



## Other Related Presentations

- **Multicast Sessions**

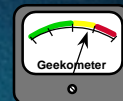
<u>Session #</u>	<u>Title</u>
303	Introduction to IP Multicast
306	PIM Protocol Concepts
314	Deploying IP Multicast
320	Advances in IP Multicast

- **MBGP Related Sessions**

<u>Session #</u>	<u>Title</u>
309	Deploying BGP
317	Advanced BGP & Troubleshooting

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

## Agenda



- **Which Mode? Sparse or Dense**
- **Multicast over Campus Networks**
- **Multicast over NBMA Networks**
- **RP Engineering**
- **Multicast Bandwidth Control**
- **Network Convergence**

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

## 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
  - Traffic engineering nearly impossible
  - Doesn't scale as well as Sparse mode

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

## Which Mode— Sparse or Dense

- **Sparse mode**
  - One additional configuration step
    - Must configure the RP
  - Very efficient
    - Uses explicit join model
    - Traffic only flows to where it's needed
    - Router state only created along flow paths
  - Traffic engineering possible
    - Use shared-trees rooted at different RP's
  - Scales better than dense mode

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

## Which Mode— Sparse or Dense

- PIMv2 changes the paradigm for defining a group as sparse or dense
- The old way was that the router determined the mode
- The new way allows for both sparse and dense groups on a router at the same time

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

## Which Mode— Sparse or Dense

- Cisco implements this change with the “ip pim sparse-dense-mode” interface command
- If a router knows about an RP the group will be sparse
- Otherwise the group will be dense

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

## Which Mode— Sparse or Dense

- Under rare situations you may want to specifically configure an interface to be sparse or dense
  - If you are an ISP
  - If you are connecting to another multicast cloud (i.e., two separate PIM clouds)

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

## Which Mode— Sparse or Dense

### Recommendation: Sparse-Dense

- Mode determined dynamically
  - IF RP exists for group, mode = Sparse
  - ELSE mode = Dense
- Use Auto-RP to define RP(s)
  - Avoids configuring RP on every router
  - Allows easy RP assignment/reassignment

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

## Agenda

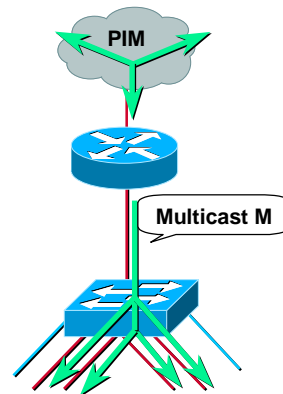
- Which Mode? Sparse or Dense
- **Multicast over Campus Networks**
- Multicast over NBMA Networks
- RP Engineering
- Multicast Bandwidth Control
- Network Convergence

www.cisco.com

## 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

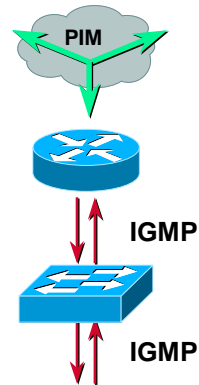


www.cisco.com

## L2 Multicast Frame Switching

### Solution 1: IGMP Snooping

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



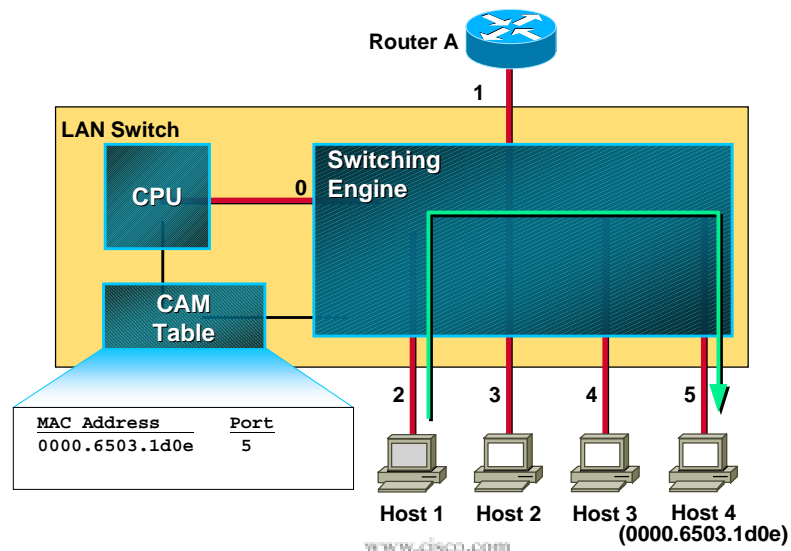
www.cisco.com

## IGMP Messages—Key Points

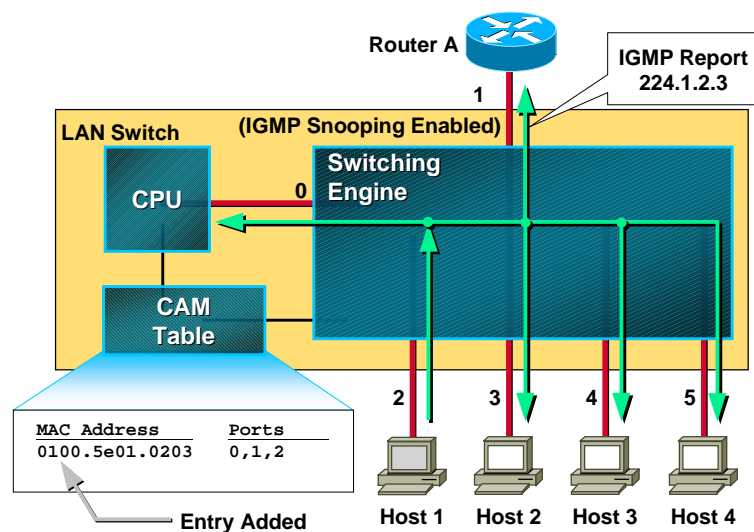
- Sent to the multicast group address
  - IGMP reports
  - IGMP leaves (old host implementations)
  - IGMP group-specific queries
- Have same L2 headers as data flow
  - IGMP traffic is indistinguishable from data at Layer 2

www.cisco.com

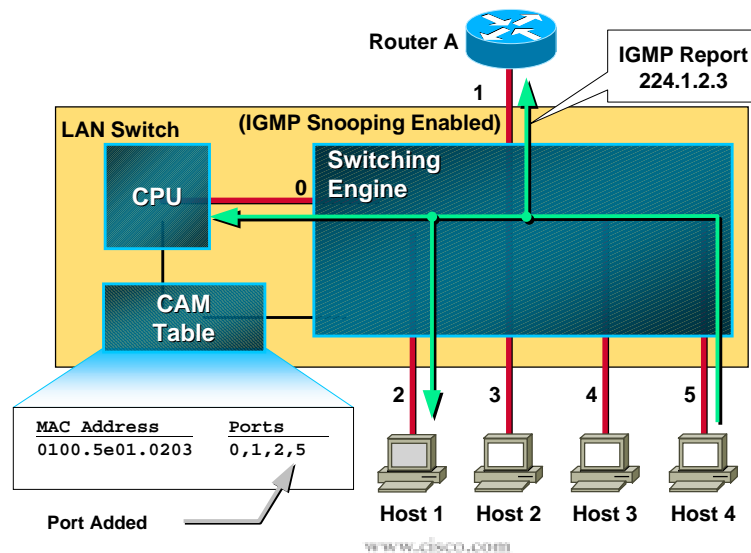
## Typical L2 Switch Architecture



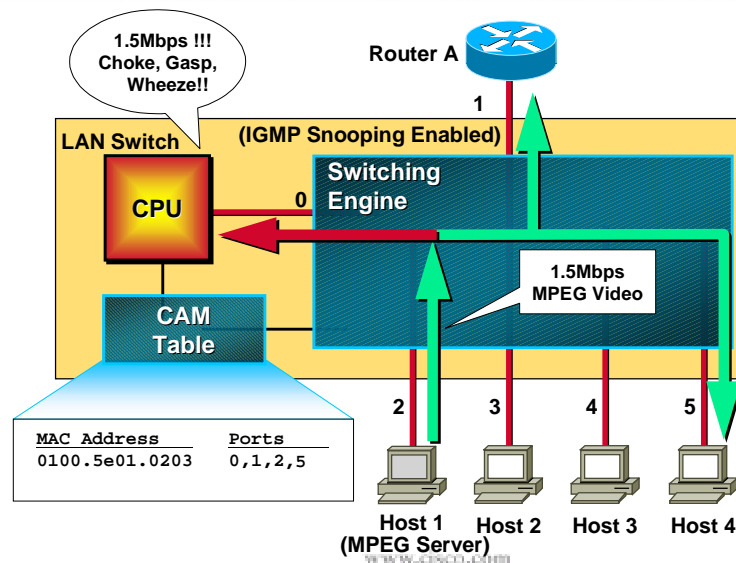
## Typical L2 Switch— 1st Join



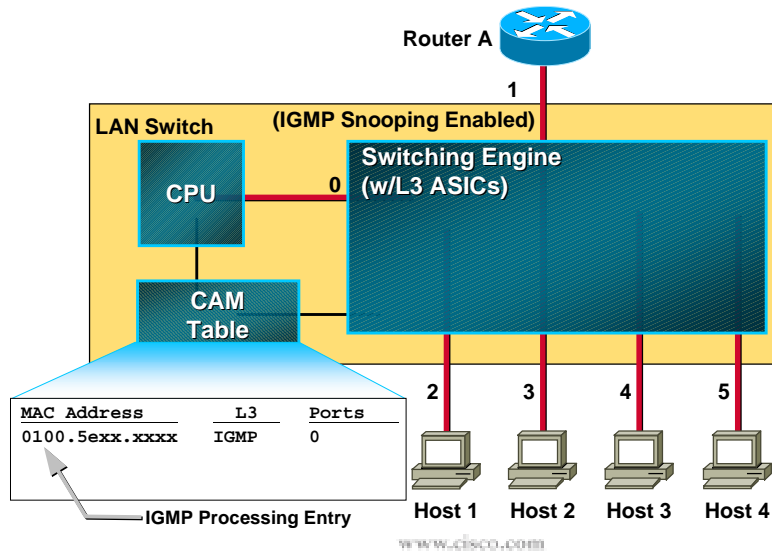
## Typical L2 Switch— 2nd Join



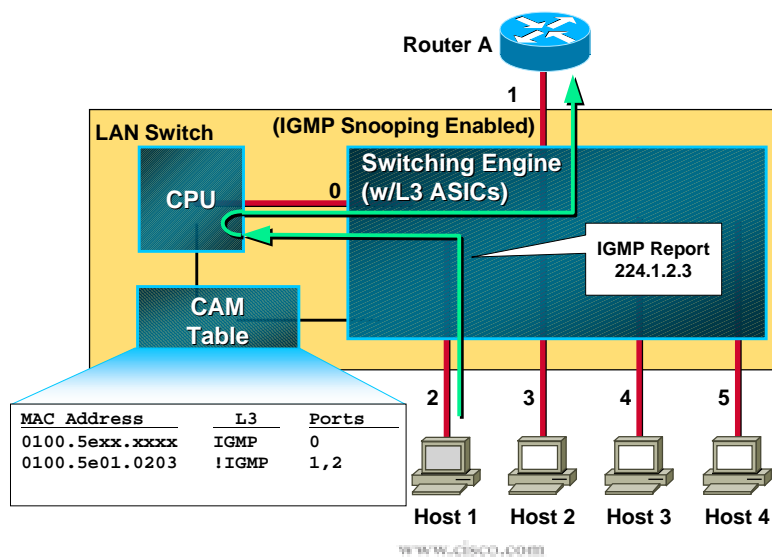
## Typical L2 Switch Meltdown!



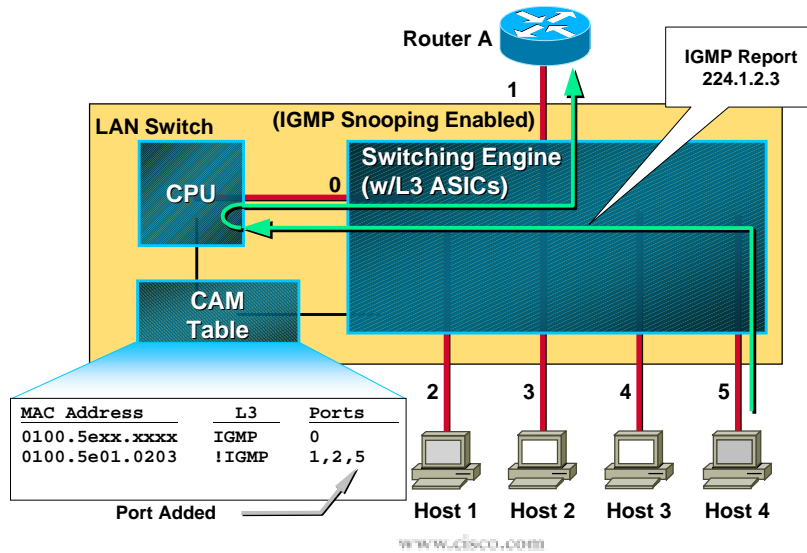
## L3 Aware Switch



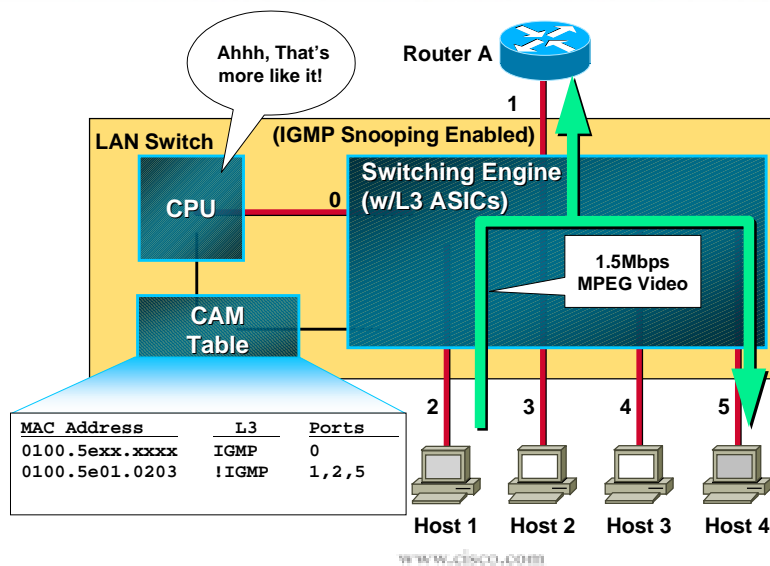
## L3 Aware Switch 1st Join



## L3 Aware Switch 2nd Join



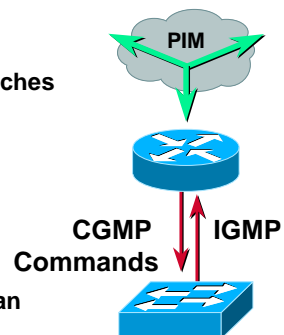
## L3 Aware Switch



## L2 Multicast Frame Switching

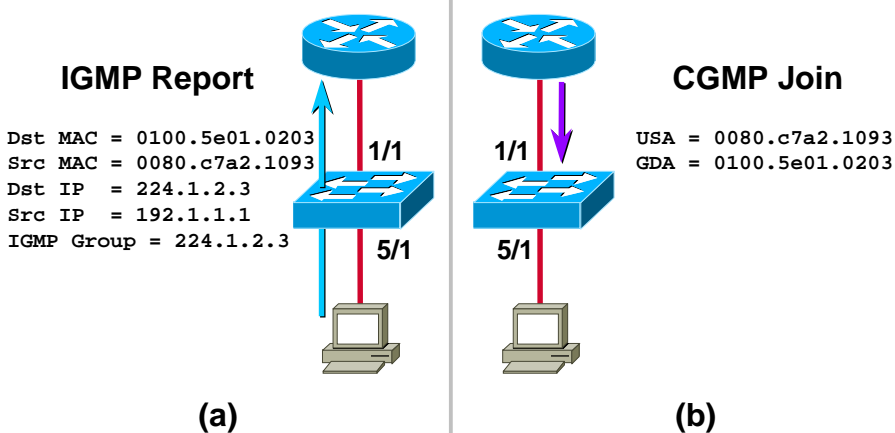
### Solution 2: CGMP—Cisco Group Multicast Protocol

- 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 an entry for a particular multicast MAC address



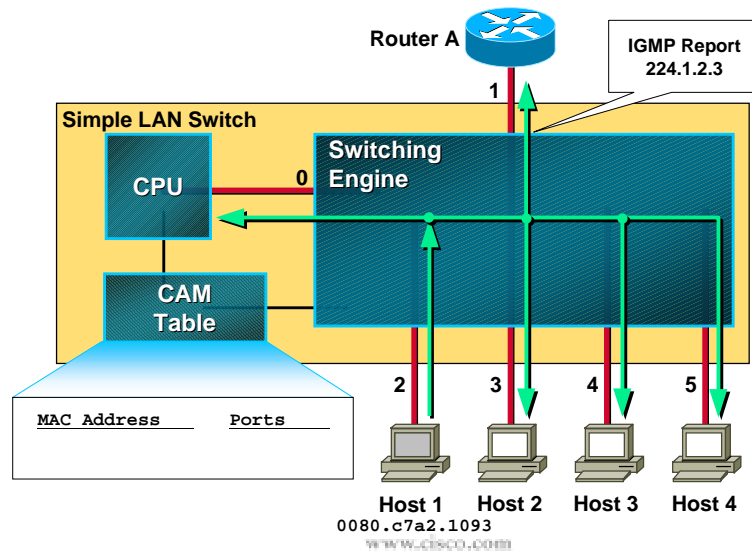
www.cisco.com

## CGMP Basics

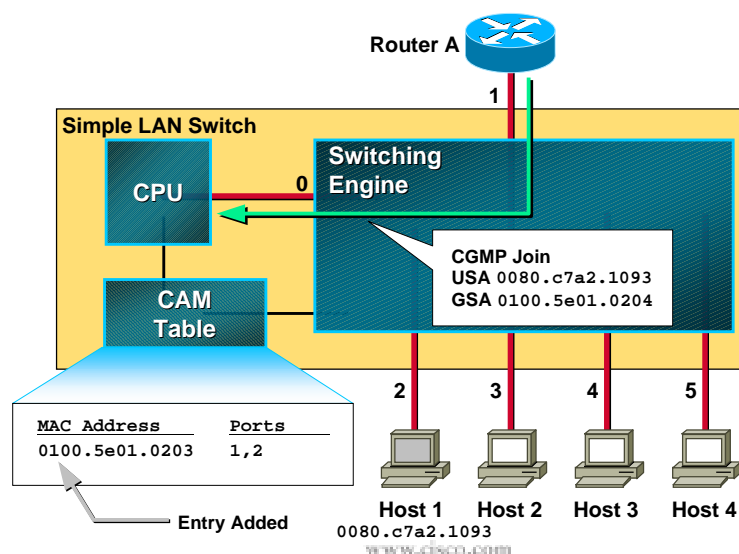


www.cisco.com

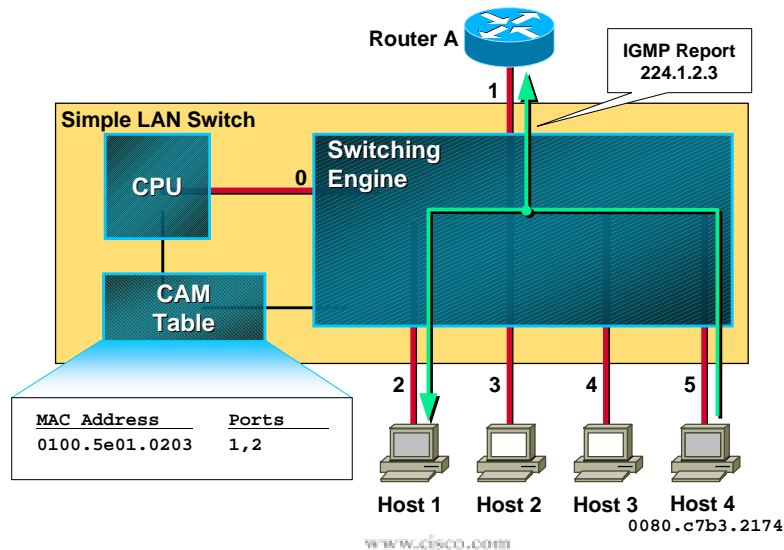
## CGMP—1st Join



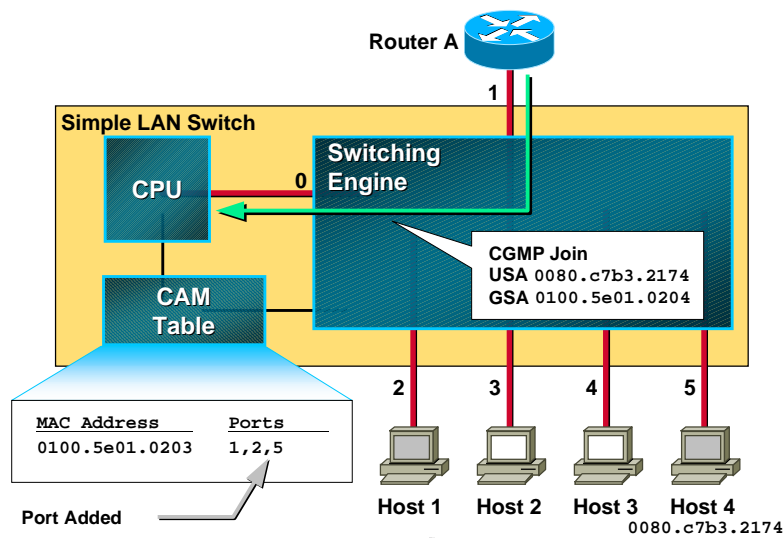
## CGMP—1st Join



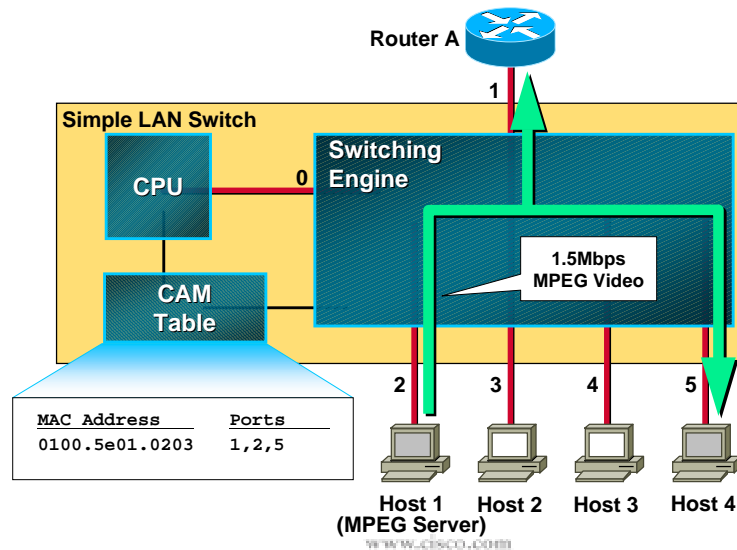
## CGMP 2nd Join



## CGMP 2nd Join



## CGMP No Load on Switch

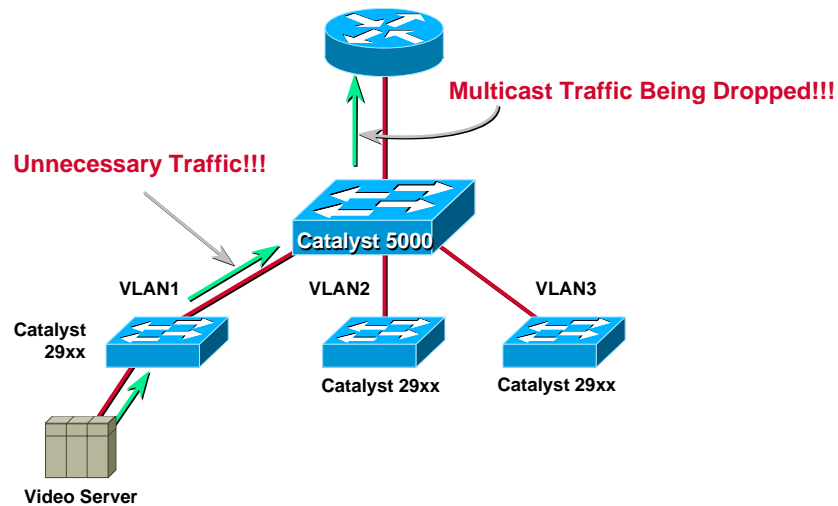


## Summary—Frame Switches

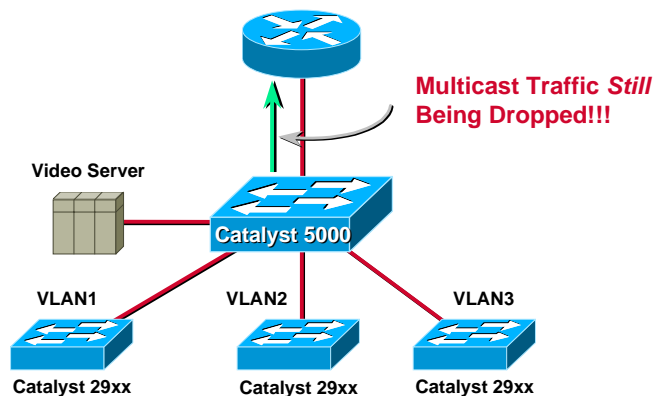
- **IGMP snooping**
  - Switches with Layer 3 aware ASICs
    - High-throughput performance maintained
    - Increases cost of switches
  - Switches without Layer 3 aware ASICs
    - Suffer serious performance degradation
- **CGMP**
  - Requires Cisco routers and switches
  - Can be implemented in low-cost switches

WWW.CISCO.COM

## Design Issue— Server Location

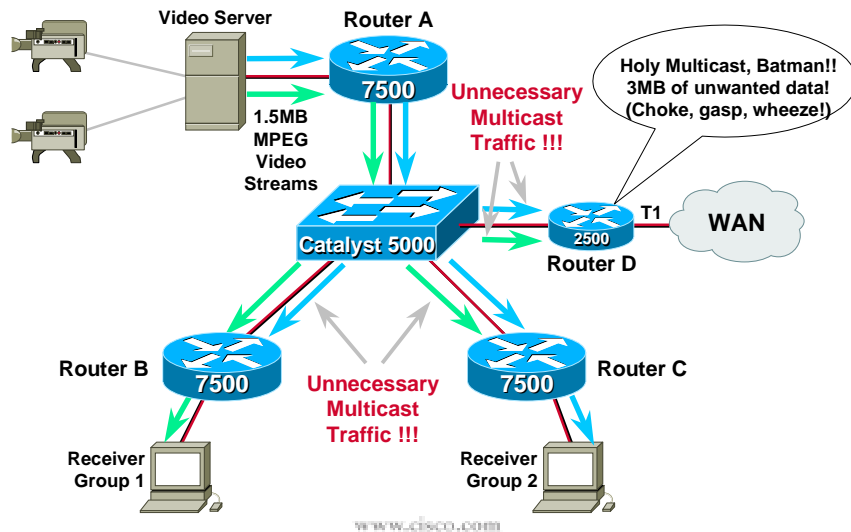


## Design Issue— Server Location

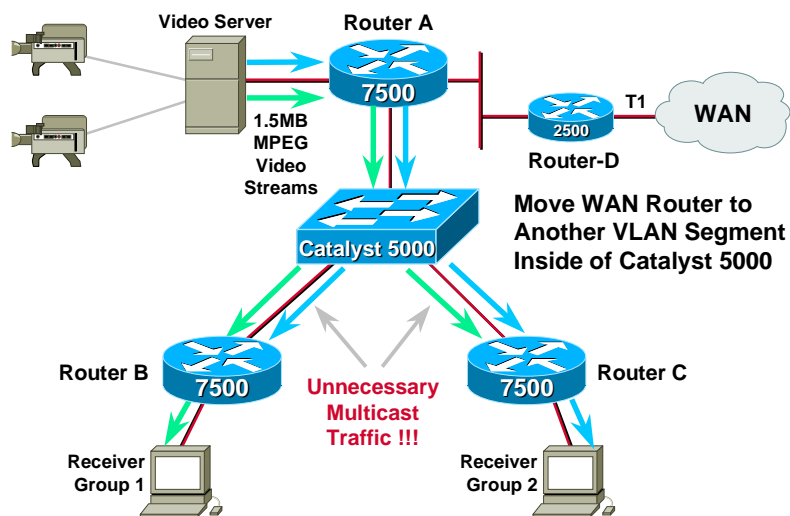


- Keep high B/W sources close to router

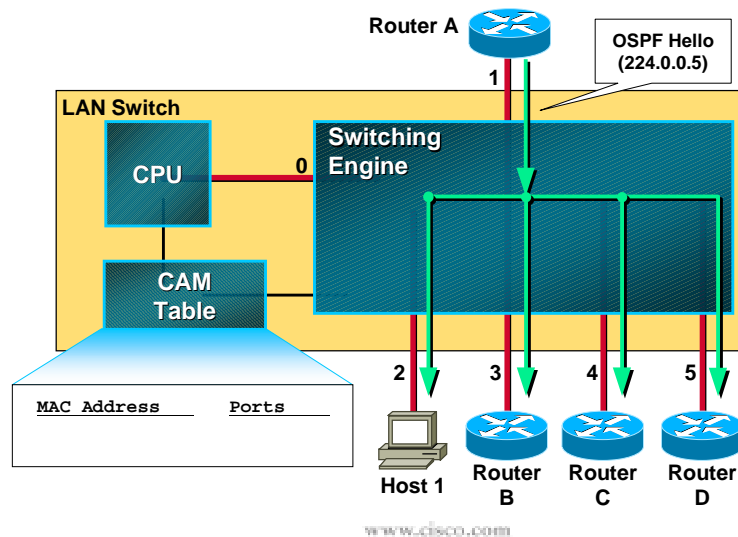
## Design Issue—Core Switch



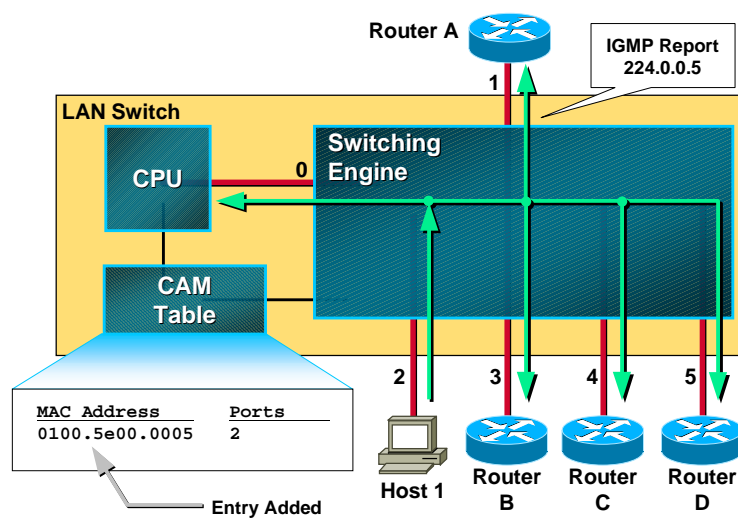
## Design Issue—Core Switch



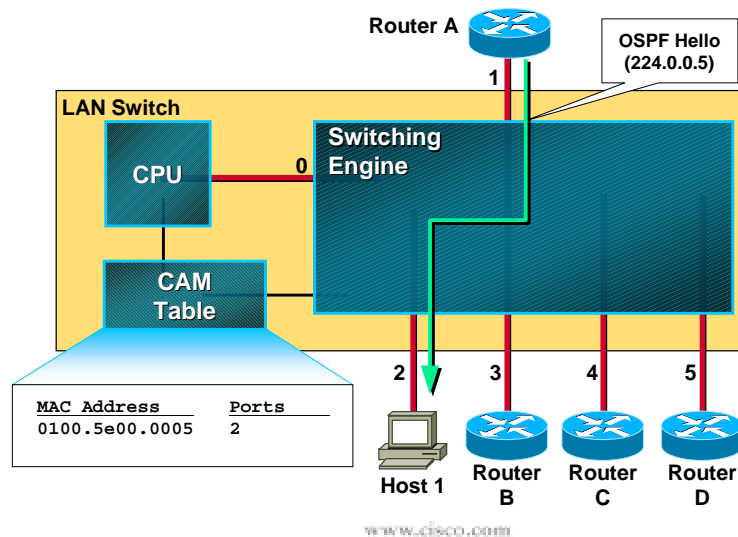
## Design Issue— 224.0.0.x Flooding



## Design Issue— 224.0.0.x Flooding

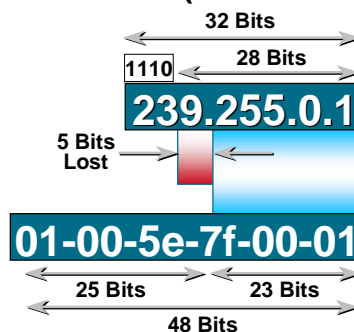


## Design Issue— 224.0.0.x Flooding



## 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**

www.cisco.com

## Design Issue— Address Overlap

### 32 - IP Multicast Addresses

224.1.1.1  
224.129.1.1  
225.1.1.1  
225.129.1.1  
:  
:  
238.1.1.1  
238.129.1.1  
239.1.1.1  
239.129.1.1

1 - Multicast MAC Address

0x0100.5E01.0101

WWW.CISCO.COM

## Design Issue— Address Overlap

### Try to Avoid Addresses that Must Be Flooded

### 32 - IP Multicast Addresses

**224.0.0.x**  
224.128.0.x  
225.0.0.x  
225.128.0.x  
:  
:  
238.0.0.x  
238.128.0.x  
239.0.0.x  
239.128.0.x

1 - Multicast MAC Address

0x0100.5E00.00xx

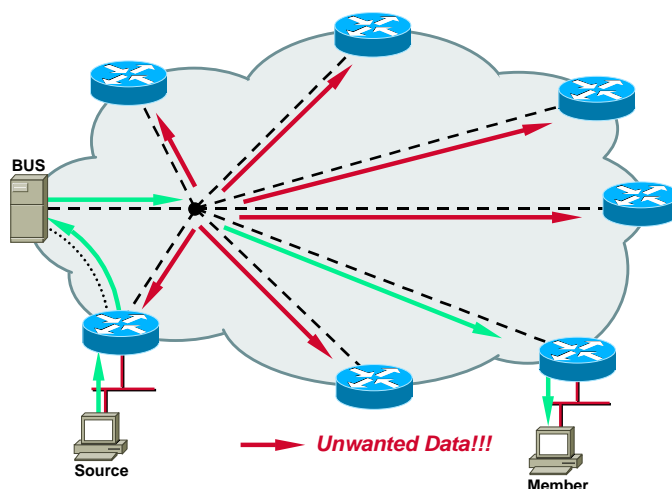
WWW.CISCO.COM

## Summary—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

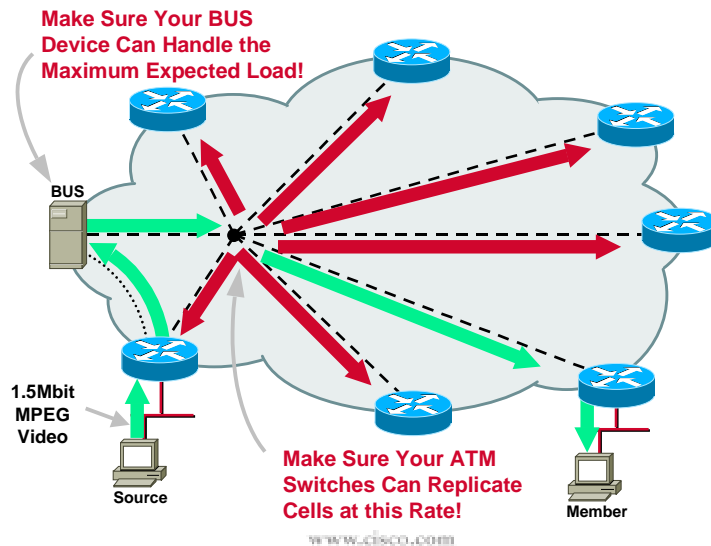
www.cisco.com

## Multicast over ATM LANE Core



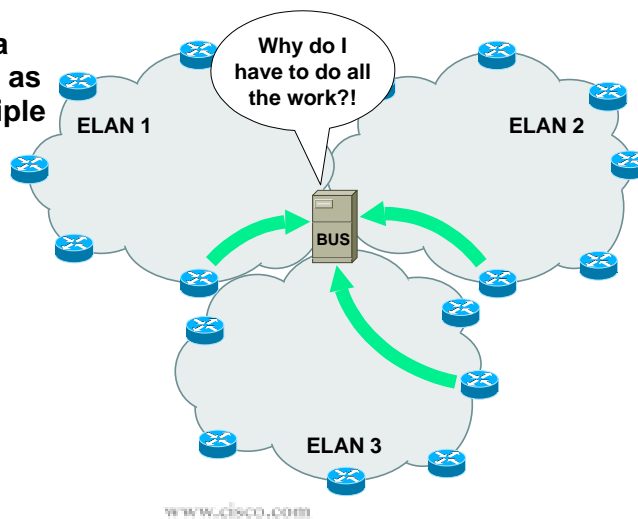
www.cisco.com

## Multicast over ATM LANE Core



## Multicast over ATM LANE Core

- Avoid using a single device as BUS for multiple ELANs



## Multicast over ATM LANE Core

- **Design issues**
  - **BUS horsepower is critical**
    - Use separate BUS device per ELAN to reduce load
    - Overloaded BUS = cell/packet loss and jitter/delay
      - Can cause problems on multimedia conferences
  - **ATM switch cell replication rate is critical**
    - Switches that replicate cells in hardware are best
  - **Add lots of bandwidth to ATM fabric**
    - Traffic will frequently be sent where it's unwanted
    - ATM core bandwidth will be wasted
- **P2MP VCs may be a better solution**

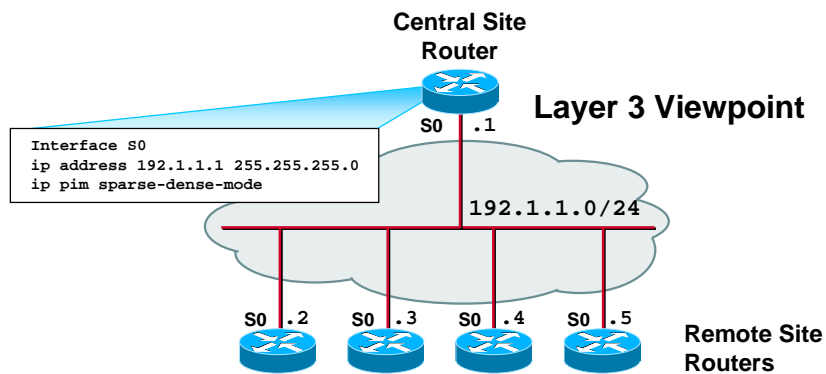
[www.cisco.com](http://www.cisco.com)

## Agenda

- Which Mode? Sparse or Dense
- Multicast over Campus Networks
- **Multicast over NBMA Networks**
- RP Engineering
- Multicast Bandwidth Control
- Network Convergence

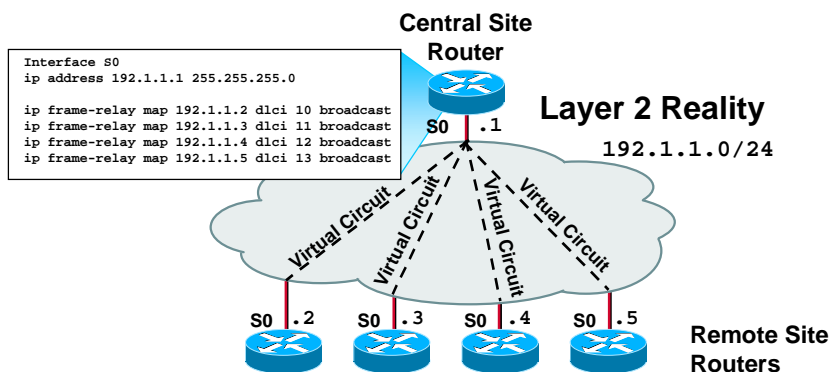
[www.cisco.com](http://www.cisco.com)

## How Routers See NBMA at L3



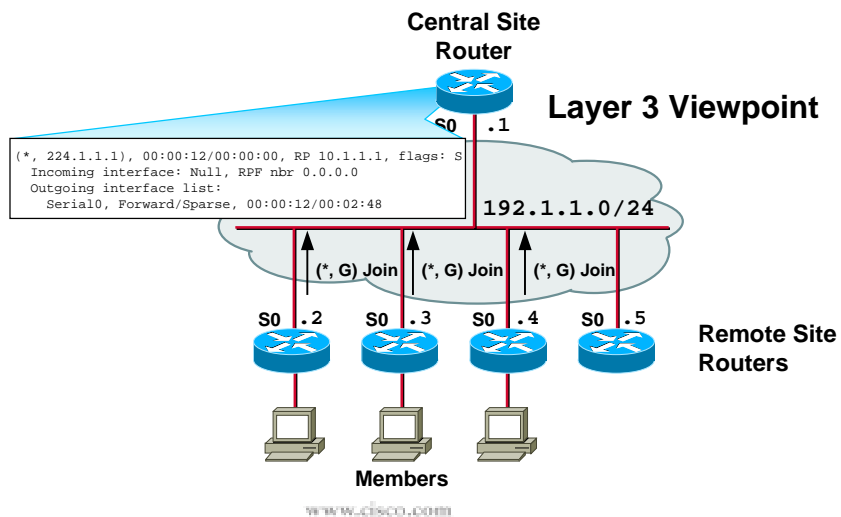
www.cisco.com

## How Routers See NBMA at L3

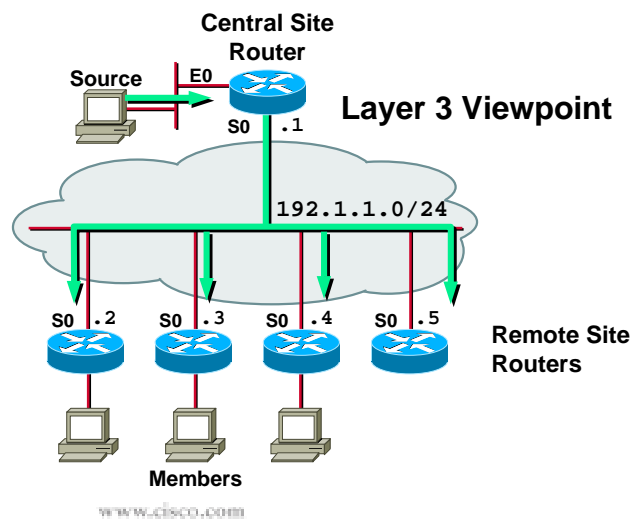


www.cisco.com

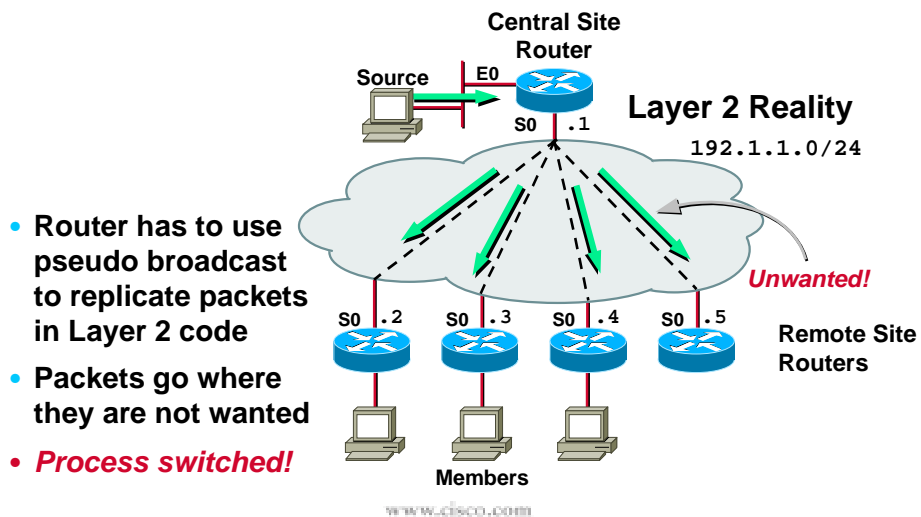
## How Routers See NBMA at L3



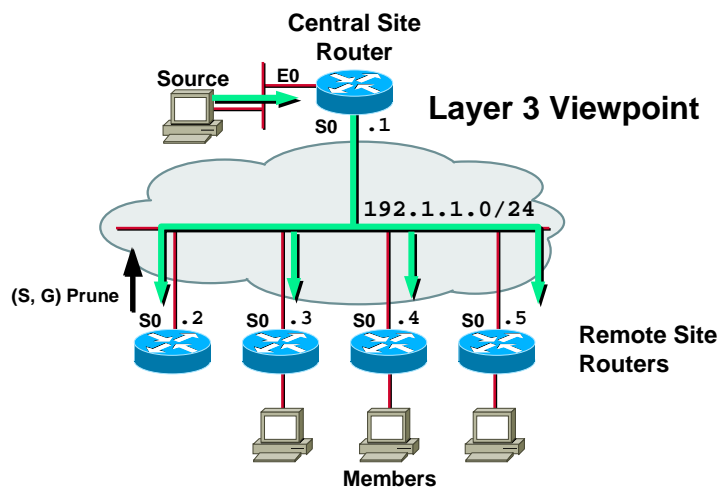
## How Routers See NBMA at L3



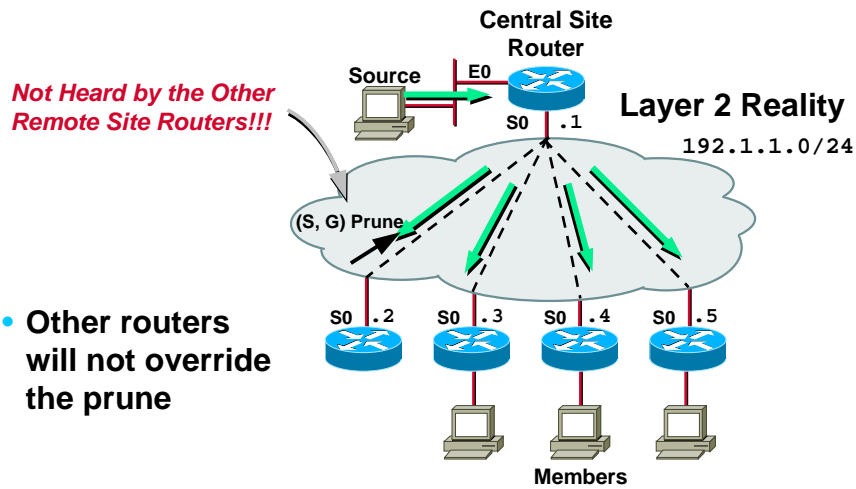
## How Routers See NBMA at L3



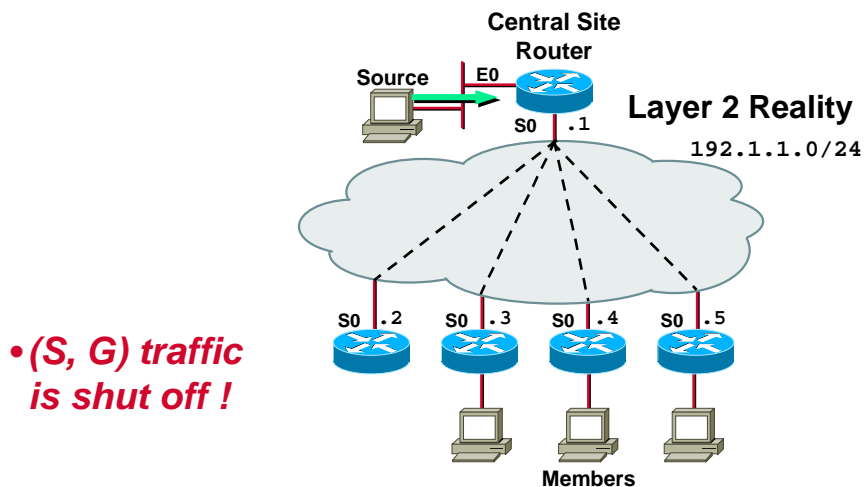
## DM Problem with Partial Mesh



## DM Problem with Partial Mesh



## DM Problem with Partial Mesh



## NBMA Mode

- **Solution: PIM-SM + NBMA mode**

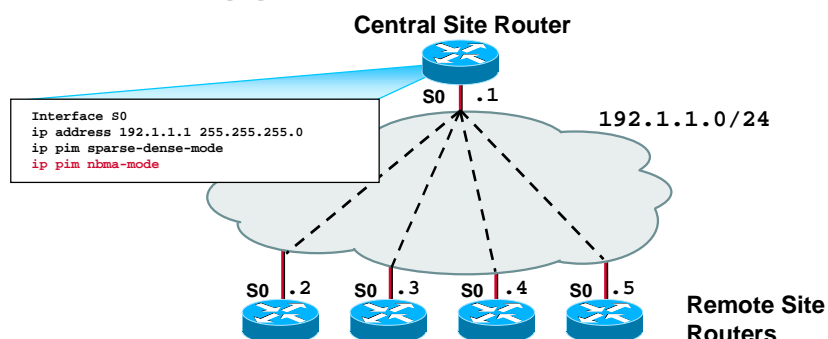
```
ip pim nbma-mode
```

- **Requires sparse mode**
- **When router receives join, it puts the interface **and** joiner in the outgoing interface list (OIL)**
- **When router receives a prune, it removes the interface/joiner from OIL**

www.cisco.com

## NBMA Mode

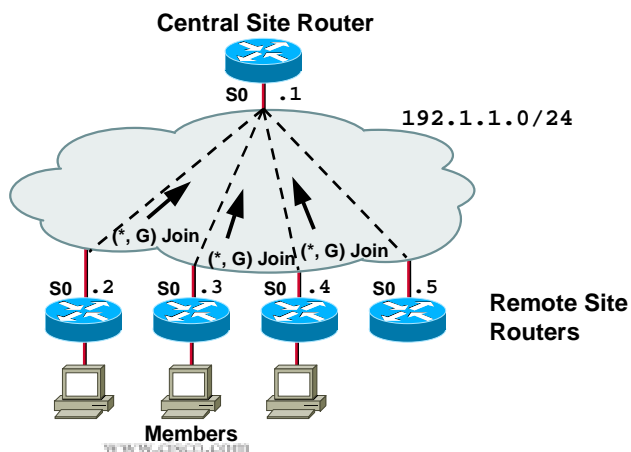
### Avoiding Pseudo Broadcast by Using 'ip pim nbma-mode'



www.cisco.com

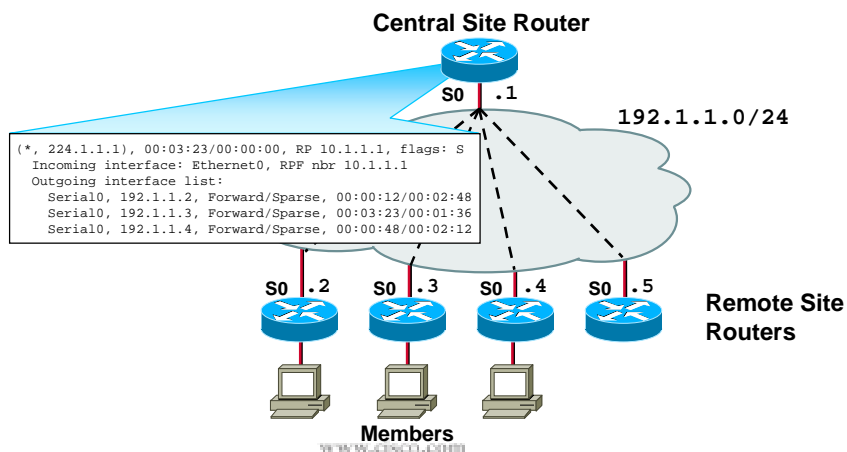
## NBMA Mode

### Avoiding Pseudo Broadcast by Using 'ip pim nbma-mode'



## NBMA Mode

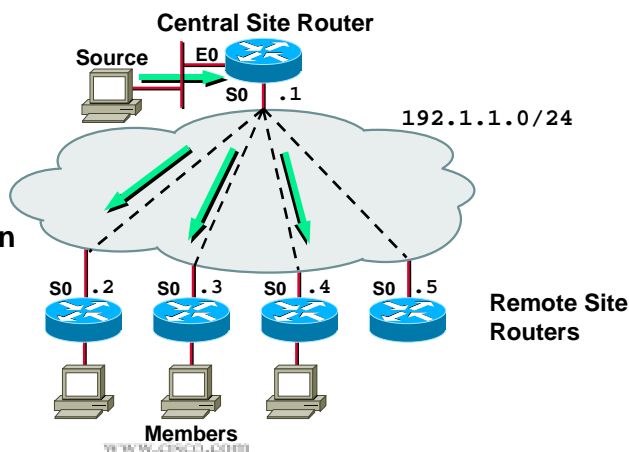
### Avoiding Pseudo Broadcast by Using 'ip pim nbma-mode'



## NBMA Mode

### Avoiding Pseudo Broadcast by Using 'ip pim nbma-mode'

- Router can now replicate packets in Layer 3 code
- Packets only go where needed
- **Fast switched!**



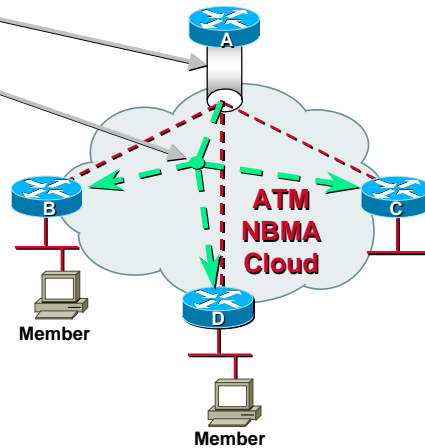
## ATM P2MP Broadcast VCs

- What if WAN media could do native multicast
  - ATM can perform the broadcast/multicast packet replication task via p2mp VCs
- Answer: ATM multipoint-signaling
  - One p2mp VC handles any and all outgoing broadcast and multicast traffic
  - Sends N copies to N neighbors out of K interested parties (not optimal)

www.cisco.com

## ATM NBMA Cloud w/P2MP Broadcast VCs

- Each PVC is p2p on a multipoint (sub)interface
- Router Backbone - p2mp VCs do all broadcast/multicast replication instead of the router
- Use any PIM mode
- Suboptimal multicast solution
- Fast switched!



Note: Only VCs for Router A shown

www.cisco.com

## ATM Multipoint Signalling

- **Commands:**
- **Requires “broadcast” keyword on all ATM map-list statements**

```
atm multipoint-signaling

atm map-list mumble
ip x.x.x.x atm-nsap xxxx.xxxx... broadcast
ip y.y.y.y atm-nsap yyyy.yyyy... broadcast
```

www.cisco.com

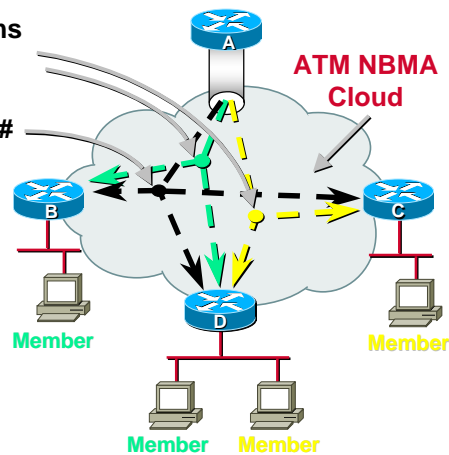
## ATM P2MP VC per Group

- What if each group had it's own p2mp VC
  - NBMA-mode solved sending K copies to K interested parties out of N neighbors
  - A p2mp VC/Group can solve sending one copy to K interested parties out of N neighbors
- Answer: Use PIM multipoint-signaling

www.cisco.com

## ATM NBMA Cloud with P2MP VC/Group

- One p2mp VC per group performs multicast replication instead of the router
- Broadcast p2mp VC used when # groups > max p2mp VC count
- Use PIM sparse mode
- p2mp VCs map group membership
- Fast Switched!



Note: Only p2mp multicast VCs for Router A shown for clarity

www.cisco.com

## ATM P2MP VC per Group (Cont.)

- **Algorithm: Similar to NBMA mode**
  - Rather than putting interface/joiner in olist, put joiner on multipoint VC
  - Received joins cause UNI signaling ADD-PARTYs
  - Received prunes cause UNI signaling DROP-PARTYs
- **Use a VC count threshold to keep down the number of VCs opened**
  - Use shared multipoint VC and fanout as tie breaker

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

## ATM P2MP VC per Group (Cont.)

- **Commands:**
  - `ip pim multipoint-signaling`
  - `ip pim vc-count <count>`
  - `ip pim minimum-vc-rate <pps>`
- **Good for ATM networks used as router core LIS**
  - ATM Network must support SVC's
  - Routers must have ATM map statements to all other routers
  - Need the shared broadcast p2mp VC
    - Otherwise uses pseudo-broadcast (ugh!)

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

## Agenda

- Which Mode? Sparse or Dense
- Multicast over Campus Networks
- Multicast over NBMA Networks
- **RP Engineering**
- Multicast Bandwidth Control
- Network Convergence

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

## 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>

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

## 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

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

## Auto-RP Fundamentals

- **All Cisco routers**

- **Join Cisco-Discovery (224.0.1.40) group**

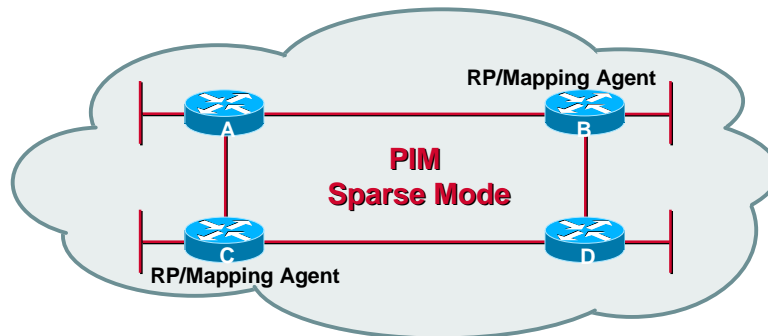
- Automatic
- No configuration necessary

- **Receive RP-Discovery messages**

- Stored in local Group-to-RP Mapping Cache
- Information used to determine RP for group range

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

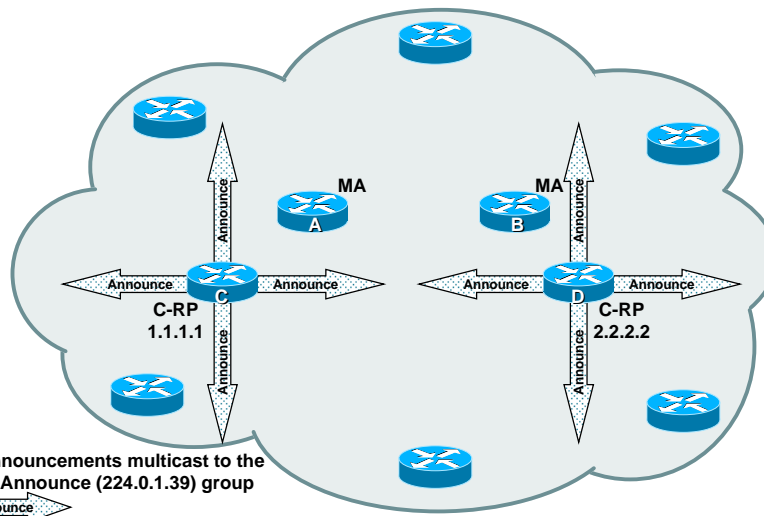
## Simple Auto-RP Configuration



On each router: `ip multicast-routing`  
 On each interface: `ip pim sparse-dense-mode`  
 On routers B and C: `ip pim send-rp-announce loopback0 scope 16`  
`ip pim send-rp-discovery scope 16`

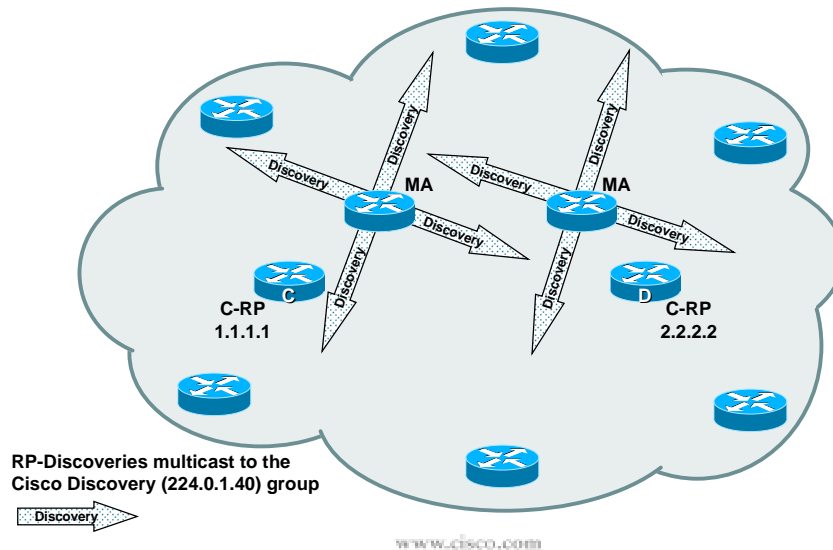
www.cisco.com

## Auto-RP—From 10,000 Feet

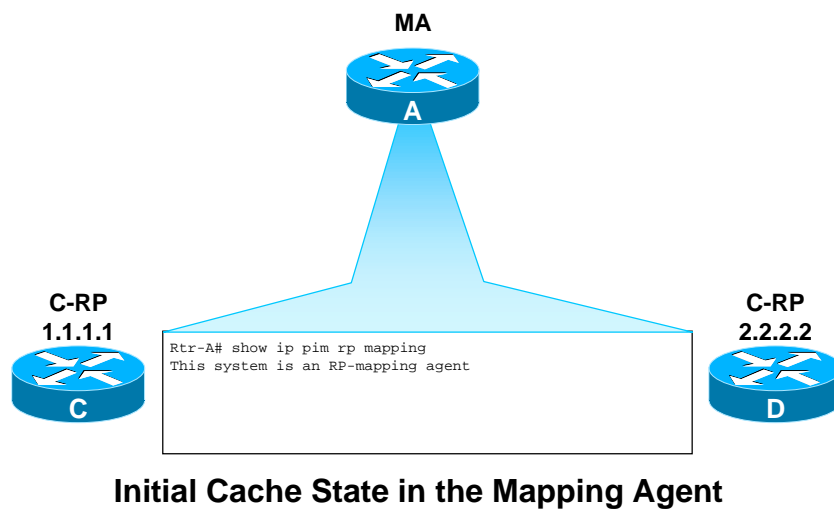


www.cisco.com

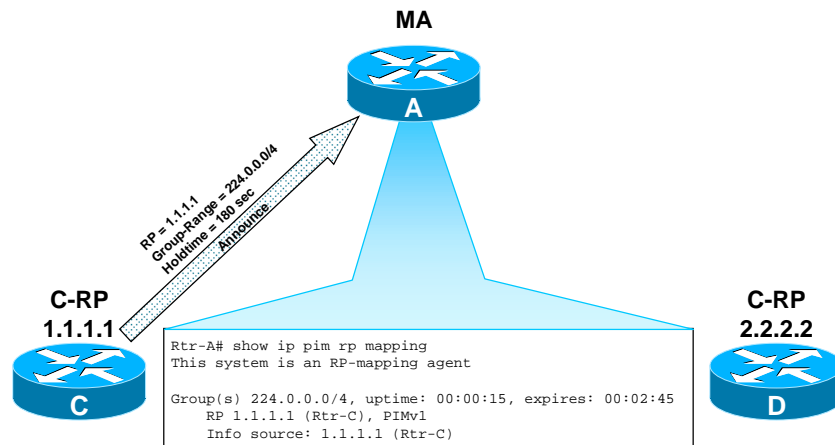
## Auto-RP—From 10,000 Feet



## Auto-RP—A Closer Look

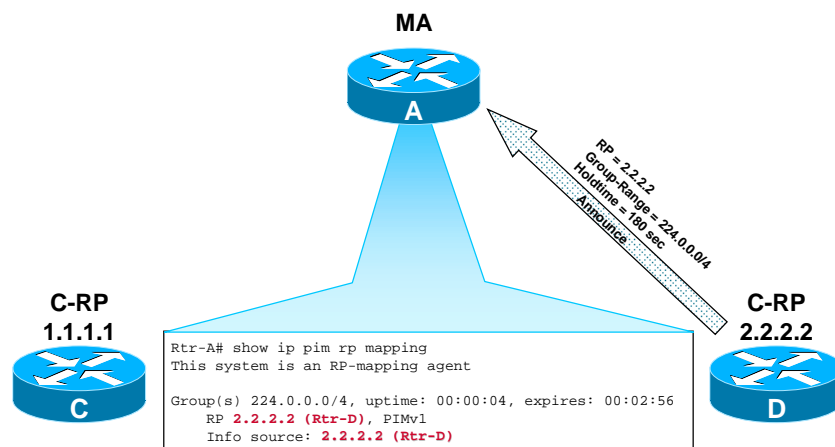


## Auto-RP—A Closer Look



www.cisco.com

## Auto-RP—A Closer Look

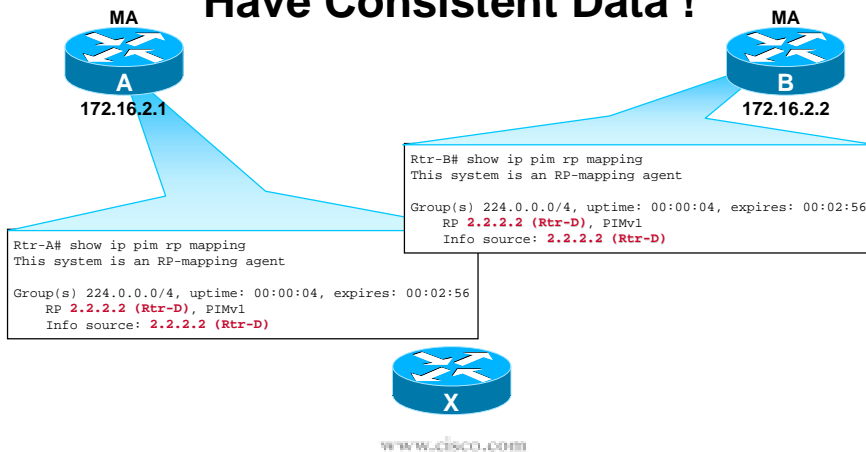


**C-RP with Highest Address Is Selected as RP and Stored in Cache**

www.cisco.com

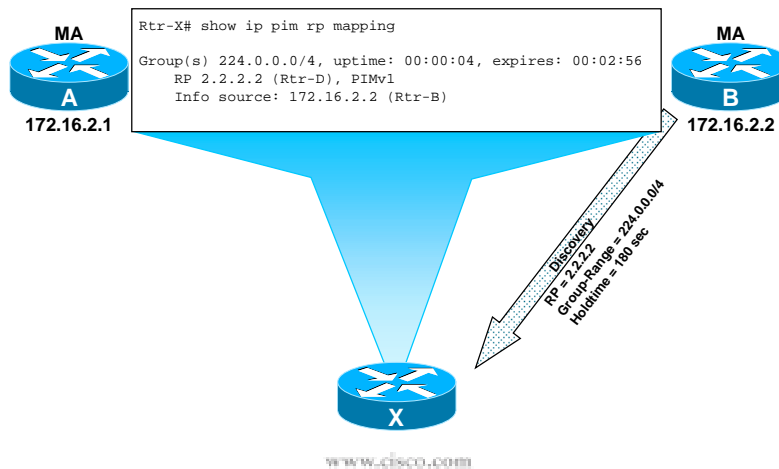
## Auto-RP—A Closer Look

All Mapping Agents **Must**  
Have Consistent Data !



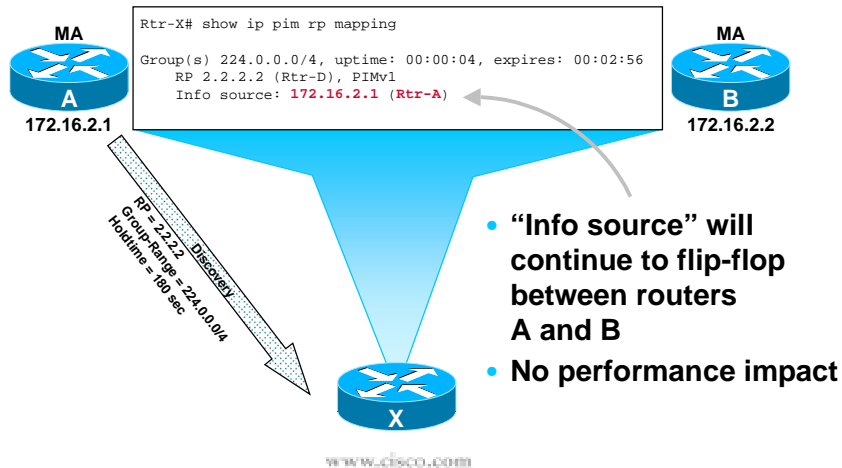
## Auto-RP—A Closer Look

Local Cache Initially Loaded from Router “B”

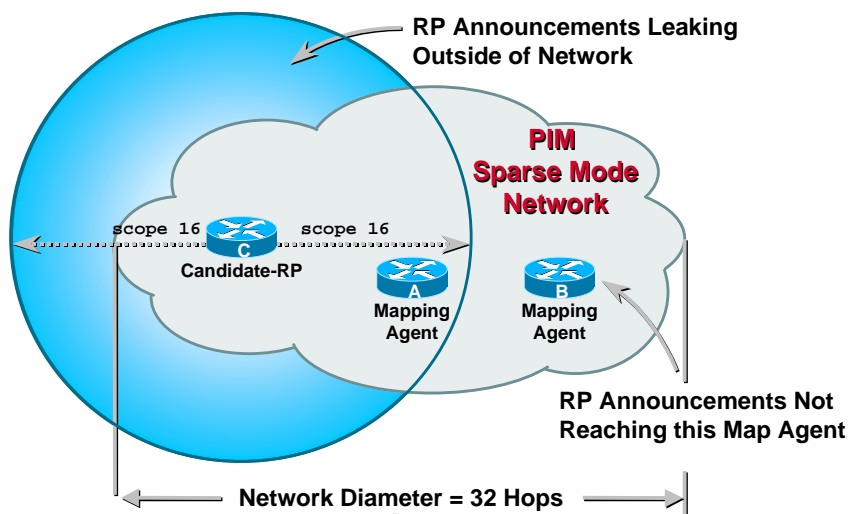


## Auto-RP—A Closer Look

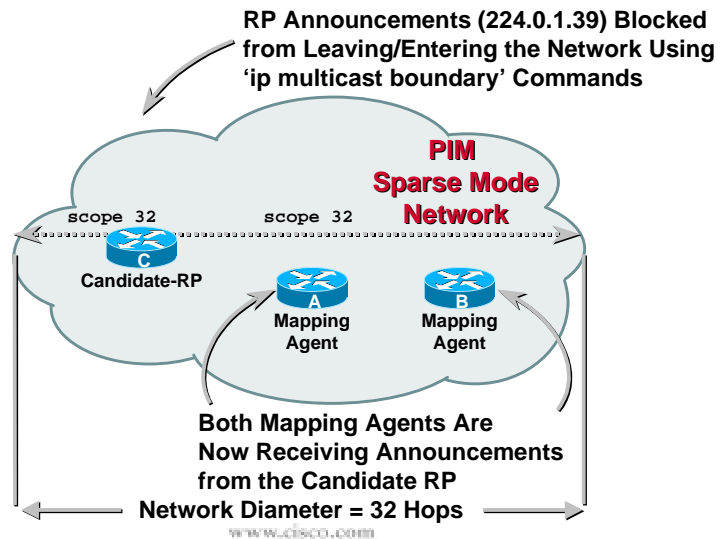
### Identical Info Received from Router “A”



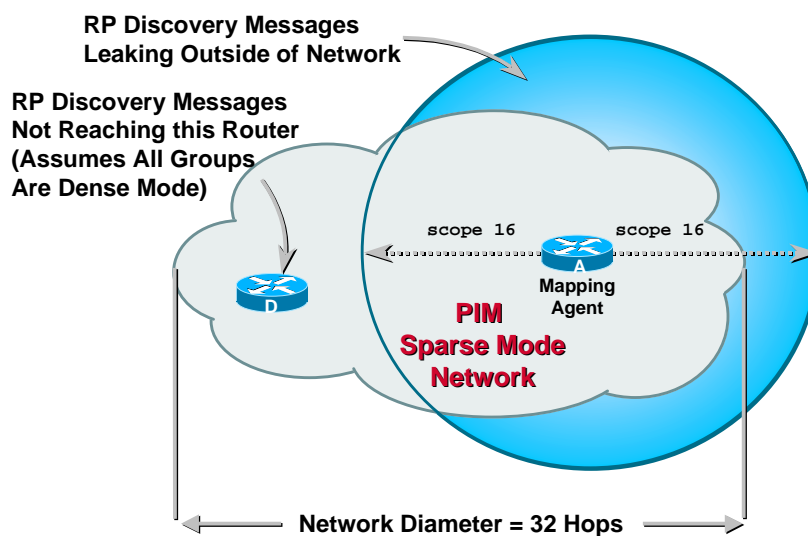
## Auto-RP Announcement Scope



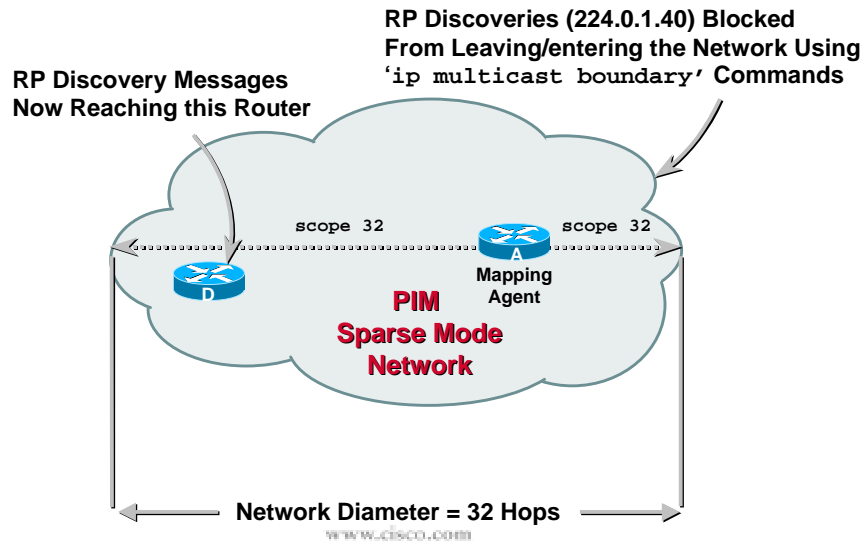
## Auto-RP Announcement Scope



## Auto-RP Discovery Scope

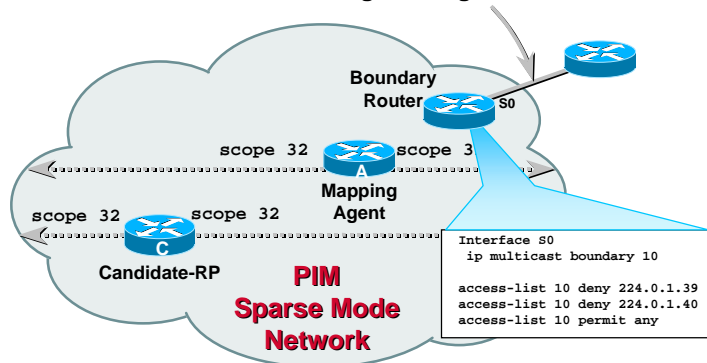


## Auto-RP Discovery Scope



## Constraining Auto-RP Messages

Need to Block Auto-RP Discovery (224.0.1.40) and Announcement (224.0.1.39) Messages from Entering/Leaving the Network



## PIMv2 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>

www.cisco.com

## PIMv2 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

www.cisco.com

## PIMv2 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

www.cisco.com

## PIMv2 BSR Fundamentals

- **BSR election mechanism**

- **C-BSR's:**

- **Begin in Candidate-BSR state**

- BSR-Timeout timer started (150 seconds)

- » If higher priority (preferred) BSR message received

- » Restart timer and forward BSR message

- Copy info to local Group-to-RP mapping cache

- » Otherwise, discard BSR message

- If timer expires, transition to Elected-BSR state

- **While in Elected-BSR state**

- Periodically originate own BSR messages

- » Include local Group-to-RP mapping cache in msg

- Return to Candidate-BSR state if preferred BSR message (higher priority) received

www.cisco.com

## PIMv2 BSR Fundamentals

- **BSR election mechanism**
  - **Non C-BSRs (i.e., all other routers):**
    - **Start in Accept-Any state**
      - Accepts first BSR message received
      - Saves BSR info and forwards BSR message
      - Transitions to Accept-Preferred state
    - **While in Accept-Preferred state**
      - Starts BSR-Timeout timer
      - Only accept and forward preferred BSR messages
        - » (i.e., BSR messages with priority > current BSR priority)
      - Otherwise, discard BSR message
      - Return to Accept-Any state if timer expires

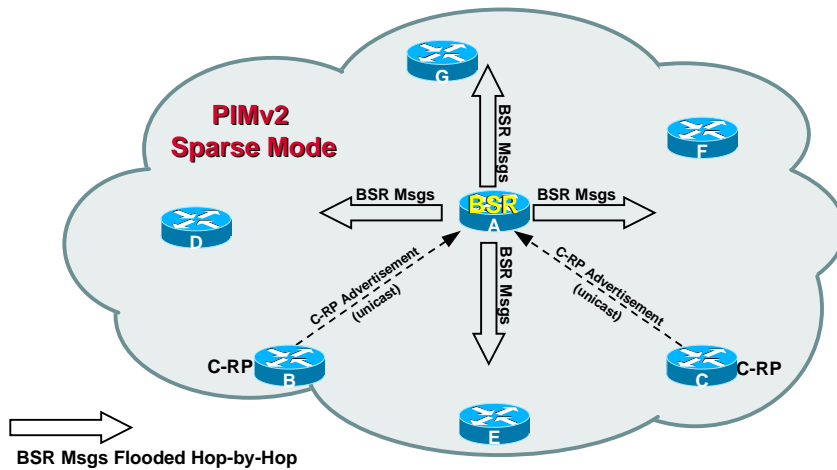
[www.cisco.com](http://www.cisco.com)

## PIMv2 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

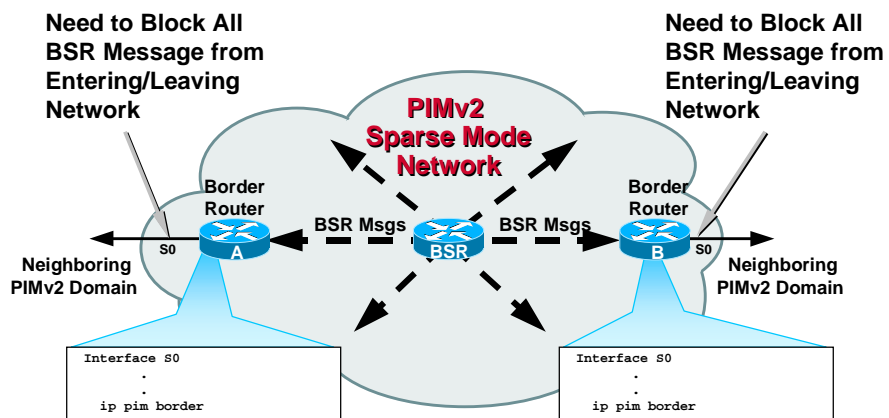
[www.cisco.com](http://www.cisco.com)

## Basic PIMv2 BSR



www.cisco.com

## Constraining BSR Messages



www.cisco.com

## Understanding Accept-RP

- **Accept-RP Command**

- **Global configuration command**

- `ip pim accept-rp <rp-address> [<acl>]`
    - `ip pim accept-rp Auto-rp [<acl>]`
    - `ip pim accept-rp 0.0.0.0 [<acl>]`

- **Multiple commands accepted**

- Command list sorted in order shown above
    - Only one Auto-RP and one 0.0.0.0 (wildcard) accepted
    - Omitting ACL implies 224.0.0.0/4 group range

- **Search Rules**

- Top down search
    - Stop on RP address match—Apply ACL and exit
    - Exception: Auto-RP denies RP/Group
      - Apply 0.0.0.0 (wildcard) entry (if it exists)

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

## Understanding Accept-RP

- **Accept-RP Command Usage**

- Case 1 — Controlling Group mode

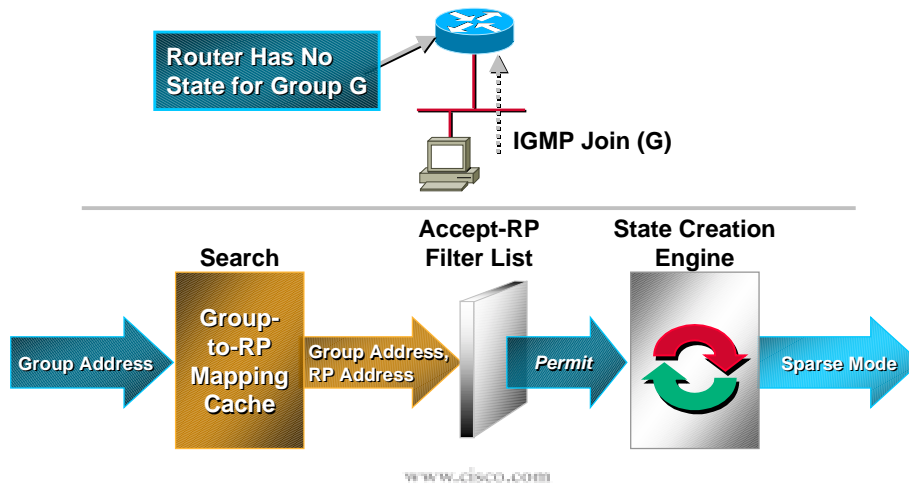
- Case 2 — Accepting (\*, G) joins

- Case 3 — Accepting PIM registers at the RP

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

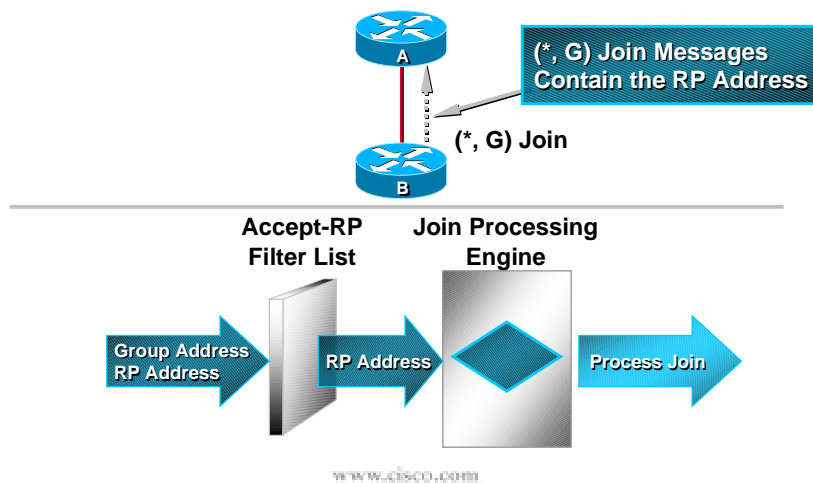
## Accept-RP—Case 1

### Controlling Group Mode



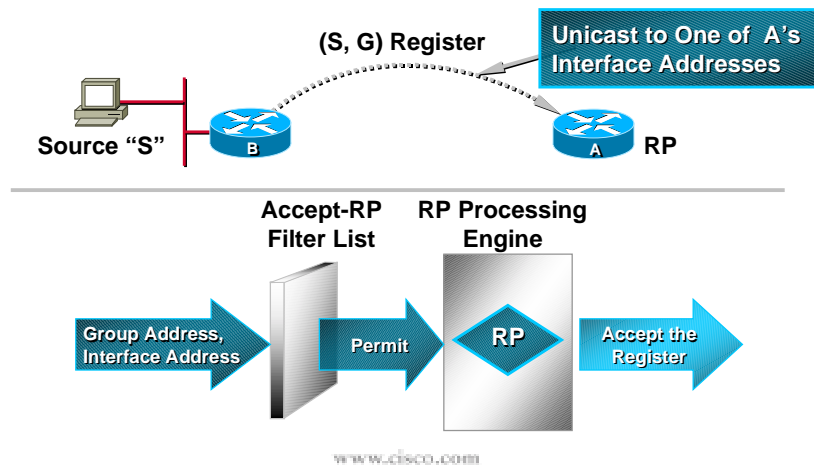
## Accept-RP—Case 2

### Accepting (\*, G) Joins



## Accept-RP—Case 3

### Accepting PIM Registers at the RP



## 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

WWW.CISCO.COM

## RP Performance Considerations

- **CPU load factors**
  - RP must process Registers
  - RP must process Shared-tree Joins/Prunes
  - RP must send periodic SPT Joins toward source
  - PIM performs RPF recalculation every 5 seconds
    - Watch the total number of mroute table entries in the RP
  - Shared-tree forwarding
    - Only when spt-threshold = infinity is in use
- **Memory load factors**
  - (\*, G) entry ~ 260 bytes + OIL size
  - (S, G) entry ~ 212 bytes + OIL size
  - Outgoing interface list (OIL) size
    - Each oil entry ~ 80 bytes

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

## Dealing with Overloaded RP's

- Increase CPU horsepower
- Increase memory
- Use SPTs if not already
- **Split RP load across several RPs!**

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

## Agenda

- Which Mode? Sparse or Dense
- Multicast over Campus Networks
- Multicast over NBMA Networks
- RP Engineering
- **Multicast Bandwidth Control**
- Network Convergence

www.cisco.com

## BW Control via Rate-Limiting

- **Rate limit interface command**  

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

  - Multiple entries may be used per interface
- **Brute-force limits**
  - Use “out” form of command on WAN links
  - Set <kbps> to desired percentage usage of link BW
  - Group and/or Source ACLs provide more granularity
- **Limiting video or whiteboard streams**
  - Add “video” or “whiteboard” keywords
  - Requires ‘ip sdr listen’ to be enabled
  - Streams identified using info from sdr cache

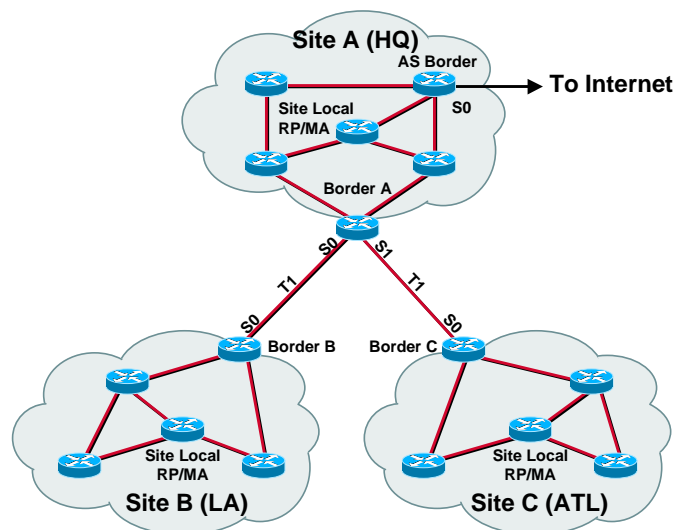
www.cisco.com

## BW Control via Admin-Scoping

- Limit high-BW source to local site
- Use administratively-scoped zones
  - Simple scoped zone example:
    - 239.255.0.0/16 = Site-Local Scope Zone
    - 239.192.0.0/10 = Org.-Local Scope Zone
    - 224.0.1.0 - 238.255.255.255 = Global scope (Internet) zone
  - High-BW sources use only site-local zone groups
  - Med.-BW, org-wide sources use org.-local zone
  - Low-Med. BW, Internet-wide sources use global zone

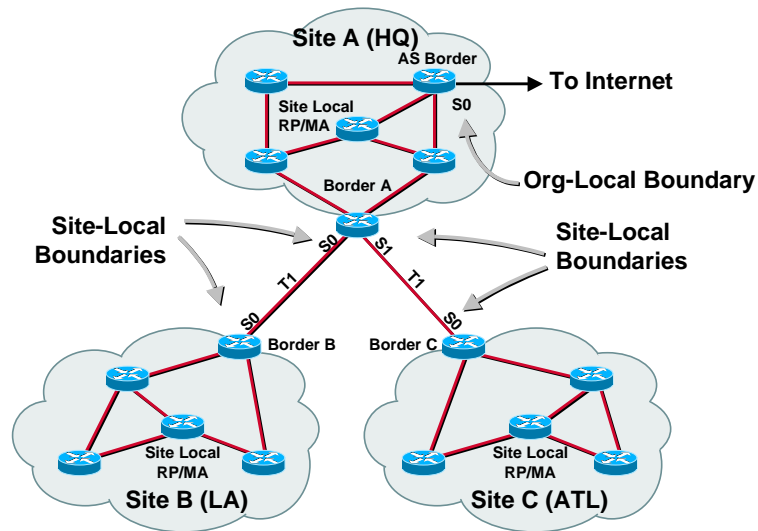
www.cisco.com

## BW Control via Admin-Scoping



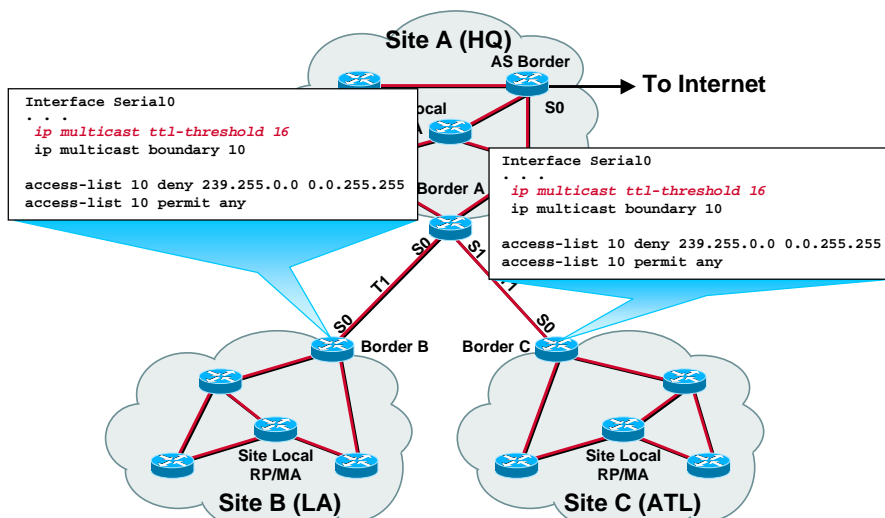
www.cisco.com

## BW Control via Admin-Scoping



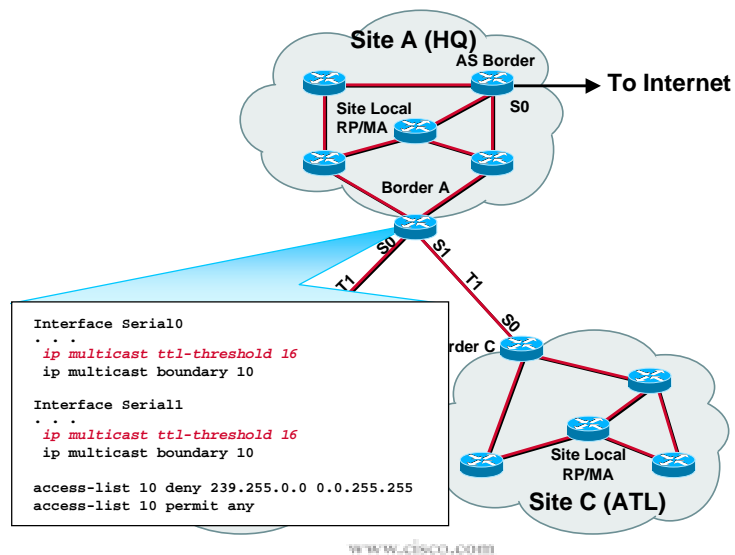
www.cisco.com

## BW Control via Admin-Scoping

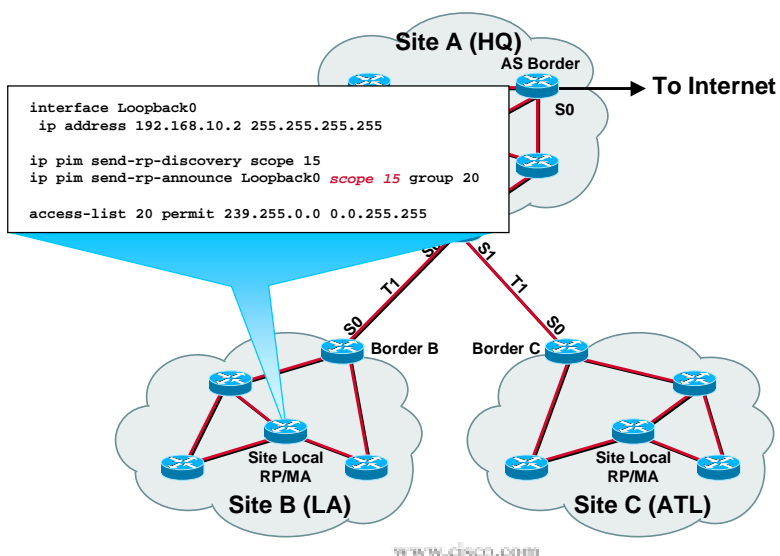


www.cisco.com

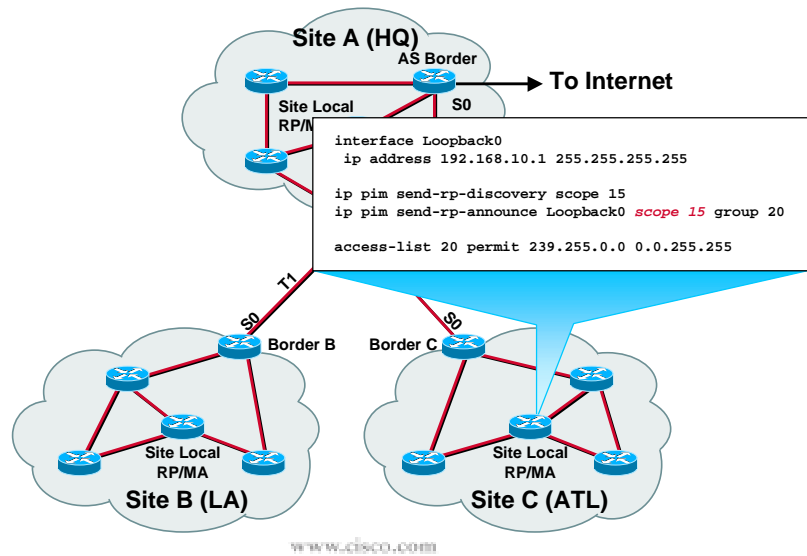
## BW Control via Admin-Scoping



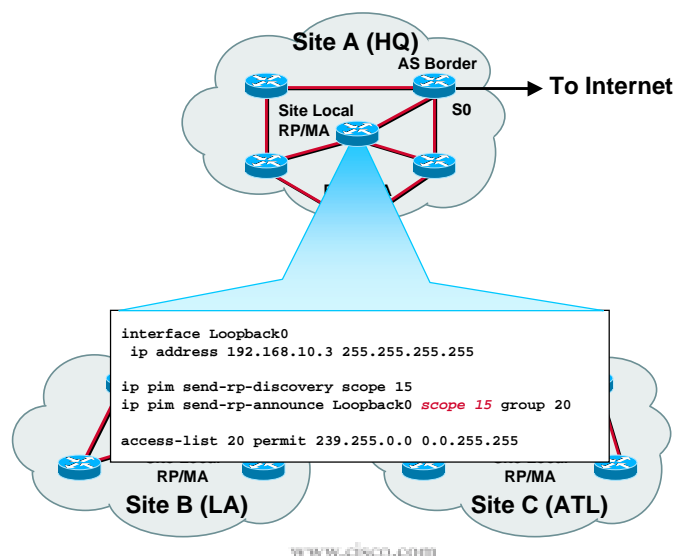
## BW Control via Admin-Scoping



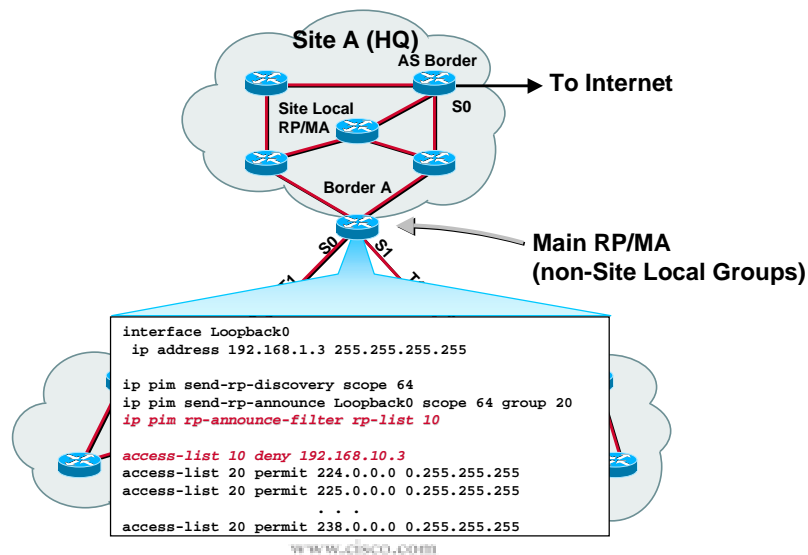
## BW Control via Admin-Scoping



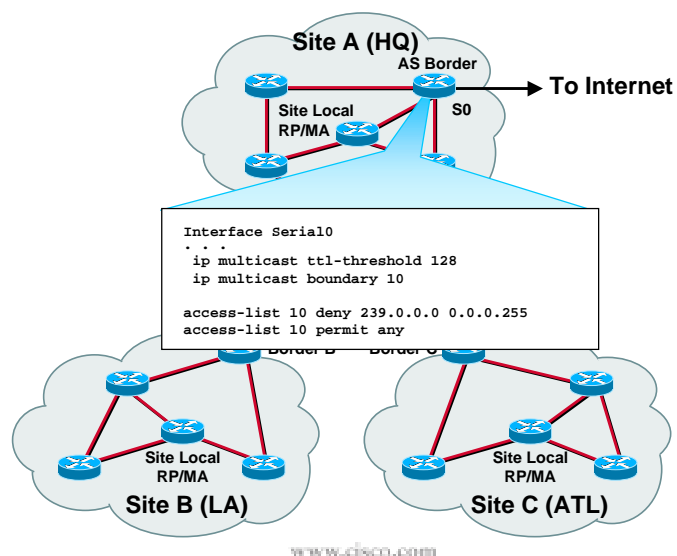
## BW Control via Admin-Scoping



## BW Control via Admin-Scoping



## BW Control via Admin-Scoping



## Agenda

- Which Mode? Sparse or Dense
- Multicast over Campus Network
- Multicast over NBMA Networks
- RP Engineering
- Multicast Bandwidth Control
- **Network Convergence**

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

## 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

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

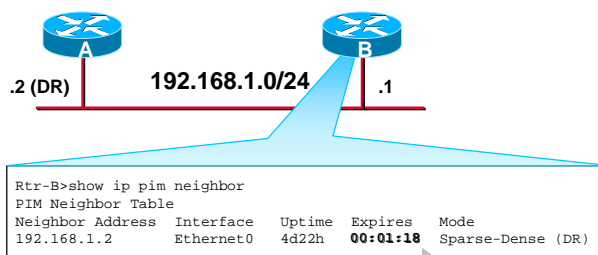
## Tuning RP-Failover

- 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

www.cisco.com

## DR Failover



- 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

www.cisco.com

## 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

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

## 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

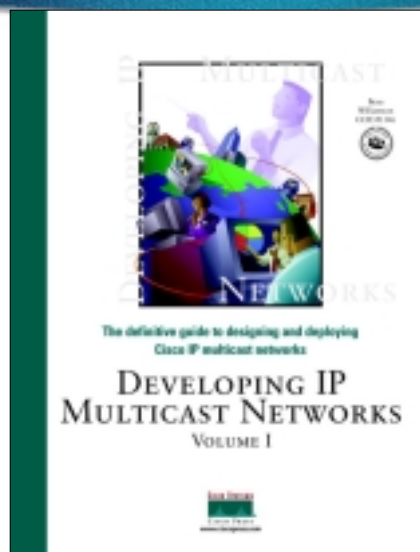
[www.cisco.com](http://www.cisco.com)

## Documentation and Contact Info

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

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

## If All Else Fails—RTFB<sup>1</sup>



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

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



# Please Complete Your Evaluation Form

**Session 314**

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

**CISCO SYSTEMS**



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

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