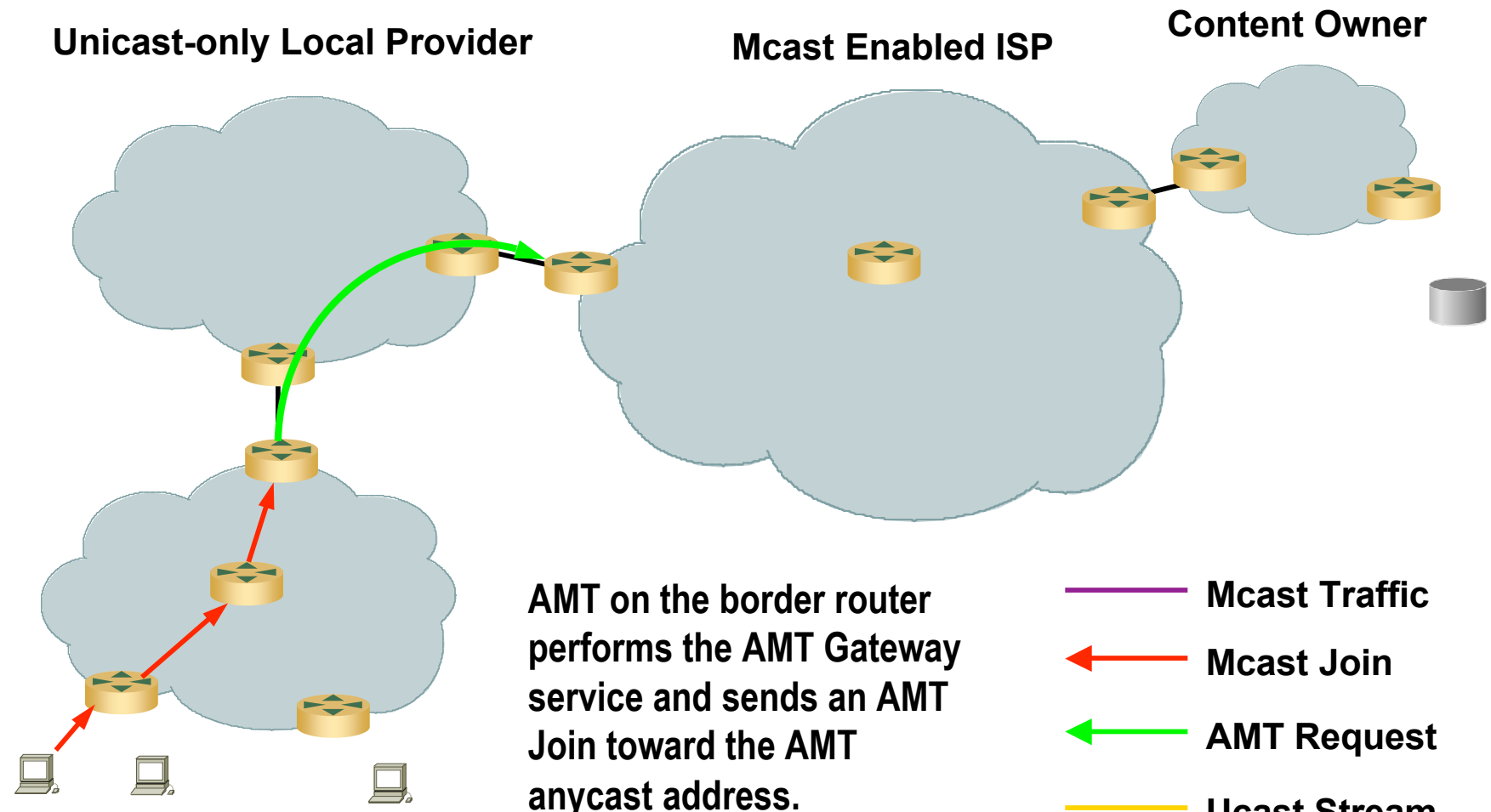
AMT
Automatic Multicast Tunneling



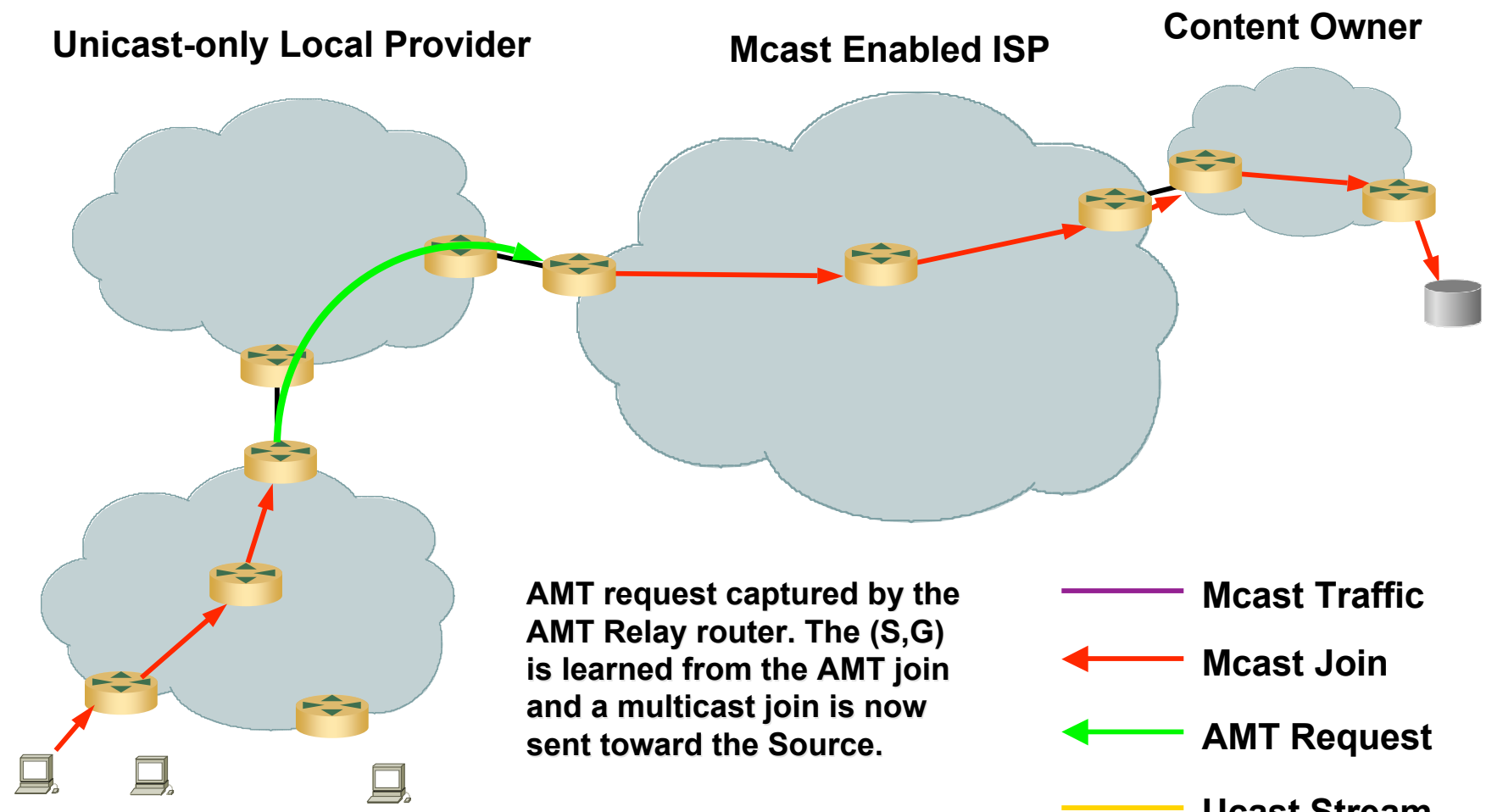
AMT
Connecting Multicast Islands



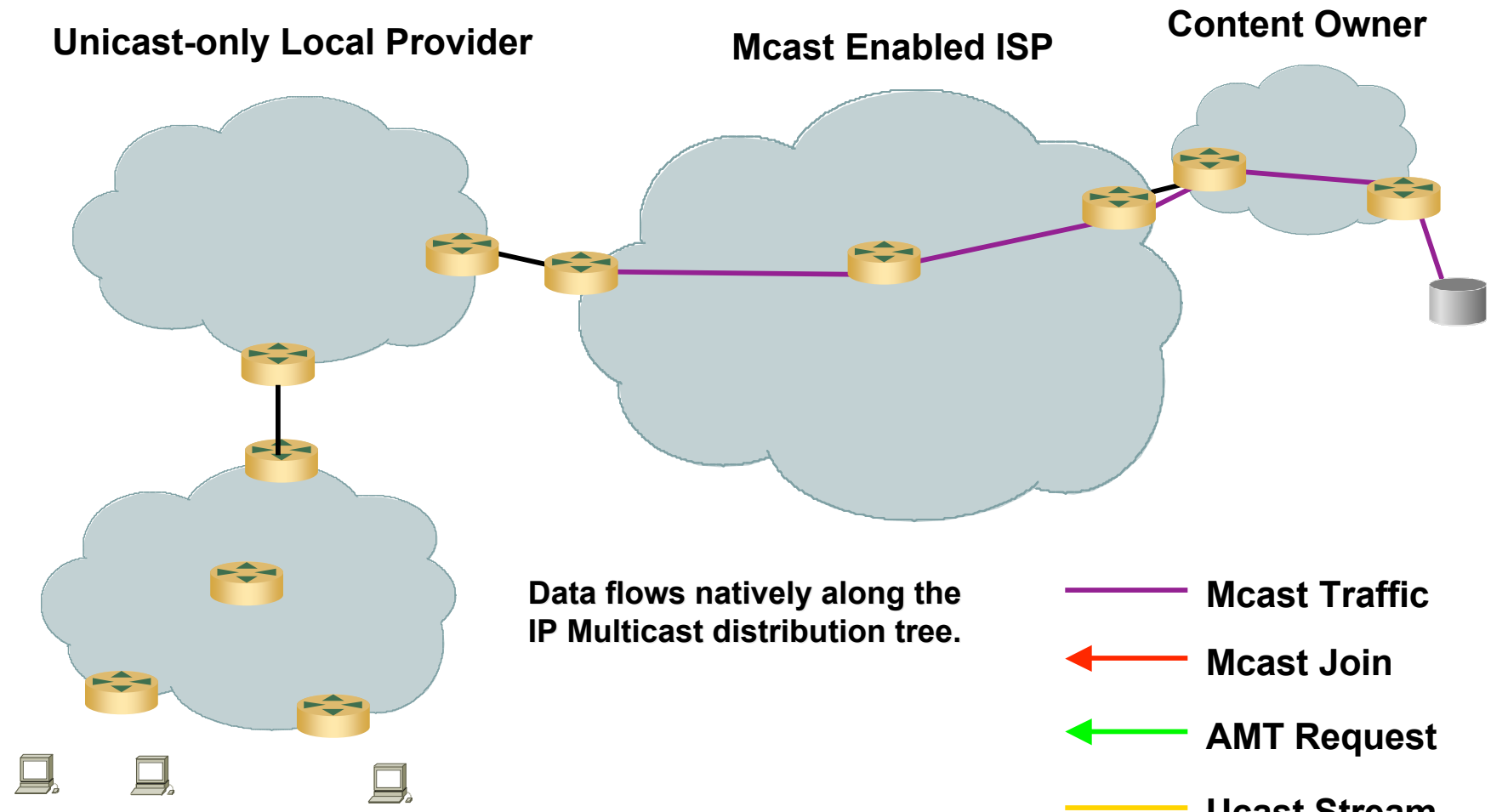
AMT
Connecting Multicast Islands



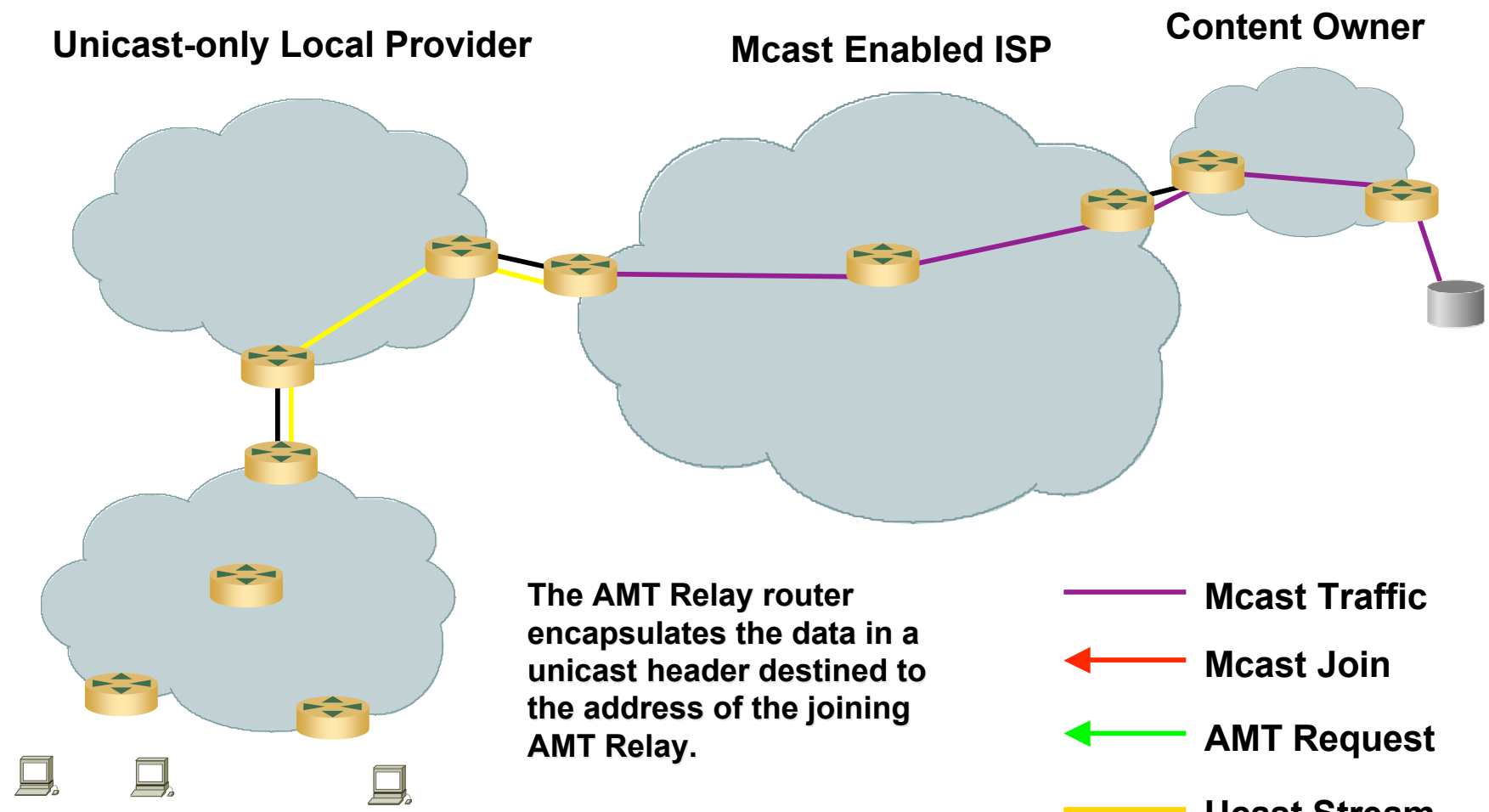
AMT
Connecting Multicast Islands



AMT
Connecting Multicast Islands



AMT
Connecting Multicast Islands



The AMT Relay router encapsulates the data in a unicast header destined to the address of the joining AMT Relay.

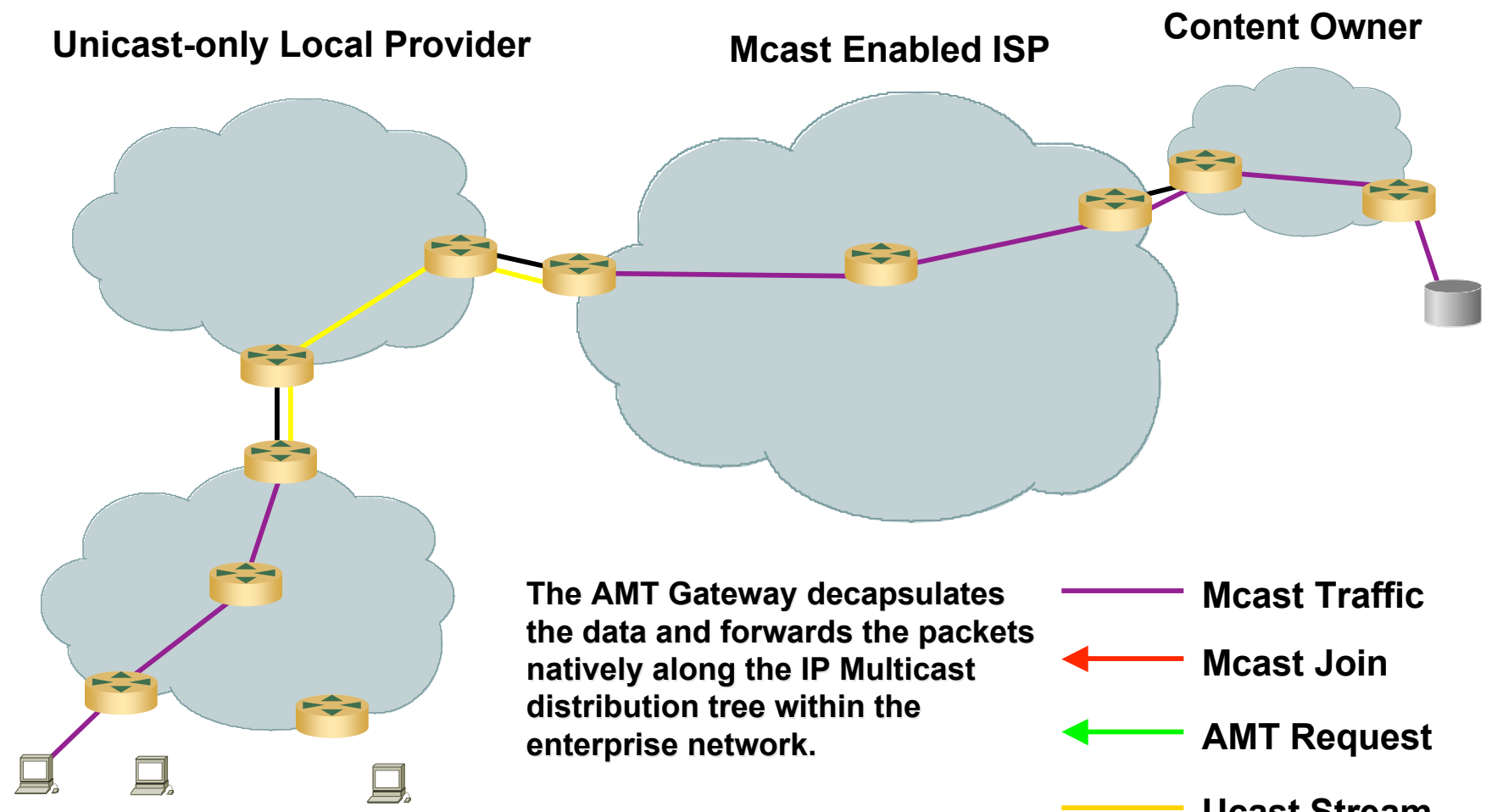
Mcast Traffic

Mcast Join

AMT Request

Ucast Stream

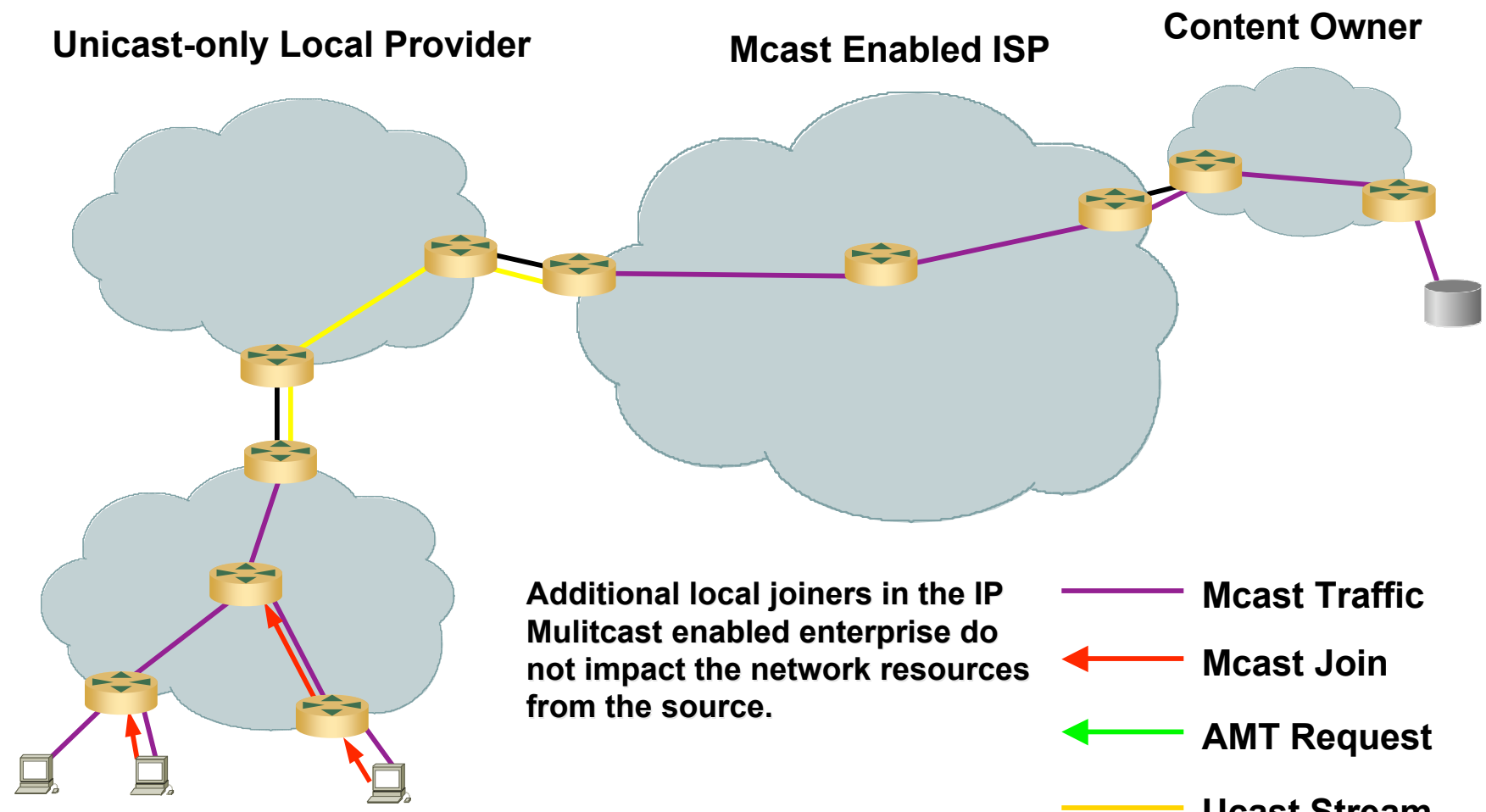
AMT
Connecting Multicast Islands



The AMT Gateway decapsulates the data and forwards the packets natively along the IP Multicast distribution tree within the enterprise network.

- Mcast Traffic
- Mcast Join
- AMT Request
- Ucast Stream

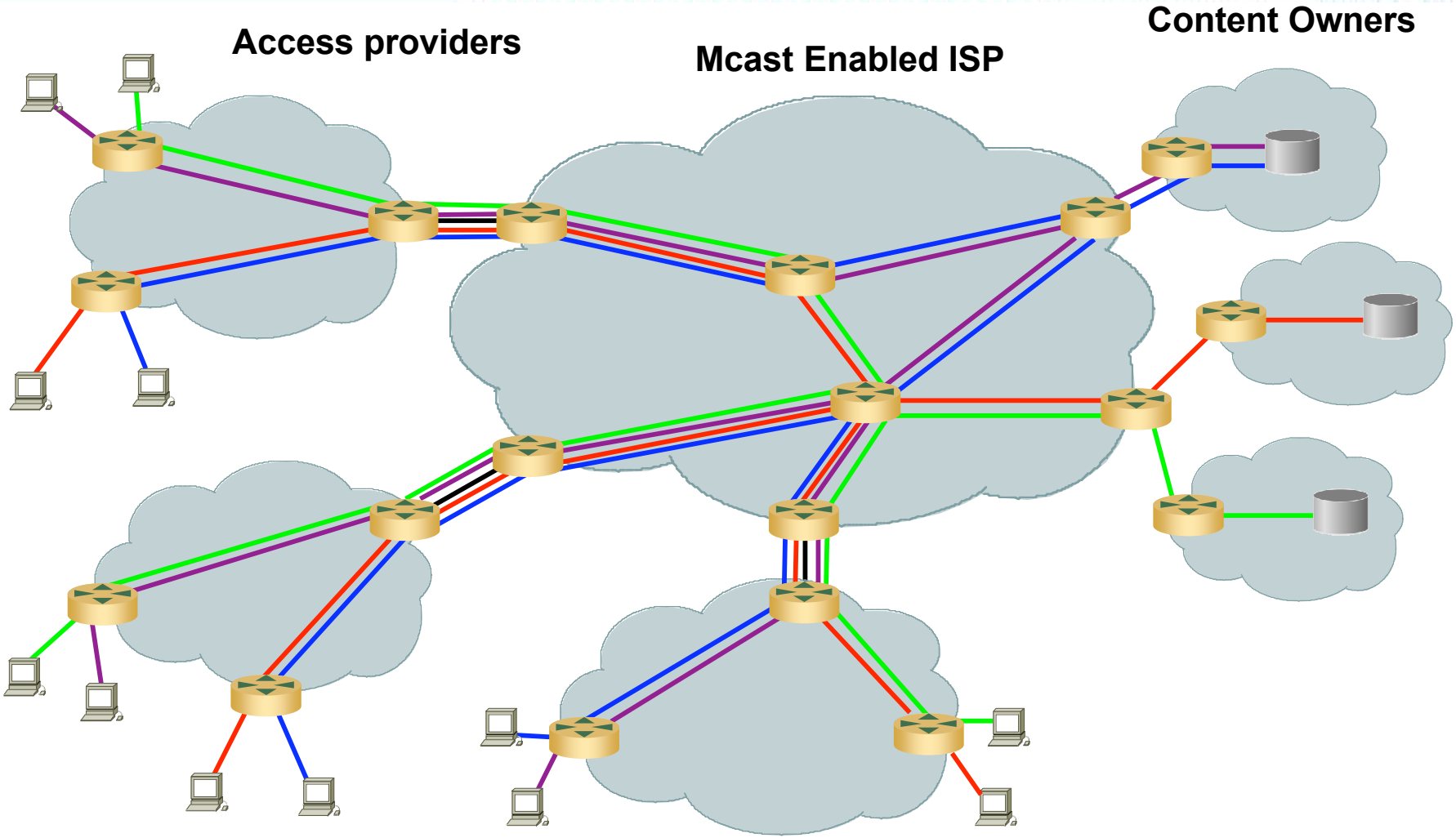
AMT
Connecting Multicast Islands



Multicast Myths

- It's too hard / to complicated
  - It's being used today in many mission-critical applications with success.
  - It just hasn't been a requirement (yet) for many people
- Provider – "If multicast catches on, my customers will stop buying big circuits."
  - Wrong (next slide)
- Vendor – "If multicast catches on, no one will need big routers and high-speed interfaces."
  - Wrong (next slide)

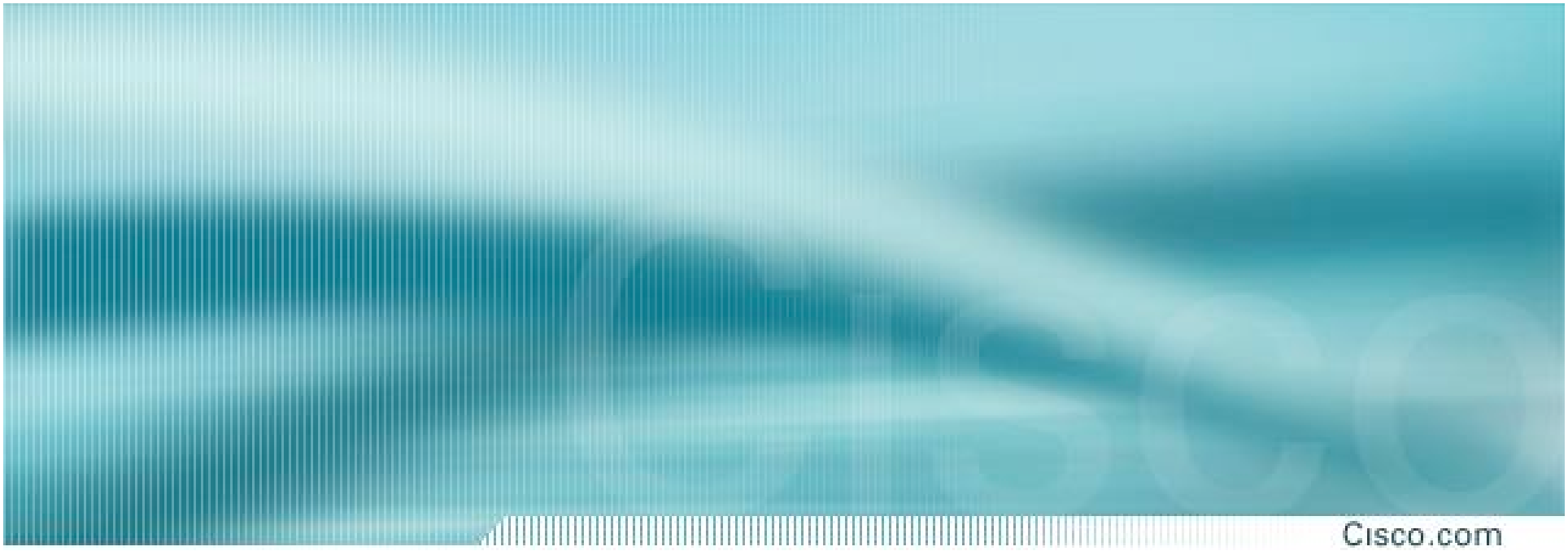
When the world deploys IPMulticast



When the world deploys IPMulticast

Cisco.com

- **A successful multicast business model makes IP profitable for content owners**
  - **Success brings MORE content**
    Higher bit-rate
    More channels
- **Access networks of tomorrow look like provider networks of today**
  - **Few large circuits upstream, many small circuits downstream.**
    (see previous slide)
- **Provider revenue model is inverted**
  - **Few small circuits (relative) from content networks, many large circuits down to access networks.**
    (see previous slide)



Thank you!

http://www.cisco.com