



2216  
1199\_05\_2000\_c3 © 2000, Cisco Systems, Inc.



2216  
1199\_05\_2000\_c3 © 2000, Cisco Systems, Inc.

# Other Related Presentations

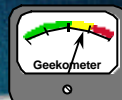
- **Multicast Sessions**

Session #	Title
2214	Introduction to IP Multicast
2215	PIM Multicast Routing
2216	Deploying IP Multicast
2217	Advanced IP Multicast Routing

- **MBGP Related Sessions**

Session #	Title
2209	Deploying BGP

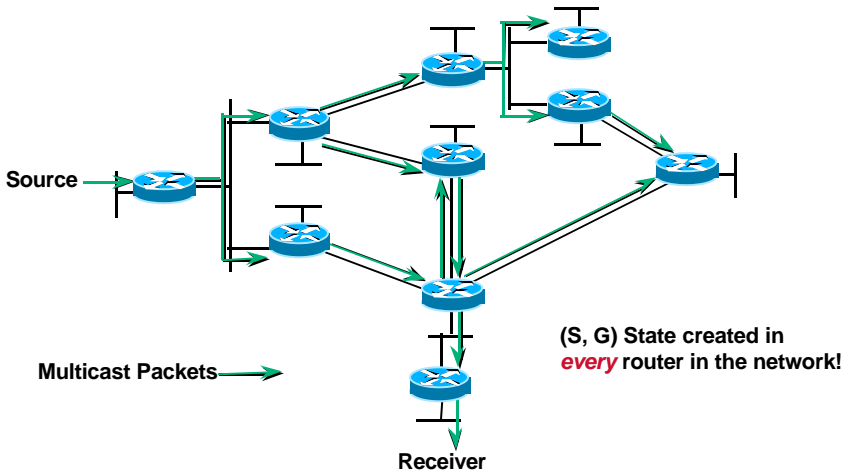
# Agenda



- **IP Multicast Review**
- Rendezvous Points (RP)
- Configuring IP Multicast
- Multicast at Layer 2
- Multicast Performance
- Multicast Traffic Engineering

# PIM-DM Flood and Prune

## Initial Flooding



2216  
1199\_05\_2000\_c3

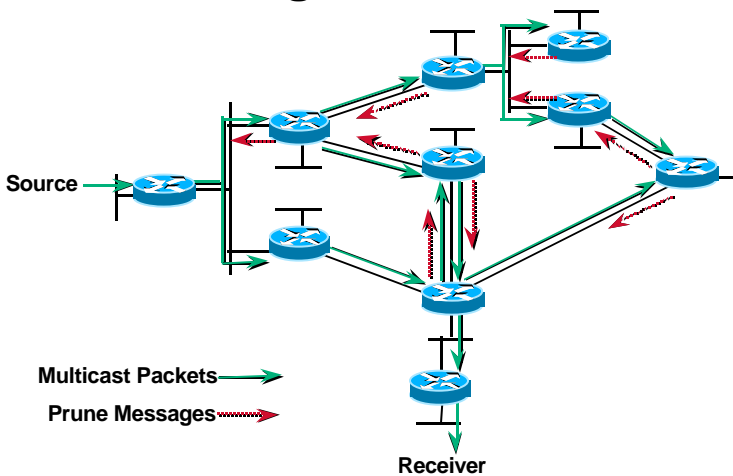
© 2000, Cisco Systems, Inc.

cisco.com

5

# PIM-DM Flood and Prune

## Pruning Unwanted Traffic



2216  
1199\_05\_2000\_c3

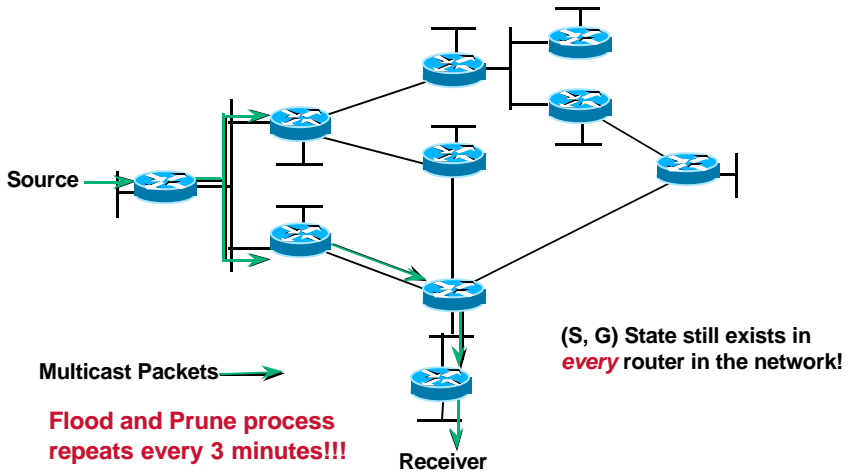
© 2000, Cisco Systems, Inc.

cisco.com

6

# PIM-DM Flood and Prune

## Results After Pruning



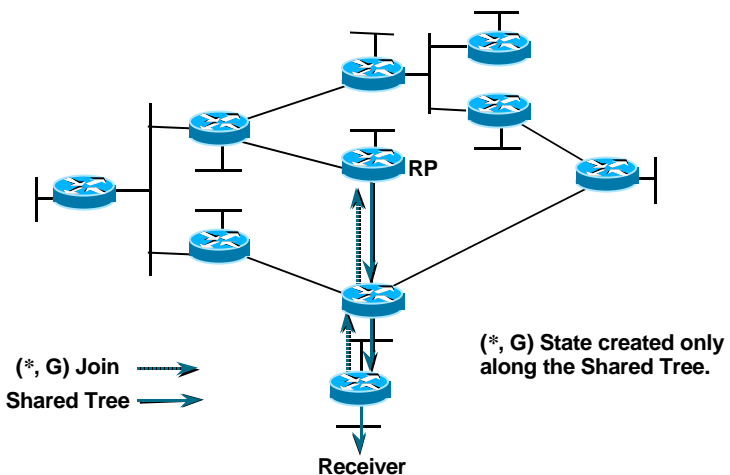
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

7

# PIM-SM Shared Tree Join



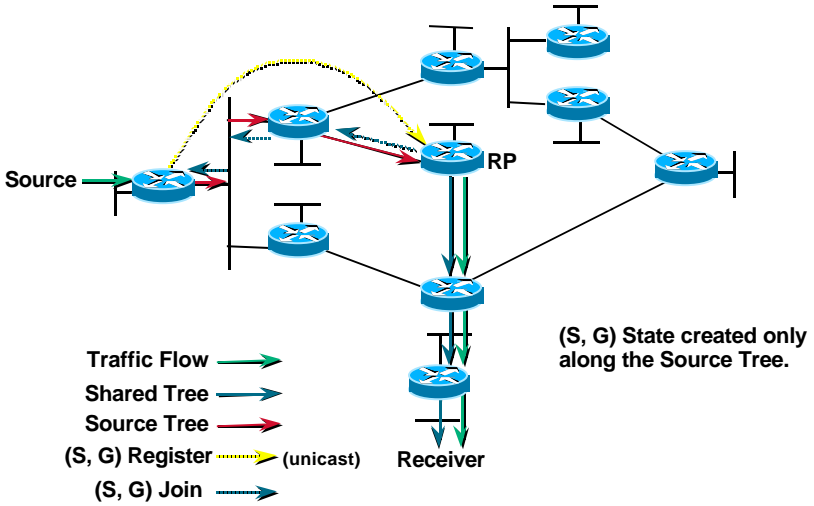
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

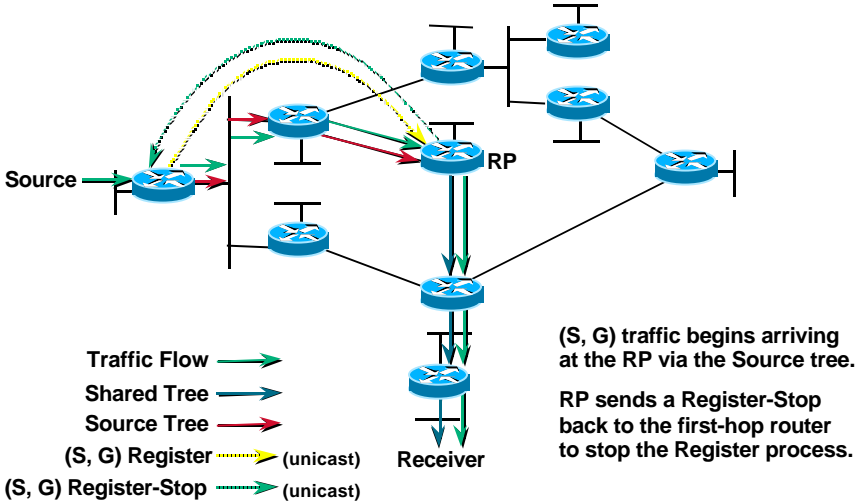
cisco.com

8

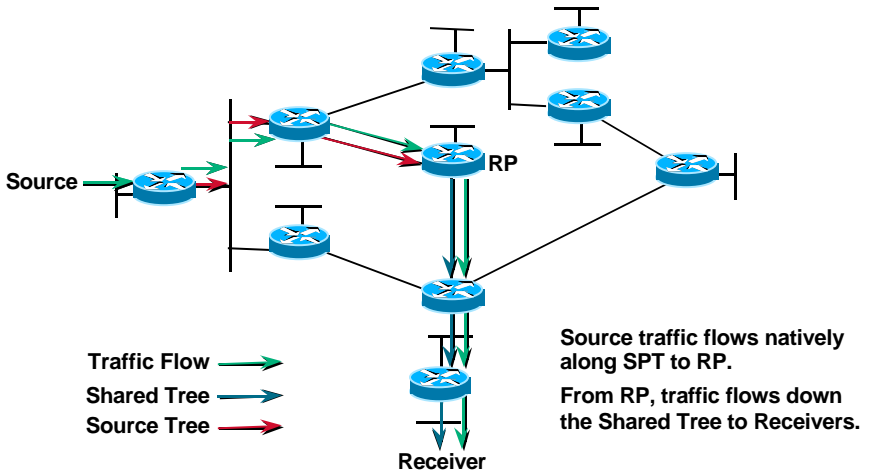
# PIM-SM Sender Registration



# PIM-SM Sender Registration



# PIM-SM Sender Registration



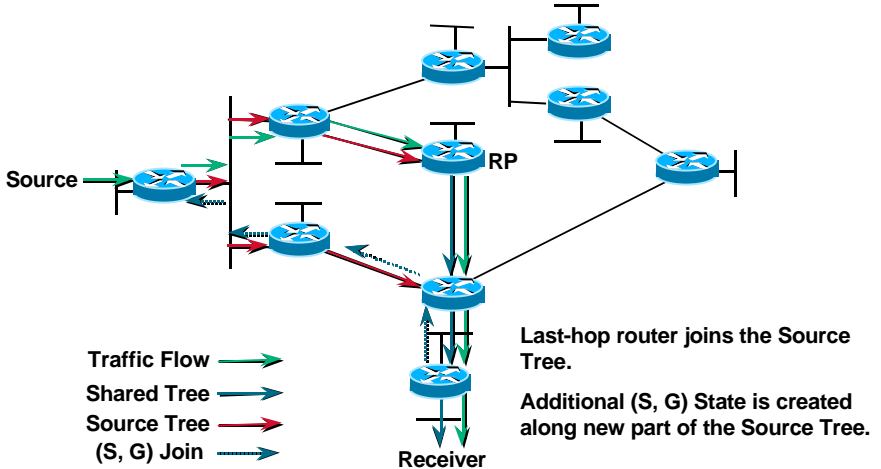
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

11

# PIM-SM SPT Switchover



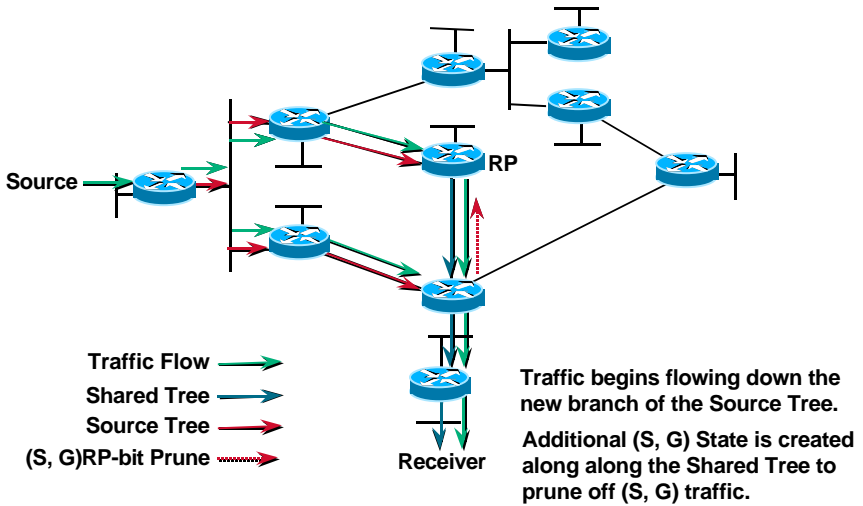
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

12

# PIM-SM SPT Switchover



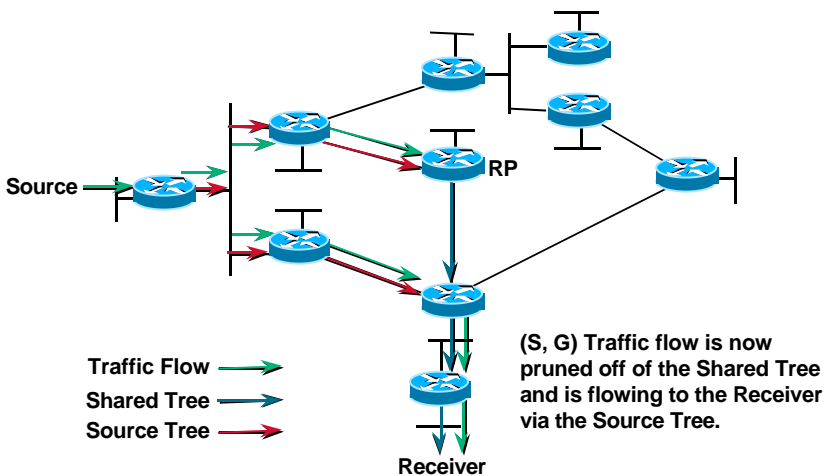
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

13

# PIM-SM SPT Switchover



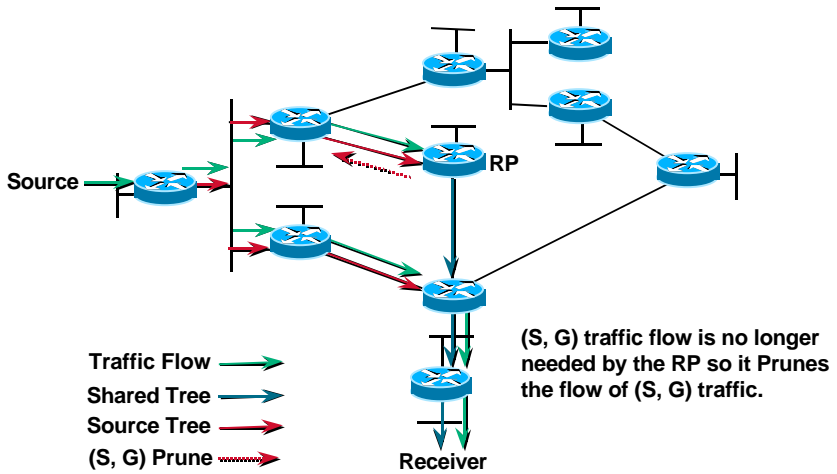
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

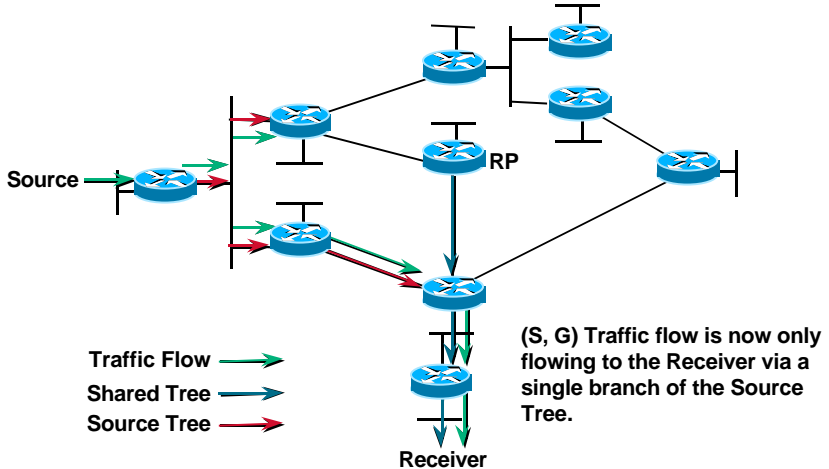
cisco.com

14

# PIM-SM SPT Switchover



# PIM-SM SPT Switchover



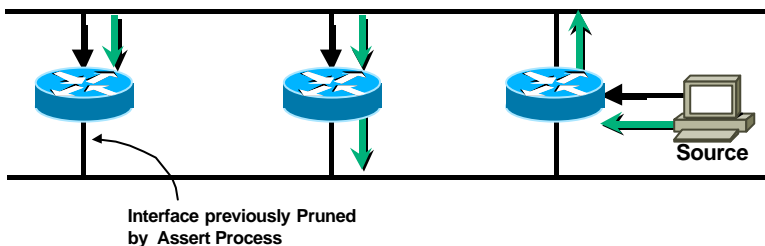
# Which Mode—Sparse or Dense

- **Dense mode**

- **Flood and Prune behavior very inefficient**
    - Can cause problems in certain network topologies
  - **Creates (S, G) state in EVERY router**
    - Even when there are no receivers for the traffic
  - **Complex Assert mechanism**
- **Mixed control and data planes**
    - Results in (S, G) state in every router in the network
    - Can result in non-deterministic topological behaviors
  - **No support for shared trees**

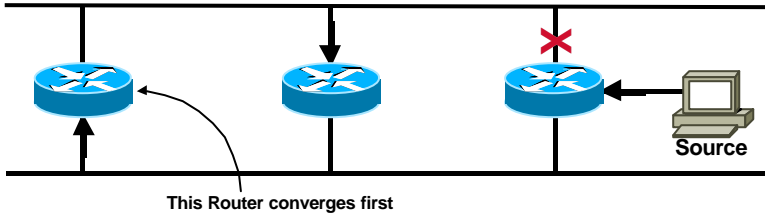
# Potential PIM-DM Route Loop

## Normal Steady-State Traffic Flow



# Potential PIM-DM Route Loop

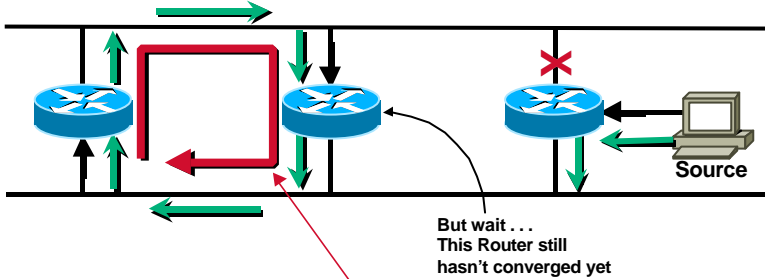
## Interface Fails



↓ RPF Interface

# Potential PIM-DM Route Loop

## New Traffic Flow

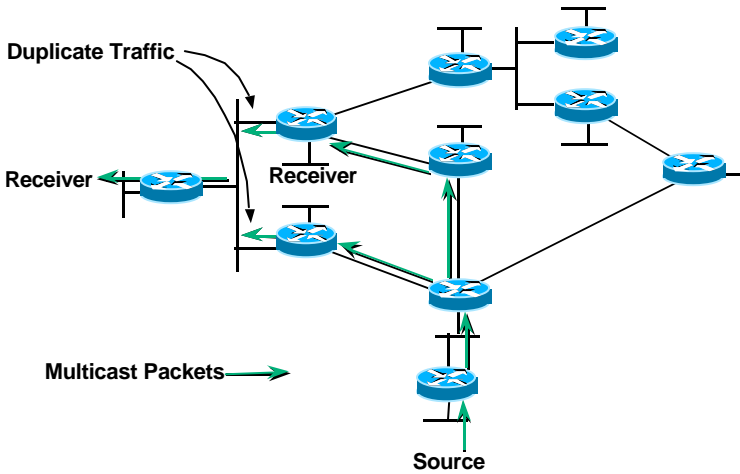


**Multicast Route Loop !!**

↓ RPF Interface

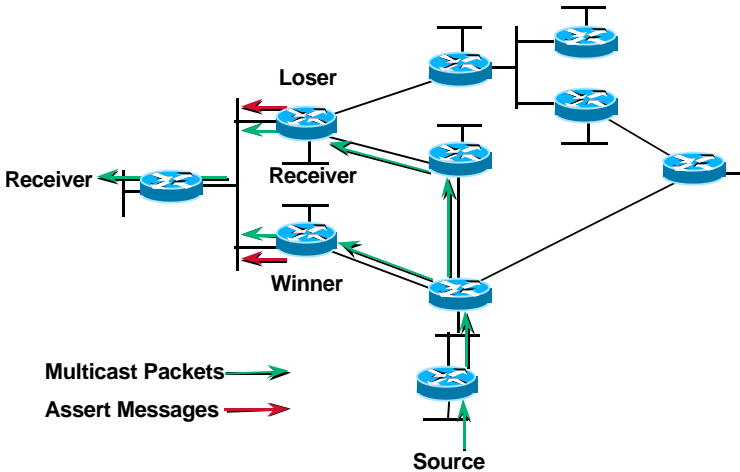
# PIM-DM Assert Problem

## Initial Flow



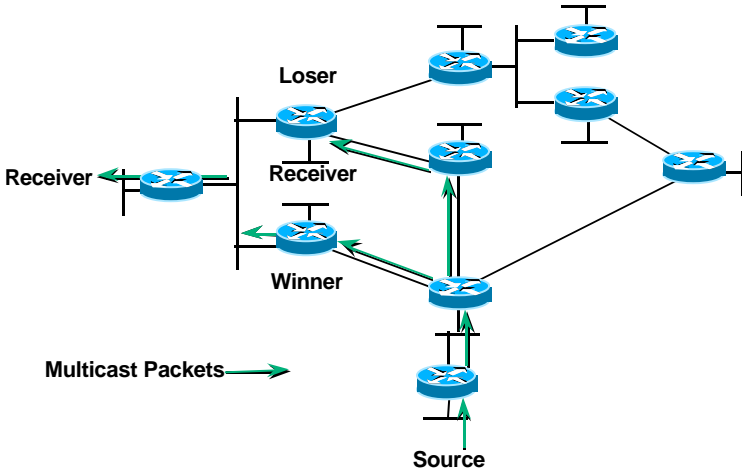
# PIM-DM Assert Problem

## Sending Asserts



# PIM-DM Assert Problem

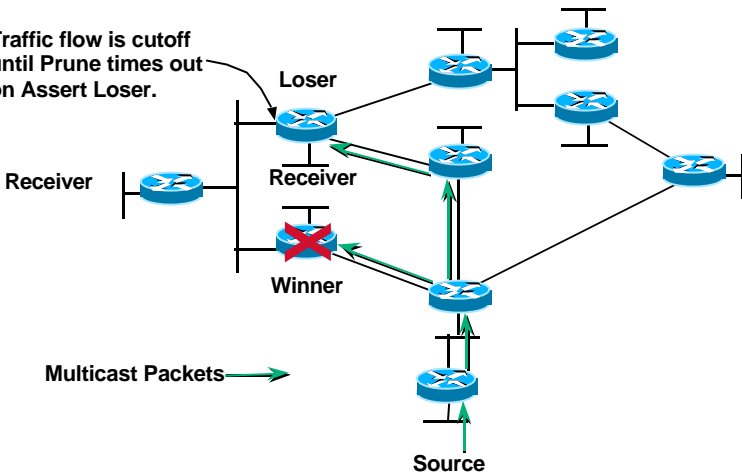
## Assert Loser Prunes Interface



# PIM-DM Assert Problem

## Assert Winner Fails

Traffic flow is cutoff until Prune times out on Assert Loser.



# Which Mode—Sparse or Dense

## • Sparse mode

- **Must configure a Rendezvous Point (RP)**
  - Statically (on every Router)
  - Using Auto-RP (Routers learn RP automatically)
  - Using BSR (Routers learn RP automatically)
- **Very efficient**
  - Uses Explicit Join model
  - Traffic only flows to where it's needed
- **Separate control and data planes**
  - Router state only created along flow paths
  - Deterministic topological behavior
- **Scales better than dense mode**
  - Works for *both* sparsely or densely populated networks

# Which Mode—Sparse or Dense

---

## **CONCLUSION**

**Virtually all production networks should be configured to run in Sparse mode!**

---

# Agenda

- IP Multicast Review
- Rendezvous Points (RP)
- Configuring IP Multicast
- Multicast at Layer 2
- Multicast Performance
- Multicast Traffic Engineering

## RP Placement

- Q: “Where do I put the RP?”
  - A: “Generally speaking, it’s not critical”
- SPT’s are normally used by default
  - RP is a place for source and receivers to meet
  - Traffic does not normally flow through the RP
  - RP is therefore not a bottleneck
- Exception: SPT-Threshold = Infinity
  - Traffic stays on the shared tree
  - RP *could* become a bottleneck

# 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 Overview

- **All routers automatically learn RP address**

- No configuration necessary except on:
  - Candidate RPs
  - Mapping Agents

- **Makes use of Multicast to distribute info**

- Two specially IANA assigned Groups used
  - Cisco-Announce - 224.0.1.39
  - Cisco-Discovery - 224.0.1.40
- Typically Dense mode is used forward these groups

- **Permits backup RP's to be configured**

- **Can be used with Admin-Scoping**

# Auto-RP Fundamentals

## • Candidate RPs

- Configured via global config command

```
ip pim send-rp-announce <intfc> scope <ttl> [group-list acl]
```

- Multicast RP-Announcement messages

- Sent to Cisco-Announce (224.0.1.39) group
- Sent every rp-announce-interval (default: 60 sec)

- RP-Announcements contain:

- Group Range (default = 224.0.0.0/4)
- Candidate's RP address
- Holdtime = 3 x <rp-announce-interval>

# Auto-RP Fundamentals

## • Mapping agents

- Configured via global config command

```
ip pim send-rp-discovery scope <ttl>
```

- Receive RP-Announcements

- Select highest C-RP IP addr as RP for group range
- Stored in Group-to-RP Mapping Cache with holdtimes

- Multicast RP-Discovery messages

- Sent to Cisco-Discovery (224.0.1.40) group
- Sent every 60 seconds or when changes detected

- RP-Discovery messages contain:

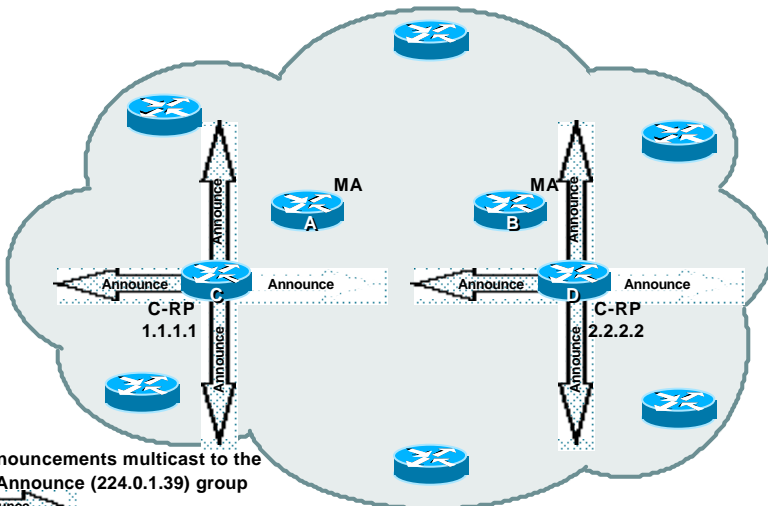
- Contents of MA's Group-to-RP Mapping Cache

# Auto-RP Fundamentals

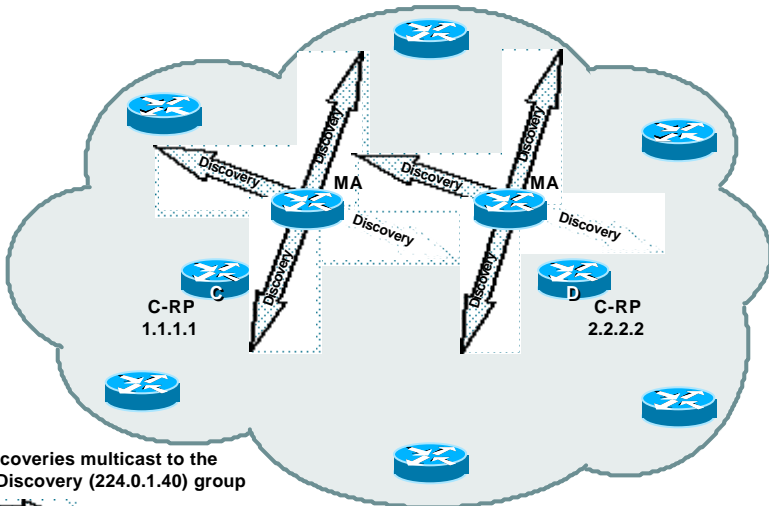
- **All Cisco routers**

- **Join Cisco-Discovery (224.0.1.40) group**
  - Automatic
  - No configuration necessary
- **Receive RP-Discovery messages**
  - Stored in local Group-to-RP Mapping Cache
  - Information used to determine RP for group range

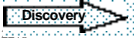
## Auto-RP—From 10,000 Feet



# Auto-RP—From 10,000 Feet



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



1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

35

## 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

2216

1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

36

# BSR Fundamentals

## • Candidate RPs

- **Configured via global config command**
  - `ip pim rp-candidate <intfc> [group-list acl]`
- **Unicast PIMv2 C-RP messages to BSR**
  - Learns IP address of BSR from BSR messages
  - Sent every `rp-announce-interval` (default: 60 sec)
- **C-RP messages contain:**
  - Group Range (default = 224.0.0.0/4)
  - Candidate's RP address
  - Holdtime = 3 x `<rp-announce-interval>`

# BSR Fundamentals

## • Bootstrap router (BSR)

- **Receive C-RP messages**
  - Accepts and stores ALL C-RP messages
  - Stored in Group-to-RP Mapping Cache w/holdtimes
- **Originates BSR messages**
  - Multicast to All-PIM-Routers (224.0.0.13) group
    - (Sent with a TTL = 1)
  - Sent out all interfaces. Propagate hop-by-hop
  - Sent every 60 seconds or when changes detected
- **BSR messages contain:**
  - Contents of BSR's Group-to-RP Mapping Cache
  - IP Address of active BSR

# BSR Fundamentals

- **Candidate bootstrap router (C-BSR)**

- **Configured via global config command**

```
ip pim bsr-candidate <intfc> <hash-length> [priority <pri>]
```

- **<intfc>**

- » Determines IP address

- **<hash-length>**

- » Sets RP selection hash mask length

- **<pri>**

- » Sets the C-BSR priority (default = 0)

- **C-BSR with highest priority elected BSR**

- **C-BSR IP address used as tie-breaker**

- » (Highest IP address wins)

- **The active BSR may be preempted**

- » New router w/higher BSR priority forces new election

# BSR Fundamentals

- **All PIMv2 routers**

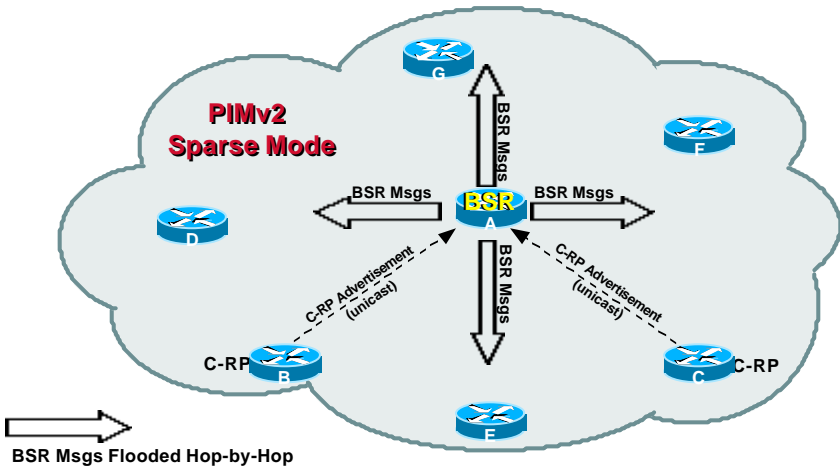
- **Receive BSR messages**

- Stored in local Group-to-RP Mapping Cache
- Information used to determine active BSR address

- **Selects RP using Hash algorithm**

- Selected from local Group-to-RP Mapping Cache
- All routers select same RP using same algorithm
- Permits RP-load balancing across group range

# BSR—From 10,000 feet



2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

41

# Auto-RP vs BSR

- **Auto-RP**

- Easy to configure
- Supports Admin. Scoped Zones
- Works in either PIMv1 or PIMv2 Cisco networks
- Cisco proprietary

- **BSR**

- Easy to configure
- Does **not** support Admin. Scoped Zones
- Non-proprietary (PIMv2 networks only)

2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

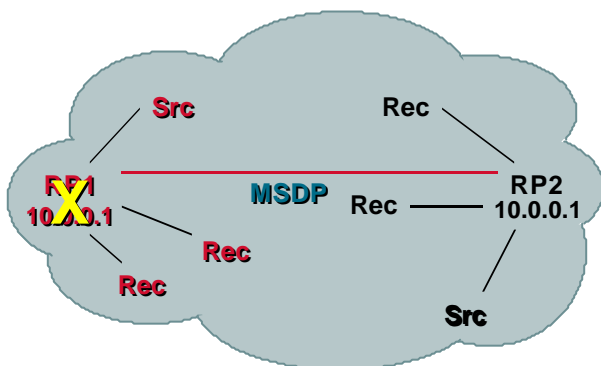
cisco.com

42

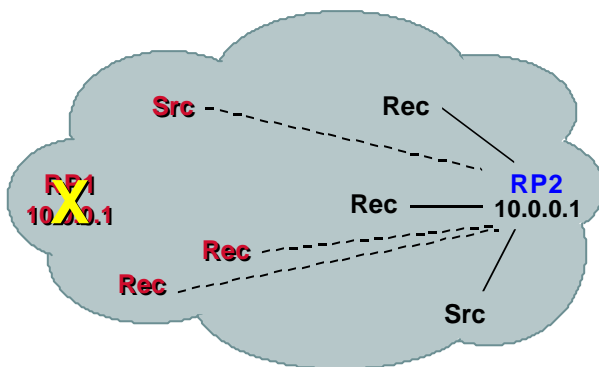
# Anycast RP—Overview

- **Uses single statically 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.
- **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—Convergence



# Anycast RP—Convergence



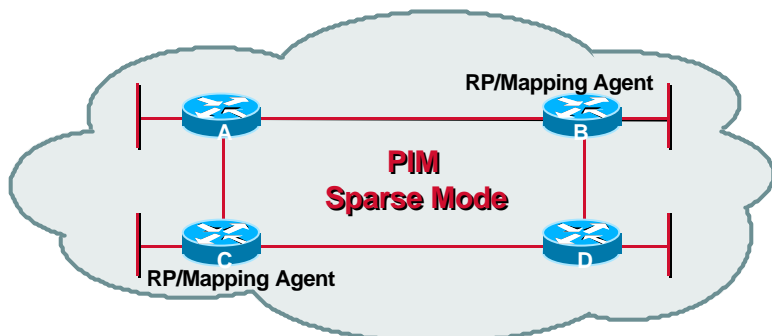
## Agenda

- IP Multicast Review
- Rendezvous Points (RP)
- **Configuring IP Multicast**
- Multicast at Layer 2
- Multicast Performance
- Multicast Traffic Engineering

# PIM Configuration Steps

- Enable Multicast Routing on **every** router
- Configure **every** interface for PIM
  - Use Sparse-Dense Mode (to support Auto-RP)
- Configure the RP (if using Sparse Mode)
  - Static RP addressing
    - RP address must be configured on every router
  - Using Auto-RP or BSR
    - Configure certain routers as Candidate RP(s)
    - All other routers automatically learn elected RP

# Fool-Proof Multicast Configuration

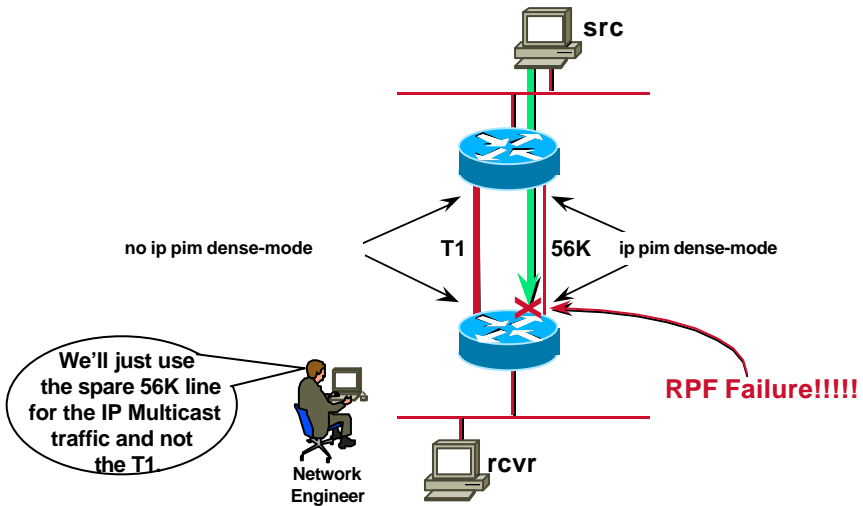


On every router: `ip multicast-routing`  
`ip pim accept-RP Auto-RP`

On every interface: `ip pim sparse-dense-mode`

On routers B and C: `ip pim send-rp-announce loopback0 scope <ttl>`  
`ip pim send-rp-discovery loopback0 scope <ttl>`

# Classic Partial Multicast Cloud Mistake



2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

49

# Beginner's Guide to IP Multicast

***JUST DO IT!***

Enable multicast on **ALL** interfaces in **ALL** routers in a given network. Failure to do so will require complex (and typically unnecessary) Multicast Traffic Engineering.

2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

50

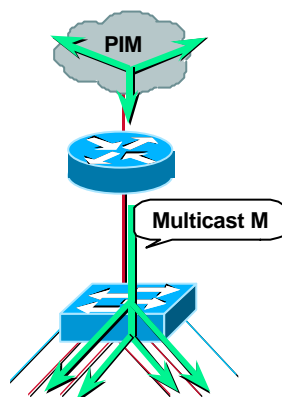
# Agenda

- IP Multicast Review
- Rendezvous Points (RP)
- Configuring IP Multicast
- **Multicast at Layer 2**
- Multicast Performance
- Multicast Traffic Engineering

## L2 Multicast Frame Switching

### **Problem:** Layer 2 Flooding of Multicast Frames

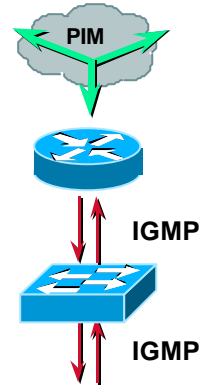
- Typical L2 switches treat multicast traffic as unknown or broadcast and must “flood” the frame to every port
- Static entries can sometimes be set to specify which ports should receive which group(s) of multicast traffic
- Dynamic configuration of these entries would cut down on user administration



# L2 Multicast Frame Switching

## Solution 1: IGMP 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!!!*



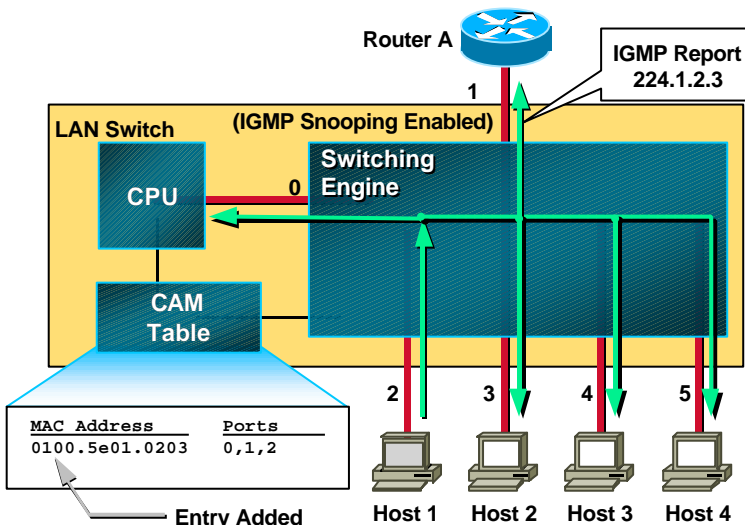
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

53

## Typical L2 Switch— 1st Join



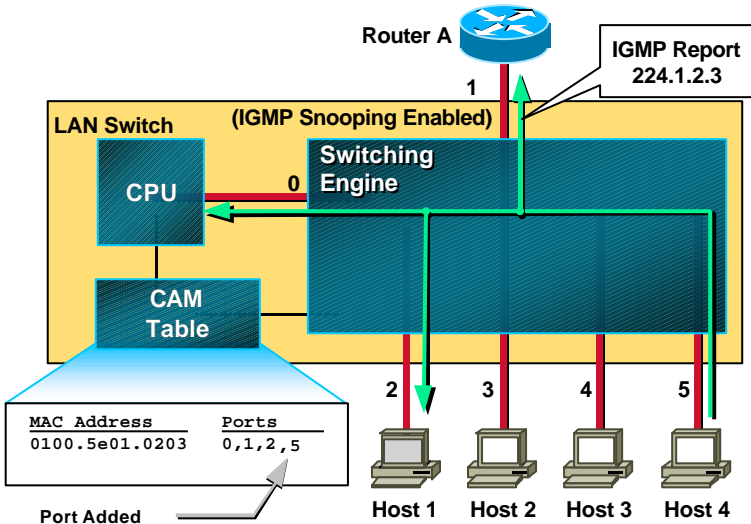
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

54

# Typical L2 Switch— 2nd Join



2216

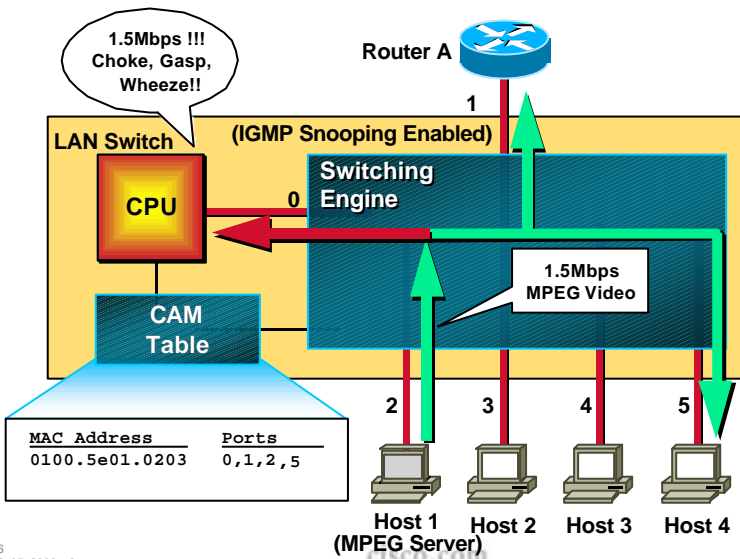
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

55

# Typical L2 Switch Meltdown!



2216

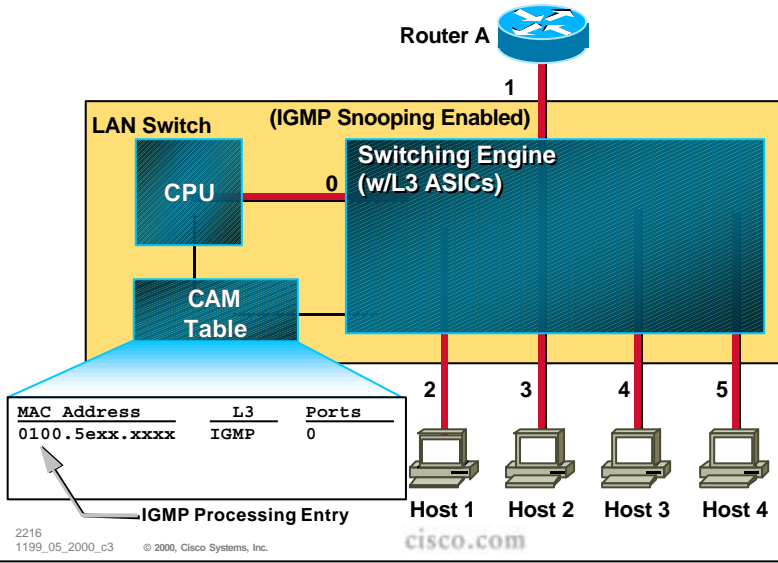
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

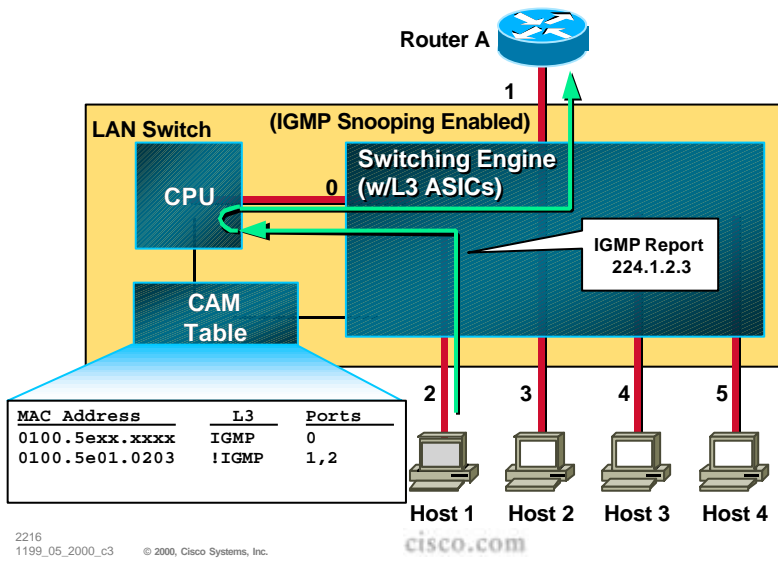
cisco.com

56

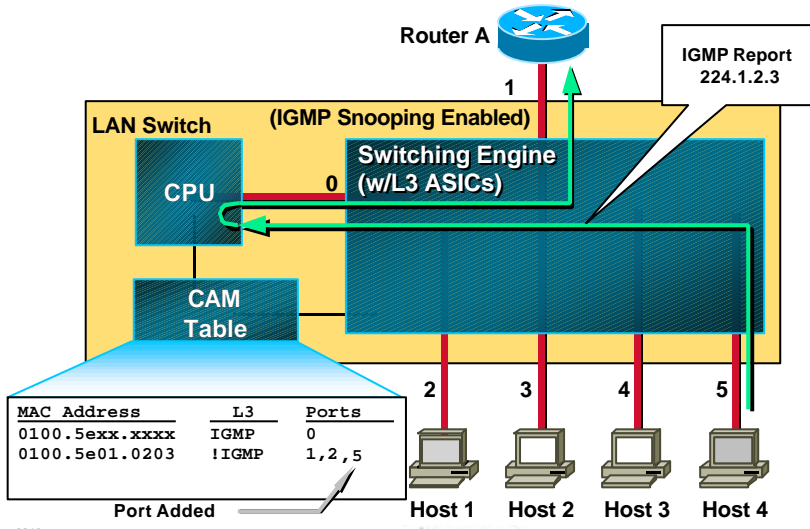
# L3 Aware Switch



# L3 Aware Switch 1st Join



# L3 Aware Switch 2nd Join



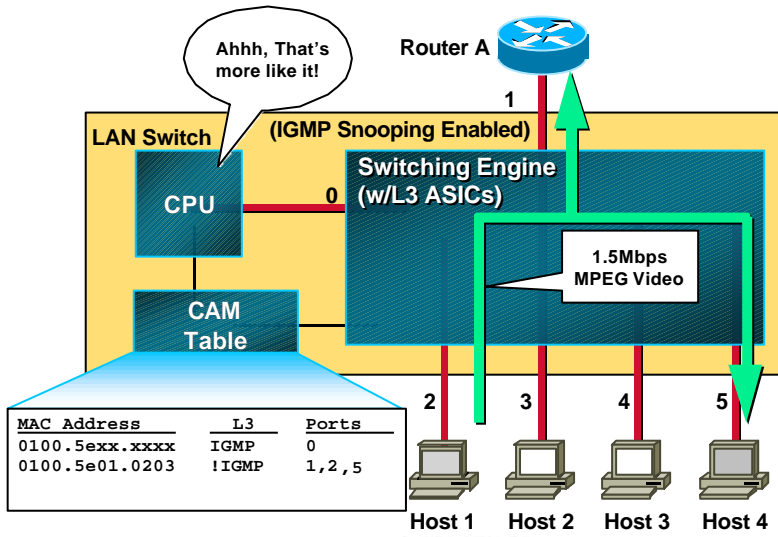
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

59

# L3 Aware Switch No Load on the CPU



2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

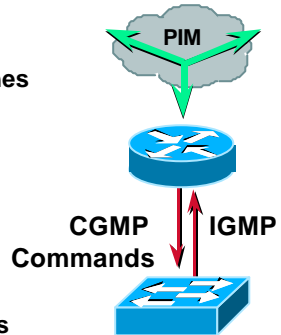
cisco.com

60

# L2 Multicast Frame Switching

## Solution 2: CGMP—Cisco Group Multicast Protocol

- Runs on both the switches and the router
- Router sends CGMP multicast packets to the switches at a well known multicast MAC address:
  - 0100.0cdd.dddd
- CGMP packet contains :
  - Type field—Join or Leave
  - MAC address of the IGMP client
  - Multicast address of the group
- Switch uses CGMP packet info to add or remove a Layer-2 entry for a particular multicast MAC address



2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

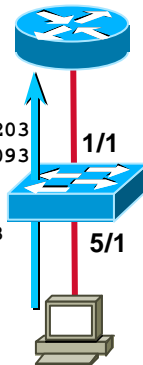
cisco.com

61

## CGMP Basics

### IGMP Report

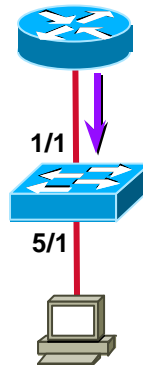
Dst MAC = 0100.5e01.0203  
Src MAC = 0080.c7a2.1093  
Dst IP = 224.1.2.3  
Src IP = 192.1.1.1  
IGMP Group = 224.1.2.3



(a)

### CGMP Join

USA = 0080.c7a2.1093  
GDA = 0100.5e01.0203



(b)

2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

62

# Summary—Frame Switches

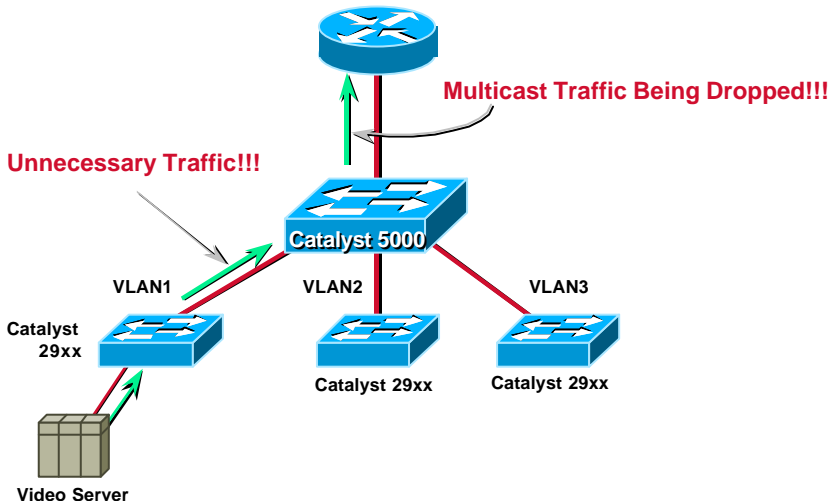
## • IGMP snooping

- 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!**

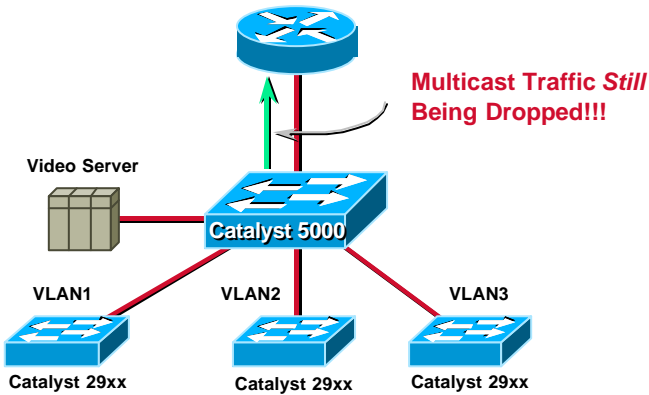
## • CGMP

- Requires Cisco routers and switches
- Can be implemented in low-cost switches

# Design Issue—Server Location

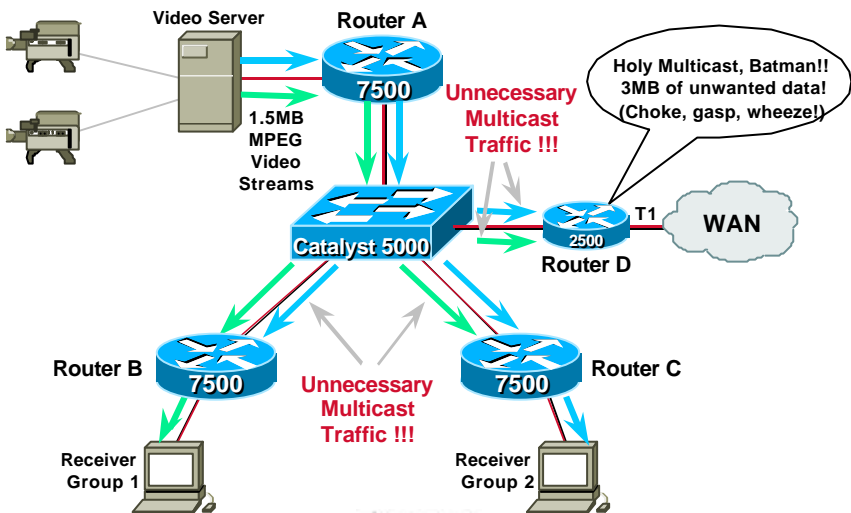


# Design Issue—Server Location

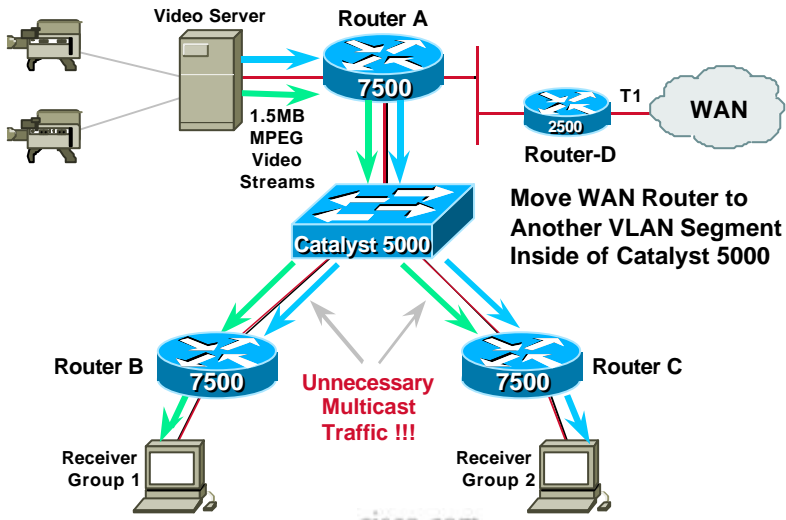


- Keep high B/W sources close to router

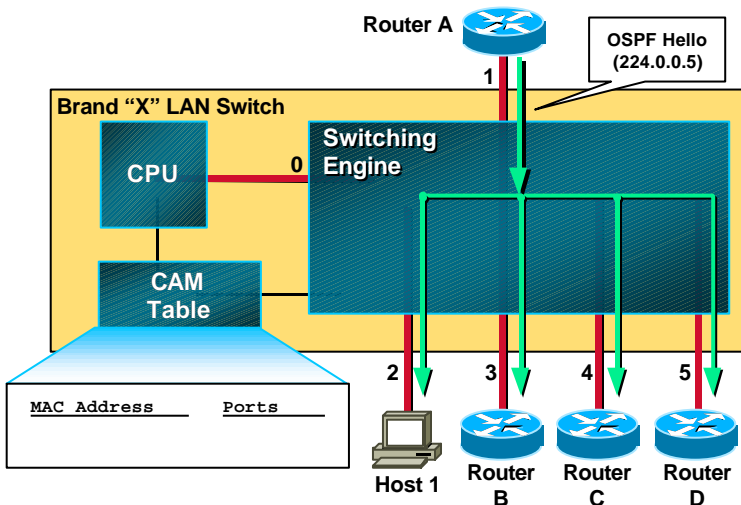
# Design Issue—Core Switch



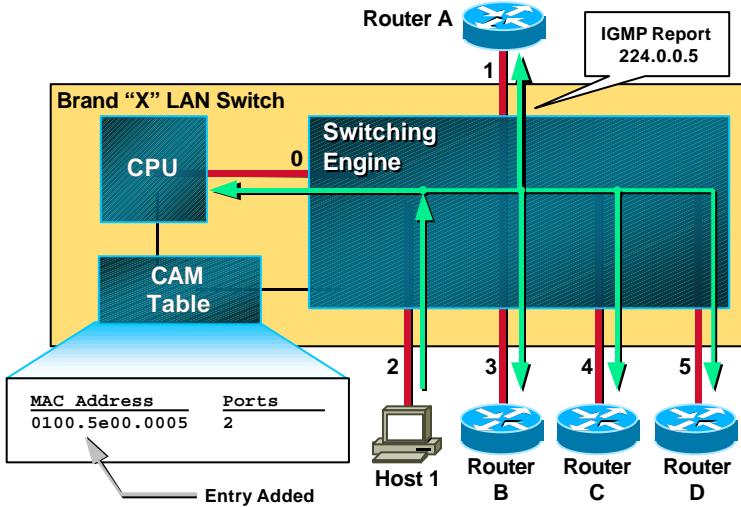
# Design Issue—Core Switch



# Design Issue—224.0.0.x Flooding



# Design Issue—224.0.0.x Flooding



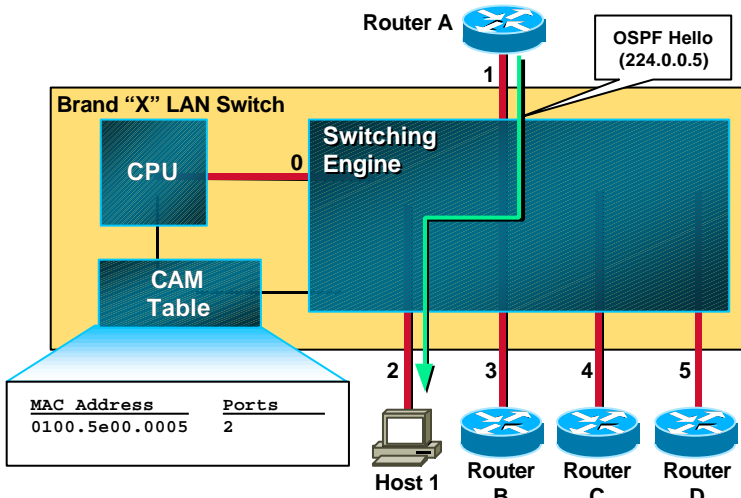
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

69

# Design Issue—224.0.0.x Flooding



2216  
1199\_05\_2000\_c3

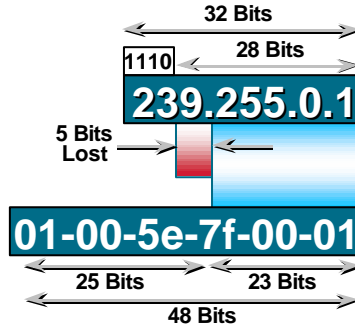
© 2000, Cisco Systems, Inc.

cisco.com

70

# Design Issue—Address Overlap

## Layer 3 IPmc Address Mapping to Layer 2 Multicast Address (FDDI and Ethernet)

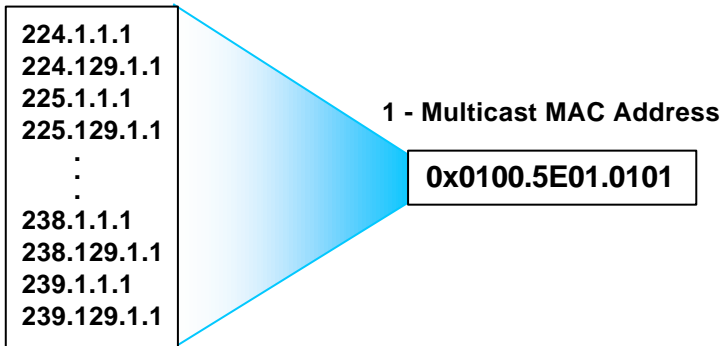


**Be Aware of the Overlap of Layer 3 Addresses to Layer 2 Addresses**

# Design Issue—Address Overlap

**Be Aware of the 32:1 Address Overlap**

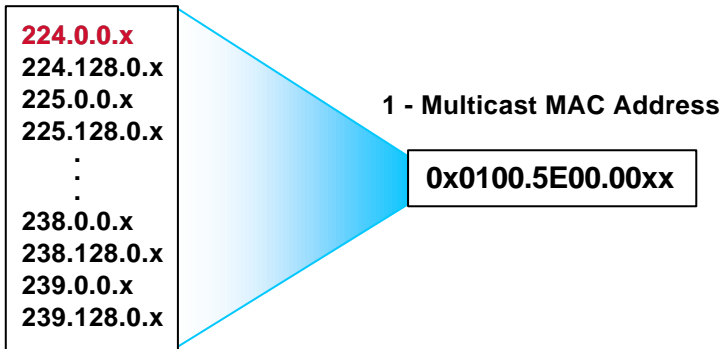
## 32 - IP Multicast Addresses



# Design Issue—Address Overlap

## Try to Avoid Addresses that Must Be Flooded

### 32 - IP Multicast Addresses



# Summary—L2 Design Issues

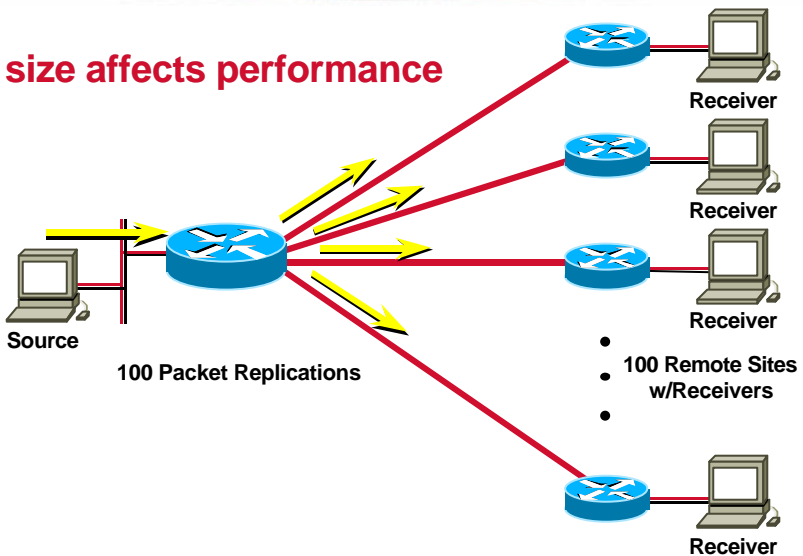
- **Pay attention to campus topology**
  - Be aware of unwanted flooding over trunks
- **Use IGMP snooping and/or CGMP**
  - Neither can solve all L2 flooding issues
  - To solve all problems requires router/switch
- **224.0.0.x flooding**
  - Watch out for switches that don't flood 224.0.0.x traffic
- **Address overlap**
  - Select group addresses to avoid L2 overlap
  - Avoid x.0.0.x group addresses when possible

# Agenda

- IP Multicast Review
- Rendezvous Points (RP)
- Configuring IP Multicast
- Multicast at Layer 2
- **Multicast Performance**
- Multicast Traffic Engineering

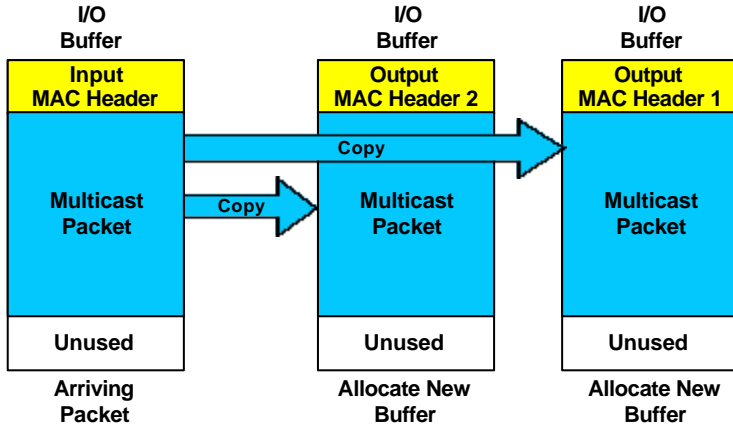
## Multicast Packet Replication

**OIL size affects performance**



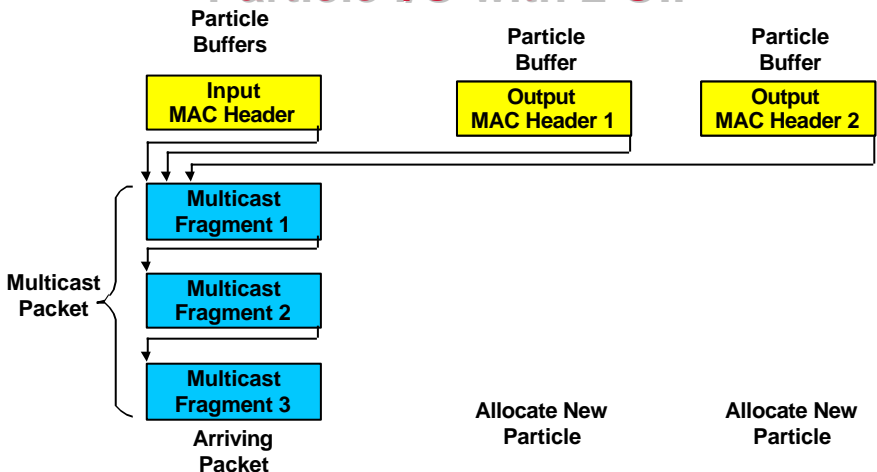
# Multicast Packet Replication

## Buffered I/O with 2 OIF



# Multicast Packet Replication

## Particle I/O with 2 OIF



# Multicast Packet Replication

- **Platforms that support Particle I/O**
  - 7200 Series Routers
  - 7100 Series Routers
  - 3600 Series Routers
  - VIP2 Line Cards
    - Local card replication only

# Multicast State Maintenance

- **CPU load factors**
  - Must send/receive Registers
  - Must send periodic Joins/Prunes
  - Must perform RPF recalculation every 5 seconds
    - Watch the total number of mroute table entries
    - Unicast route table size impacts RPF recalculation
- **Memory load factors**
  - (\*, G) entry ~ 380 bytes + OIL size
  - (S, G) entry ~ 220 bytes + OIL size
  - Outgoing interface list (OIL) size
    - Each oil entry ~ 150 bytes

# RP-Failover

- **RP failover time**

- **Function of ‘Holdtime’ in RP-Announcement**
  - Holdtime = 3 x <rp-announce-interval>
  - Default <rp-announce-interval> = 60 seconds
  - Worst-case (default) Failover ~ 3 minutes

- **Minimizing impact of RP failure**

- **Use SPTs to reduce impact**
  - Traffic on SPTs not affected by RP failure
  - Immediate switch to SPTs is on by default
  - New and/or bursty sources still a problem

# Tuning RP-Failover

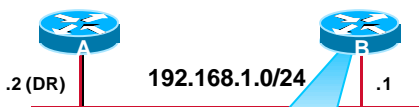
- **Tune Candidate RPs**
- **New ‘interval’ clause added for C-RPs**

```
ip pim send-rp-announce <intfc> scope <ttl>  
                        [group-list acl]  
                        [interval <seconds>]
```



- **Allows rp-announce-interval to be adjusted**
- **Smaller intervals = Faster RP failover**
- **Smaller intervals increase amount Auto-RP traffic**
- **Increase is usually insignificant**
- **Total RP failover time reduced**
- **Min. failover ~ 3 seconds**

# DR Failover



```
Rtr-B>show ip pim neighbor
PIM Neighbor Table
Neighbor Address  Interface  Uptime   Expires   Mode
192.168.1.2      Ethernet0  4d22h    00:01:18  Sparse-Dense (DR)
```

- Depends on neighbor expiration time
- Expiration Time sent in PIM query messages
  - Expiration time = 3 x <query-interval>
  - Default <query-interval> = 30 seconds
  - DR Failover ~ 90 seconds (worst case) by default

# DR Failover

- Tune PIM query interval
  - Use interface configuration command  
`ip pim query-interval <seconds>`
  - Permits DR failover to be adjusted
    - Min. DR failover ~ 3 seconds (worst case)
    - Smaller intervals increase PIM query traffic
      - Increase is usually insignificant

# Network Topology Changes

- Unicast routing must converge first
- PIM converges ~ 5 seconds after unicast
- PIM convergence algorithm
  - Entire mroute table scanned every 5 seconds
  - RPF interface recalculated for every (\*, G) and (S, G)
  - Joins/prunes/grafts triggered as needed

# Agenda

- IP Multicast Review
- Rendezvous Points (RP)
- Configuring IP Multicast
- Multicast at Layer 2
- Multicast Performance
- **Multicast Traffic Engineering**

# Tunneling Multicast

- Tunnels are used when multicast routers don't have contiguous connectivity
- Support two types of encapsulations for IP multicast traffic
  - DVMRP tunnels (IP protocol number 4)
  - GRE tunnel (IP protocol number 47)
- Both are supported by fast switching

# DVMRP Tunnels

- DVMRP tunnels may be used between a Cisco router and another DVMRP router (mrouted)
- DVMRP tunnels **cannot** be used between Cisco routers
- Commands

```
interface tunnel0
ip unnumbered ethernet0
ip pim sparse-dense
tunnel mode dvmrp
tunnel source ethernet0
tunnel destination <ip-address>
```

- Control messages are sent directly between endpoints, data gets another IP header

# GRE Tunnels

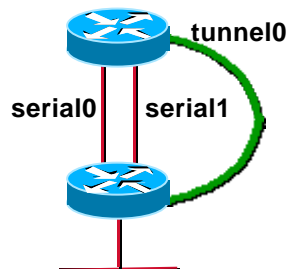
- May be used between two Cisco routers
- Looks like a P2P link for all protocols in the box
  - All packets get an extra IP header
  - Provides data sequencing and security
- Commands

```
interface tunnel0
ip unnumbered ethernet0
ip pim sparse-dense
tunnel mode gre ip
tunnel source ethernet0
tunnel destination <ip-address>
```

# Load Splitting Using Tunnels

- We use tunnels and load split across different (S,G) entries
  - Per packet when process level switching
- When doing MAC level rewrite, select among a set of equal-cost paths to the tunnel endpoint

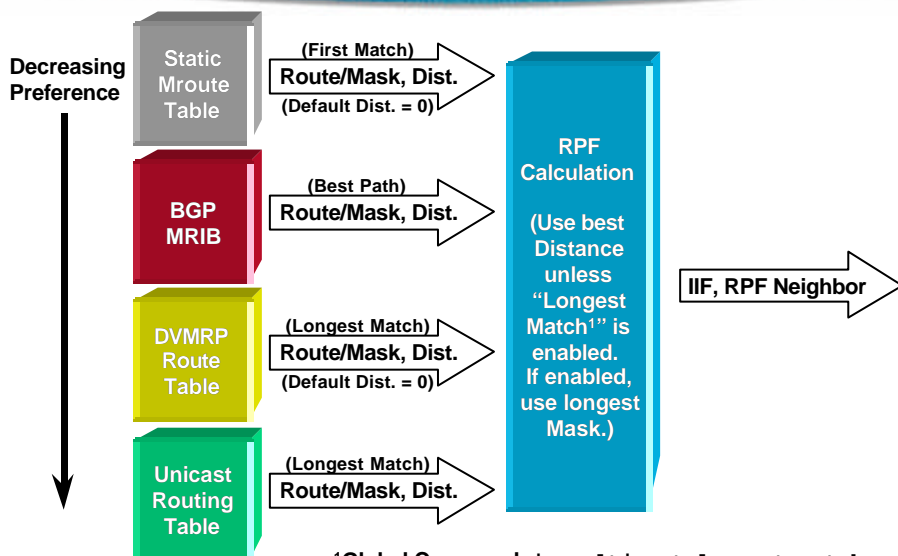
(S1, G), oif = tu0  
rewrite = serial0  
(S2, G), oif = tu0  
rewrite = serial1



# Non-Congruent Networks

- **Why would you have non-congruent unicast and multicast networks?**
  - Multicast is not enabled on all paths in the network
  - Tunnels are used to bypass normal unicast routing
  - You have policy reasons for making them different
  - You want to use idle links for multicast traffic
- **Non-congruent unicast/multicast networks**
  - RPF Calculation cannot use unicast route table
  - Other source of RPF information must be used

## PIM RPF Calculation Details



<sup>1</sup>Global Command: `ip multicast longest-match`

# Alt. Path Routing with Static Mroutes

- **Statically configured using command:**

```
ip mroute <source> <mask> [<protocol>][route-map <map>]  
                <rpf-nbr> | <interface> [<distance>]
```

- **Multiple mroutes may be specified**

- Searched in order of configuration
- Search stops on first match and route is used
- Admin distance of mroute compared to other routes

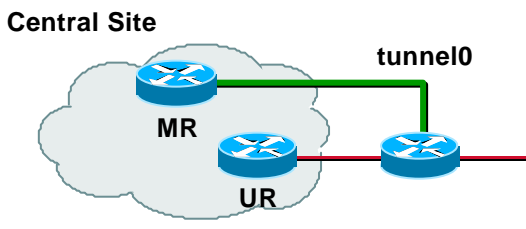
- **Mroutes have a default distance of zero**

- Preferred over all other routes by default

# Alt. Path Routing with Static Mroutes

- **A stub connection where you have a tunnel for multicast access**

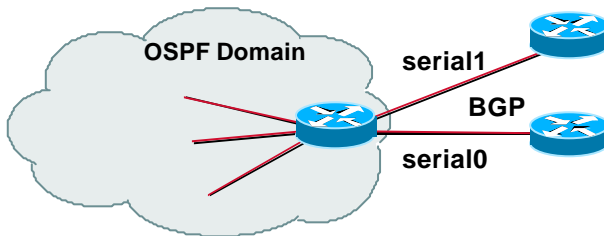
```
ip mroute 0.0.0.0 0.0.0.0 tunnel0
```



# Alt. Path Routing with Static Mroutes

- You want to tailor RPF for many routes

```
ip mroute 0.0.0.0 0.0.0.0 ospf 1 null0 255
ip mroute 0.0.0.0 0.0.0.0 serial1
```

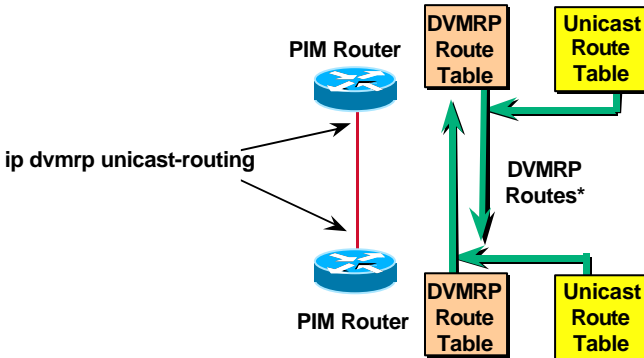


# Alt. Path Routing with DVMRP

- Use DVMRP routes for RPF Check
  - Permits separate unicast and multicast topologies
  - Can use some unicast routes and some routes from the DVMRP table
  - DVMRP routes are preferred by default
    - Default DVMRP Distance = 0
- **Warning!**
  - Care must be used to prevent route redistribution problems

# Alt. Path Routing with DVMRP

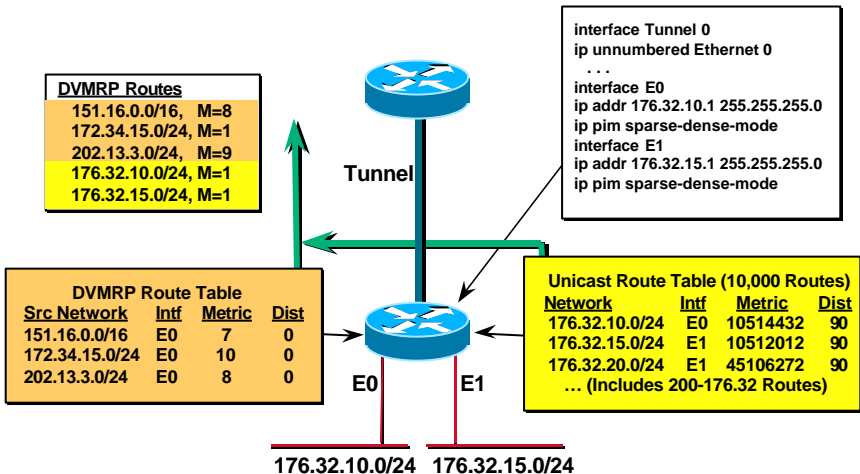
“ip dvmrp unicast-routing” causes DVMRP routes to be exchanged between two Cisco routers.



\* Routes injected from the Unicast Route Table depend no configuration options.

# Alt. Path Routing with DVMRP

Only “Connected” Unicast Routes Are Advertised by Default



# Alt. Path Routing with DVMRP

- **Command for additional route injection:**

```
ip dvmrp metric <metric>    [list <acl>]
                             [<protocol> | dvmrp]
                             [route-map <map>]
```

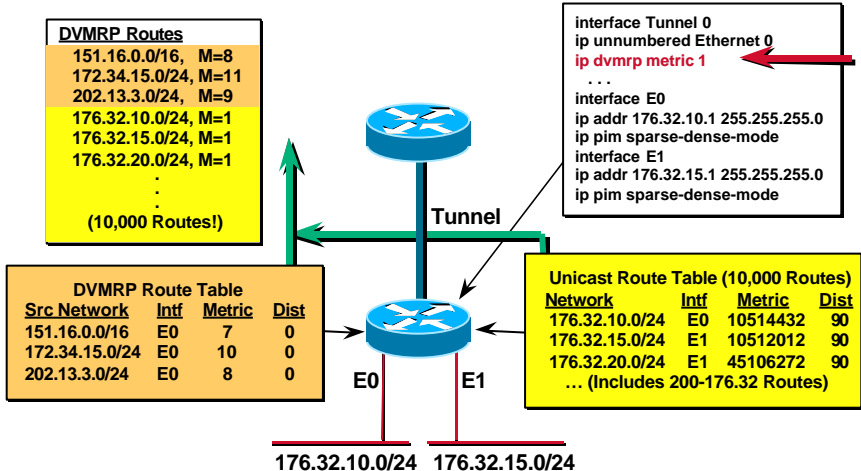
- Metric 0 means don't inject
- Multiple commands may be configured

- **Can select routes based on:**

- Routing protocol
- Route-map specification
- Enumeration using access-lists

# Alt. Path Routing with DVMRP

## Injecting **ALL** routes using the 'ip dvmrp metric' command

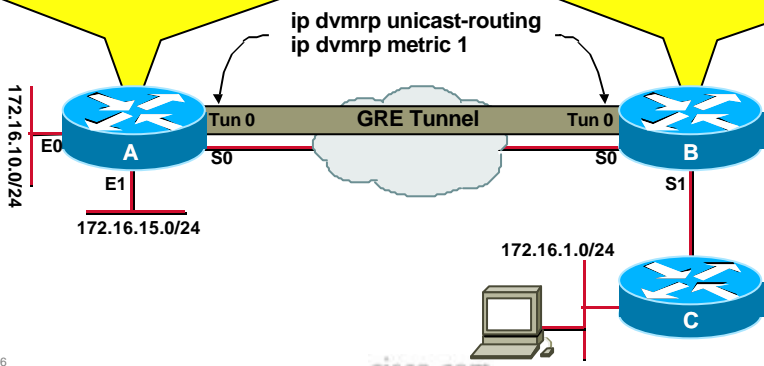


**Always Use an Access-List with the "ip dvmrp metric" Command**

# DVMRP Redistribution Problem

Network	Intf	Metric	Dist
172.16.10.0/24	E0	10514432	90
172.16.15.0/24	E1	10512012	90
172.16.1.0/24	S0	45106272	90
172.16.20.0/24	S0	45126319	90
...			

Network	Intf	Metric	Dist
172.16.10.0/24	E0	10514432	90
172.16.15.0/24	E1	10512012	90
172.16.1.0/24	S1	45034510	90
172.16.20.0/24	S1	45085628	90
...			

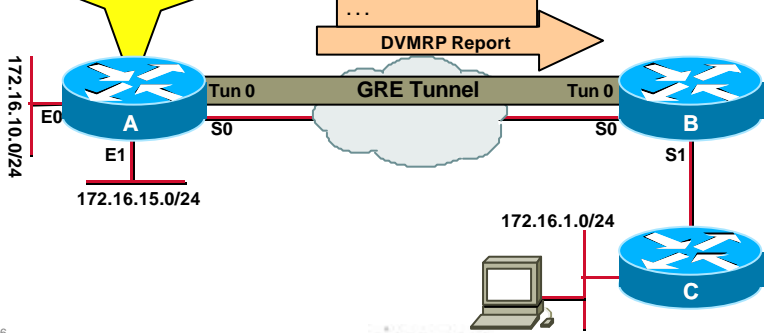


# DVMRP Redistribution Problem

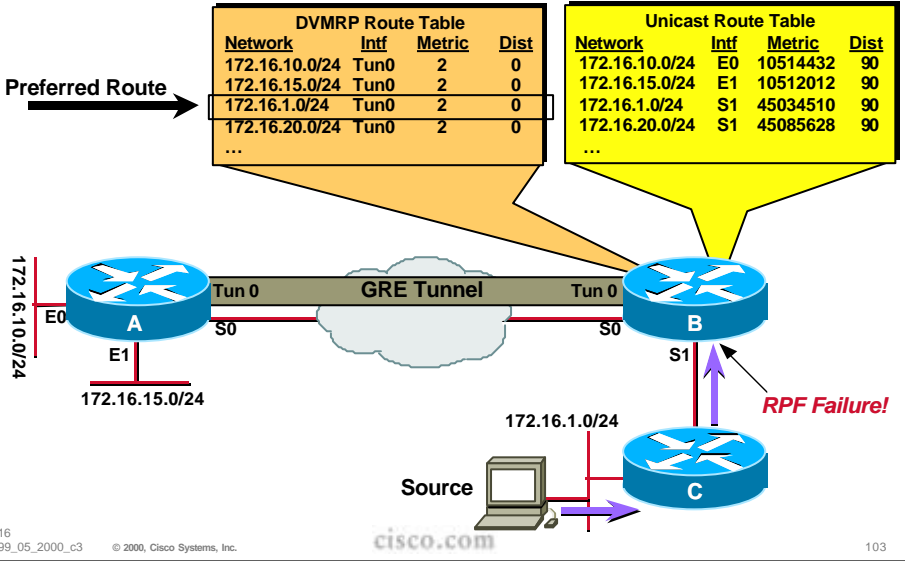
Network	Intf	Metric	Dist
172.16.10.0/24	E0	10514432	90
172.16.15.0/24	E1	10512012	90
172.16.1.0/24	S0	45106272	90
172.16.20.0/24	S0	45126319	90
...			

172.16.10.0/16,	M=1
172.16.15.0/24,	M=1
172.16.1.0/24,	M=1
172.16.20.0/24	M=1
...	

All Unicast routes are advertised as DVMRP routes as a result of the "ip dvmrp metric 1" command.



# DVMRP Redistribution Problem

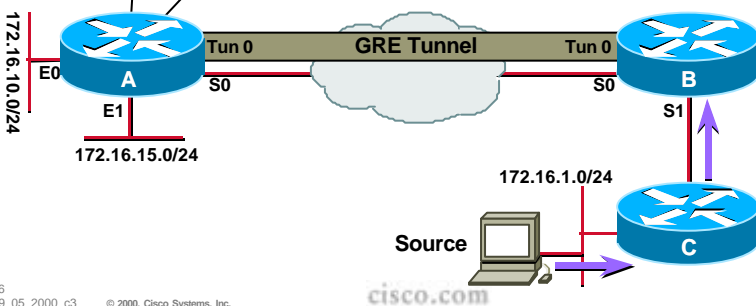


# DVMRP Redistribution Problem

## Correct Configuration

```
interface Tunnel0
ip address <address> <mask>
ip dvmrp unicast-routing
ip dvmrp metric 1 list 10

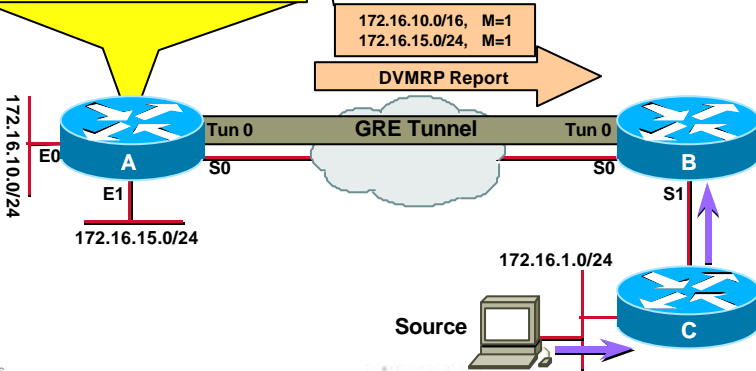
access-list 10
permit 172.16.15.0 0.0.0.255
permit 172.16.10.0 0.0.0.255
```



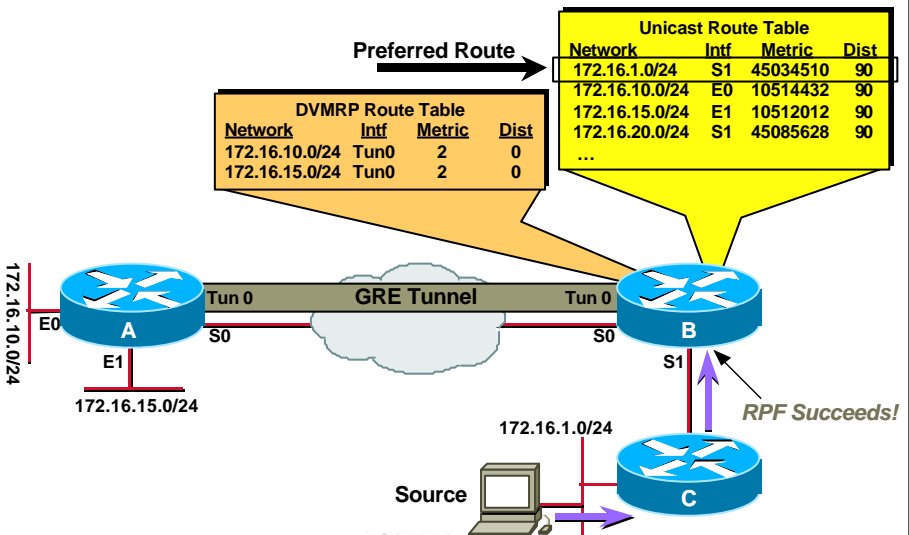
# DVMRP Redistribution Problem

Network	Intf	Metric	Dist
172.16.10.0/24	E0	10514432	90
172.16.15.0/24	E1	10512012	90
172.16.1.0/24	S0	45106272	90
172.16.20.0/24	S0	45126319	90
...			

Only selected Unicast routes are advertised as DVMRP routes as a result of the new acl on the "ip dvmrp metric 1" command.



# DVMRP Redistribution Problem



## Alt. Path Routing with DVMRP

- **Must use ACL's everywhere**
  - To prevent route loops and RPF problems.
  - Complex problem to administer.
  - ACLs may prevent network from converging after a failure.

## Alt. Path Routing with MBGP

- **MBGP: Multiprotocol BGP**  
(aka **Multicast BGP** in multicast networks)
  - Defined in RFC 2283 (extensions to BGP)
  - Can carry different types of routes
    - Unicast
    - Multicast
  - May be carried in same BGP session
  - Does not propagate multicast state info
    - Still need PIM to build Distribution Trees
  - Same path selection and validation rules
    - AS-Path, LocalPref, MED, ...

# Alt. Path Routing with MBGP

- **Separate BGP tables maintained**
  - Unicast Routing Information Base (U-RIB)
  - Multicast Routing Information Base (M-RIB)
  - New BGP 'nlri' keyword specifies which RIB
  - Allows different unicast/multicast topologies or policies
- **Unicast RIB (U-RIB)**
  - Contains unicast prefixes for unicast forwarding
- **Multicast RIB (M-RIB)**
  - Contains unicast prefixes for RPF checking

2216  
1199\_05\_2000\_c3

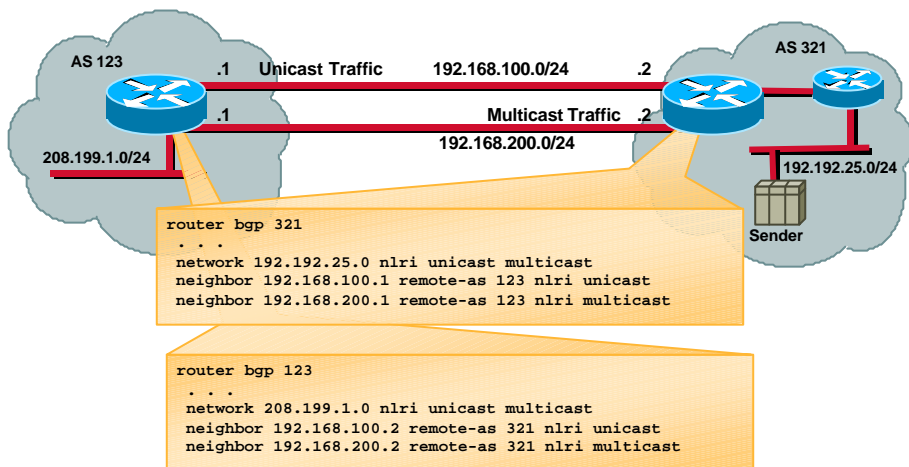
© 2000, Cisco Systems, Inc.

cisco.com

109

# Alt. Path Routing with MBGP

## Incongruent Topologies



2216  
1199\_05\_2000\_c3

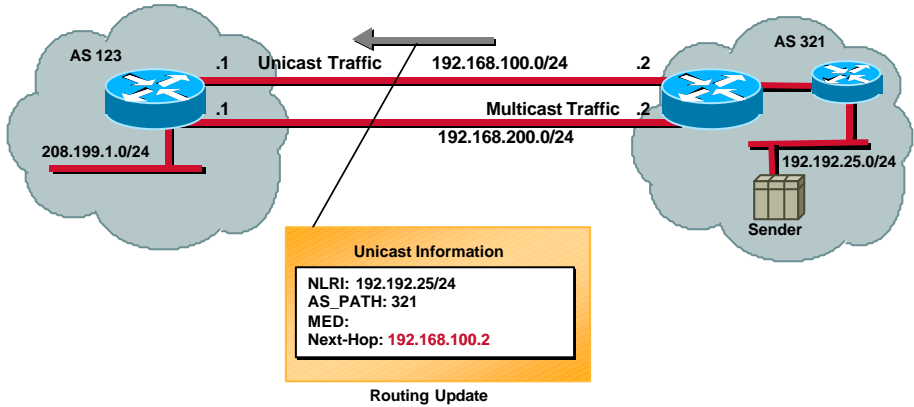
© 2000, Cisco Systems, Inc.

cisco.com

110

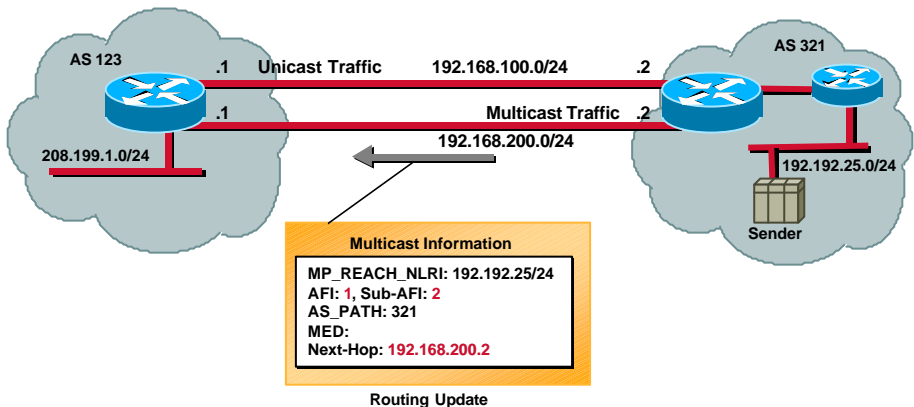
# Alt. Path Routing with MBGP

## Incongruent Topologies



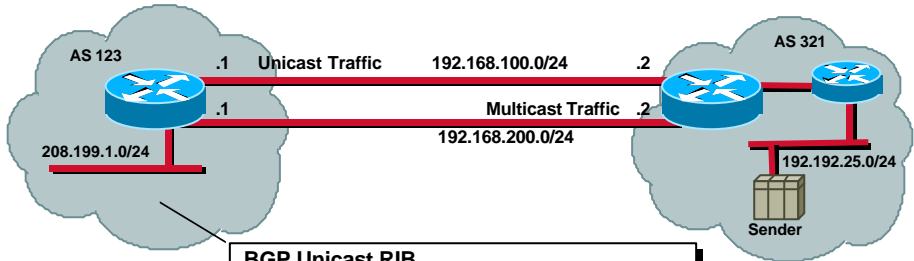
# Alt. Path Routing with MBGP

## Incongruent Topologies



# Alt. Path Routing with MBGP

## Incongruent Topologies



### BGP Unicast RIB

Network	Next-Hop	Path
192.192.25.0/24	192.168.100.2	321

### BGP Multicast RIB

Network	Next-Hop	Path
192.192.25.0/24	192.168.200.2	321

## Alternate Path Routing

### **CONCLUSION**

**Multicast Alternate Path Routing is very complex to implement and administer.**

**Avoid doing it if you can!**

# SPT Thresholds

- **Shared trees are good for router state savings**
  - When delay and frequency is not an issue
- **Source trees are good for low delay paths**
  - At the expense of router state
- **SPT thresholds allow you to use both tree types—you can tailor when you switch from shared to source trees**

# SPT Thresholds

- **How to configure SPT Thresholds**

```
ip pim spt-threshold <kbps> | infinity  
[group-list <acl>]
```
- **When you want only shared trees**

```
ip pim spt-threshold infinity
```

  - Also useful when shared trees and source trees overlap (stub networks)
- **Default SPT-Threshold is ZERO**
  - i.e. Switch to SPT Immediately

## SPT Thresholds

- **Example for globally scoped groups**

```
ip pim spt-threshold infinity group-list 1
access-list 1 permit 224.2.0.0 0.0.255.255
```

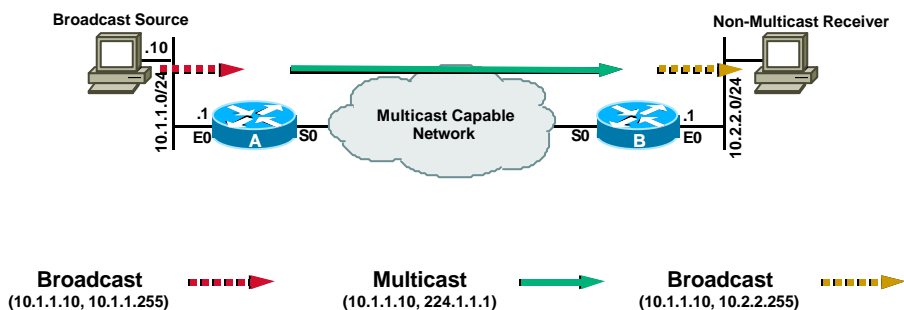
- **Must be configured in all leaf routers**

- Just configuring threshold in RP doesn't work.
- *Usage Guideline:* Configure SPT-Thresholds the same in all routers in the network.

## IP Multicast Helper Maps

- **Problem:** Hosts take longer than routers to get IP multicast deployed
- **Issue:** There are host applications deployed that use UDP broadcast transmission
- **Solution:** Have routers map broadcast address to multicast address
  - To make use of IP multicast in the infrastructure as soon as possible

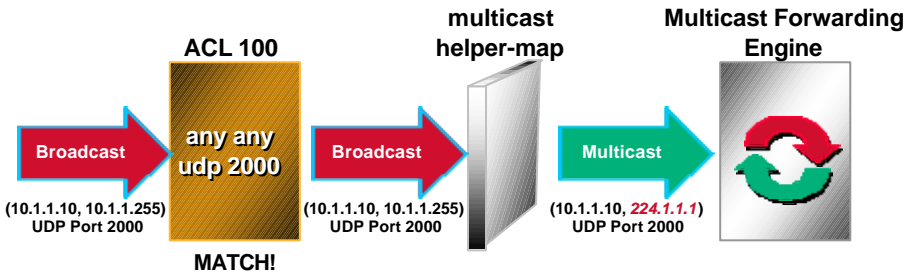
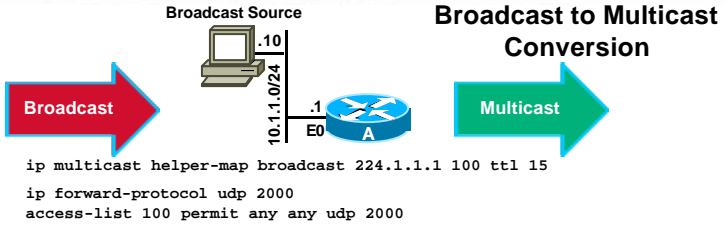
# IP Multicast Helper Maps



# IP Multicast Helper Maps

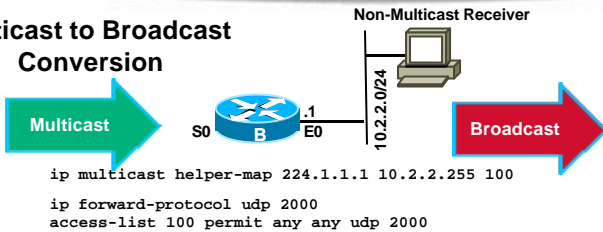
- **Mapping from broadcast to multicast**  
`ip multicast helper-map broadcast <group-address> <acl>`
- **Mapping from multicast to broadcast**  
`ip multicast helper-map <group-address> <bcast-address> <acl>`
- **Router automatically joins group on receiving end**  
`ip igmp join-group <group-address>`  
– Command automatically added to configuration

# IP Multicast Helper Maps



# IP Multicast Helper Maps

## Multicast to Broadcast Conversion



## Multicast Forwarding Engine



# BW Control via Rate-Limiting

- **IP multicast traffic can be rate-limited**
  - Any data over the limit is discarded
  - Rate-limit is on per second time slots
  - Can rate-limit on input as well as output
- **Designed to**
  - Deal with misbehaving sources
  - Sharing bandwidth with unicast traffic

# BW Control via Rate-Limiting

- **Interface-Based Rate-Limiting**
  - Limits **total** rate of all multicast flows in/out of an interface
- **Flow-Based Rate-Limiting**
  - Limits rate of each **individual** (S, G) or (\*,G) flow in/out of an interface

**Note: Both Interface and Flow-based limits may not be used on an interface at the same time!**

# BW Control via Rate-Limiting

## • Rate limit interface command

```
ip multicast rate-limit in | out { [video] | [whiteboard] }  
  [group-list <acl>] [source-list <acl>] [<kbps>]
```

- An Interface-based rate limit is defined when the optional Group and/or Source ACLs are not used.
- A Flow-based rate limit is defined when the optional Group and/or Source ACLs are used
  - Multiple Flow-based entries may be used per interface
  - Flow-based and Interface-based limits may not be used at the same time

## • Typical Rate-Limit Application

- Use “out” form of command on WAN links
- Set <kbps> to desired percentage usage of link BW

# BW Control via Rate-Limiting

## • Limiting video or whiteboard streams

- Add “video” or “whiteboard” keywords
- Requires ‘ip sdr listen’ to be enabled
- Streams identified using info from sdr cache

### – Example:

- Listening to IETF broadcast behind a 128kbps link
- They’re sending video at 128kbps and audio at 64kbps

### – Requirements

- Want crystal clear audio
- Want good response to data actions (interactive)
- Marginal video acceptable

### – Configuration

```
interface serial 0  
  ip multicast rate-limit out video 48
```

- Router will differentiate UDP port numbers for the same group

# Debugging Rate-Limits

```
Rtr-A> show ip mroute 224.1.1.1
(*, 224.1.1.1), 7w0d/00:03:29, RP 171.69.10.13, flags: S
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 04:23:22/00:03:25, Int Limit 512 kbps
    Serial1, Forward/Sparse-Dense, 04:23:28/00:03:29, Limit 56 kbps
    Serial3, Forward/Sparse-Dense, 04:23:28/00:02:37,

(128.9.160.238, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, Int limit 512 kbps
    Serial1, Forward/Sparse-Dense, 00:00:48/00:03:29, Limit 56 kbps
    Serial3, Forward/Sparse-Dense, 00:00:48/00:03:25,

(*, 224.2.2.2), 00:00:38/00:02:51, RP 171.68.20.1, flags: S
  Incoming interface: Ethernet0, RPF nbr 171.70.100.1, Int Limit 1000 kbps
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:38/00:02:51, Int limit 512 kbps
    Serial1, Forward/Sparse-Dense, 00:00:38/00:03:29,
    Serial3, Forward/Sparse-Dense, 00:00:38/00:03:25,
. . .
```

Interface-based Output Rate-limit

- **Total** output multicast traffic rate on Serial0 will not exceed 512 Kbps.

# Debugging Rate-Limits

```
Rtr-A> show ip mroute 224.1.1.1
(*, 224.1.1.1), 7w0d/00:03:29, RP 171.69.10.13, flags: S
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 04:23:22/00:03:25, Int Limit 512 kbps
    Serial1, Forward/Sparse-Dense, 04:23:28/00:03:29, Limit 56 kbps
    Serial3, Forward/Sparse-Dense, 04:23:28/00:02:37,

(128.9.160.238, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, Int limit 512 kbps
    Serial1, Forward/Sparse-Dense, 00:00:48/00:03:29, Limit 56 kbps
    Serial3, Forward/Sparse-Dense, 00:00:48/00:03:25,

(*, 224.2.2.2), 00:00:38/00:02:51, RP 171.68.20.1, flags: S
  Incoming interface: Ethernet0, RPF nbr 171.70.100.1, Int Limit 1000 kbps
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:38/00:02:51, Int limit 512 kbps
    Serial1, Forward/Sparse-Dense, 00:00:38/00:03:29
    Serial3, Forward/Sparse-Dense, 00:00:38/00:03:25
. . .
```

Interface-based Input Rate-limit

- **Total** input multicast traffic rate on Ethernet0 will not exceed 1 Mbps.

# Debugging Rate-Limits

```

Rtr-A> show ip mroute 224.1.1.1
(*, 224.1.1.1), 7w0d/00:03:29, RP 171.69.10.13, flags: S
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 04:23:22/00:03:25, Int Limit 512 kbps
    Serial1, Forward/Sparse-Dense, 04:23:28/00:03:29, Limit 56 kbps
    Serial3, Forward/Sparse-Dense, 04:23:28/00:02:37,

(128.9.160.238, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, Int limit 512 kbps
    Serial1, Forward/Sparse-Dense, 00:00:48/00:03:29, Limit 56 kbps
    Serial3, Forward/Sparse-Dense, 00:00:48/00:03:25,

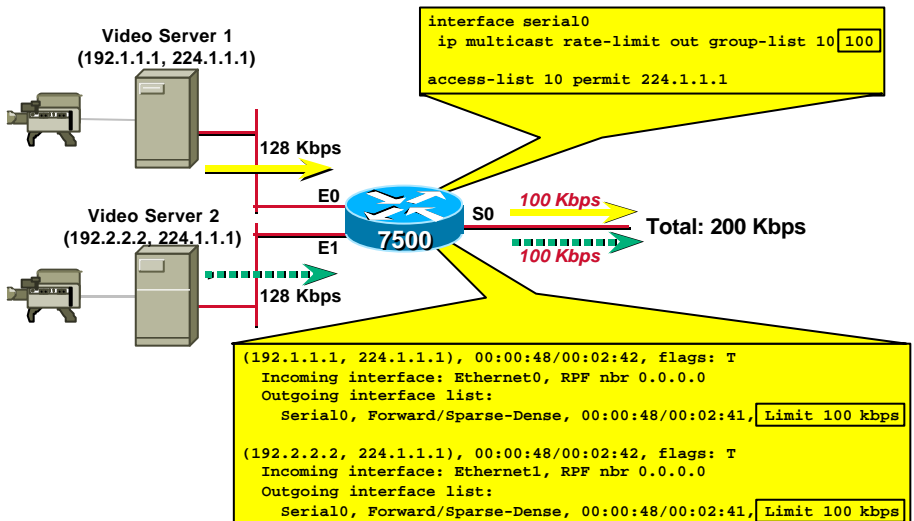
(*, 224.2.2.2), 00:00:38/00:02:51, RP 171.68.20.1, flags: S
  Incoming interface: Ethernet0, RPF nbr 171.70.100.1, Int Limit 1000 kbps
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:38/00:02:51, Int limit 512 kbps
    Serial1, Forward/Sparse-Dense, 00:00:38/00:03:29
    Serial3, Forward/Sparse-Dense, 00:00:38/00:03:25
  . . .
    
```

Flow-based Output Rate-limit

- Each **individual** output multicast flow on Serial1 will not exceed 56 Kbps.
- The **total** output on Serial1 is the sum of all flows and **can** exceed 56Kbps.

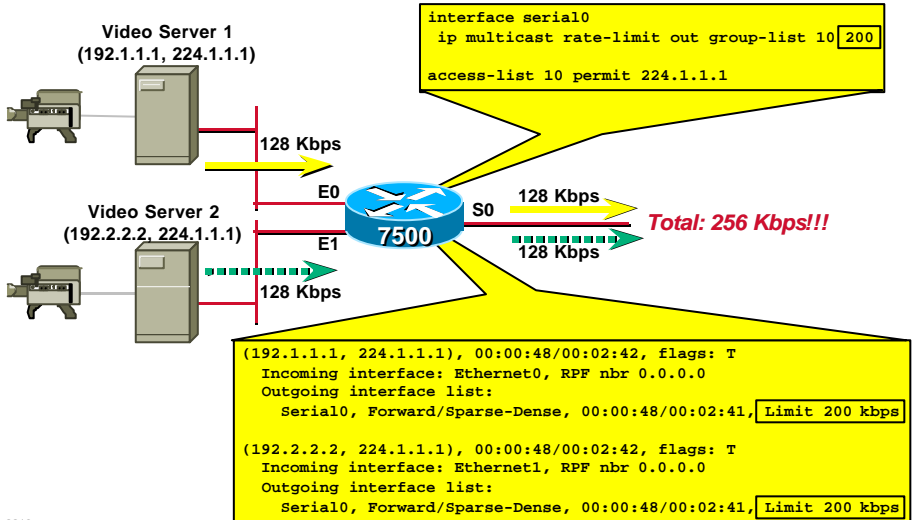
# BW Control via Rate-Limiting

## Example: Flow-based Rate-Limiting



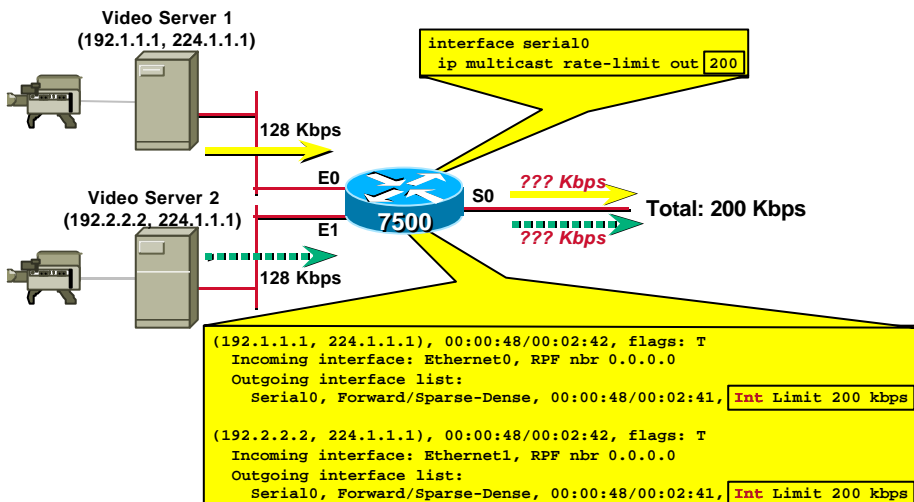
# BW Control via Rate-Limiting

## Example: Flow-based Rate Limiting



# BW Control via Rate-Limiting

## Example: Interface-based Rate Limiting



# BW Control via Rate-Limiting

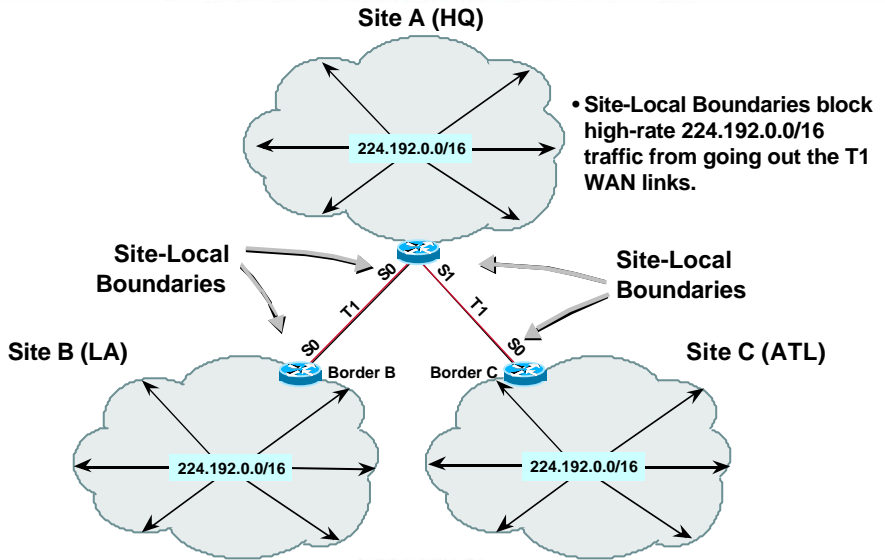
## Summary

Flow-based rate-limits do not limit the *total* aggregate of all the matching flows. Therefore, the use of interface-based rate-limits are recommended when an upper bound on multicast traffic rates is desired.

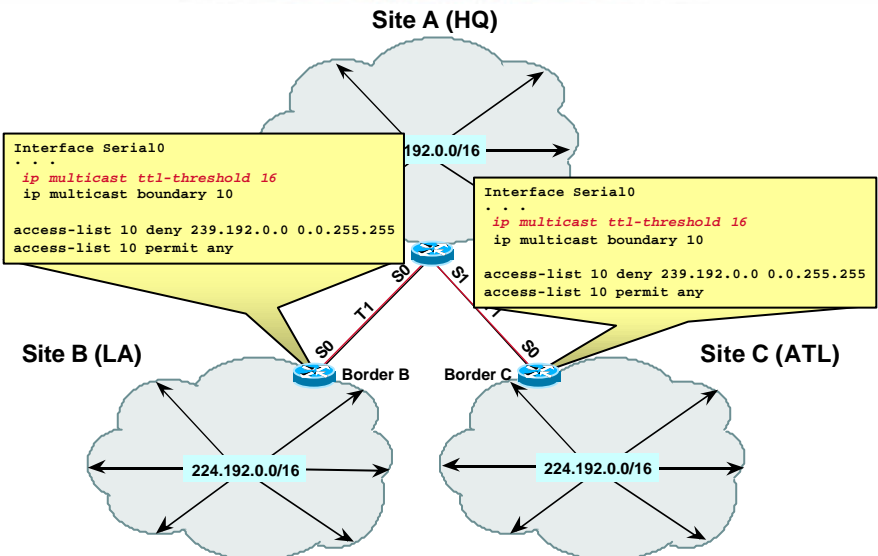
# BW Control via Admin-Scoping

- **Limit high-BW sources to local site**
- **Use administratively-scoped zones**
  - **Simple scoped zone example:**
    - 239.192.0.0/16 = Site-Local Scope Zone
    - 224.0.1.0 - 238.255.255.255 = Global scope (Internet) zone
  - **High-BW sources use only site-local zone groups**
  - **Low-Med. BW, Internet-wide sources use global zone**

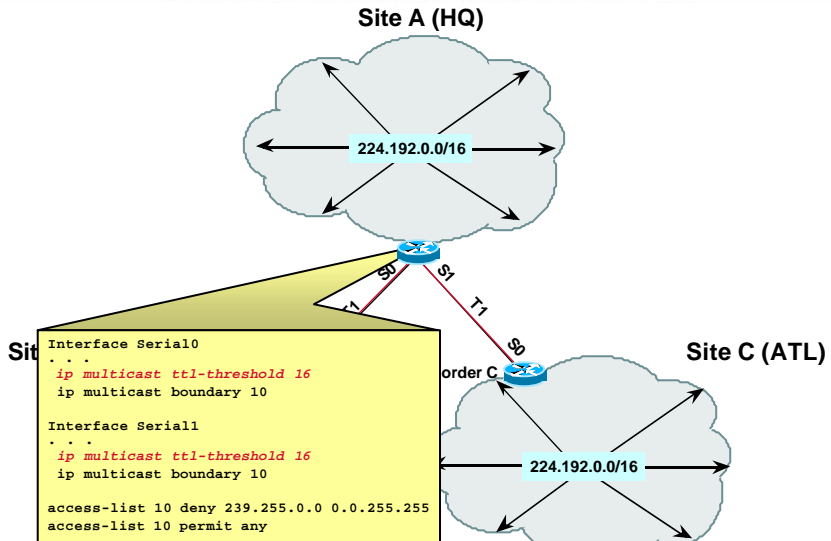
# BW Control via Admin-Scoping



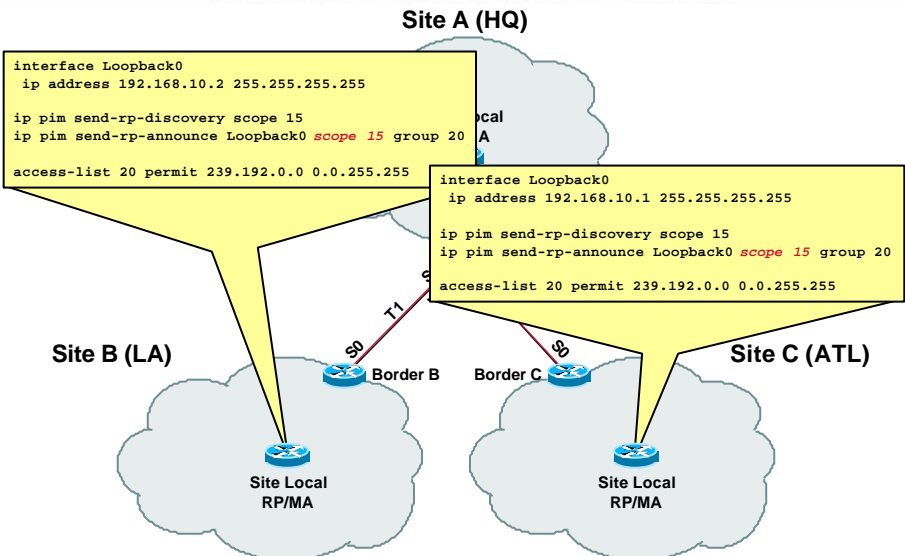
# BW Control via Admin-Scoping



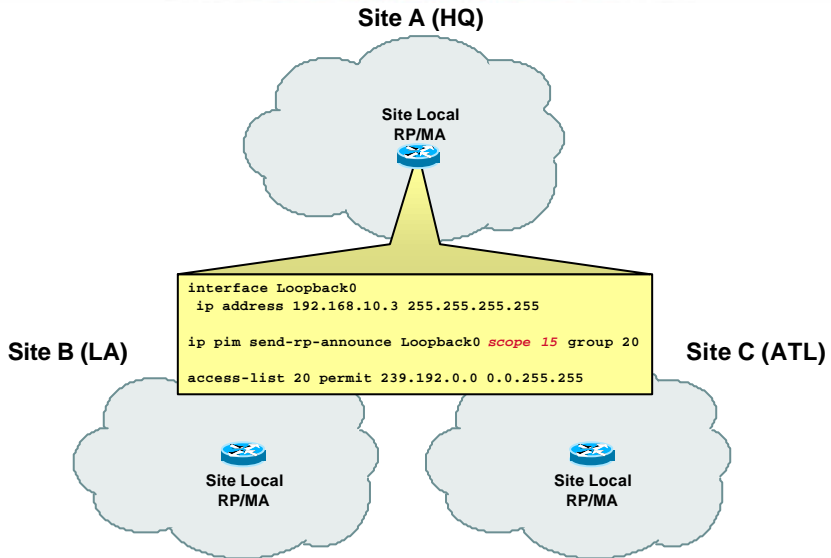
# BW Control via Admin-Scoping



# BW Control via Admin-Scoping



# BW Control via Admin-Scoping



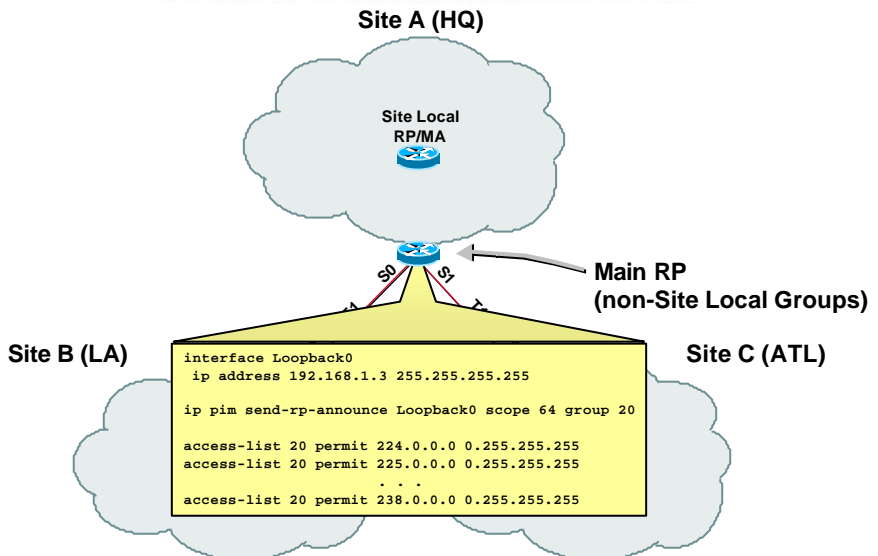
2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

139

# BW Control via Admin-Scoping



2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

cisco.com

140

# Documentation and Contact Info

- **EFT/Beta Site Web Page:**  
<ftp://ftpeng.cisco.com/ipmulticast.html>
- **TAC Support Mailing List:**  
[tac@cisco.com](mailto:tac@cisco.com)
- **Customer Support Mailing List:**  
[cs-ipmulticast@cisco.com](mailto:cs-ipmulticast@cisco.com)

# If All Else Fails—RTFB<sup>1</sup>



<sup>1</sup> Read this fine book



# Deploying IP Multicast

## Session 2216

2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

[cisco.com](http://cisco.com)

143



# Please Complete Your Evaluation Form

## Session 2216

2216  
1199\_05\_2000\_c3

© 2000, Cisco Systems, Inc.

[cisco.com](http://cisco.com)

144

# CISCO SYSTEMS



EMPOWERING THE  
INTERNET GENERATION<sup>SM</sup>