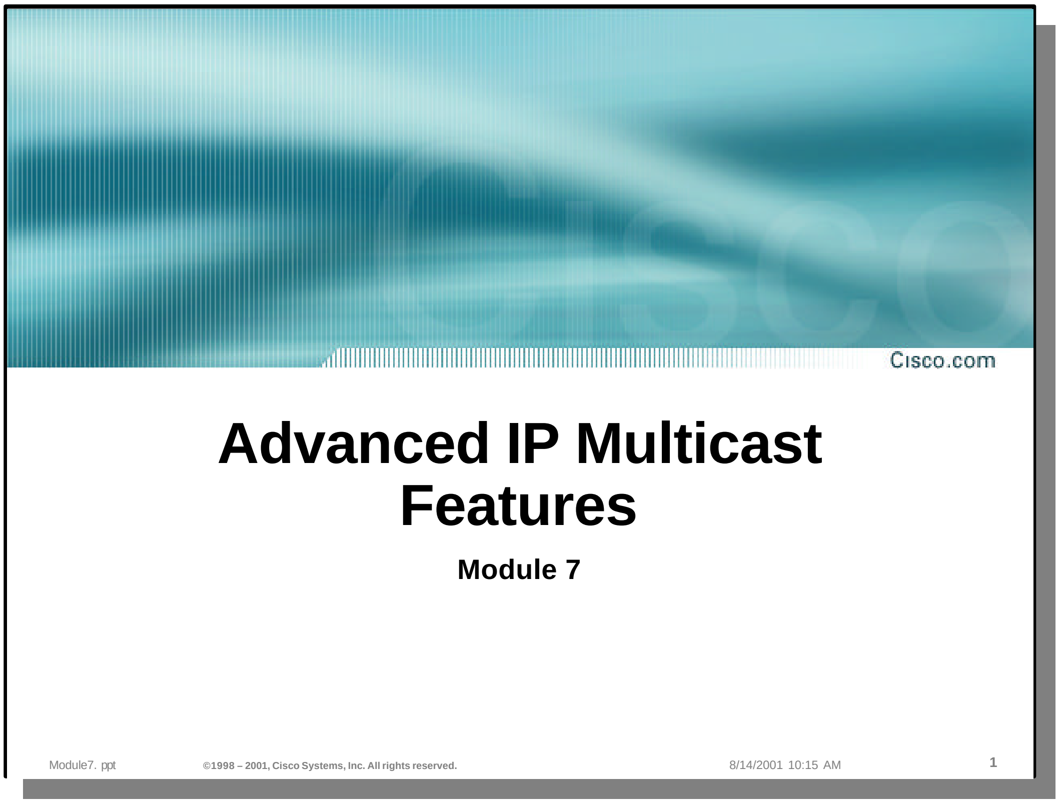
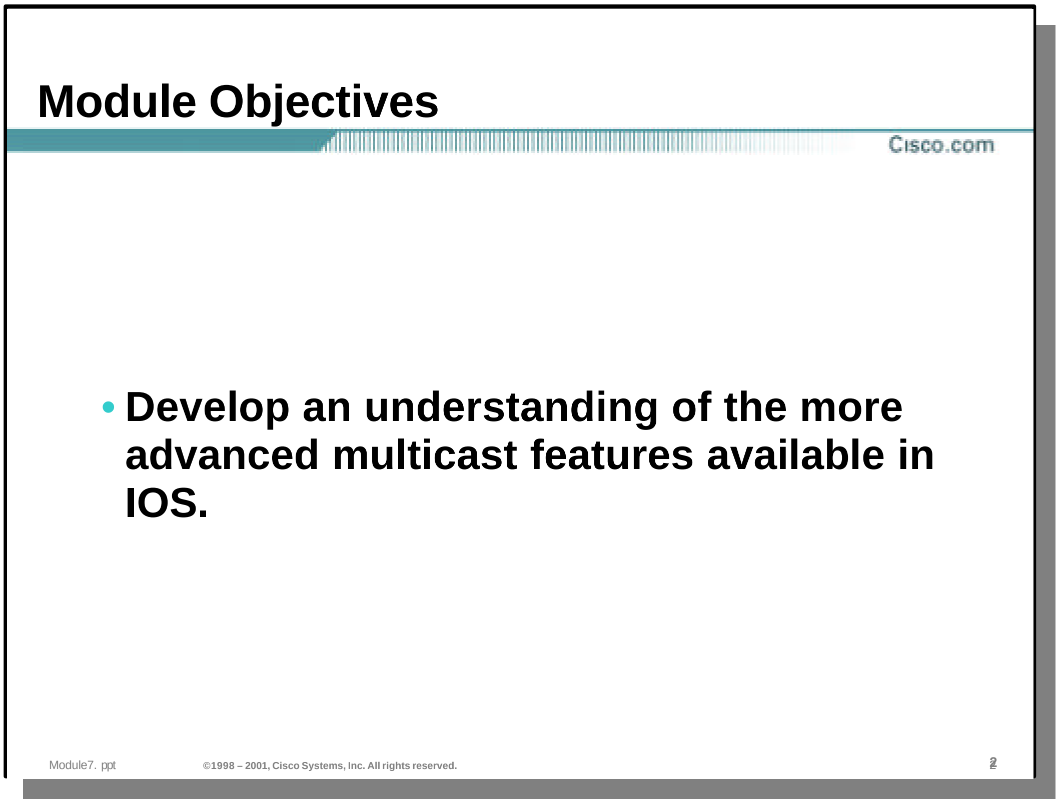


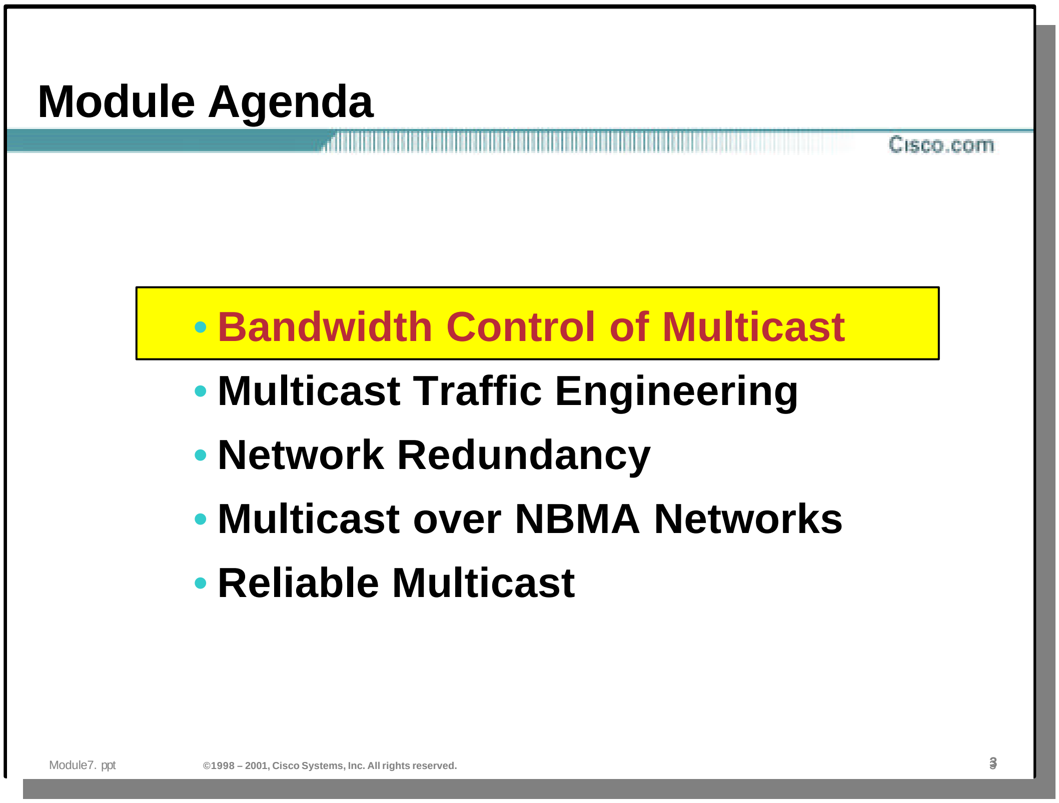
Advanced IP Multicast Features

Module 7



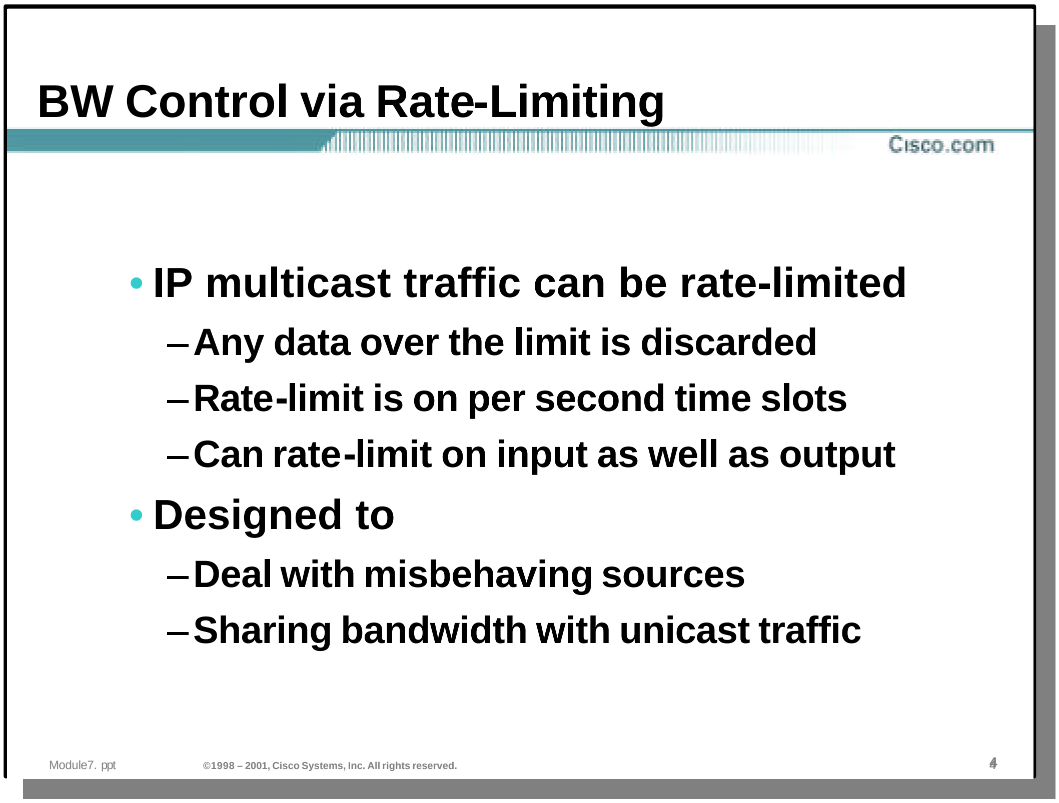
Module Objectives

- Develop an understanding of the more advanced multicast features available in IOS.



Module Agenda

- Bandwidth Control of Multicast
- Multicast Traffic Engineering
- Network Redundancy
- Multicast over NBMA Networks
- Reliable Multicast

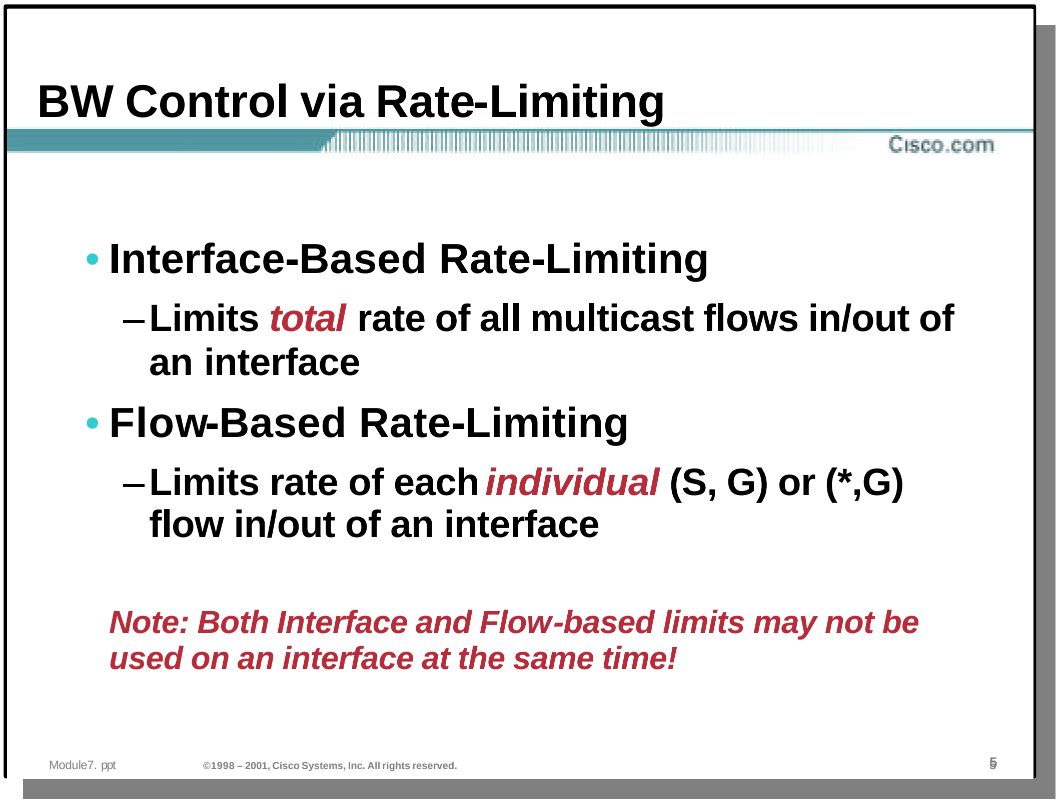


BW Control via Rate-Limiting

Cisco.com

- 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

- **Bandwidth Control via Rate-Limiting**
  - In general, concern over bandwidth utilization by multicast traffic is often more of an issue of FUD (Fear, Uncertainty, Doubt) than a real issue. However, like many other traffic types, multicast traffic can be rate-limited.
    - Rates are measured over 1 second windows and compared to configured limits.
    - Once the configured limit has been reached, further data that would exceed this limit is discarded.
    - Rate-limiting may be applied to either incoming or outgoing traffic.
  - Rate-limiting provides protection against:
    - Misbehaving sources that are consuming too much bandwidth.
    - Multicast consuming all of the available bandwidth.



BW Control via Rate-Limiting

Cisco.com

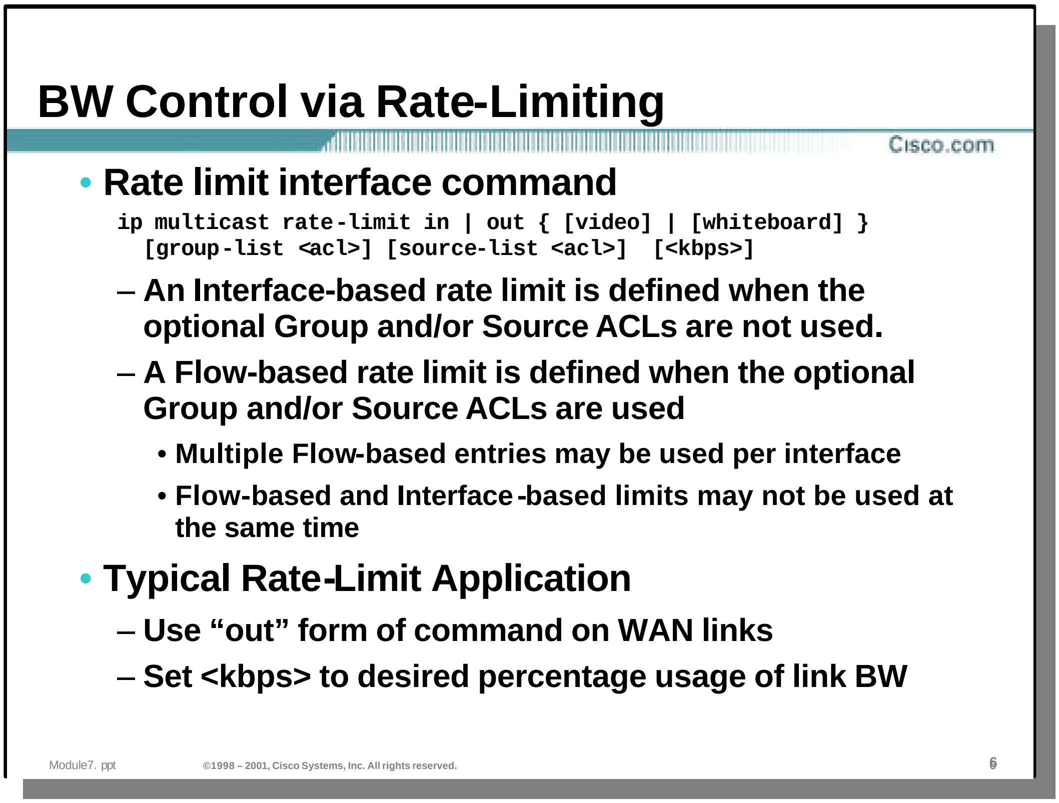
- Interface-Based Rate-Limiting
  - Limits *total* rate of all multicast flows in/out of an interface
- Flow-Based Rate-Limiting
  - Limits rate of each *individual* (S, G) or (*,G) flow in/out of an interface

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

- **Interface-Based Rate-Limiting**
  - Rate limits may be applied to the total overall rate of multicast traffic flowing into or out of an interface. This is the most commonly used form of multicast rate limiting, particularly for outgoing traffic. This permits an upper bound to be set on the bandwidth consumed by multicast traffic on an interface.
- **Flow-Based Rate-Limiting**
  - Flow-based rate limits may be applied to an interface on a “Group” or “Source/Group” basis. When flow-based rate limits are defined, it causes the rate to be applied to the incoming or outgoing interface of matching (*, G) or (S, G) entries in the mroute table.

    For example, if an outgoing 50Kbps flow-based rate limit has been configured on interface **Serial0** for (*, 224.1.1.1), then a 50Kbps rate limit will be set on **Serial0** whenever it appears in the OIL of any (*, 224.1.1.1) or (S, 224.1.1.1) mroute table entries. This will limit the rate of each these flows to a maximum of 50Kbps.

    Note: Flow-based rate limiting is applied independently to each individual flow. In the above example, this would limit each *individual* matching flow out **Serial0** to 50Kbps. It would not rate limit the *total* flow of 224.1.1.1 traffic out **Serial0** to 50Kbps. Therefore, if there were several sources for the 224.1.1.1 group, the total flow can easily exceed 50Kbps.
  - Keywords such as “video” and “whiteboard” may be used to further identify media specific flows on a UDP port basis. However, for this to work, ‘ip sdr listen’ must be configured so that the router can obtain the necessary SDR session information to identify which flows are video and which are whiteboard.



BW Control via Rate-Limiting

Cisco.com

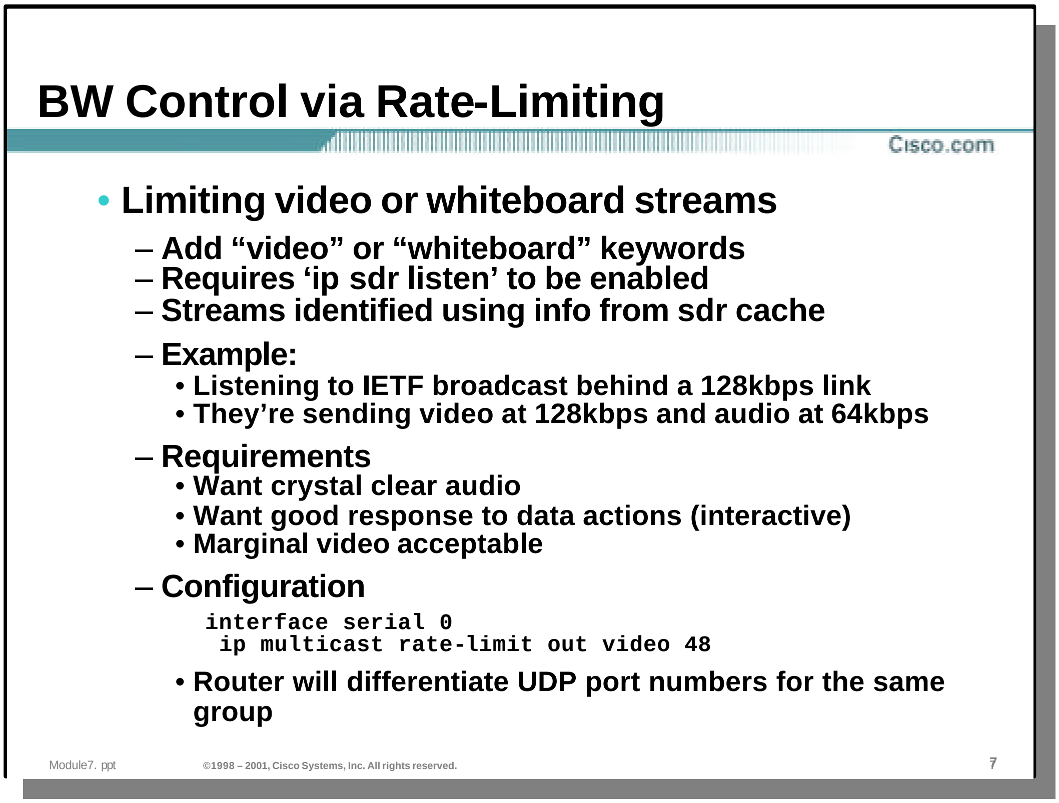
- **Rate limit interface command**

```
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**

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 6

- **Rate-Limit Interface Command**
  - The format of the rate-limit interface command is shown above.
    - The “in” and “out” keywords are used to specify if the rate-limit is to be applied to incoming or outgoing multicast traffic, respectively.
    - An interface-based rate-limit is defined when the optional group-list and/or source-list ACL’s are NOT used. Only one “in” and one “out” interface-based rate limit may be defined on an interface.
    - A flow-based rate-limit is defined when either the group-list or source-list ACL’s are specified. Multiple flow-based rate-limit commands may be configured on an interface
  - When an interface is added to the outgoing interface list or becomes the incoming interface, the list of rate-limits configured for that interface are searched for a match as follows:
    - All rate-limits configured on an interface are maintained and searched in the order in which they were entered.
    - If the interface is being used as the incoming interface for the mroute table entry the list is searched for an “in” limit. If the interface is being added to the outgoing interface list of the mroute entry, the list is search for an “out” limit.
    - The list is searched for the first rate-limit that matches the optional group and source ACL. (if there was no group or source ACL specified, then the limit is an interface-based limit and the limit matches unconditionally.)
    - The appropriate limit (interface or flow) is configured for the interface.
  - The most typical usage for rate-limits is to configure “out” interface-based limits on WAN links where the value of <kbps> is set to some desired percentage of the overall WAN link bandwidth. This prevents misbehaving multicast sessions from consuming all of the link bandwidth.



BW Control via Rate-Limiting

Cisco.com

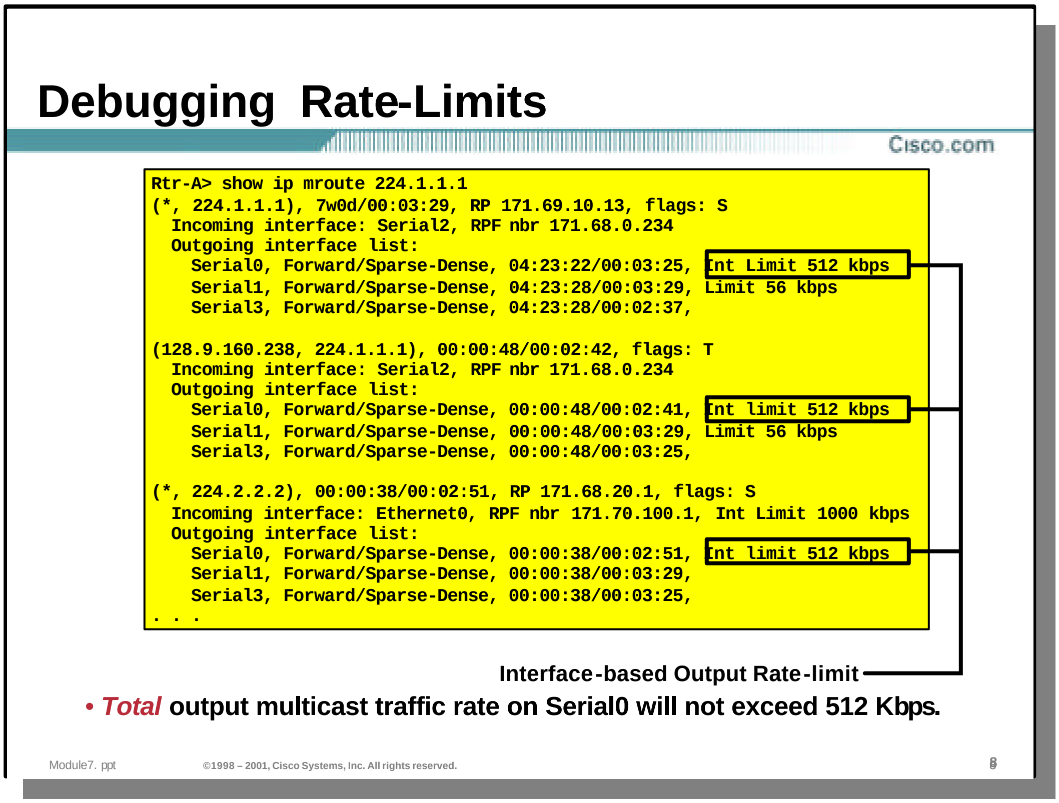
- **Limiting video or whiteboard streams**
  - **Add “video” or “whiteboard” keywords**
  - **Requires ‘ip sdr listen’ to be enabled**
  - **Streams identified using info from sdr cache**
  - **Example:**
    - **Listening to IETF broadcast behind a 128kbps link**
    - **They’re sending video at 128kbps 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**

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 7

- **Limiting Video or Whiteboard Streams**
  - Keywords such as “video” and “whiteboard” may be used to further identify media specific flows on a UDP port basis.
    - For this to work, ‘ip sdr listen’ must be configured so that the router can obtain the necessary SDR session information to identify which flows are video and which are whiteboard.
    - The router actually uses the SDR session information to identify the multicast group and UDP port number that is being used to send video or whiteboard data.
  - In the example above, the upstream serial interface is configured so that video can only consume 48kbps of the 128kbps ISDN line. This permits the 64kbps audio to be sent without experiencing any loss due to the video stream over subscribing the line.

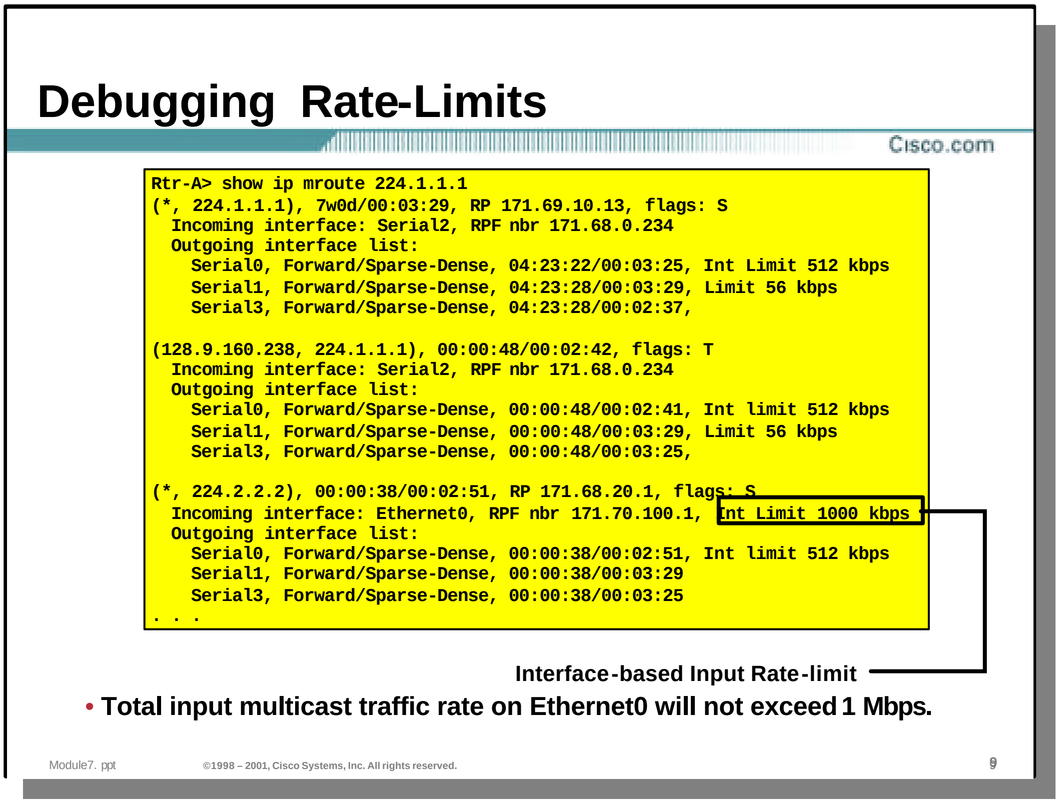


Debugging Rate-Limits

```
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,
. . .
```

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

- **Debugging Rate-Limits**
  - Rate limits may be displayed via the ‘show ip mroute’ command. Output interface-based rate-limits are shown on the entries in the outgoing interface list (OIL) of each mroute table entry. The text following the OIL will have the word Int preceding the rate limit value indicating that this is an Interface-based limit.
  - In the example above, an output interface-based rate-limit of 512kbps has been configured on interface **Serial0**.



Debugging Rate-Limits

Cisco.com

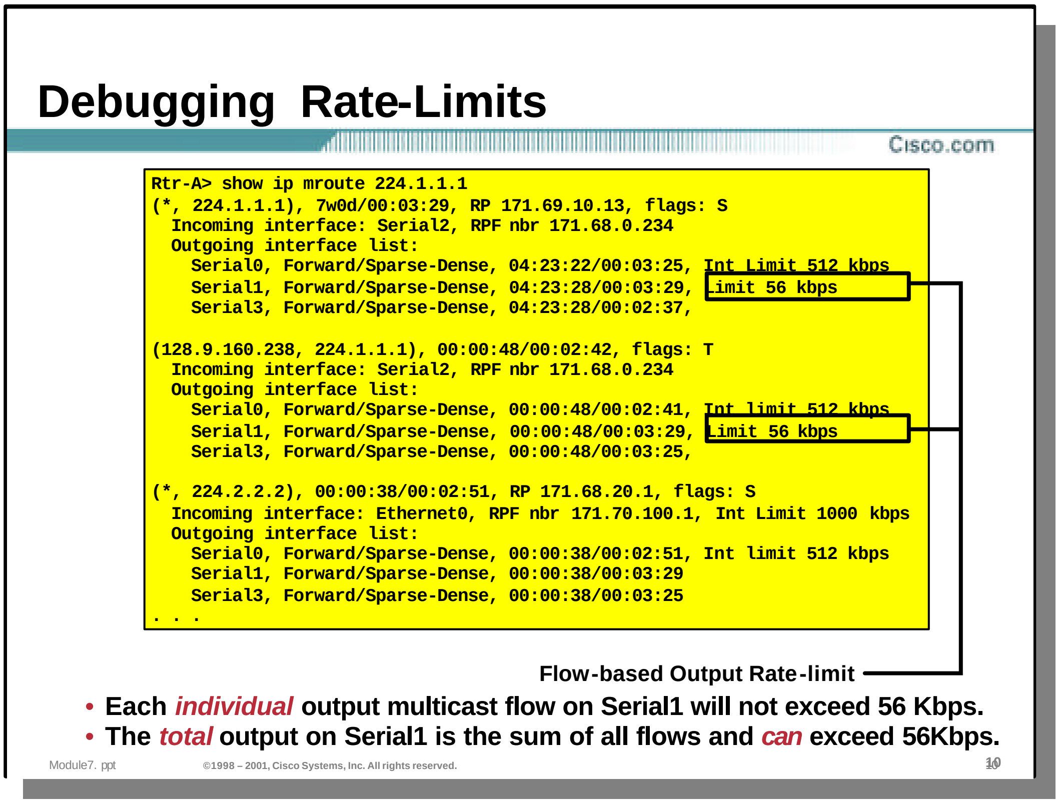
```
Rtr-A> show ip mroute 224.1.1.1
(*, 224.1.1.1), 7w0d/00:03:29, RP 171.69.10.13, flags: S
  Incoming interface: Serial2, RPF nbr 171.68.0.234
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 04:23:22/00:03:25, Int Limit 512 kbps
    Serial1, Forward/Sparse-Dense, 04:23:28/00:03:29, Limit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

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 9

- **Debugging Rate-Limits**
  - Rate-limits may be displayed via the ‘show ip mroute’ command. Input interface-based rate-limits are shown on the incoming interface of each mroute table entry. The text following the incoming interface information will have the word Int preceding the rate limit value indicating that this is an Interface-based limit.
  - In the example above, an input interface-based rate-limit of 1 Mbps has been configured on interface **Ethernet0**.



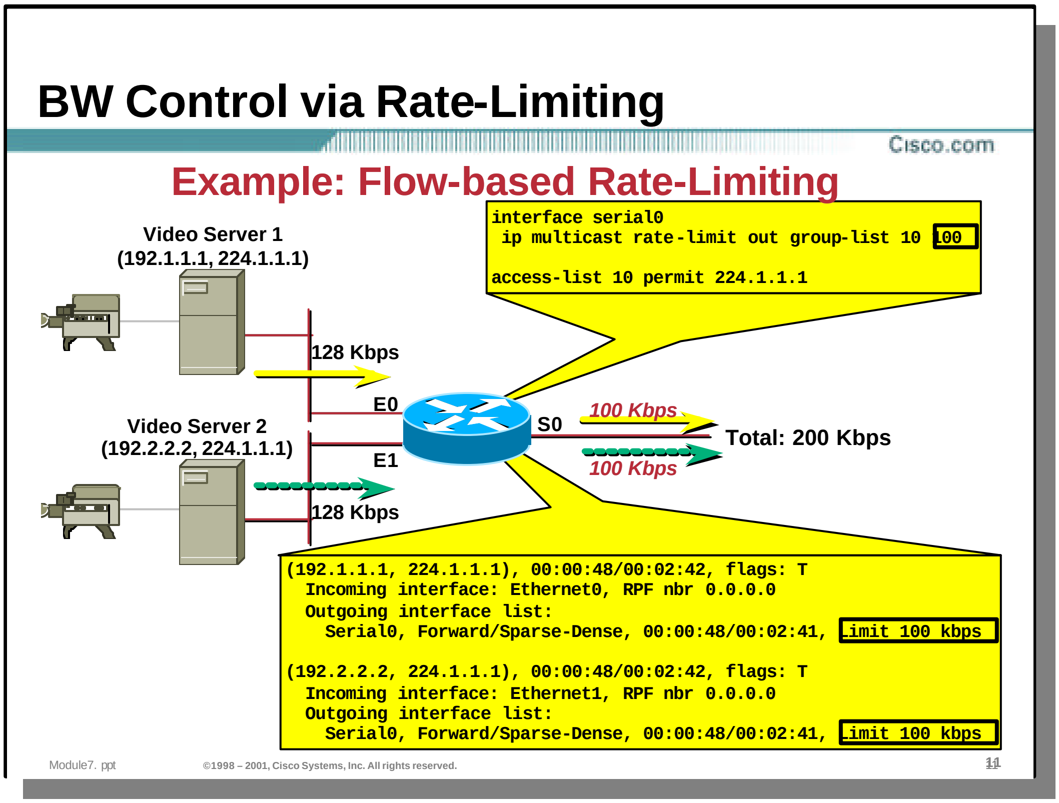
Debugging Rate-Limits

```
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.**

- **Debugging Rate-Limits**
  - Rate limits may be displayed via the ‘show ip mroute’ command. Output flow-based rate-limits are shown on the entries in the outgoing interface list (OIL) of each mroute table entry. The text following the OIL will be missing the word Int preceding the rate limit value. This indicates that this is an flow-based limit.
  - In the example above, an output flow-based rate-limit of 56kbps has been configured on interface **Serial1**.
    - NOTE: Each individual matching *flow* will be rate-limited to 56kbps, not the total aggregate of all matching flows. This means that if there are 10 active flows that match the configured flow-based rate-limit, the total aggregate rate out **Serial1** could be as high as 560kbps!



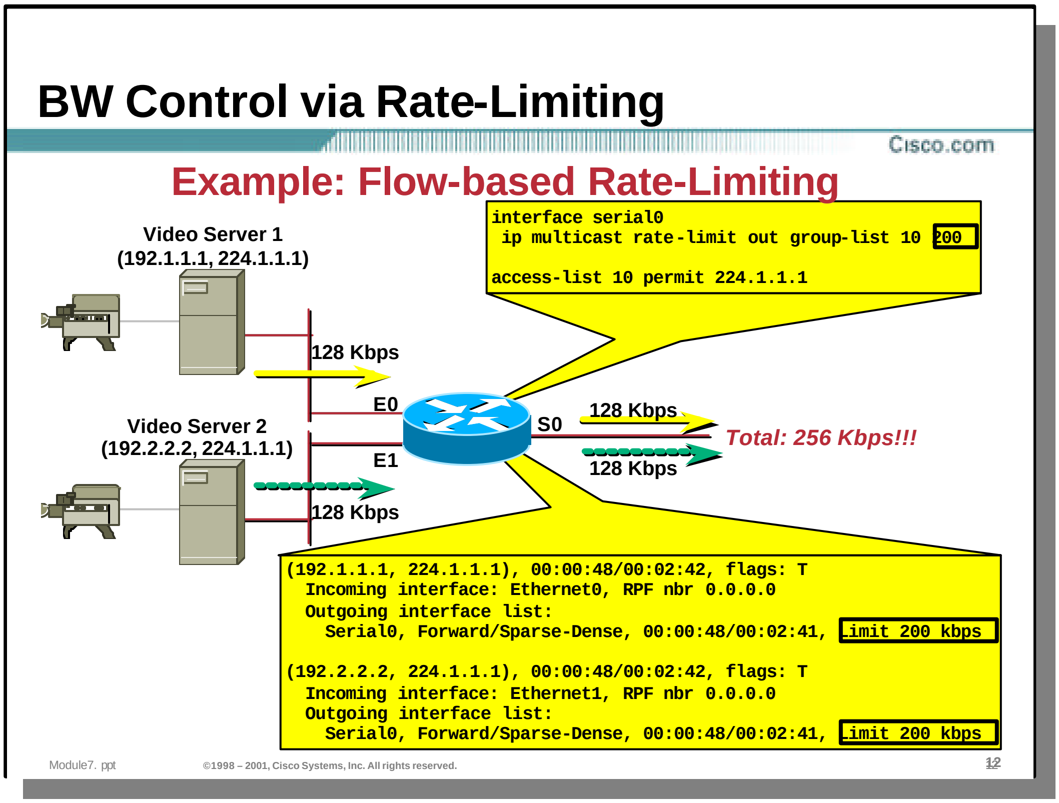
BW Control via Rate-Limiting

Example: Flow-based Rate-Limiting

```
interface serial0
 ip multicast rate-limit out group-list 10 100

access-list 10 permit 224.1.1.1
```

```
(192.1.1.1, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Ethernet0, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, limit 100 kbps

(192.2.2.2, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Ethernet1, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, limit 100 kbps
```

- **Flow-based Rate-Limiting Example**
  - In the above example, there are two 128kbps sources of video being sent to group 224.1.1.1 which in turn, flow through the 7500 router and out **Serial0** to downstream receivers (not shown).
  - Interface **Serial0** is configured with an output flow-based rate limit of 100kbps (as shown in the configuration excerpt).
  - The output of a 'show ip mroute' command clearly shows that this output flow-based rate limit has been applied to **Serial0** for both sources.
  - The results will be that *each* video flows will be rate-limited to a maximum of 100kbps, thereby causing drops in the multicast video streams.



BW Control via Rate-Limiting

Cisco.com

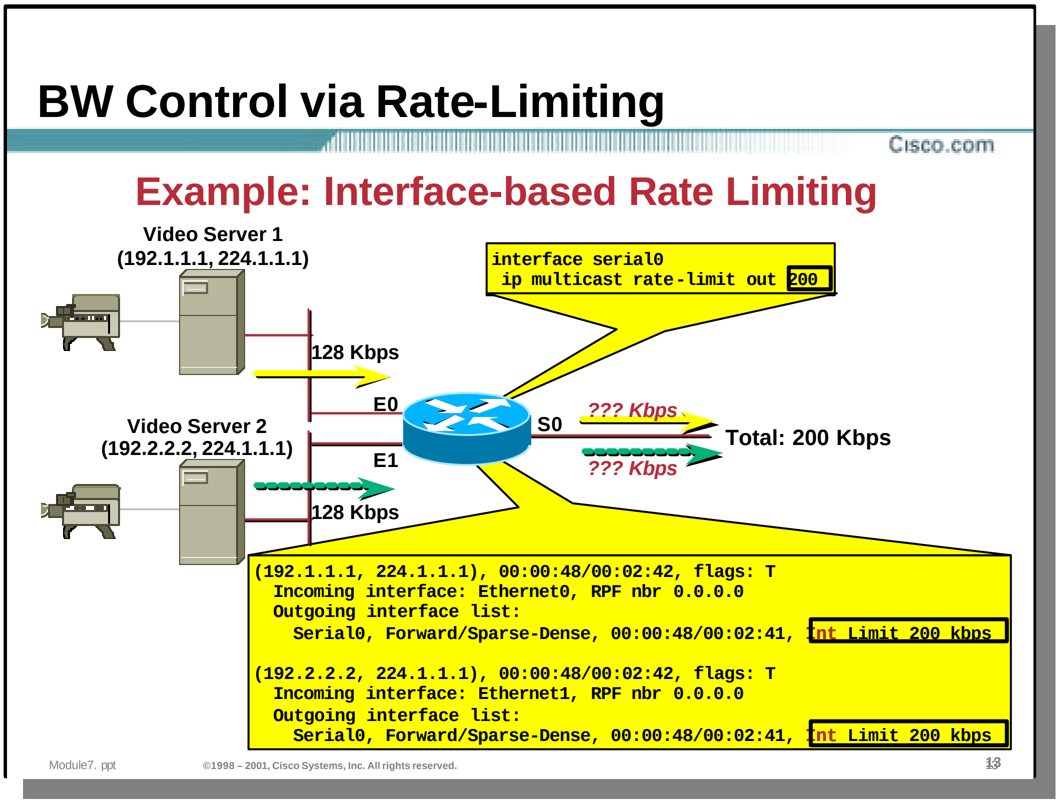
Example: Flow-based Rate-Limiting

```
interface serial0
 ip multicast rate-limit out group-list 10 200

access-list 10 permit 224.1.1.1
```

```
(192.1.1.1, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Ethernet0, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, limit 200 kbps

(192.2.2.2, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Ethernet1, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, limit 200 kbps
```

Module7. ppt 12

- **Flow-based Rate-Limiting Example**
  - Continuing with the same example, interface **Serial0** is now reconfigured with an output flow-based rate limit of 200kbps (as shown in the configuration excerpt).
  - The output of a ‘show ip mroute’ command clearly shows that this output flow-based rate limit has been applied to **Serial0** for both sources.
  - The results will be that *neither* video flow will be rate-limited since their maximum streaming speed is only 128kbps. This results in a total aggregate rate of 256kbps for both flows. Note that this rate-limit will *NOT* limit the total aggregate flow rate of these matching flows to a maximum of 200kbps as might be expected by some network engineers.



BW Control via Rate-Limiting

Example: Interface-based Rate Limiting

Video Server 1
(192.1.1.1, 224.1.1.1)

128 Kbps

E0

Video Server 2
(192.2.2.2, 224.1.1.1)

E1

128 Kbps

S0

??? Kbps

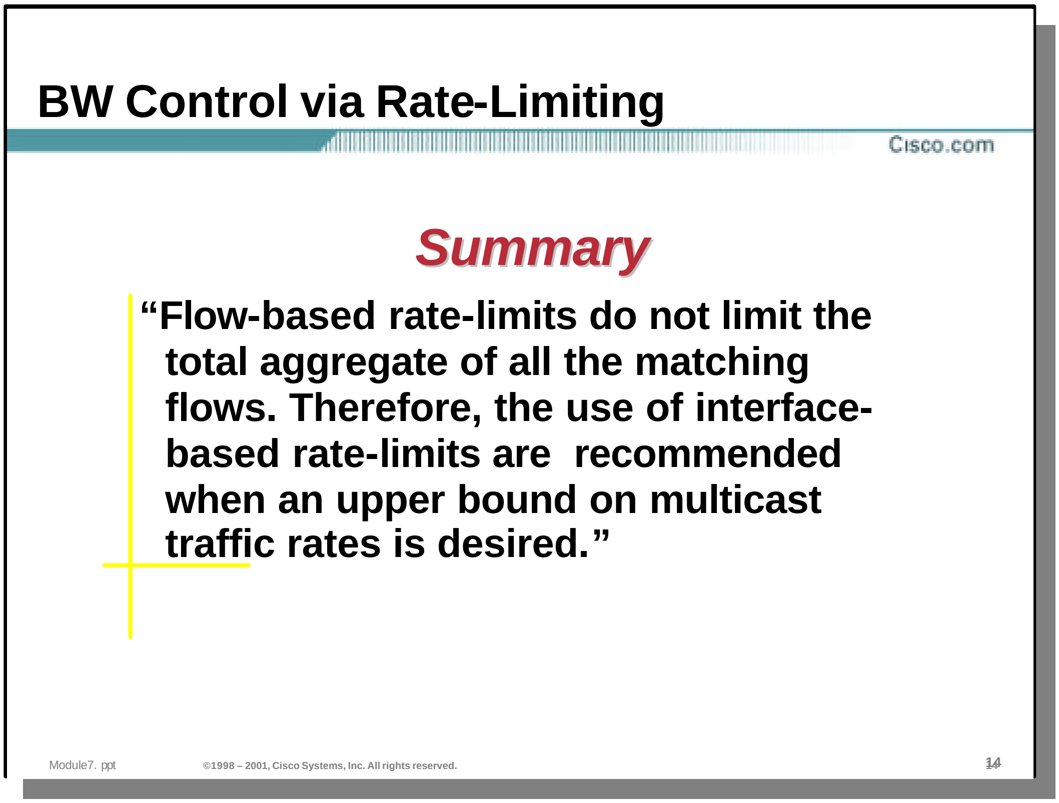
??? Kbps

Total: 200 Kbps

```
interface serial0
 ip multicast rate-limit out 200
```

```
(192.1.1.1, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Ethernet0, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, Int Limit 200 kbps

(192.2.2.2, 224.1.1.1), 00:00:48/00:02:42, flags: T
  Incoming interface: Ethernet1, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse-Dense, 00:00:48/00:02:41, Int Limit 200 kbps
```

- **Interface-based Rate-Limiting Example**
  - Finally, interface **Serial0** is once again reconfigured with an output interface-based rate limit of 200kbps (as shown in the configuration excerpt).
  - The output of a 'show ip mroute' command clearly shows that this output interface-based rate limit has been applied to **Serial0** for both sources.
  - The results will be that *all* multicast traffic (including these two video flows) will be rate-limited to a total aggregate rate of 200kbps. Note that this rate-limit results in potential loss for both flows since the combined rate of the two flows exceeds the 200kbps output interface limit.

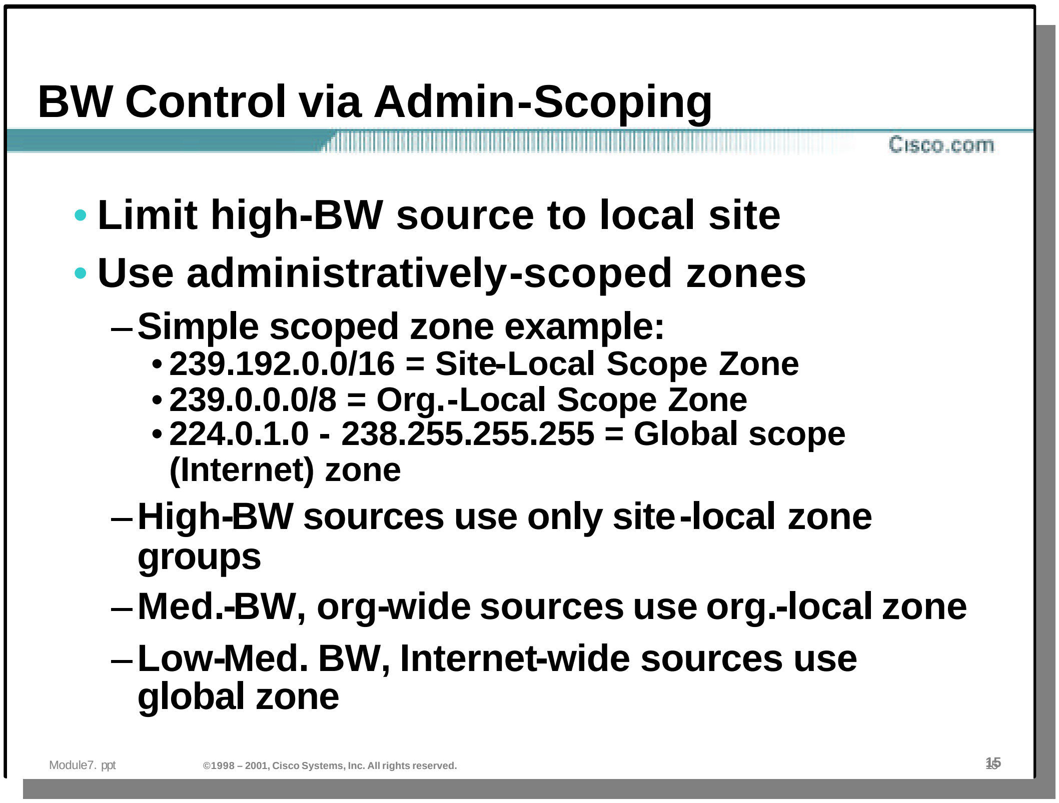


BW Control via Rate-Limiting

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."

- **Summary**
  - Flow-based rate-limits generally do not provide the sort of rate-limiting that is very useful in real-world networks. As a result, only interface-based rate-limits are normally used when an upper bound on multicast traffic is desired.



BW Control via Admin-Scoping

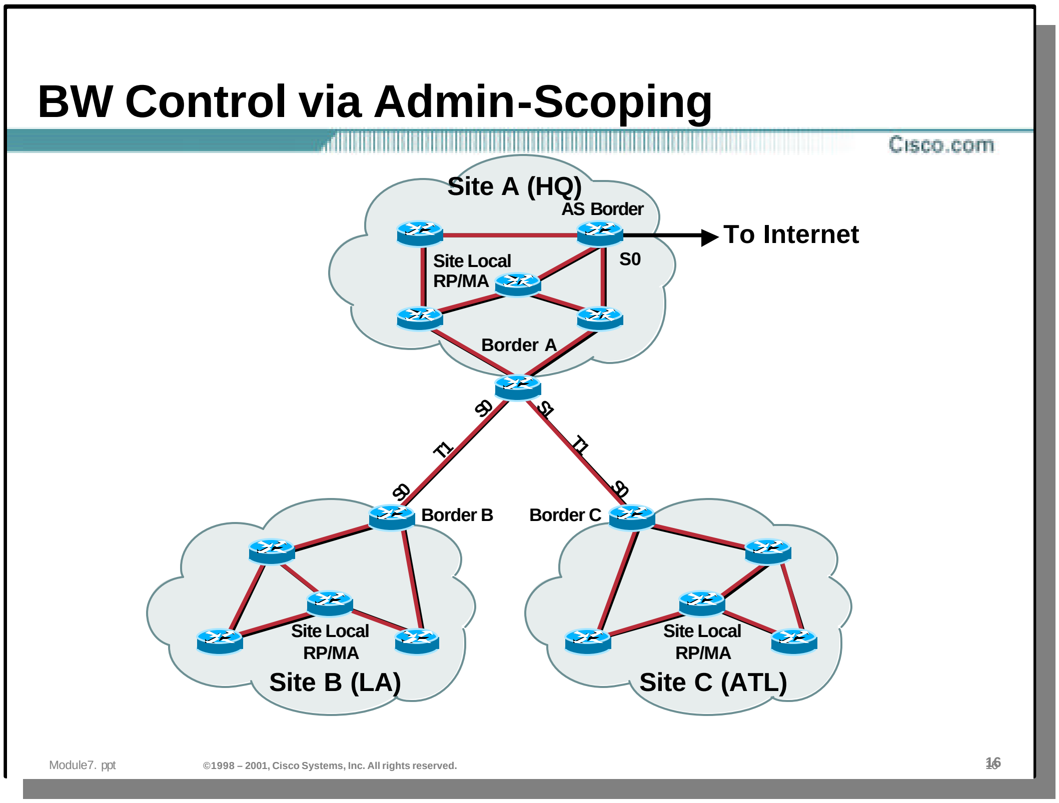
- Limit high-BW source to local site
- Use administratively-scoped zones
  - Simple scoped zone example:
    - 239.192.0.0/16 = Site-Local Scope Zone
    - 239.0.0.0/8 = Org.-Local Scope Zone
    - 224.0.1.0 - 238.255.255.255 = Global scope (Internet) zone
  - High-BW sources use only site-local zone groups
  - Med.-BW, org-wide sources use org.-local zone
  - Low-Med. BW, Internet-wide sources use global zone

Module7. ppt 15

- **BW Control via Admin Scoping**
  - Another method that can be used to control the BW usage by multicast traffic is to employ Admin-Scoped zones along with corresponding multicast boundaries. This allows one to restrict high-rate multicast flows from leaving certain geographical boundaries where bandwidth is plentiful, and going out over bandwidth constrained WAN links.
  - In the above example, the following Admin Scoped ranges are in use

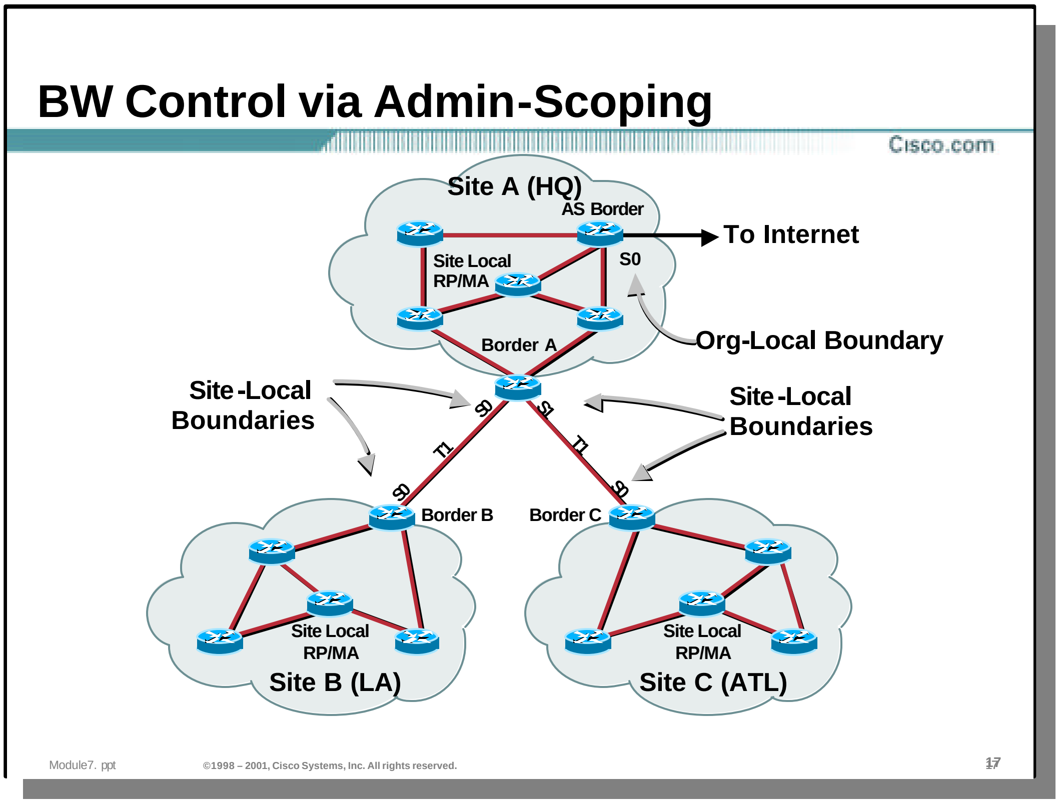
• Site-Local Scope	- 239.192.0.0/16
• Organization-Local Scope (Company?)	- 239.0.0.0/8
• Global Scope (Internet)	- 224.0.1.0 - 238.255.255.255

  - High BW, Site-Local sources should always use a group address in the Site-Local Scope.
  - Medium BW, Organization-Local sources should always use a group address in the Organization-Local Scope.
  - Sources that wish to transmit to the Internet should use group addresses that do not fall in either the Site-Local or Organization-Local scopes.



BW Control via Admin-Scoping

- **Admin-Scoping Example**
  - In this example, two remote sites (Los Angeles and Atlanta) are linked to the company HQ site via T1 lines.
    - Each of the three sites has its own Mapping Agent and a Site-Local RP that serves the Site-Local group range 239.192.0.0/16 that was described in the previous slide. This is necessary since the goal is to keep all Site-Local traffic within each site and therefore each site must have it’s own independent RP for this group range.



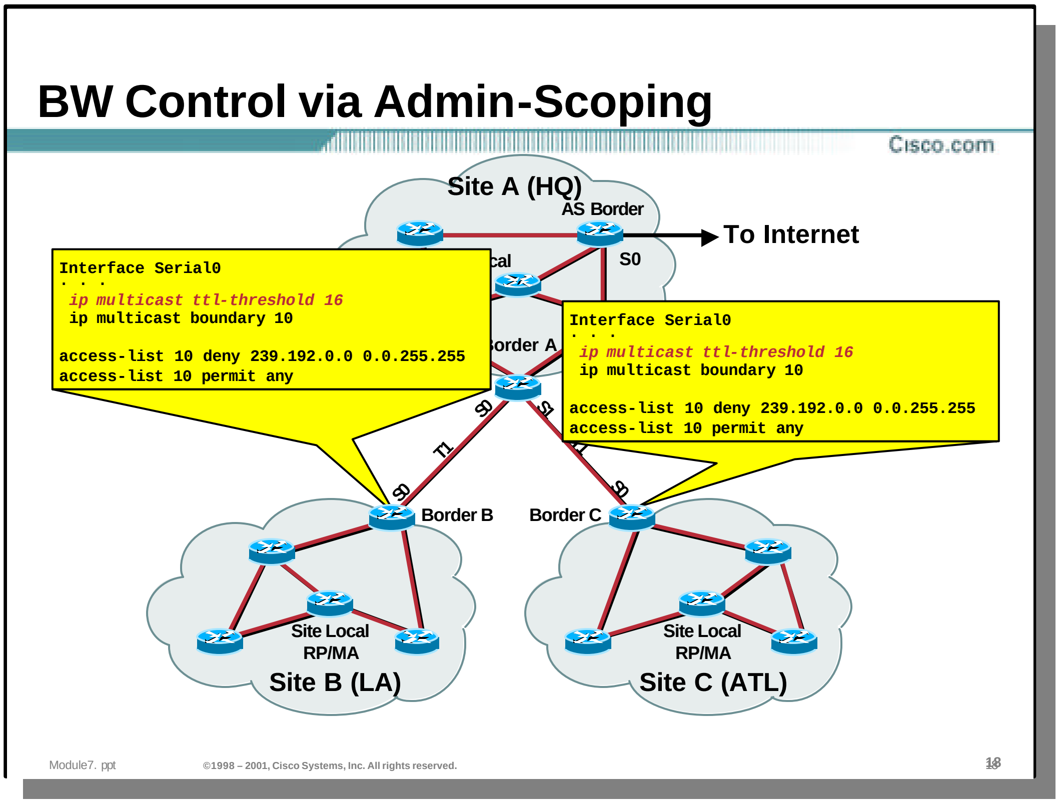
BW Control via Admin-Scoping

Cisco.com

Module7. ppt

17

- **Admin-Scoping Example**
  - The first step is to establish multicast boundaries that prevent Site-Local and Organization-Local multicast traffic from crossing certain boundaries.
    - In the case of Site-Local traffic, these boundaries are established on the T1 links on each of the three border routers, A, B and C. The Site-Local boundaries are implemented using the ‘ip multicast boundary’ interface command along with an access-control list that “denies” multicast traffic in the Site-Local group range (239.192.0.0/16) from crossing this boundary.
    - The Organization-Local boundary is established on the link to the internet and is also implemented using the ‘ip multicast boundary’ interface command. The access-control list for this boundary “denies” multicast traffic in the Organization-Local group range (239.0.0.0/8) from crossing this boundary.



BW Control via Admin-Scoping

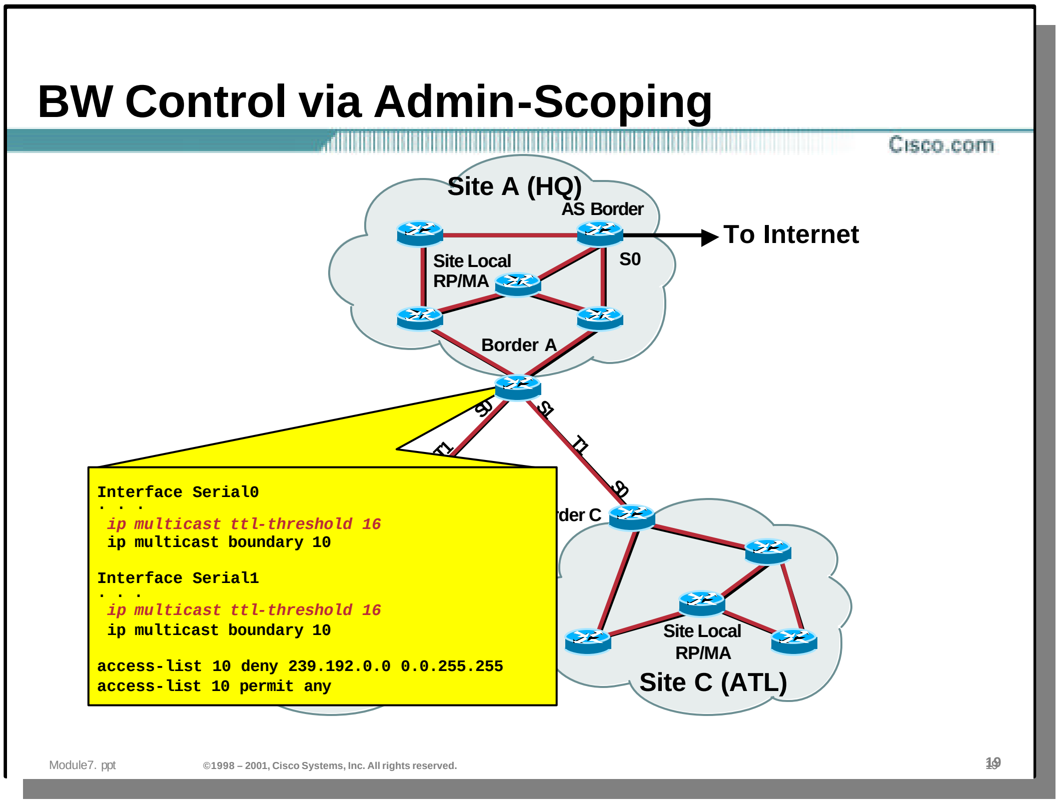
Cisco.com

```
Interface Serial0
. . .
 ip multicast ttl-threshold 16
 ip multicast boundary 10

access-list 10 deny 239.192.0.0 0.0.255.255
access-list 10 permit any
```

```
Interface Serial0
. . .
 ip multicast ttl-threshold 16
 ip multicast boundary 10

access-list 10 deny 239.192.0.0 0.0.255.255
access-list 10 permit any
```

Module7. ppt

- **Admin-Scoping Example**
  - The configuration commands necessary to establish the Site-Local boundaries on border routers B and C are listed in the above drawing.
    - In both configurations, the Site-Local boundary is established on interface **Serial0** via the '`ip multicast boundary 10`' interface command. Access-control list 10 is constructed to "deny" the passage of any traffic in the Site-Local group range (239.192.0.0/16) while all other multicast groups are "permitted" to cross the interface.
    - Pay particular attention to the '`ip multicast ttl-threshold 16`' command also configured on **Serial0**. The requirement for this command will be discussed in a later slide.



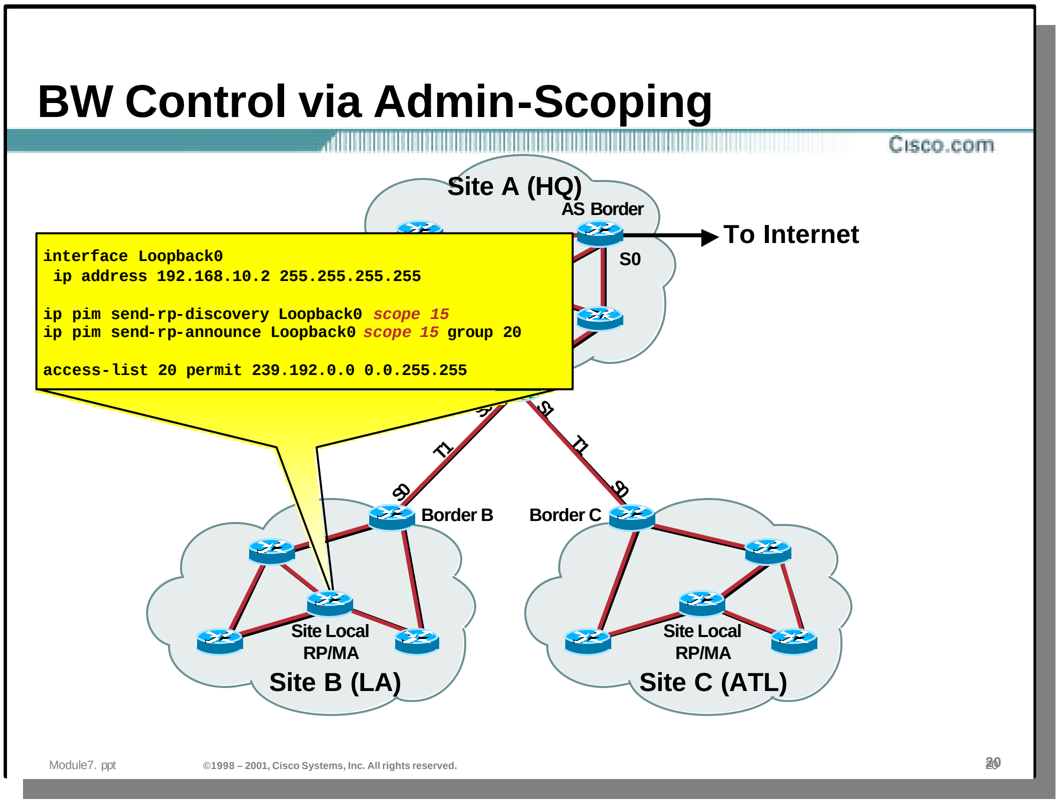
BW Control via Admin-Scoping

```
Interface Serial0
. . .
 ip multicast ttl-threshold 16
 ip multicast boundary 10

Interface Serial1
. . .
 ip multicast ttl-threshold 16
 ip multicast boundary 10

access-list 10 deny 239.192.0.0 0.0.255.255
access-list 10 permit any
```

Module7. ppt

19

- **Admin-Scoping Example**
  - The configuration commands necessary to establish the Site-Local boundaries on border router A are listed in the above drawing.
    - In this case, the Site-Local boundary is established on both interface **Serial0** and **Serial1** via the ‘`ip multicast boundary 10`’ interface command. Access-control list 10 is constructed to “deny” the passage of any traffic in the Site-Local group range (239.192.0.0/16) while all other multicast groups are “permitted” to cross the interface.
    - Again notice the ‘`ip multicast ttl-threshold 16`’ command also configured on **Serial0** and **Serial1**. The requirement for this command will be discussed in a later slide.



BW Control via Admin-Scoping

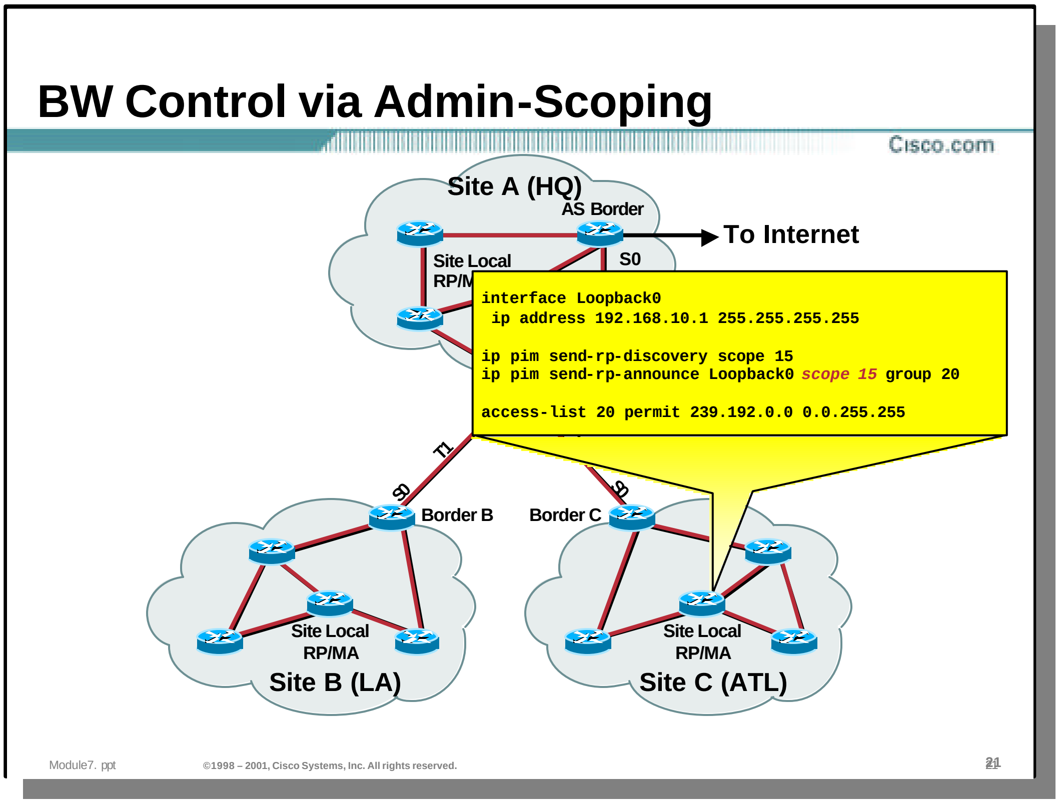
```
interface Loopback0
 ip address 192.168.10.2 255.255.255.255

ip pim send-rp-discovery Loopback0 scope 15
ip pim send-rp-announce Loopback0 scope 15 group 20

access-list 20 permit 239.192.0.0 0.0.255.255
```

Module7. ppt

20

- **Admin-Scoping Example**
  - The next step is to configure the independent Mapping Agents for each site as well as the independent Site-Local group RP. In the drawing above, the configuration for the RP/Mapping Agent in Site B is shown.
    - In this case, interface **Loopback0** has been configured for use as the interface of choice for the Mapping Agent and RP. A loopback interface is often used for this purpose as it provides more flexibility in the management of the Mapping Agent as well as the selection of the RP when multiple candidate RP's are define. (The highest candidate IP address is chosen as the active RP by Mapping Agents.)
    - The '`ip pim send-rp-discovery`' global command defines the router as a Mapping Agent for Site B with an IP address of the loopback interface. Note that a TTL scope of 15 is used on this command so that the Discovery messages sourced by this Mapping Agent will not exit the Site. (This is accomplished by the use of a ttl-threshold of 16 on **Serial0** that was configured in the previous slides.)
    - The '`ip pim send-rp-announce`' global command along with access-control list 20, defines the router as a Candidate RP for the Site-Local group range. Note that a TTL scope of 15 is used on this command so that the Announce messages sourced by this Candidate-RP will not exit the Site. This is accomplished by the use of a ttl-threshold of 16 on **Serial0** that was configured in the previous slides.)

      Note: It is crucial that the Candidate-RP Announce messages do not leak outside of this site and into other sites. If this were to occur, the Mapping Agent(s) in the other site(s) might select the Candidate-RP in Site B as the currently active RP for the Site-Local group. This would break Site-Local multicast in that site.



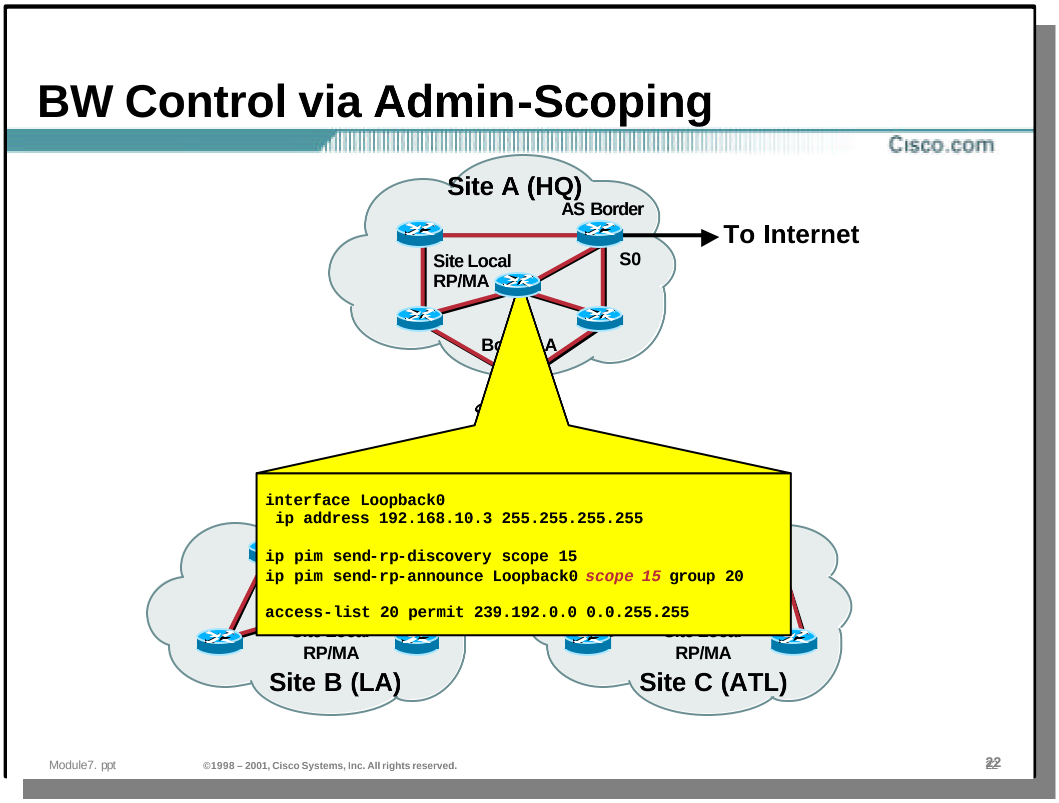
BW Control via Admin-Scoping

```
interface Loopback0
 ip address 192.168.10.1 255.255.255.255

ip pim send-rp-discovery scope 15
ip pim send-rp-announce Loopback0 scope 15 group 20

access-list 20 permit 239.192.0.0 0.0.255.255
```

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 21

- **Admin-Scoping Example**
  - In the drawing above, the configuration for the RP/Mapping Agent in Site C is shown.
    - Interface **Loopback0** has also been configured for use as the interface of choice for the Mapping Agent and RP.
    - The ‘`ip pim send-rp-discovery`’ global command defines the router as a Mapping Agent for Site C with an IP address of the loopback interface. Note once again that a TTL scope of 15 is used on this command so that the Discovery messages sourced by this Mapping Agent will not exit the Site. (This is accomplished by the use of a ttl-threshold of 16 on **Serial0** that was configured in the previous slides.)
    - The ‘`ip pim send-rp-announce`’ global command along with access-control list 20, defines the router as a Candidate RP for the Site-Local group range. Note that a TTL scope of 15 is used on this command so that the Announce messages sourced by this Candidate-RP will not exit the Site. This is accomplished by the use of a ttl-threshold of 16 on **Serial0** that was configured in the previous slides.)

      Note: It is crucial that the Candidate-RP Announce messages do not leak outside of this site and into other sites. If this were to occur, the Mapping Agent(s) in the other site(s) might select the Candidate-RP in Site C as the currently active RP for the Site-Local group. This would break Site-Local multicast in that site.

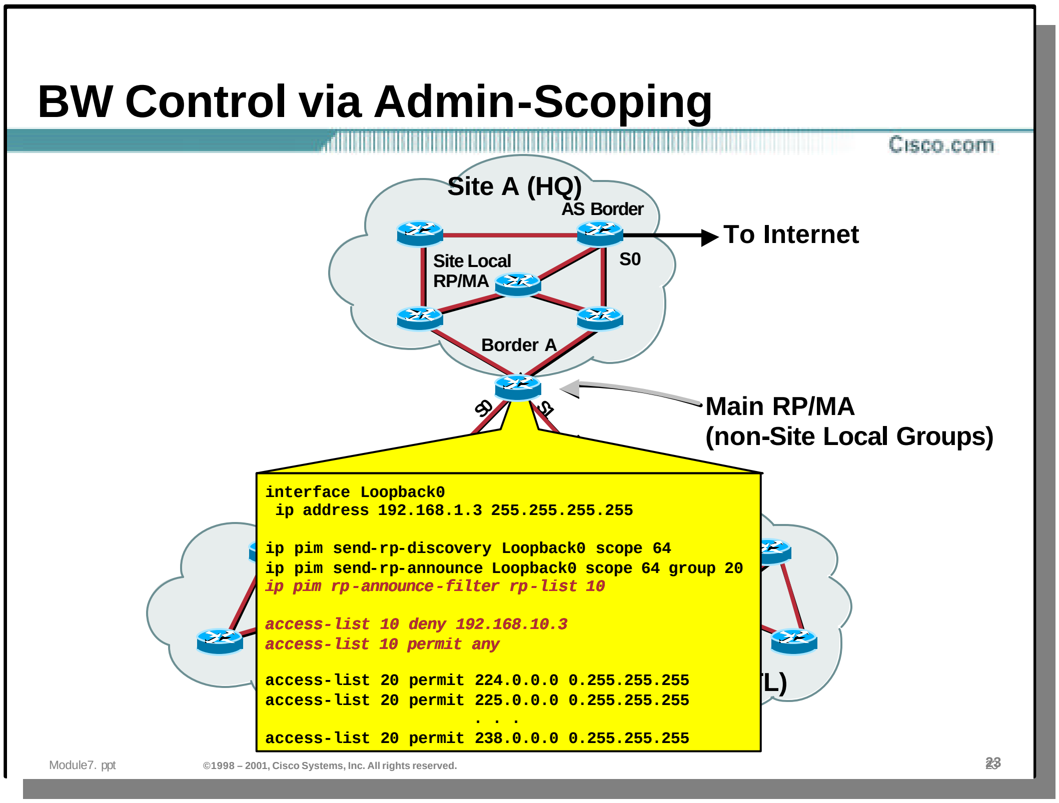


BW Control via Admin-Scoping

```
interface Loopback0
 ip address 192.168.10.3 255.255.255.255

ip pim send-rp-discovery scope 15
ip pim send-rp-announce Loopback0 scope 15 group 20

access-list 20 permit 239.192.0.0 0.0.255.255
```

- **Admin-Scoping Example**
  - In the drawing above, the configuration for the RP/Mapping Agent in Site A is shown.
    - Interface **Loopback0** has also been configured for use as the interface of choice for the Mapping Agent and RP.
    - The ‘`ip pim send-rp-discovery`’ global command defines the router as a Mapping Agent for Site A with an IP address of the loopback interface. Note once again that a TTL scope of 15 is used on this command so that the Discovery messages sourced by this Mapping Agent will not exit the Site. (This is accomplished by the use of a ttl-threshold of 16 on **Serial0** that was configured in the previous slides.)
    - The ‘`ip pim send-rp-announce`’ global command along with access-control list 20, defines the router as a Candidate RP for the Site-Local group range. Note that a TTL scope of 15 is used on this command so that the Announce messages sourced by this Candidate-RP will not exit the Site. This is accomplished by the use of a ttl-threshold of 16 on **Serial0** that was configured in the previous slides.)

    Note: It is crucial that the Candidate-RP Announce messages do not leak outside of this site and into other sites. If this were to occur, the Mapping Agent(s) in the other site(s) might select the Candidate-RP in Site A as the currently active RP for the Site-Local group. This would break Site-Local multicast in that site.

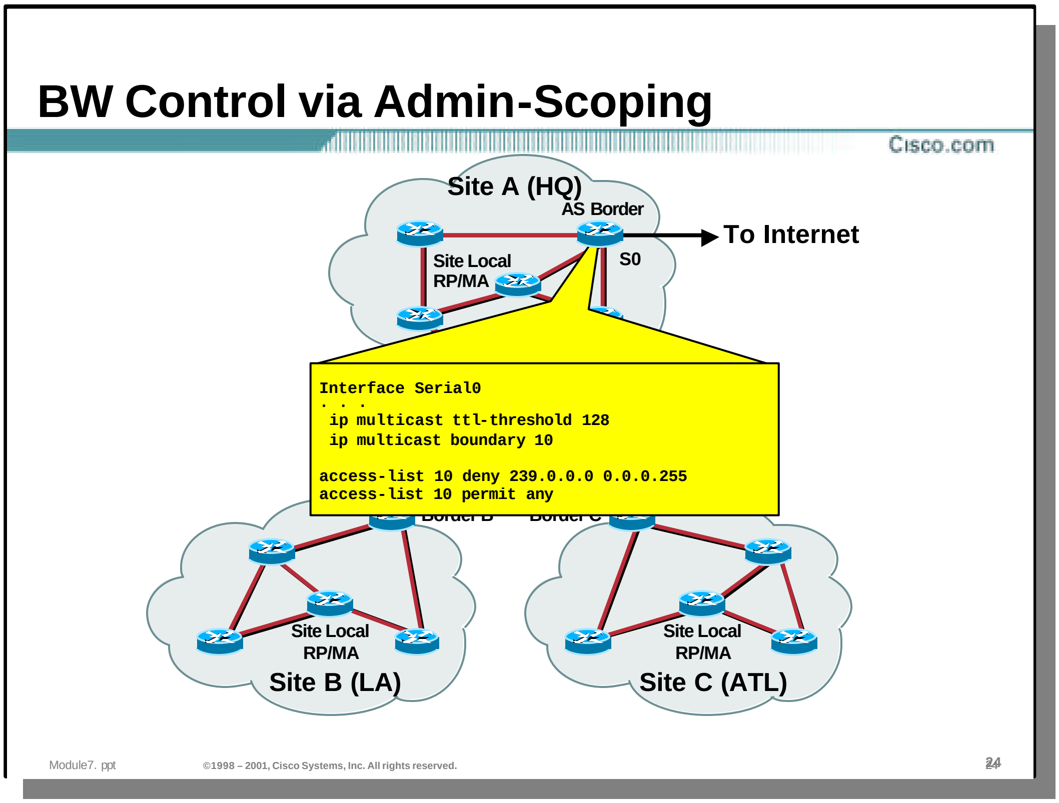


BW Control via Admin-Scoping

```
interface Loopback0
 ip address 192.168.1.3 255.255.255.255

ip pim send-rp-discovery Loopback0 scope 64
ip pim send-rp-announce Loopback0 scope 64 group 20
ip pim rp-announce-filter rp-list 10

access-list 10 deny 192.168.10.3
access-list 10 permit any

access-list 20 permit 224.0.0.0 0.255.255.255
access-list 20 permit 225.0.0.0 0.255.255.255
                        . . .
access-list 20 permit 238.0.0.0 0.255.255.255
```

- **Admin-Scoping Example**
  - Next, an RP for all the non-Site-Local multicast groups must be configured. Border router A has been chosen for this task. Additionally, router A is configured as a Mapping Agent with sufficient scope to cover the entire network. (Note: The independent Mapping Agents for each site make this step unnecessary as they can handle the Mapping Agent functionality for their site.)
    - Once again, interface **Loopback0** has been configured for use as the interface of choice for the Mapping Agent and RP.
    - The '`ip pim send-rp-announce`' global command along with access-control list 20, defines the router as a Candidate RP for all non-Site-Local groups. Note that a TTL scope of 64 is used on this command so that the Announce messages sourced by this Candidate-RP will reach all routers in all Sites.
    - The '`ip pim send-rp-discovery`' global command defines the router as a Mapping Agent for the entire network. Note that a TTL scope of 64 is also used on this command so that the Discovery messages sourced by this Mapping Agent reach all routers in all Sites.
    - Because the scope of the Discovery messages are 64, they will reach all routers in all sites. Therefore, care must be taken to insure that the C-RP information from the Site-Local C-RP in the HQ site is not accepted and inadvertently advertised in Discovery messages by the Mapping Agent. If this were to occur, the routers in the other site(s) might select the Candidate-RP in the HQ Site as the currently active RP for the Site-Local group. This would break Site-Local multicast in that site. To prevent this from happening, the '`ip pim rp-announce-filter`' command along with access-control list 10 is used to filter out C-RP Announcement messages from the Site-Local C-RP in the HQ Site.

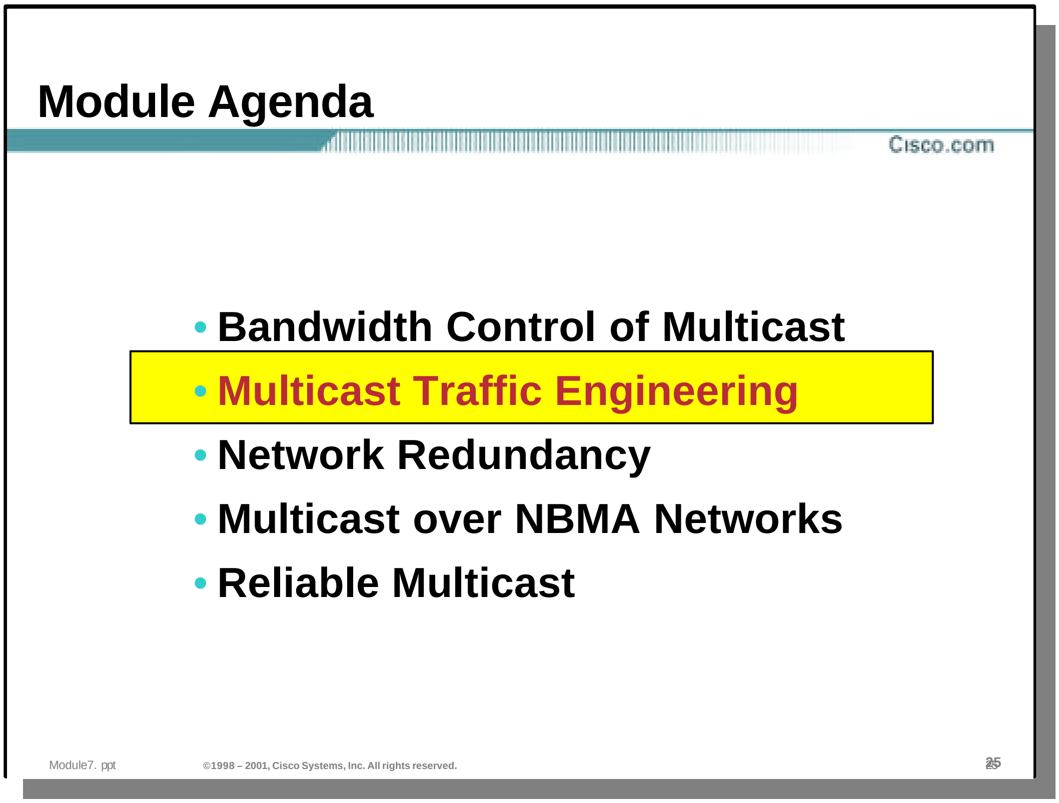
      Note: This problem can be avoided if router A is *not* configured as a Mapping Agent.



BW Control via Admin-Scoping

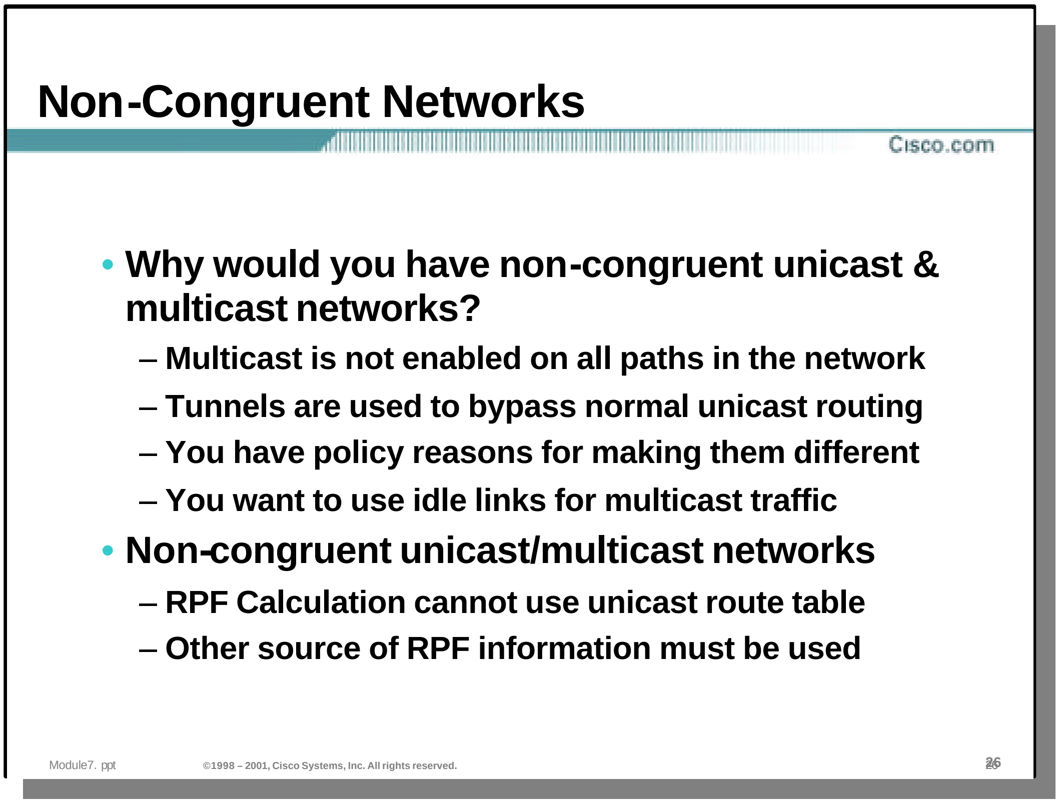
```
Interface Serial0
. . .
 ip multicast ttl-threshold 128
 ip multicast boundary 10

access-list 10 deny 239.0.0.0 0.0.0.255
access-list 10 permit any
```

- **Admin-Scoping Example**
  - Finally, the AS Border router must be configured with a multicast boundary so that all locally scoped multicast traffic in the 239.0.0.0/8 range is blocked from entering or leaving the company.
    - In this configuration, the multicast boundary is established on interface **Serial0** via the '`ip multicast boundary 10`' interface command. Access-control list 10 is constructed to "deny" the passage of any traffic in the Admin-Scoped group range (239.0.0.0/8) while all other multicast groups are "permitted" to cross the interface.
    - Although it is not necessary to implement Admin-Scoping, the '`ip multicast ttl-threshold 128`' command is also configured on **Serial0** of the AS border router. This is often used to provide TTL scoping of traffic inside of the company. Sources that do not wish to have their multicast traffic leave the company can transmit with a TTL less than 128 and be insured that the traffic will not be forwarded into the Internet.



Module Agenda

- Bandwidth Control of Multicast
- Multicast Traffic Engineering
- Network Redundancy
- Multicast over NBMA Networks
- Reliable Multicast



Non-Congruent Networks

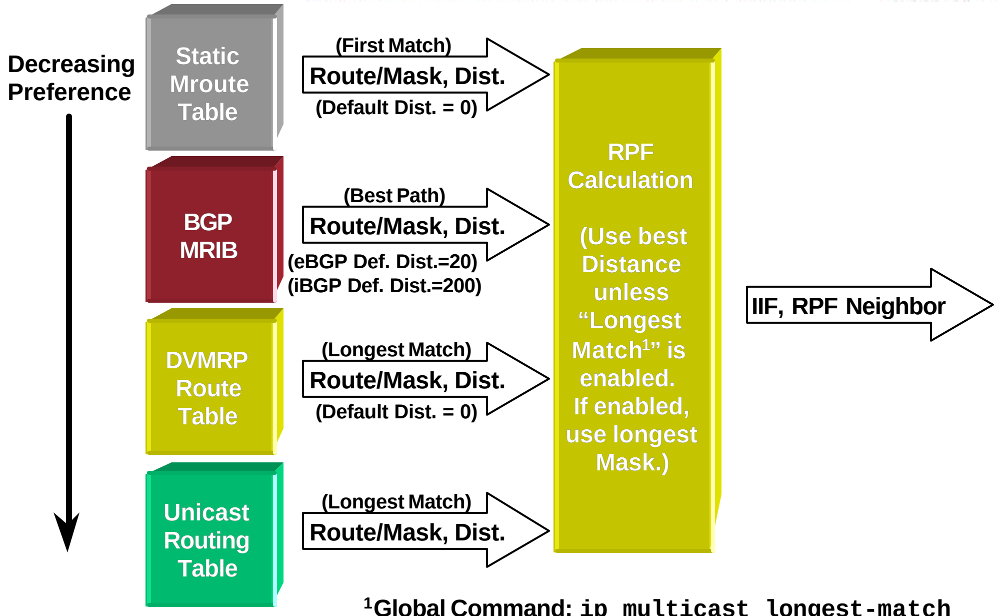
- Why would you have non-congruent unicast & multicast networks?
  - – Multicast is not enabled on all paths in the network
  - – Tunnels are used to bypass normal unicast routing
  - – You have policy reasons for making them different
  - – You want to use idle links for multicast traffic
- Non-congruent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

Module7. ppt 26

- **Non-congruent Multicast/Unicast Networks**
  - – There are several cases were it might be desirable for Unicast and Multicast traffic to follow separate paths through a network. Some of these cases are:
    - When multicast is not enabled on all interfaces of a router.
    - When tunnels are in use to bypass unicast-only sections of a network.
    - There exists some policy that dictates that multicast traffic follow different paths.
    - It is desirable for multicast traffic to flow over idle/backup links for better load balancing.
  - – When the unicast and multicast networks are not congruent, certain limitations come into play, such as:
    - The RPF calculation cannot use the unicast routing table since that would cause the unicast and multicast networks to be congruent by default.
    - Some other source of information than the unicast routing table must be used. This imposes additional configuration and administration requirements that (in many cases) are non-trivial.

PIM RPF Calculation Details

Cisco.com



Decreasing Preference

Static Mroute Table

(First Match)
Route/Mask, Dist.
(Default Dist. = 0)

BGP MRIB

(Best Path)
Route/Mask, Dist.
(eBGP Def. Dist.=20)
(iBGP Def. Dist.=200)

DVMRP Route Table

(Longest Match)
Route/Mask, Dist.
(Default Dist. = 0)

Unicast Routing Table

(Longest Match)
Route/Mask, Dist.

RPF Calculation

(Use best Distance unless “Longest Match[1]” is enabled. If enabled, use longest Mask.)

IIF, RPF Neighbor

[1] Global Command: `ip multicast longest-match`

- **PIM RPF Calculations**

  Cisco IOS permits other sources of information to be used in the RPF calculation other than the unicast routing table. In general, these other sources are preferred based on their Admin. Distance. If Admin Distance values are equal, the sources are preferred in the order listed below:

  – Static Mroute Table

  Static Mroutes may be defined that are local to the router on which they are defined. If a matching Static Mroute is defined, its default Admin. Distance is zero and is therefore preferred over other sources. (If another source also has a distance of zero, the Static Mroute takes precedence.)

  – BGP Multicast RIB (M-RIB)

  If MBGP is in use and a matching prefix exists in the MBGP M-RIB, it will be used as long as its Admin. Distance is the lowest of the other sources. (MBGP M-RIB prefixes are preferred over DVMRP or Unicast routes if the Admin Distances are the same.)

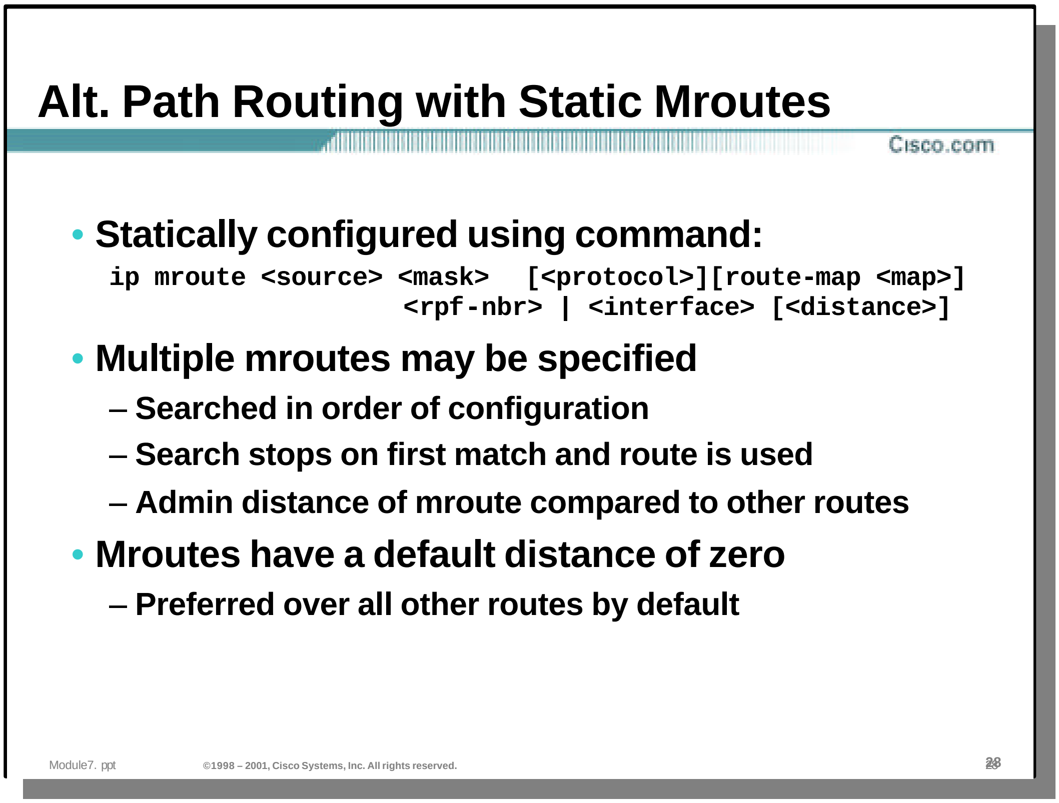
  – DVMRP Route Table

  If DVMRP routes are being exchanged and there exists a matching route in the DVMRP route table, the default Admin. Distance of this route is zero. DVMRP routes are preferred over Unicast routes if their Admin. Distances are equal.

  – Unicast Route Table

  This is least preferred source of information. If no other source has a matching route with a lower Admin. Distance, then this information is used.

  Note: The above behavior can be modified so that the longest match route is used from the available sources. This is configured with the ‘ip multicast longest-match’ hidden command.



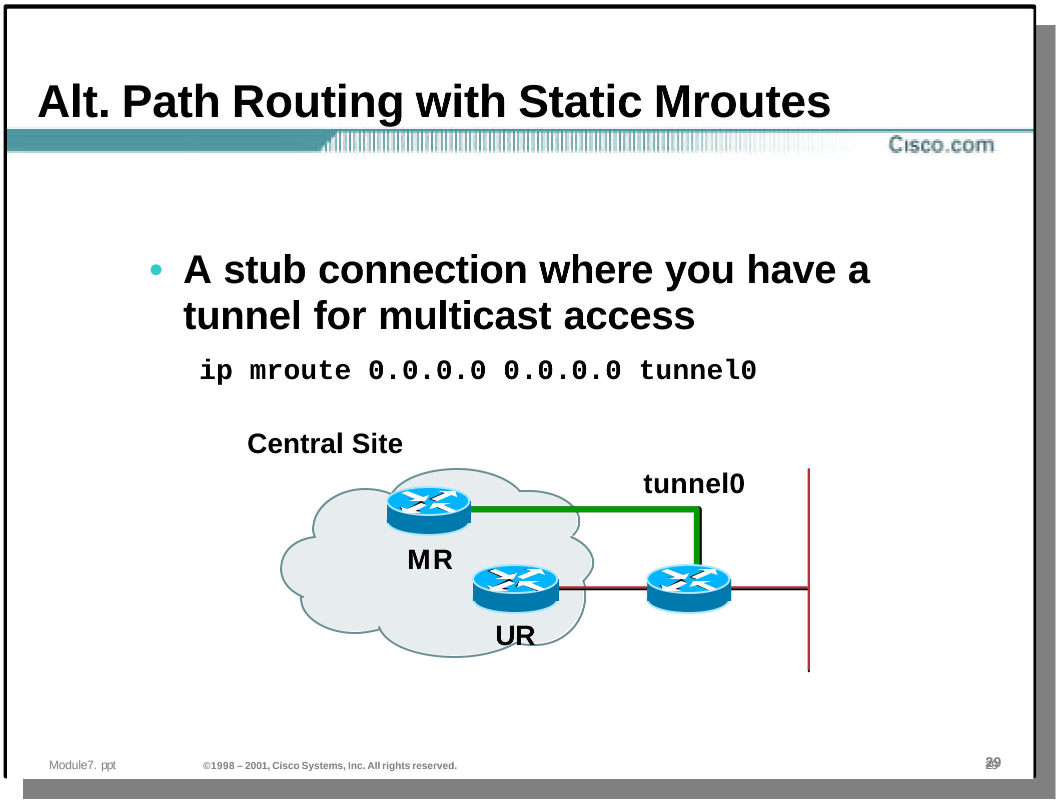
Alt. Path Routing with Static Mroutes

- **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**

- **Static Mroutes**
  - A Static Mroute may be configured using the command listed above to match based on the ***<source>*** and ***<mask>*** parameters.
    - Either an ***<rpf-nbr>*** address or a specific ***<interface>*** can be configured as the next-hop.
    - An optional ***<protocol>*** may be configured which requires that a matching route must exist in the specified unicast routing protocol's database for the Static Mroute to match.
    - An optional **route-map <*map*>** clause may be configured to constrain the match to the qualifications specified in the route-map in order for the Static Mroute to match.
    - The default Admin. Distance of a Static Mroute is zero. This value may be overridden by the use of the ***<distance>*** parameter.
  - If multiple Static Mroutes are configured, they are searched for a match in the order in which they were configured. If a match is found, the search terminates and the Static Route is used if it has an Admin. Distance less than or equal to any other source of RPF information.
  - **Note:** Unlike their unicast counterparts, Static Mroutes only have significance on the router on which they are defined and cannot be redistributed.



Alt. Path Routing with Static Mroutes

Cisco.com

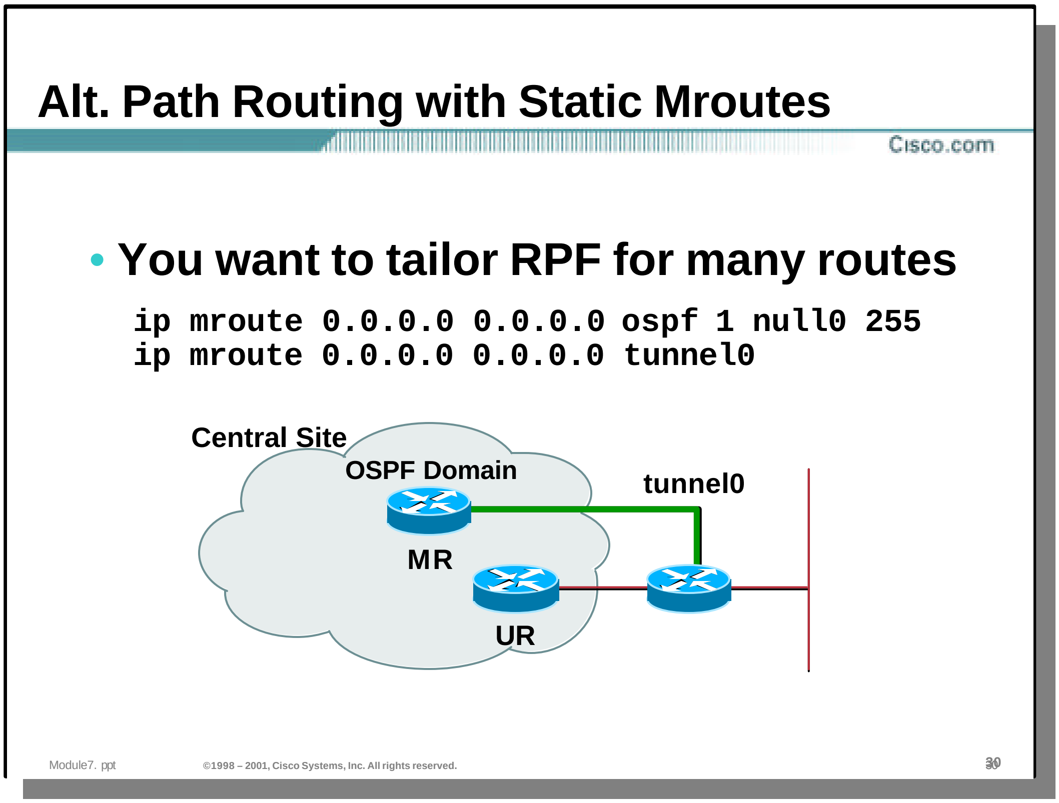
- **A stub connection where you have a tunnel for multicast access**

```
ip mroute 0.0.0.0 0.0.0.0 tunnel0
```

Central Site
tunnel0
MR
UR

- **Static Mroute: Example 1**
  – In this example, a multicast router (MR) has been configured with a tunnel to a multicast router outside of the network so that it can receive multicast traffic.
    - The ‘`ip mroute 0.0.0.0 0.0.0.0 Tunnel0`’ command instructs router “MR” to RPF to **Tunnel0** for all multicast sources instead of using the unicast routing information. Therefore, any multicast traffic arriving via **Tunnel0** will be RPF correctly using this Static Mroute. If this Static Mroute was not used, router “MR” would us the unicast routing information which would cause it to RPF to router “UR” instead of **Tunnel0** for traffic arriving from the outside of the network.

      Note: While this is a simple way to *force* router “MR” to RPF to the tunnel for traffic arriving from outside of the network, traffic arriving at “MR” from sources *inside* the network will RPF fail. This is because the source/mask covers *all* multicast sources, both inside and outside of the network. Obviously, a more sophisticated solution is required. (See next slide.)



Alt. Path Routing with Static Mroutes

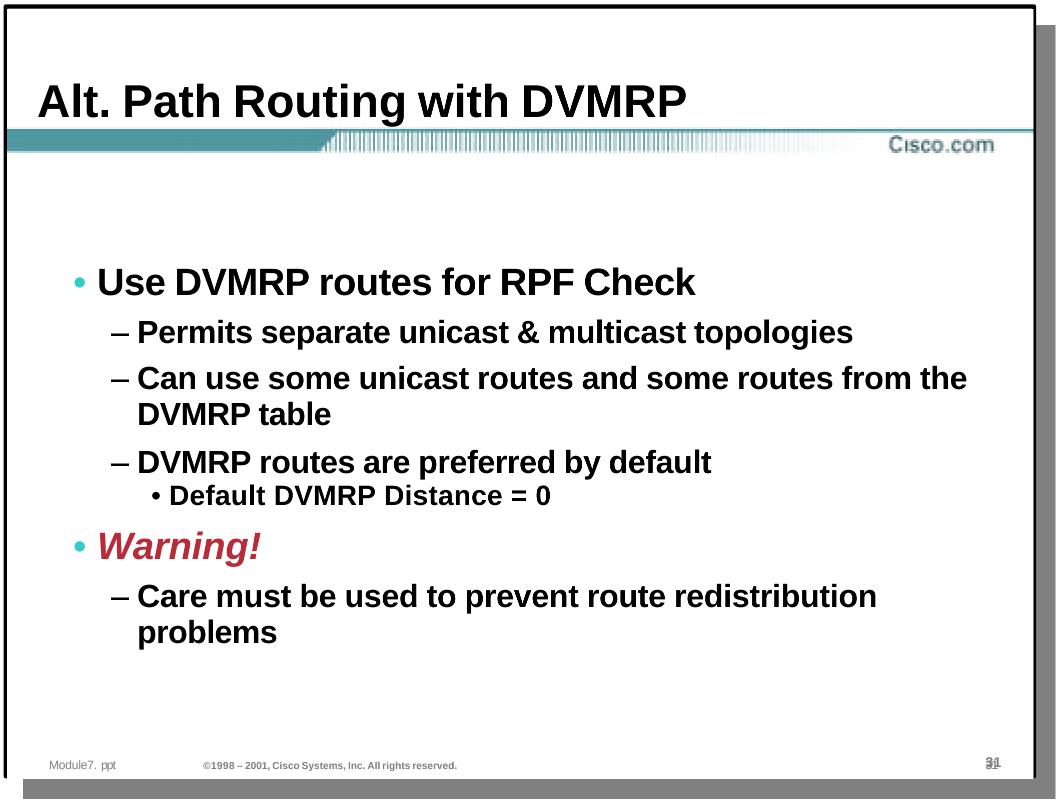
Cisco.com

- 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 tunnel0
```

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 30

- **Static Mroute: Example 2**
  - In this example, the multicast router (MR) has been reconfigured so that only traffic arrving from outside the network will RPF to **Tunnel0** while traffic arriving from sources inside the OSPF domain will RPF correctly.
    - An additional Static Mroute command '`ip mroute 0.0.0.0 0.0.0.0 ospf 1 null0 255`' is configured *ahead of* the original Static Mroute used in the previous example. The command instructs router "MR" to match on any multicast source that has a corresponding matching entry in the *OSPF 1* process database. For these matching sources, the RPF interface is set to **Null0** and the Admin. Distance of the Static Mroute is set to **255**.
    - The second (and original) '`ip mroute 0.0.0.0 0.0.0.0 Tunnel0`' command instructs router "MR" to RPF to **Tunnel0** for all multicast sources just as was done in the previous example. However, since this command appears second in the configuration, it will only be reached if the first command fails to match (i.e. the source is *not* inside the OSPF domain).
  - Exactly how these two commands combine to achieve the desired result may not be immediately obvious and is therefore described below:
    - **Sources outside the OSPF domain:** Will not match on the first Static Mroute in the list since there is no matching route in the OSPF database. However, the second Static Mroute *will* match with an RPF interface of **Tunnel0** and an Admin. Distance of zero. Therefore, these sources will RPF to **Tunnel0**.
    - **Sources inside the OSPF domain:** Will match on the first Static Mroute because there is a matching route in the OSPF database and the search of the Static Mroutes terminates with an RPF interface of **Null0** and an Admin. Distance of **255**. However, since the unicast routing table will also have a matching OSPF route with a lower (better) Admin. Distance than **255**, the route in the unicast routing table will be used. Therefore, the router will RPF to the correct interface for this source inside the OSPF domain.

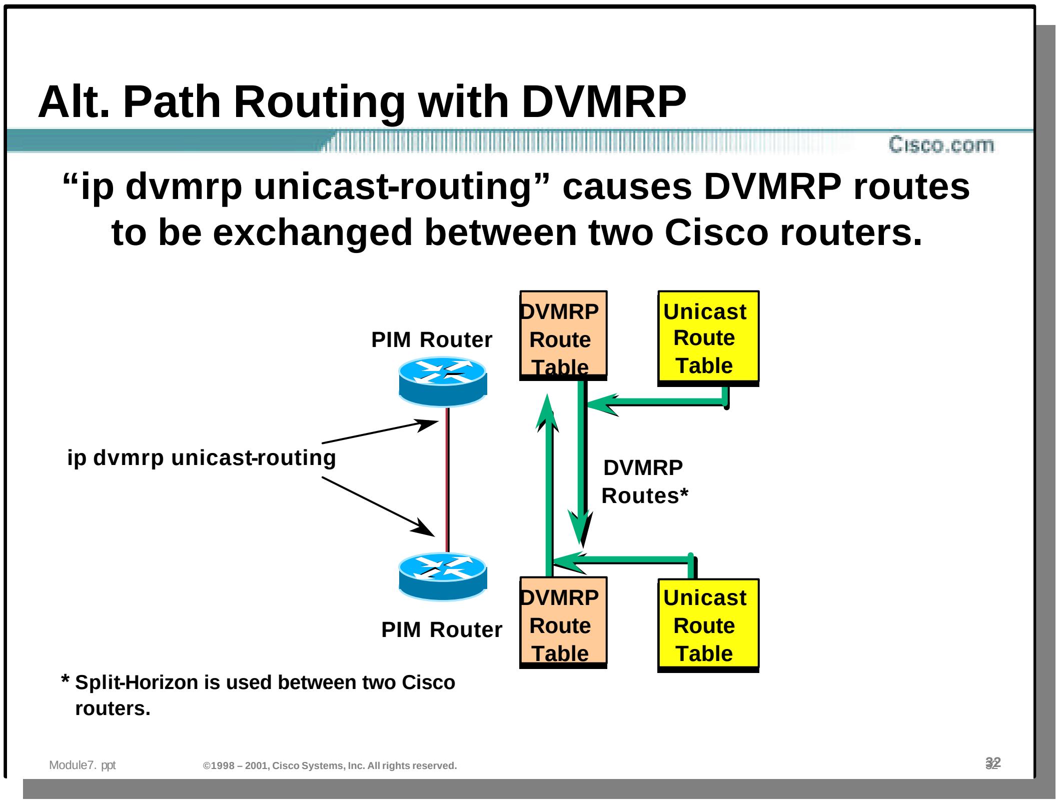


Alt. Path Routing with DVMRP

- **Use DVMRP routes for RPF Check**
  - **Permits separate unicast & 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**

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 31

- **Using DVMRP Routes**
  - As shown in the slide on RPF Details, DVMRP routes can be used as an alternate source of RPF information which, in turn, can be used to provide separate unicast and multicast topologies.
  - The default Admin. Distance of DVMRP routes (zero) makes them preferred over routes in the unicast route table. (In the case of a tie in Admin. Distance values, a DVMRP route is preferred over a unicast route for RPF calculation.)
  - Just as in unicast redistribution scenarios, care must be taken to avoid route loops from occurring when exchanging DVMRP routes between Cisco routers. This is due to the fact that unicast routes are automatically injected into DVMRP by default. This redistributing can cause multicast route loops or RPF failures to occur.



Alt. Path Routing with DVMRP

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

PIM Router

DVMRP Route Table

Unicast Route Table

ip dvmrp unicast-routing

DVMRP Routes*

PIM Router

DVMRP Route Table

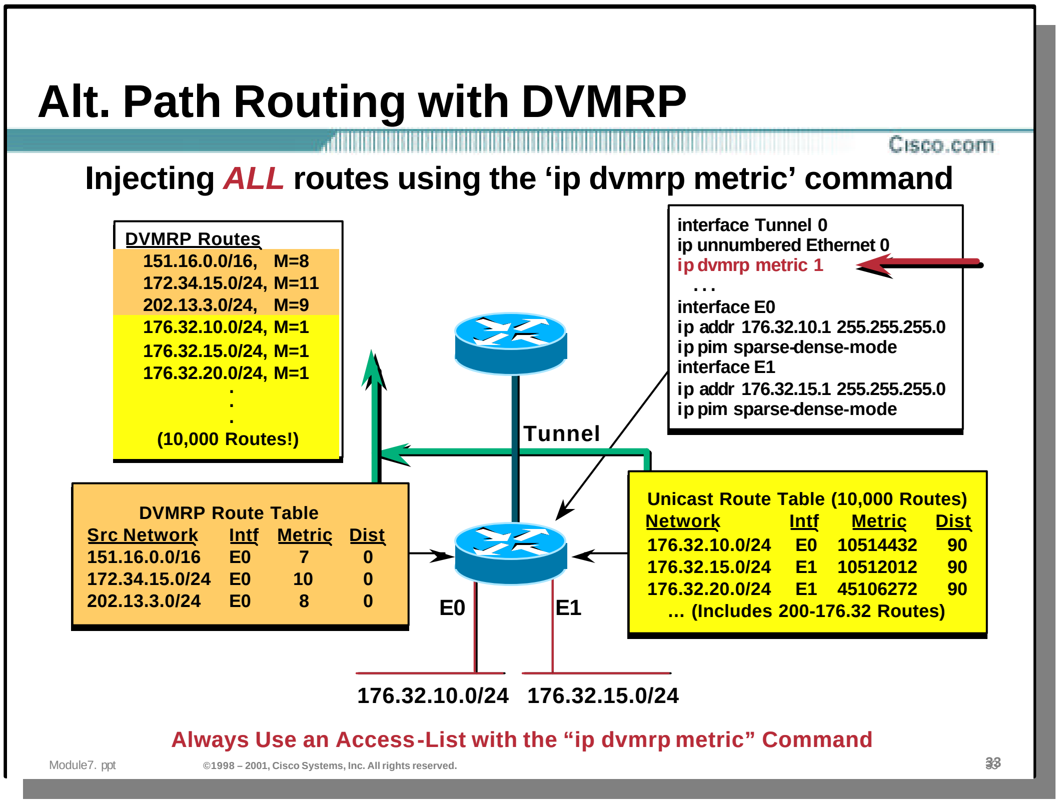
Unicast Route Table

* Split-Horizon is used between two Cisco routers.

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 32

- **Using DVMRP Routes**
  - The ‘`ip dvmrp unicast-routing`’ interface command instructs a Cisco router to begin sending and receiving DVMRP routes on the interface.
  - When two Cisco routers are connected via interfaces that have this command configured (such as in the example above), they will:
    - Receive DVMRP route reports and install them in their DVMRP route table using standard DVMRP metrics.
    - Send DVMRP route reports from the routes contained in their DVMRP route table.
    - Inject certain selected routes from the unicast routing table in DVMRP route reports. By default, only “connected” routes are injected in these DVMRP route reports. However, additional routes may be injected from the unicast routing table. (See next slide.)

  Note: When two Cisco routers are exchanging DVMRP route reports, the normal DVMRP Poison-Reverse mechanism is not used. Instead, Split-Horizon is used so that DVMRP routes are not advertised back out interface from which they were received.



Alt. Path Routing with DVMRP

Cisco.com

Injecting *ALL* routes using the ‘ip dvmrp metric’ command

DVMRP Routes
151.16.0.0/16, M=8
172.34.15.0/24, M=11
202.13.3.0/24, M=9
176.32.10.0/24, M=1
176.32.15.0/24, M=1
176.32.20.0/24, M=1
.
.
.
(10,000 Routes!)

```
interface Tunnel 0
ip unnumbered Ethernet 0
ip dvmrp metric 1
  . . .
interface E0
ip addr 176.32.10.1 255.255.255.0
ip pim sparse-dense-mode
interface E1
ip addr 176.32.15.1 255.255.255.0
ip pim sparse-dense-mode
```

DVMRP Route Table

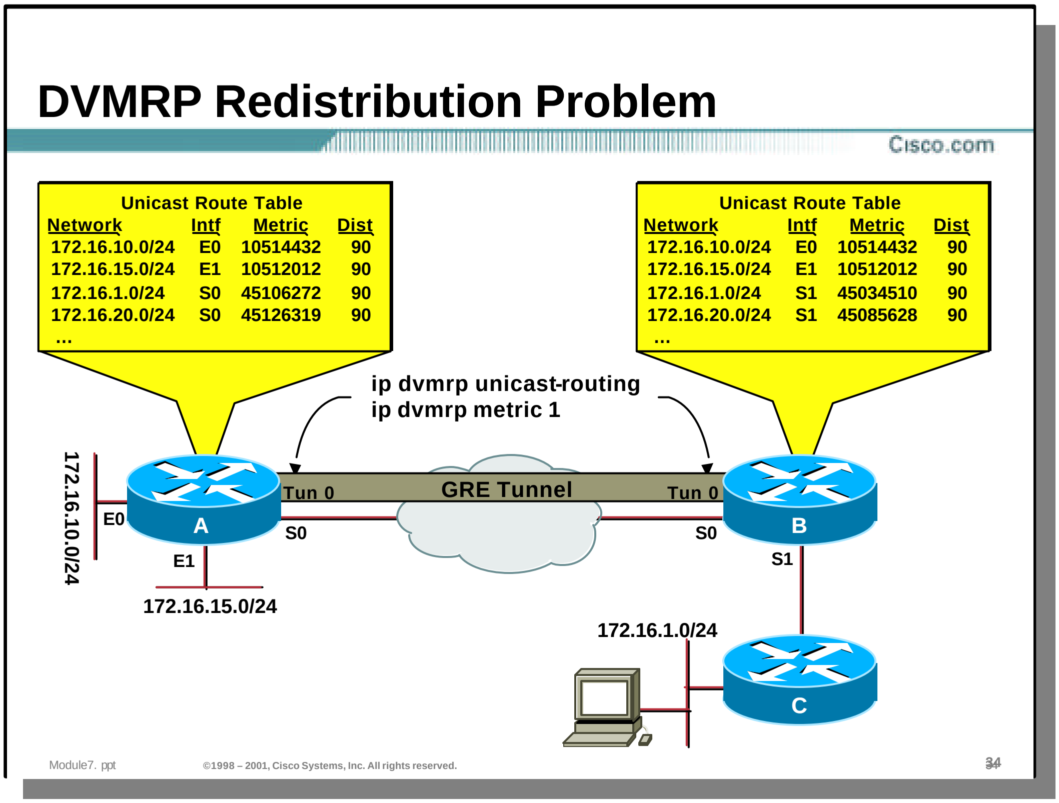
Src Network	Intf	Metric	Dist
151.16.0.0/16	E0	7	0
172.34.15.0/24	E0	10	0
202.13.3.0/24	E0	8	0

Unicast Route Table (10,000 Routes)

Network	Intf	Metric	Dist
176.32.10.0/24	E0	10514432	90
176.32.15.0/24	E1	10512012	90
176.32.20.0/24	E1	45106272	90
… (Includes 200-176.32 Routes)			

Always Use an Access-List with the “ip dvmrp metric” Command

- **Using DVMRP Routes - Injecting unicast routes**
  - The default behavior of a Cisco router is to inject only “connected” routes. However, the ‘`ip dvmrp metric`’ interface command can be used to modify this behavior. When this command is used without an access-control list, *ALL* routes in the unicast routing table will be injected into DVMRP route reports.
  - An example of this is shown In the drawing above. The ‘ip dvmrp metric 1’ command results in the entire unicast routing table being injected into DVMRP route reports with a metric of 1. In most cases, this is not desirable and an access-control list should be used to limit which routes are injected.



DVMRP Redistribution Problem

Unicast Route Table

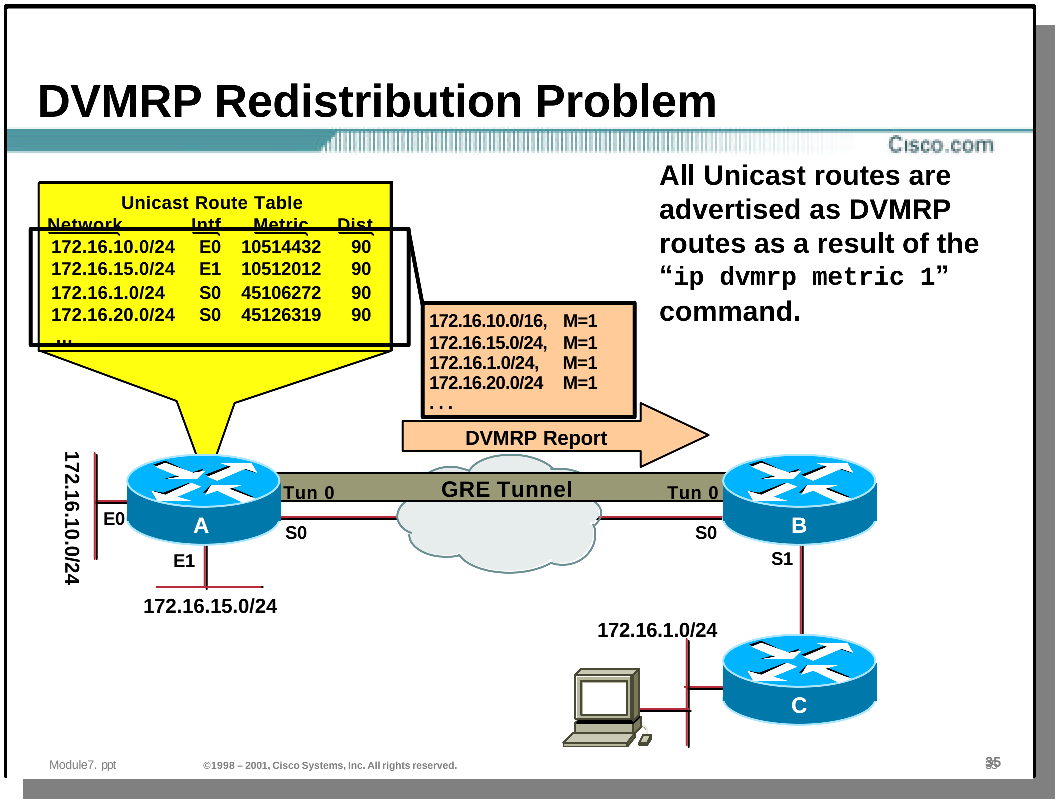
Network	Intf	Metric	Dist
172.16.10.0/24	E0	10514432	90
172.16.15.0/24	E1	10512012	90
172.16.1.0/24	S0	45106272	90
172.16.20.0/24	S0	45126319	90
…			

Unicast Route Table

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
…			

```
ip dvmrp unicast-routing
ip dvmrp metric 1
```

- **DVMRP Redistribution Problem**
  - As previously stated, care must be taken to avoid redistribution problems when DVMRP routes are being exchanged between Cisco routers. This example demonstrates what can happen if careful attention to route injection is not observed.
  - The drawing above shows two Cisco routers (along with their unicast route table) connected via a GRE Tunnel which is being used to tunnel through a unicast-only cloud. At each end of the tunnel, '`ip dvmrp unicast-routing`' has been enabled and the entire set of routes from the unicast routing table is being injected into DVMRP route reports with a metric of 1 by the use of the '`ip dvmrp metric 1`' command.



DVMRP Redistribution Problem

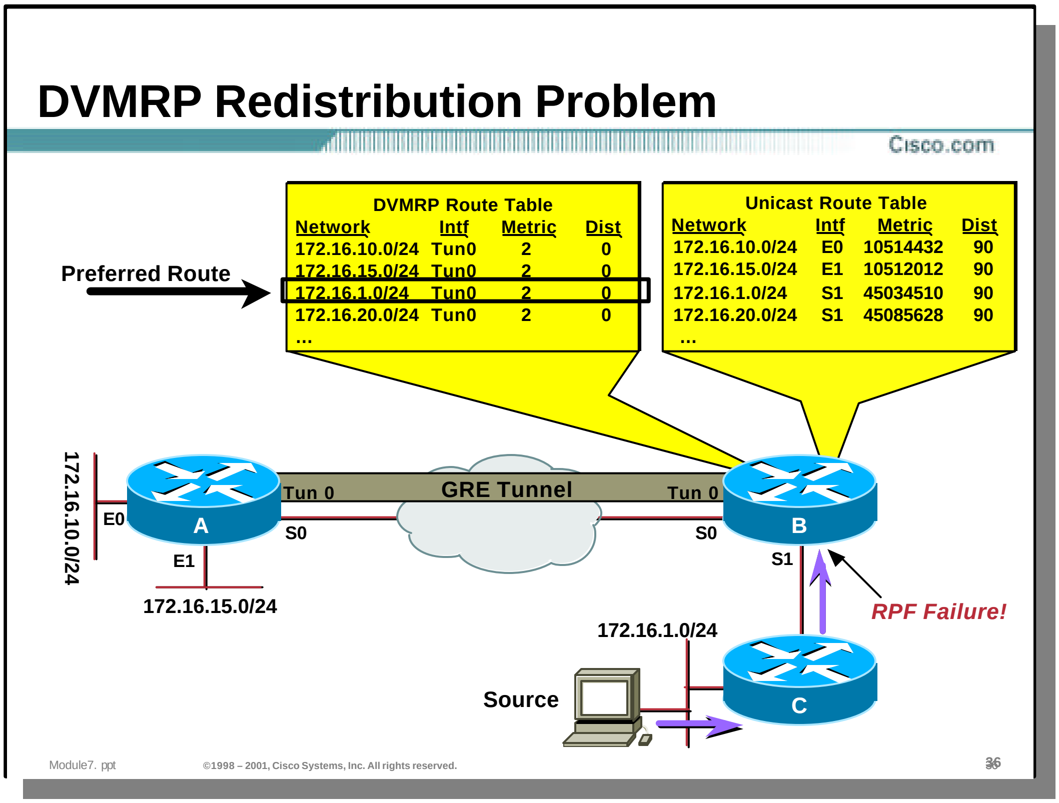
Unicast Route Table

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**
  - The drawing above shows router A injecting its entire set of unicast routes into a DVMRP report, each with a metric of 1.



DVMRP Redistribution Problem

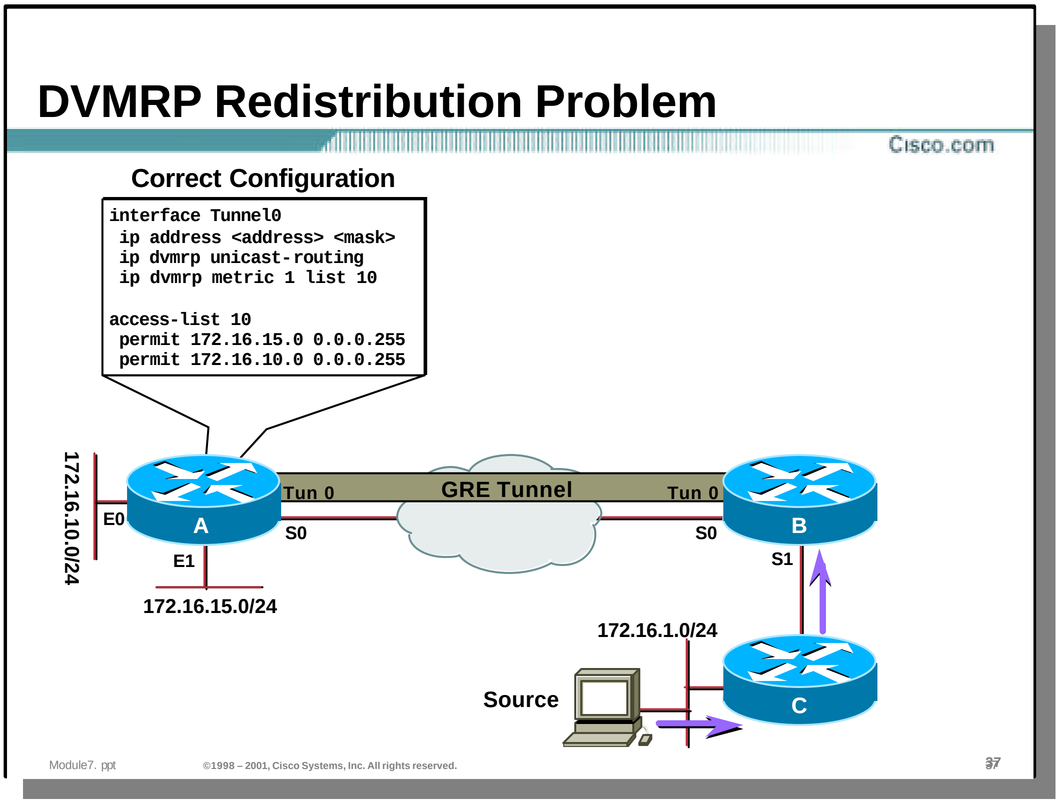
DVMRP Route Table

Network	Intf	Metric	Dist
172.16.10.0/24	Tun0	2	0
172.16.15.0/24	Tun0	2	0
172.16.1.0/24	Tun0	2	0
172.16.20.0/24	Tun0	2	0
…			

Unicast Route Table

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**
  - As a result, router B has installed these DVMRP routes in its DVMRP route table with the appropriate metrics.
  - Router B will now prefer the DVMRP route for network 172.16.1.0/24 over the same unicast route since the DVMRP route has an Admin Distance of zero.
  - Now assume that the source begins to send multicast traffic which arrives at router B via interface **Serial1**. Unfortunately, this traffic will RPF Fail because the preferred DVMRP route indicates that the correct RPF interface for this traffic is **Tunnel0**.

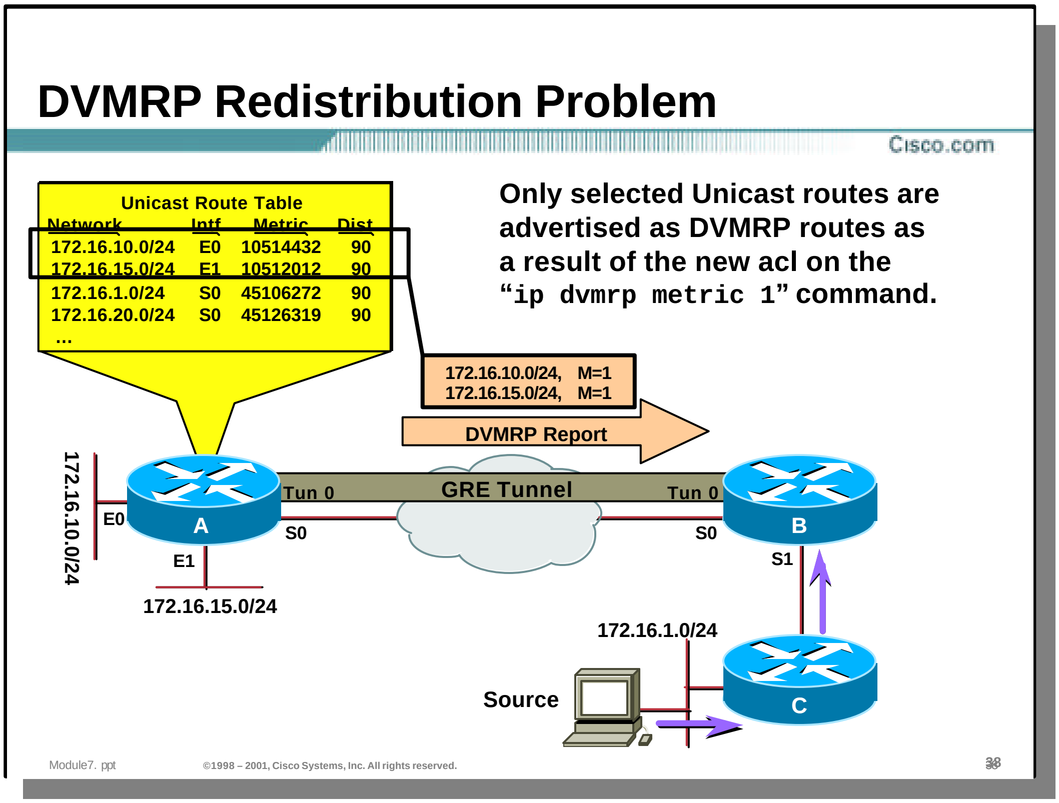


DVMRP Redistribution Problem

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
```

- **Solution to DVMRP Redistribution Problem**
  - The correct way to configure router A is to use an access-control list on the ‘ip dvmrp metric’ command that specifies only those networks behind router A. (In this case, networks 172.16.15.0/24 and 172.16.10.0/24.)



DVMRP Redistribution Problem

Cisco.com

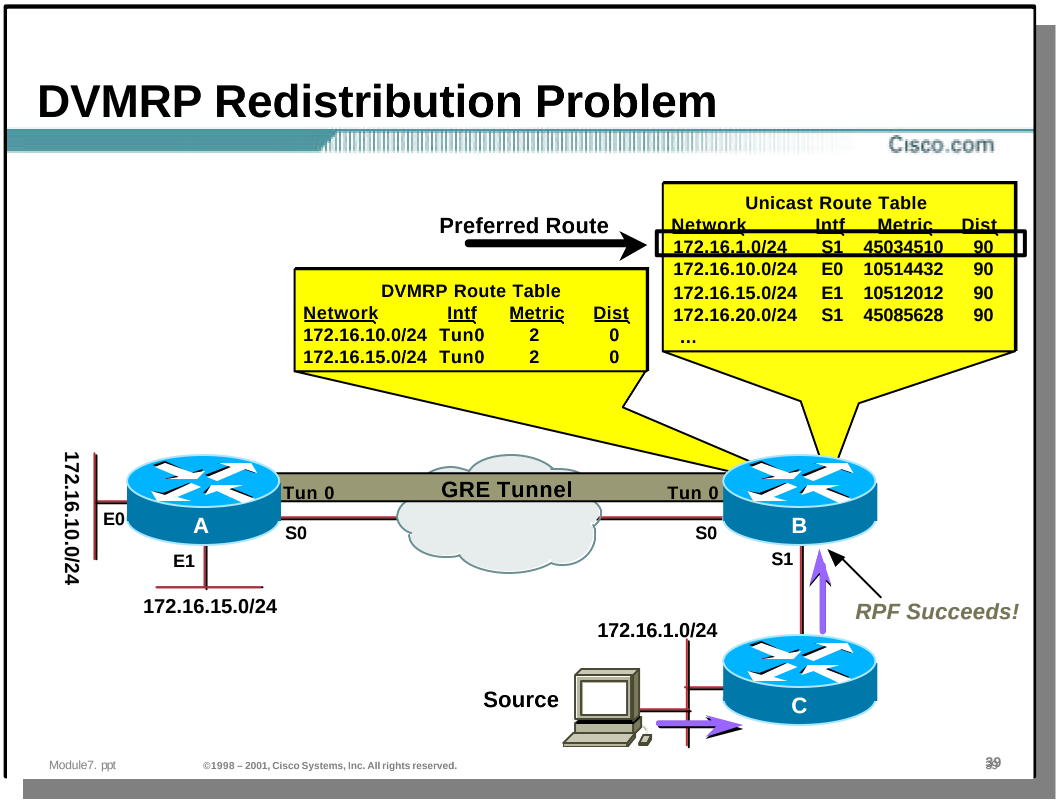
Only selected Unicast routes are advertised as DVMRP routes as a result of the new acl on the “`ip dvmrp metric 1`” command.

Unicast Route Table

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
…			

172.16.10.0/24, M=1
172.16.15.0/24, M=1

- **Solution to DVMRP Redistribution Problem**
  - Using this new configuration, we now see that router A is only injecting networks 172.16.10.0/24 and 172.16.15.0/24 in DVMRP route reports that are sent to router B.



DVMRP Redistribution Problem

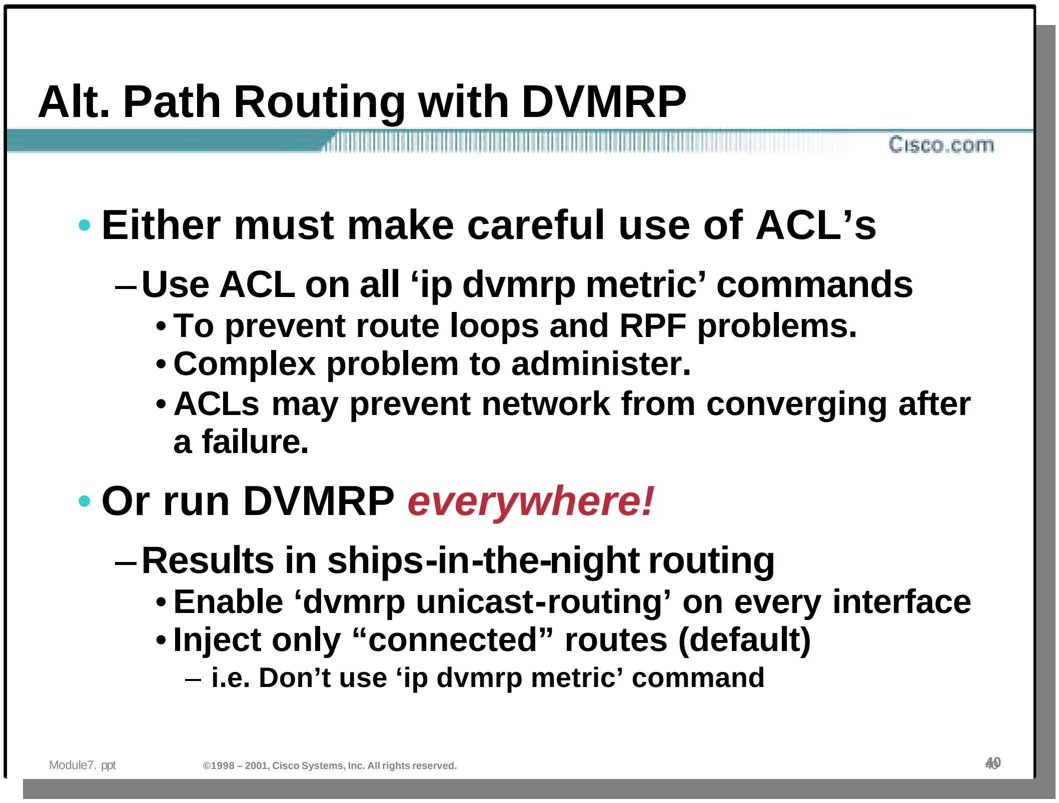
DVMRP Route Table

Network	Intf	Metric	Dist
172.16.10.0/24	Tun0	2	0
172.16.15.0/24	Tun0	2	0

Unicast Route Table

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

- **Solution to DVMRP Redistribution Problem**
  - As a result, router B has installed only these two DVMRP routes in its DVMRP route table with the appropriate metrics.
  - Now when router B receives traffic from the source on network 172.16.1.0/24, it will only find this route in its unicast routing table and will use this information to correctly calculate the RPF interface as **Serial1**. This will allow the RPF check to succeed for traffic arriving from the source.



Alt. Path Routing with DVMRP

Cisco.com

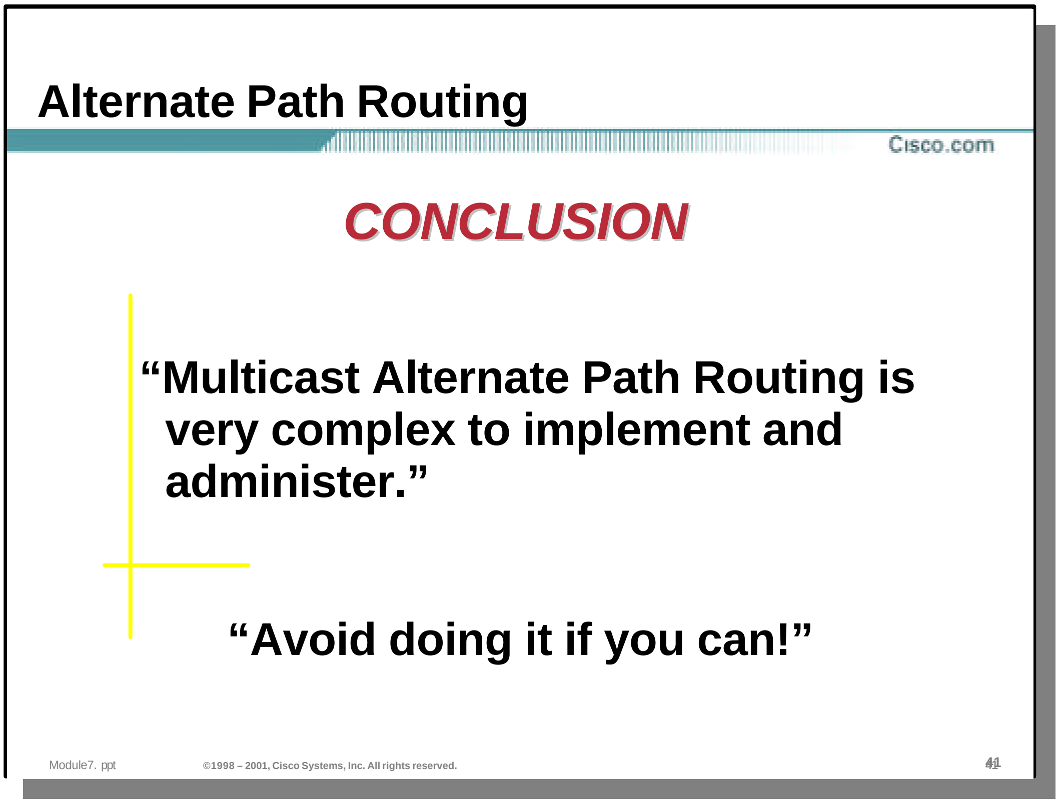
- Either must make careful use of ACL’s
  - Use ACL on all ‘ip dvmrp metric’ commands
    - To prevent route loops and RPF problems.
    - Complex problem to administer.
    - ACLs may prevent network from converging after a failure.
- Or run DVMRP *everywhere!*
  - Results in ships-in-the-night routing
    - Enable ‘dvmrp unicast-routing’ on every interface
    - Inject only “connected” routes (default)
      - i.e. Don’t use ‘ip dvmrp metric’ command

Module7. ppt 40

- **Using DVMRP Routes — Summary**

  The use of DVMRP for alternate path routing must be done using careful planning and configuration.

  - Access-control lists must be used on all ‘`ip dvmrp metric`’ commands if route loops and RPF problems are to be avoided. The administration problems can grow quite large for networks where DVMRP alternate path routing is used on a large scale basis. Furthermore, because the access-control lists used to control which unicast routes are injected into DVMRP are static, changes in topology can result in the network failing to route around failed links.
  - The other alternative is to use a “Ships-in-the-night” approach where DVMRP routes are exchanged *over every interface in the network!*
    - When this approach is used, only “connected” networks should be injected on every interface by every router in the network. (This is the default behavior if the ‘`ip dvmrp metric`’ command is *not* used when ‘`ip dvmrp unicast-routing`’ is enabled on an interface.)
    - The advantage of this method is that no ACL’s are necessary and the network will re-converge around failed links. The disadvantage is that the network is now running both DVMRP and a separate unicast routing protocol which pass each other like “Ships-in-the-night” at every point in the network. In most cases, this is generally not desirable.



Alternate Path Routing

Cisco.com

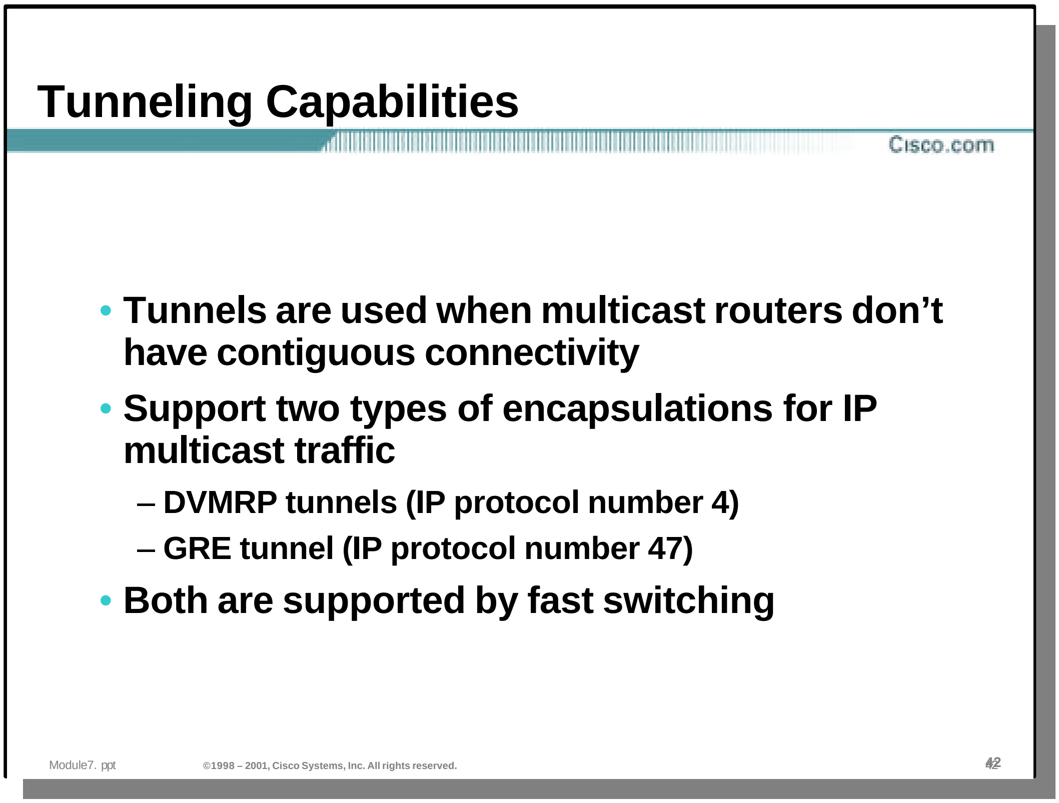
CONCLUSION

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

“Avoid doing it if you can!”

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 41

- **Alternate Path Routing — Conclusion**
  - The use of alternate path routing for multicast traffic generally results in complex and difficult to administer networks. Until new tools become available to simplify this task, Network administrators are advised to use this only as a last resort.



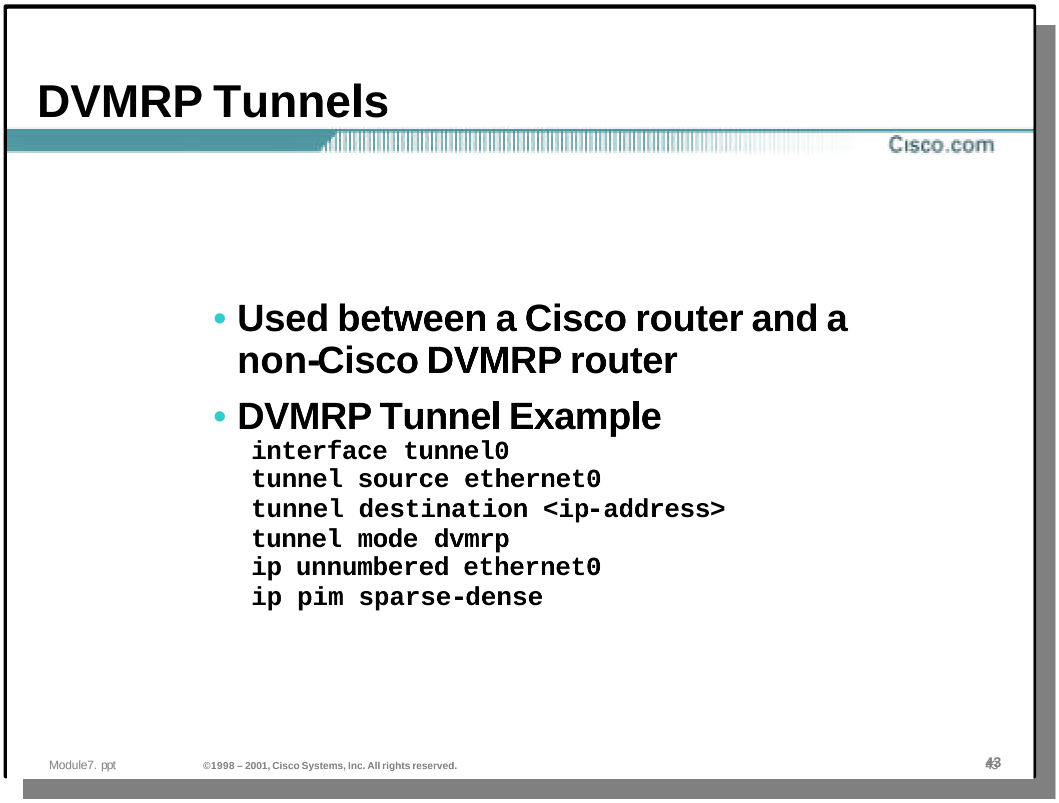
Tunneling Capabilities

Cisco.com

- 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

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 42

- **Tunneling Multicast**
  - – Cisco IOS supports two types of tunnels for IP Multicast traffic. Both of these tunnel modes are fast-switched.
    - **DVMRP Tunnels** - This is actually an IP-in-IP tunnel and uses the protocol number of 4. DVMRP tunnels are not supported between two Cisco routers. It is solely intended for use between a Cisco router and a non-Cisco router running DVMRP.
    - **GRE Tunnels** - IP Multicast traffic may be tunneled between two Cisco routers using GRE tunnels (protocol 47).



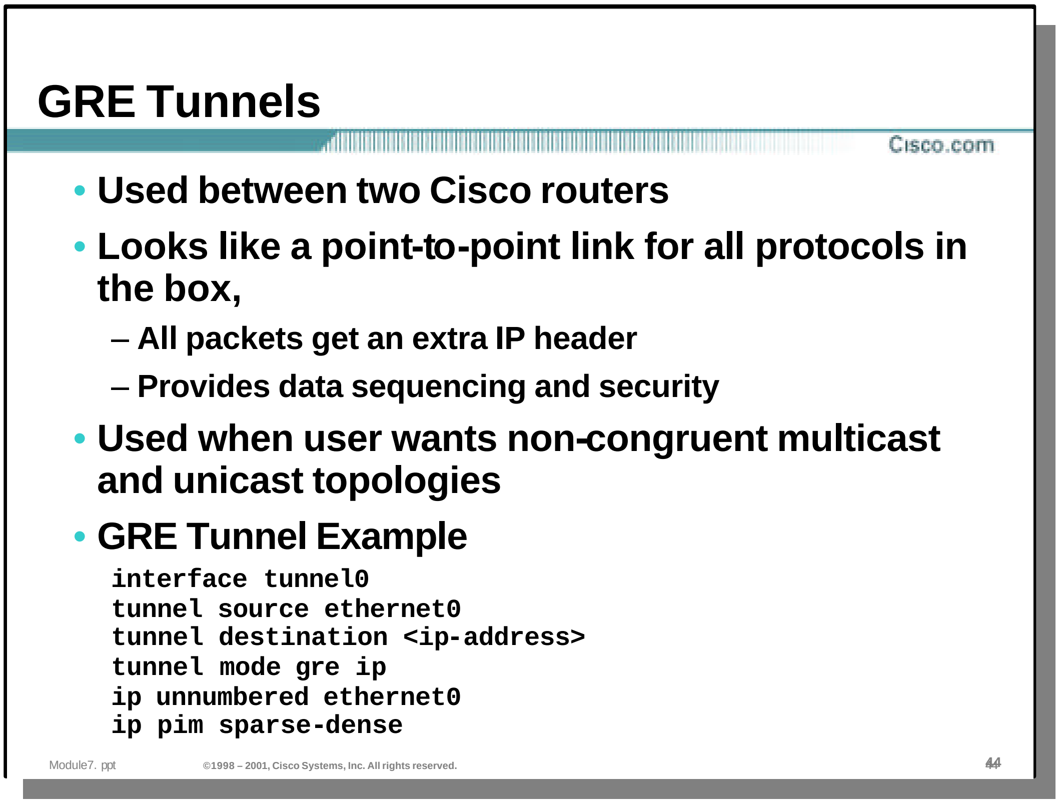
DVMRP Tunnels

Cisco.com

- Used between a Cisco router and a non-Cisco DVMRP router
- DVMRP Tunnel Example

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

- **DVMRP Tunnels**
  - DVMRP tunnels are only used between Cisco routers and non-Cisco DVMRP routers. *DVMRP tunnels cannot be used between two Cisco routers!*
  - A normal tunnel configuration (such as the one shown above) is used with the “tunnel mode” set to “**`dvmrp`**”.



GRE Tunnels

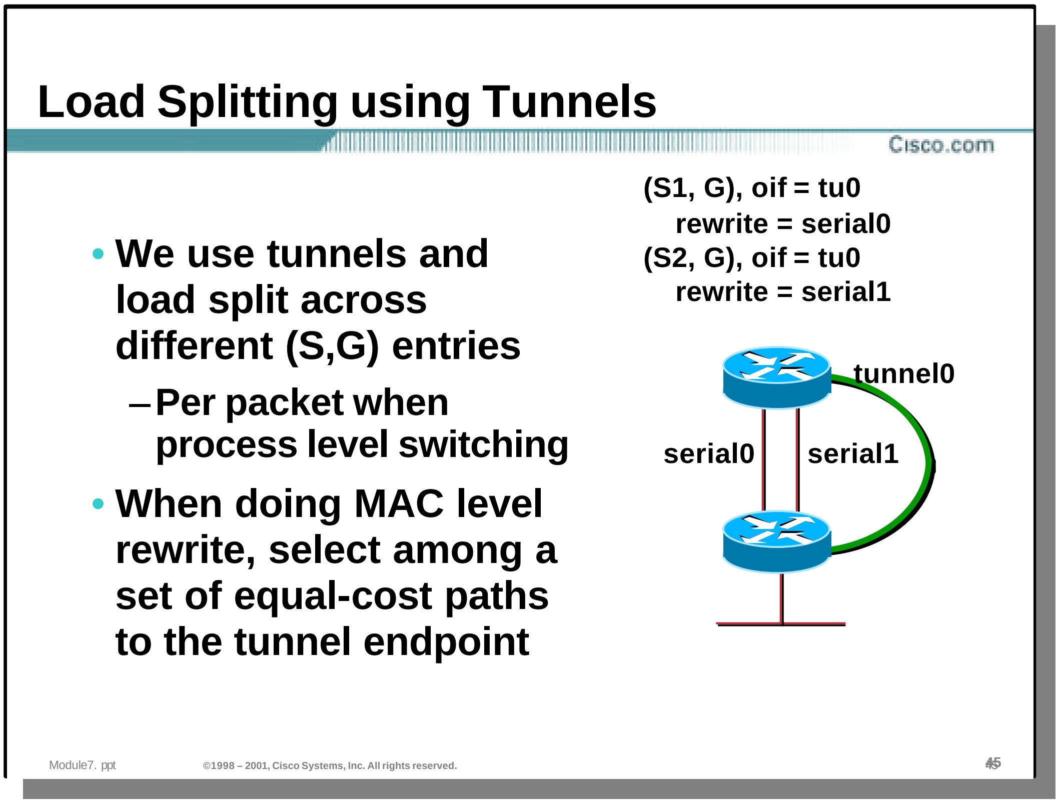
Cisco.com

- Used between two Cisco routers
- Looks like a point-to-point link for all protocols in the box,
  - All packets get an extra IP header
  - Provides data sequencing and security
- Used when user wants non-congruent multicast and unicast topologies
- GRE Tunnel Example

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

Module7. ppt 44

- **GRE Tunnels**
  - When it is necessary to tunnel multicast traffic between two Cisco routers, a GRE Tunnel must be used.
  - The GRE Tunnel appears as a point-to-point link to both ends. Each packet sent down the tunnel gets an extra IP header. GRE tunnels also provide data sequencing and some degree of security.
  - A normal tunnel configuration (such as the one shown above) is used with the "tunnel mode" set to "**`gre ip`**".



Load Splitting using Tunnels

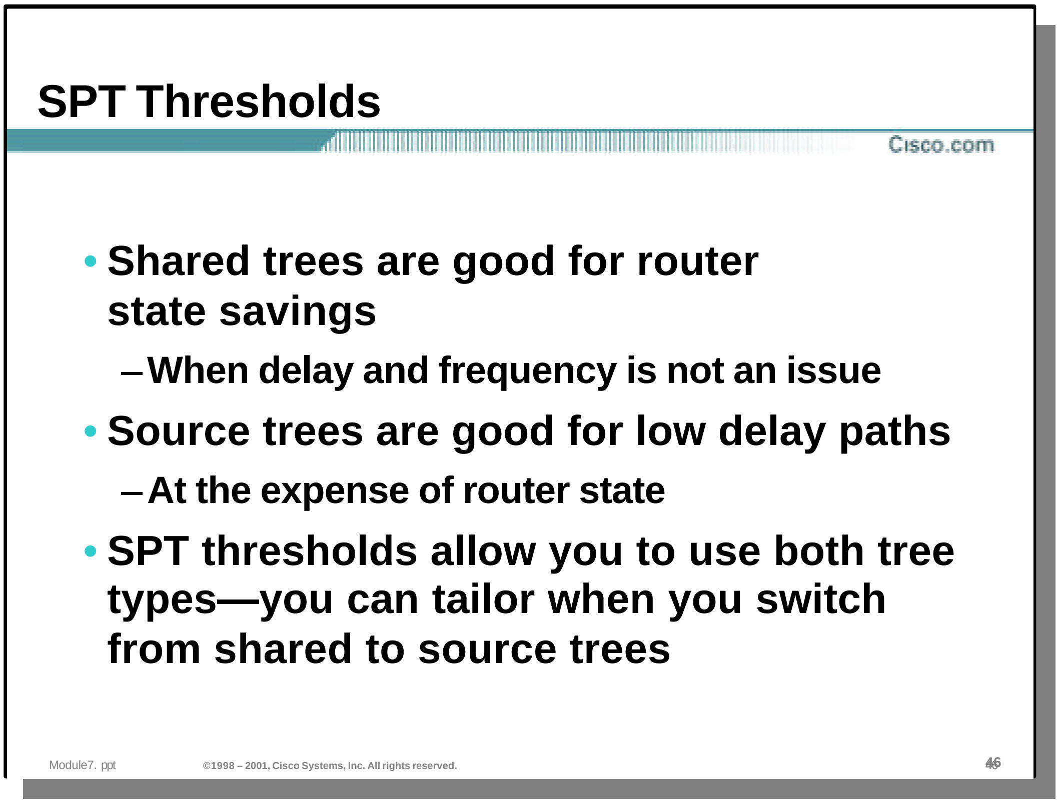
Cisco.com

- 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

tunnel0
serial0
serial1

- **Load Splitting using Tunnels**
  - Normally, a multicast traffic flow can have only one RPF interface. Even if multiple equal-cost routes exist in the unicast routing table, only the highest IP address is used. This normally precludes any load balancing of multicast traffic. However, the router can be made to see two equal-cost paths as a single interface if a Tunnel is used.
  - The example above show how a tunnel may be used to accomplish a limited degree of load balancing.
    - Tunnel0 is configured between the two routers.
    - Multicast is enabled on **Tunnel0** but not **Serial0** and **Serial1**.
    - Static Mroutes or some other form of alternate path routing is used so that the routers will RPF to **Tunnel0** for all traffic arriving from the other router. (**Warning:** Care must be taken when alternate path routing is used to avoid route loops or black holes.)
  - The load balancing method will depend on whether **Tunnel0** is process or fast-switching.
    - **Fast Switching** - Load balancing will be on a (*,G) or (S,G) flow basis.
    - **Process Switching** - Load balancing will be on a per packet basis albeit at the reduced through-puts associated with Process Switching.



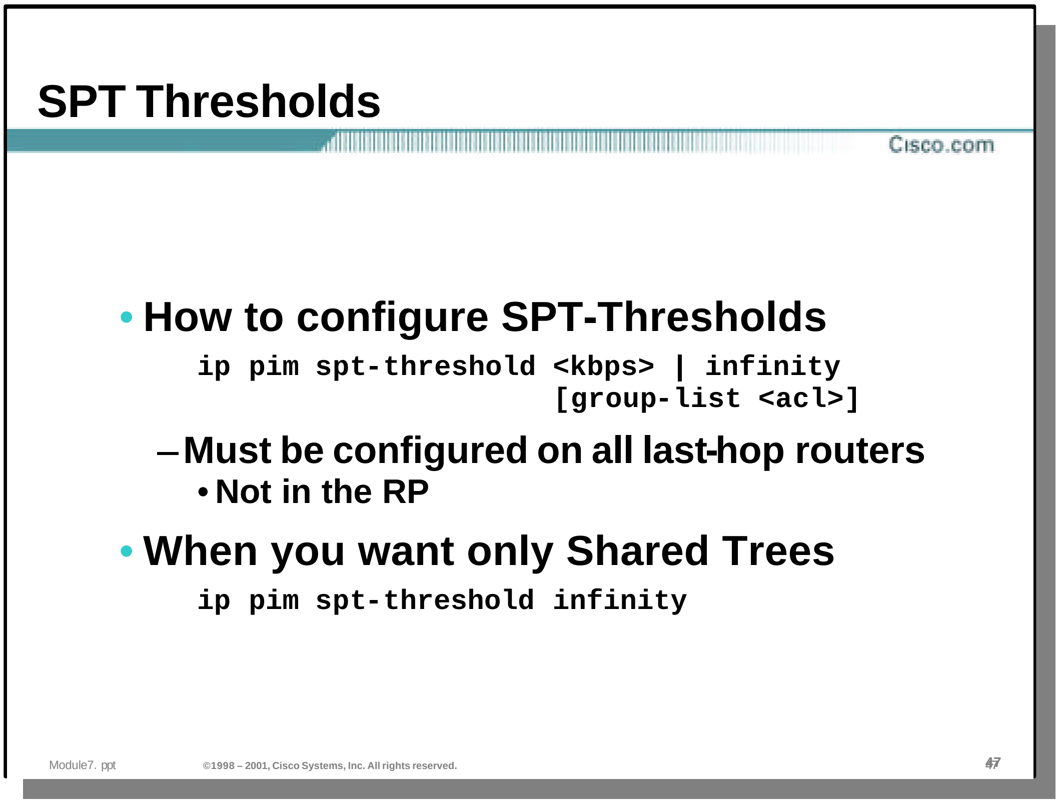
SPT Thresholds

Cisco.com

- 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

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 46

- **SPT Thresholds**
  - The key advantage of a Shared Tree is that all multicast flows in the group may use the Shared Tree thereby reducing the amount of multicast forwarding state in the routers in the network. However, the (normally) sub-optimal paths of the Shared Tree introduce additional delay and possible points of congestion along the single common tree.
  - Switching to Sources Trees (aka Shortest-Path Trees or SPT for short) is the default behavior of Cisco’s PIM implementation. The advantage of Source Trees is that multicast flows via the shortest path from the sources to the receivers which reduces latency and the potential for congestion. However, this is accomplished at the expense of more multicast forwarding state in the routers in the network.
  - SPT Thresholds permit the network engineer to tune at what point in terms of kbps a last-hop router switches to the SPT.



SPT Thresholds

Cisco.com

- How to configure SPT-Thresholds

```
ip pim spt-threshold <kbps> | infinity
                     [group-list <acl>]
```

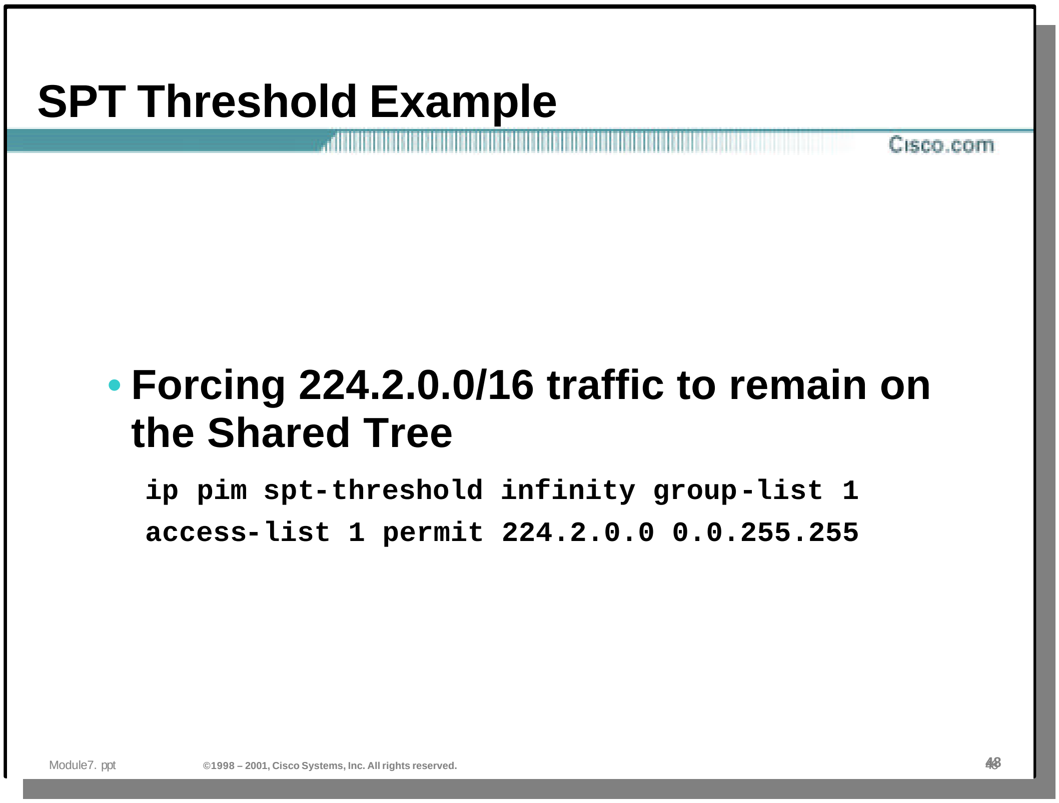
  - –Must be configured on all last-hop routers
    - Not in the RP

- When you want only Shared Trees

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

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 47

- **SPT Thresholds**
  - – SPT Thresholds may be configured on a router using the global configuration command shown above.
    - ***<kbps>* | infinity -** Defines at what rate in kbps a last-hop router switches to the SPT. If the value of **infinity** is used the router will *never* switch to the SPT and traffic will only flow down the Shared Tree.
    - **group-list <*acl*>** - Option ACL that defines the groups for which the SPT-threshold is applicable. If this ACL is not specified, all multicast groups are assumed.
  - – SPT-Thresholds must be configured on each individual router in the network. It will not have the desired affect if it is only configured on the RP. (This is because the RP does not communicate this value to the routers in the network.)



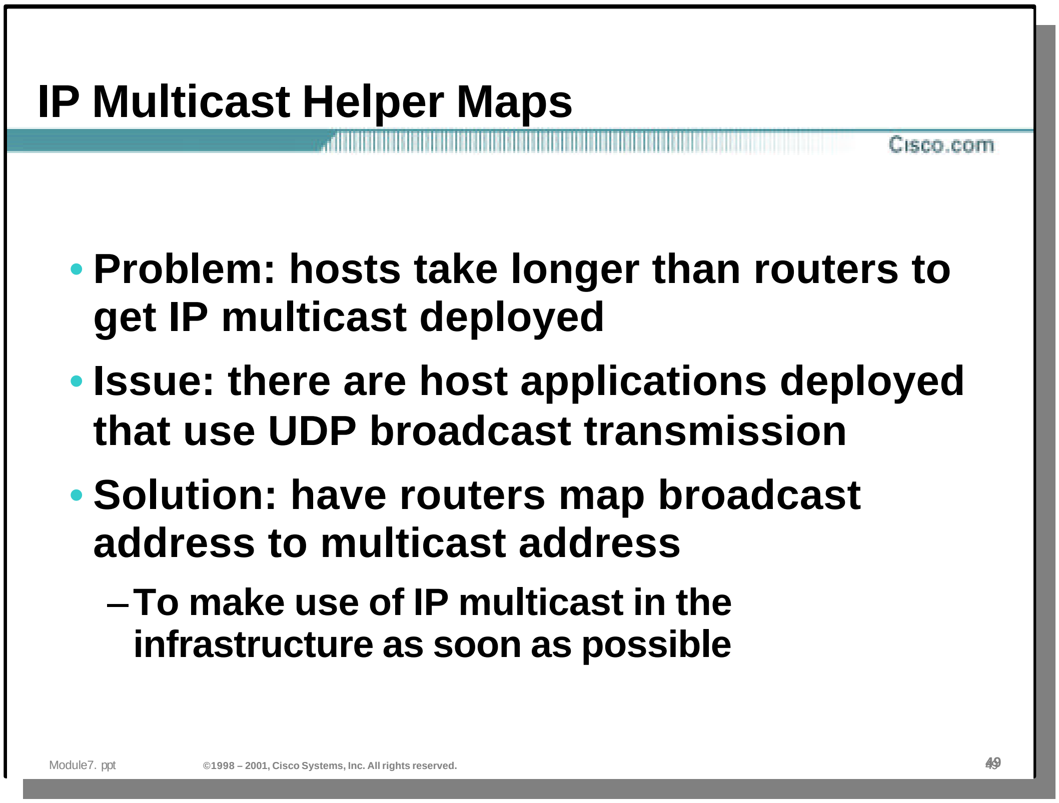
SPT Threshold Example

Cisco.com

- **Forcing 224.2.0.0/16 traffic to remain on the Shared Tree**

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

- **SPT-Threshold Example**
  - In the above example, the network engineer desires to force all multicast traffic in the 224.2.0.0/16 group range to *never* switch to the SPT. This will help reduce the amount of multicast forwarding state in the network for this group at the expense of sub-optimal routing paths.



IP Multicast Helper Maps

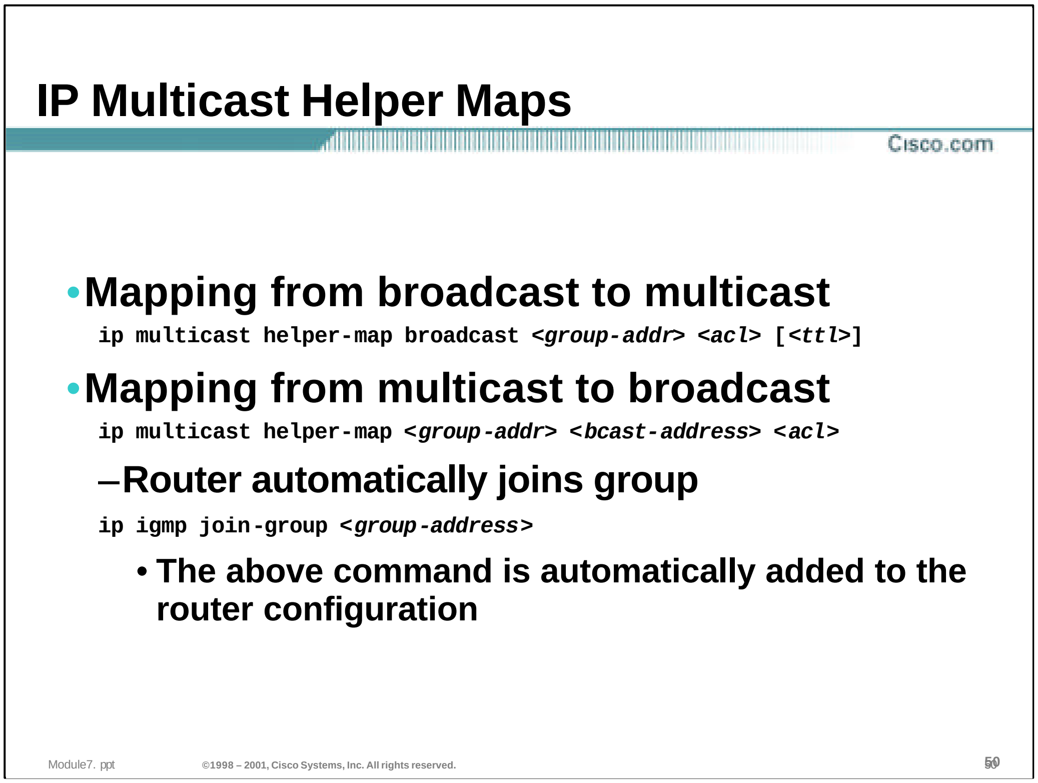
Cisco.com

- **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**

Module7. ppt

49

- **IP Multicast Helper Maps**
  - Although most of today’s modern IP stacks support IGMP and multicast, there are still cases where multicast is not supported. For example, there are several cases where Tandem or IBM hosts are used to provide Stock Market ticker information that is sent via IP in unicast or (ugh) UDP broadcast.
  - When it is desired to send this UDP broadcast traffic across a routed network, problems arise. UDP Flooding has often been used but is difficult to maintain and does not scale. The solution is to have the first-hop router convert the UDP Broadcast into multicast. This is a straight-forward process requiring the following steps:
    - Packets of the UDP broadcast are identified by UDP port number.
    - The destination broadcast IP address of the packet is rewritten with the desired multicast group address.
    - A new checksum for the IP header is recalculated.
    - The (now) multicast packet is forwarded as any other multicast packet.
  - If it is also the case that the receivers do not support multicast, the last-hop routers can also convert specific multicast flows into local subnet broadcasts which can be received by these brain damaged hosts. This is also a straight-forward process that requires the following steps:
    - The router joins the desired multicast group.
    - The packets desired UDP flow in this group are identified by UDP port number.
    - The destination multicast group address is rewritten with the specified local subnet broadcast address.
    - A new checksum for the IP header is recalculated.
    - The (now) UDP broadcast packet is forwarded onto the local subnet.



IP Multicast Helper Maps

Cisco.com

- **Mapping from broadcast to multicast**

```
ip multicast helper-map broadcast <group-addr> <acl> [<ttl>]
```

- **Mapping from multicast to broadcast**

```
ip multicast helper-map <group-addr> <bcast-address> <acl>
```

  – **Router automatically joins group**

```
ip igmp join-group <group-address>
```

    - **The above command is automatically added to the router configuration**

Module7. ppt

50

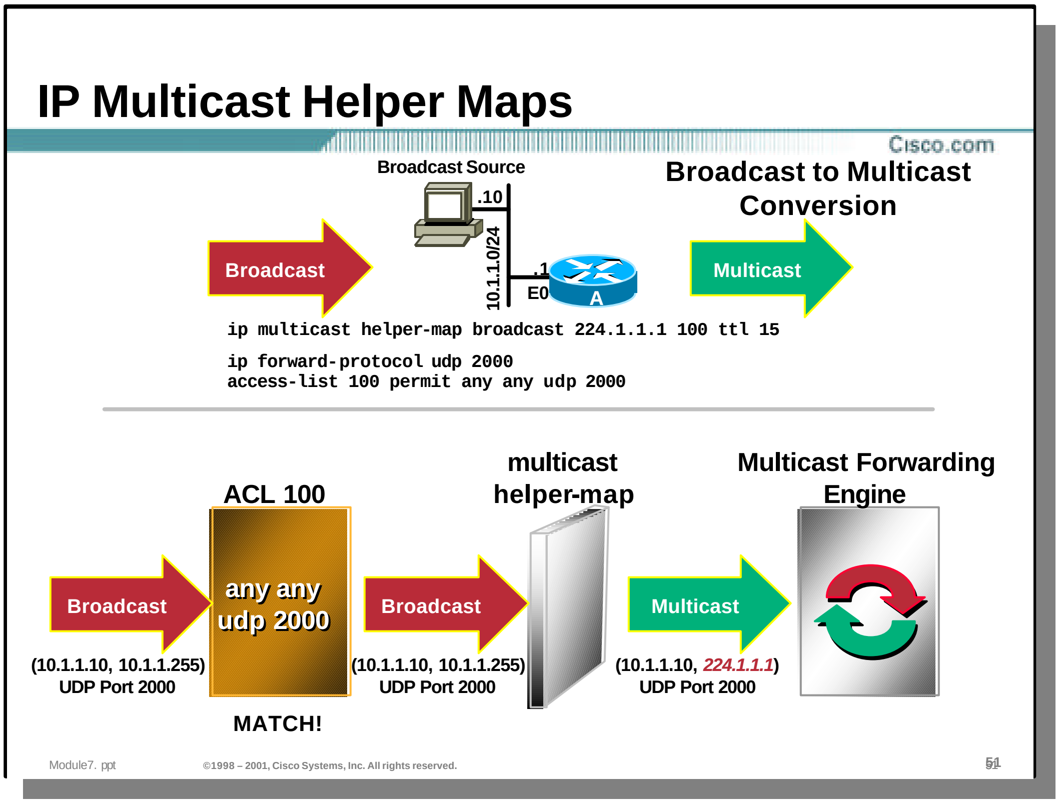
- **IP Multicast Helper Maps**
  - Mapping from UDP Broadcast to Multicast is accomplished using the following command syntax:

    ```
    ip multicast helper-map broadcast <group-addr> <acl>
    ```

    - ***<group-addr>*** - is the multicast group address that the router will use to replace the destination broadcast address in the received broadcast packet.
    - ***<acl>*** - is an extended access-control list that is used to identify which UDP broadcast flow is to be converted to multicast.
    - ***<ttl>*** **-** is an optional parameter that can be used to specify the TTL value of the multicast packet. This may be important if TTL Scoping is in use.
  - Mapping from multicast back to broadcast is accomplished using the following command syntax:

    ```
    ip multicast helper-map <group-addr> <bcast-addr> <acl>
    ```

    - ***<group-addr>*** - is the group address of the multicast flow that the router will convert back to a local broadcast.
    - ***<bcast-addr>*** - is the destination broadcast address that the router will use to replace the above multicast group address when it rewrites the packet. (This address also identifies which interface/sub-net that the rewritten packet is to be sent.)
    - ***<acl>*** - is an extended access-control list that is used to identify the UDP port number of the multicast flow that is to be converted to broadcast.
  - When mapping from multicast back to broadcast, the router joins the group specified by the ***<group-addr>*** parameter by automatically adding an ‘`ip igmp join-group`’ command to the configuration.



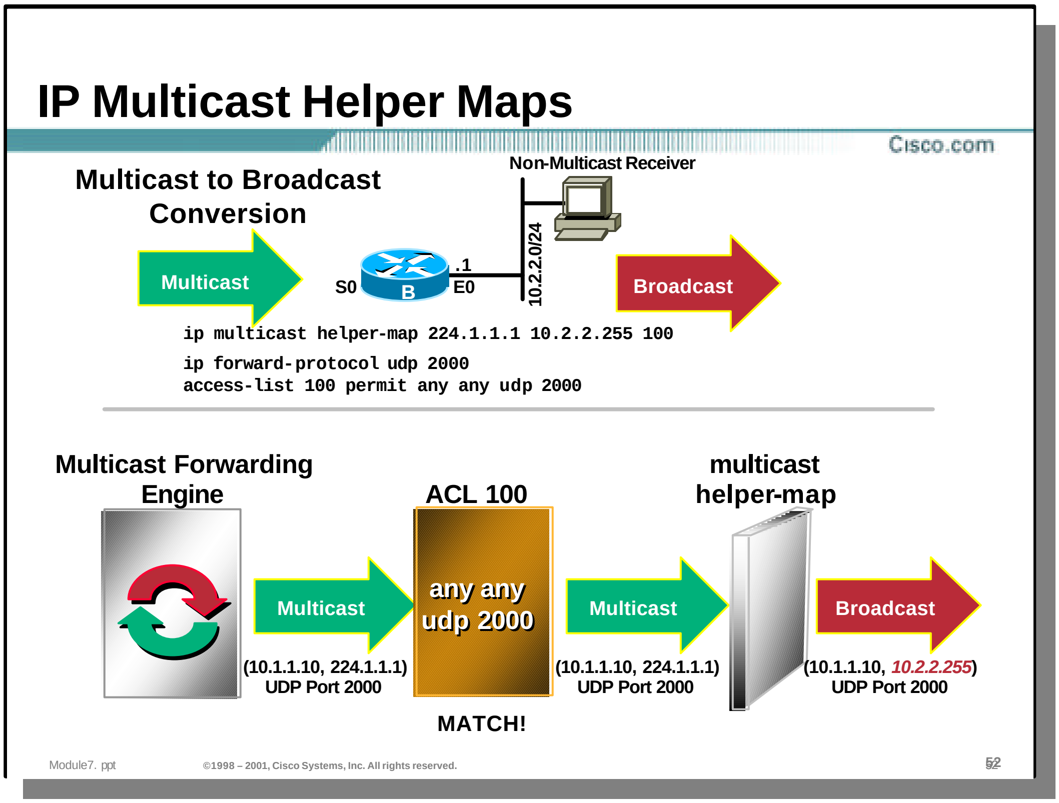
IP Multicast Helper Maps

Cisco.com
Broadcast Source
.10
10.1.1.0/24
.1
E0
A
Broadcast
Broadcast to Multicast Conversion
Multicast

```
ip multicast helper-map broadcast 224.1.1.1 100 ttl 15

ip forward-protocol udp 2000
access-list 100 permit any any udp 2000
```

ACL 100
multicast helper-map
Multicast Forwarding Engine
Broadcast
any any udp 2000
Broadcast
Multicast
(10.1.1.10, 10.1.1.255) UDP Port 2000
(10.1.1.10, 10.1.1.255) UDP Port 2000
(10.1.1.10, 224.1.1.1) UDP Port 2000
MATCH!

- **Broadcast to Multicast Conversion Details**
  - In the top half of the above drawing, router A is configured to convert arriving UDP broadcast packets with a port number of 2000 into multicast packets with a destination group address of 224.1.1.1 and a TTL of 15. Extended access-list 100 is used to identify the UDP broadcast flow to UDP port 2000. (Note that it is necessary to configure the '`ip forward-protocol udp 2000`' command to prevent the router from immediately discarding broadcast packets to UDP 2000.)
  - The bottom half of the drawing show the steps that are taken by the router to convert the packet to multicast.
    - A UDP broadcast packet with a UDP port of 2000 is received from host 10.1.1.10, sent to the subnet broadcast address 10.1.1.255.
    - The packet is checked against extended access-list 100 to see if it matches the specified flow. (In this case, it does.)
    - The matching packet is then sent to the "multicast-helper" front-end which simply replaces the destination broadcast address with the specified multicast group address (224.1.1.1) and recalculates a new IP header checksum.
    - The (now) multicast packet is handed off to the router's Multicast Forwarding Engine which processes the packet like any other arriving multicast packet. (Note: As far as the Multicast Forwarding Engine is concerned, the packet was sent by the original host, in this case 10.1.1.10.)

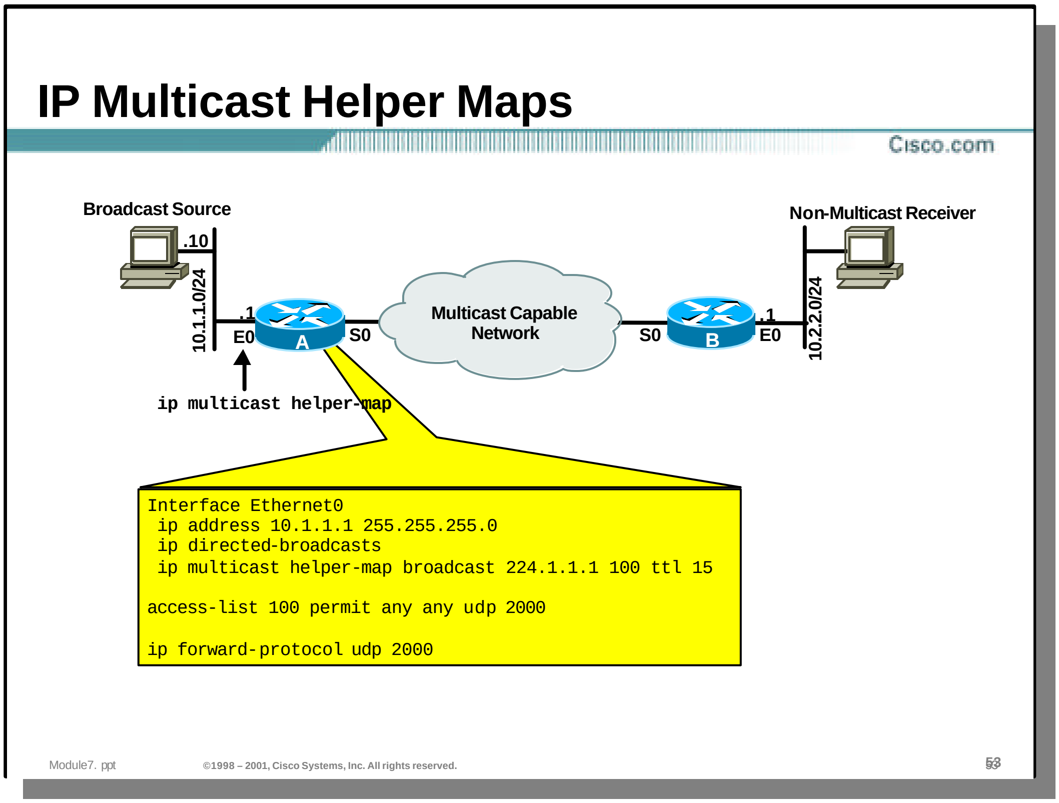


IP Multicast Helper Maps

Multicast to Broadcast Conversion

```
ip multicast helper-map 224.1.1.1 10.2.2.255 100
ip forward-protocol udp 2000
access-list 100 permit any any udp 2000
```

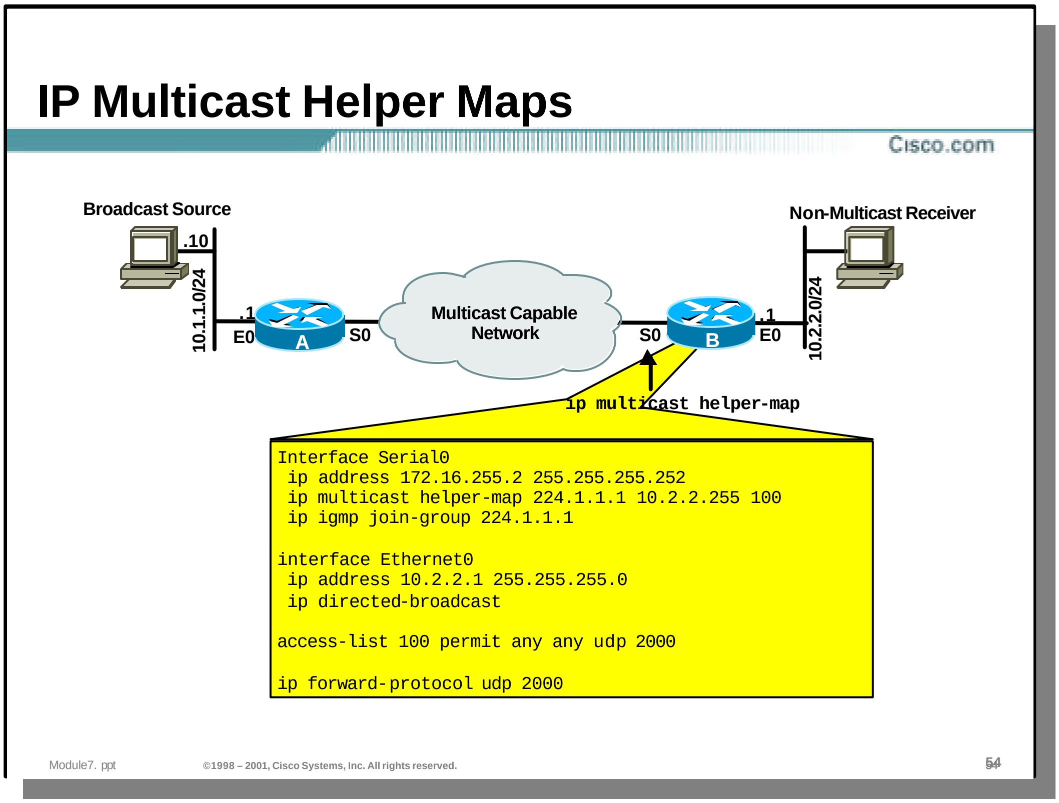
- **Multicast to Broadcast Conversion Details**
  - In the top half of the above drawing, router A is configured to convert arriving group 224.1.1.1 multicast packets with a port number of 2000 into UDP broadcast packets with a destination subnet broadcast address of 10.2.2.255. Extended access-list 100 is used to identify UDP packets addressed to UDP port 2000 within the group 224.1.1.1 multicast flow.
  - The bottom half of the drawing show the steps that are taken by the router to convert the packet to multicast.
    - The Multicast Forwarding Engine identifies arriving multicast packets for group 224.1.1.1 and checks them against extended access-list 100 to see if it matches the specified flow. (In this case, it does.)
    - The matching packet is then sent to the “multicast-helper” back-end which simply replaces the destination multicast group address (224.1.1.1) with the specified broadcast address (10.2.2.255) and recalculates a new IP header checksum.
    - The (now) UDP broadcast packet is forwarded to the appropriate subnet. (Note that it is necessary to configure the ‘`ip forward-protocol udp 2000`’ command so the the router will forward the broadcast packets to the destination subnet.)



IP Multicast Helper Maps

```
Interface Ethernet0
 ip address 10.1.1.1 255.255.255.0
 ip directed-broadcasts
 ip multicast helper-map broadcast 224.1.1.1 100 ttl 15

access-list 100 permit any any udp 2000

ip forward-protocol udp 2000
```

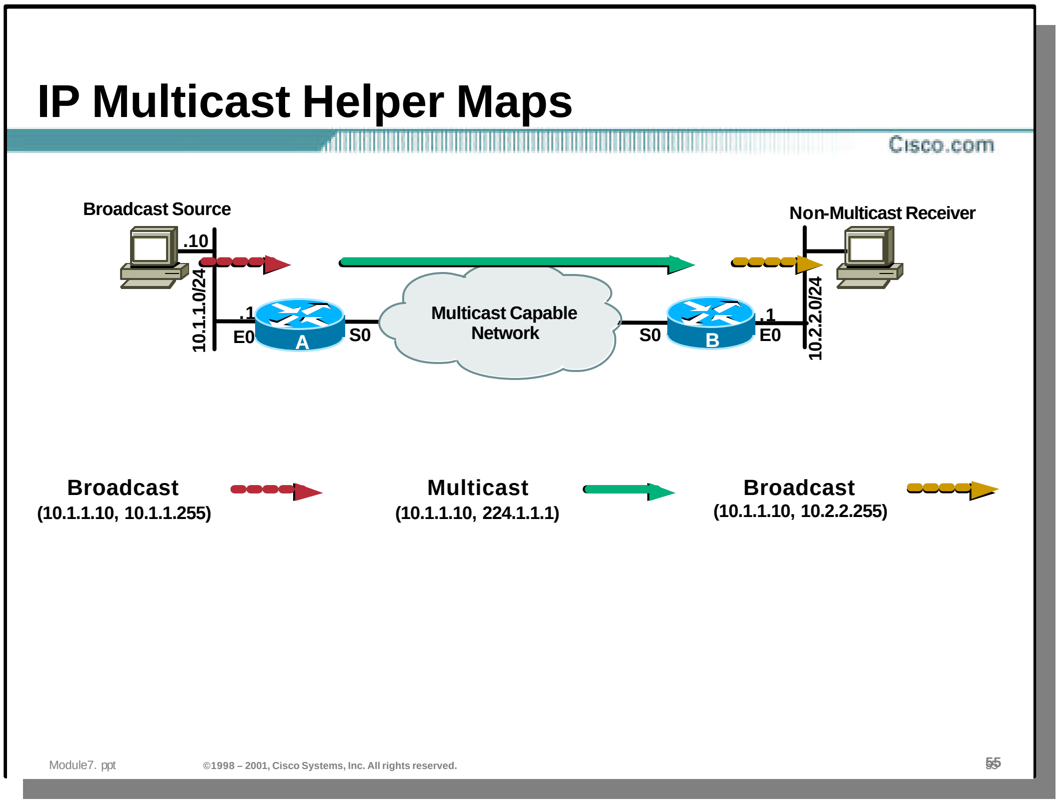
- **IP Multicast Helper Maps - Example**
  - This is an example configuration where the UDP broadcast traffic from a broadcast source is converted to IP Multicast, travels across the multicast enable network and is converted back to a directed subnet broadcast at the far end.
  - The Router A configuration necessary to convert from broadcast to multicast is shown above. Notice the following:
    - ‘`ip directed-broadcasts`’ must be enabled.
    - ‘`ip forward-protocol udp 2000`’ must be configured to prevent the router from ignoring the UDP broadcast packets.
    - Extended access-list 100 is used to identify the UDP broadcast stream that is to be converted.
    - The ‘`ip multicast helper-map`’ command causes the UDP broadcast stream to be converted to a 224.1.1.1 multicast stream with a TTL of 15.



IP Multicast Helper Maps

```
Interface Serial0
 ip address 172.16.255.2 255.255.255.252
 ip multicast helper-map 224.1.1.1 10.2.2.255 100
 ip igmp join-group 224.1.1.1

interface Ethernet0
 ip address 10.2.2.1 255.255.255.0
 ip directed-broadcast

access-list 100 permit any any udp 2000

ip forward-protocol udp 2000
```

- **IP Multicast Helper Maps - Example**
  - The Router B configuration necessary to convert from multicast back to broadcast is shown above. Notice the following:
    - Extended access-list 100 is used to identify the UDP multicast stream within group 224.1.1.1 that is to be converted.
    - The ‘`ip multicast helper-map`’ command causes the UDP broadcast stream to be converted to a 224.1.1.1 multicast stream with a TTL of 15.
    - The router has automatically configured the ‘`ip igmp join-group 224.1.1.1`’ command.
    - ‘`ip directed-broadcasts`’ must be enabled on Ethernet0.
    - ‘`ip forward-protocol udp 2000`’ must be configured to get the router to forward the UDP broadcast packets.



IP Multicast Helper Maps

Cisco.com

Broadcast Source

Non-Multicast Receiver

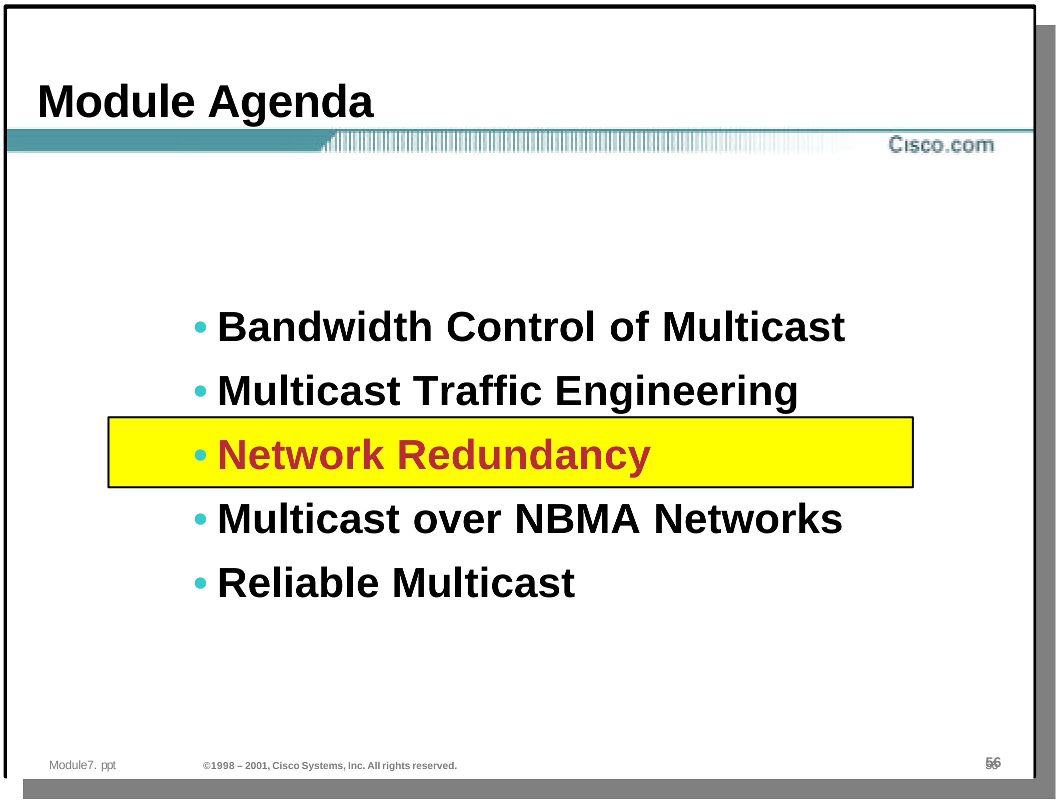
Multicast Capable Network

Broadcast (10.1.1.10, 10.1.1.255)

Multicast (10.1.1.10, 224.1.1.1)

Broadcast (10.1.1.10, 10.2.2.255)

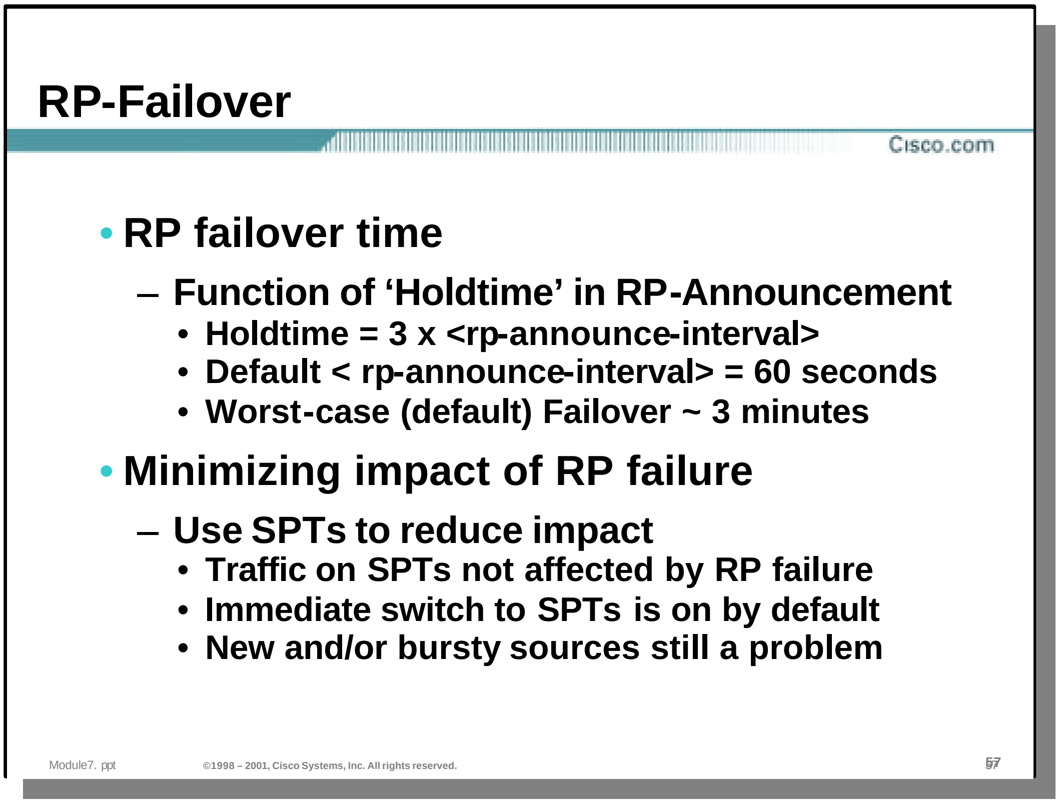
- **IP Multicast Helper Maps - Example**
  - The results of this configuration are show above.
    - UDP port 2000 broadcast packets arriving at Ethernet0 on router A have their IP destination address rewritten to multicast group 224.1.1.1.
    - These 224.1.1.1 multicast packets flow across the multicast enabled network to router B.
    - UDP port 2000 Multicast flows to group 224.1.1.1arriving on Serial0 at router B have their IP destination address rewritten to the subnet broadcast address of 10.2.2.255.
    - Router B forwards these 10.2.2.255 subnet broadcast packets to Ethernet0 which is subnet 10.2.2.0.



Module Agenda

- Bandwidth Control of Multicast
- Multicast Traffic Engineering
- **Network Redundancy**
- Multicast over NBMA Networks
- Reliable Multicast

Module7. ppt

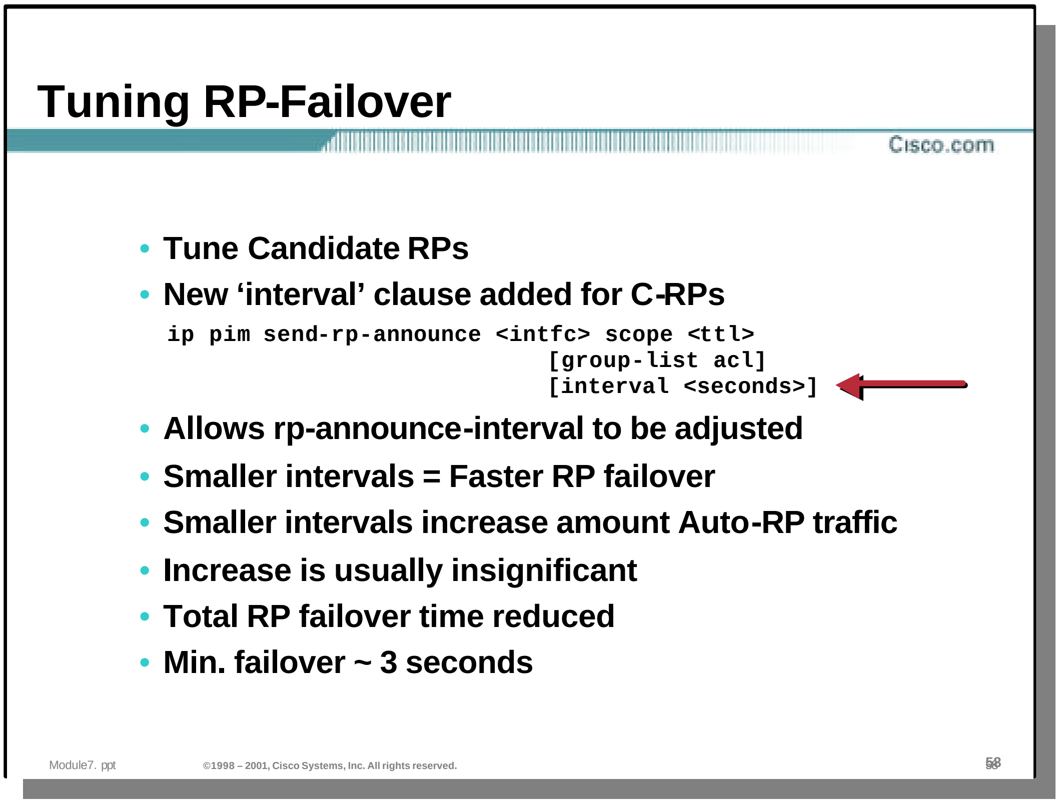


RP-Failover

- 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

- **RP Failover**
  - The time it takes for the network to detect the failure of the active RP and switch to a backup RP depends on the value of the “Holdtime” field in the RP-Announcement.
    - The “Holdtime” field indicates when the RP will timeout and assumed to be down if another RP-Announcement is not received by the Mapping Agent.
    - “Holdtime” is computed as 3 times the “RP-Announce-Interval” which has a default value of 60 seconds. This results in Holdtime values of 3 minutes which is the worst case failover time.
  - The use of SPT’s will reduce the affect of an RP failure since the Shared Tree is not being used to deliver multicast traffic. Therefore, if the RP fails, traffic from *currently active sources* will continue to flow to *currently active receivers.* However, new sources and or new receivers will not be able to register or join the Shared Tree until the RP failure has been detected and a switch to a backup RP occurs.

    Note: There is no failover mechanism built in to PIM for Static-RP’s. In order to use backup RP’s, either Auto-RP or BSR must be used. (A special configuration called “Anycast RP” can be used if multiple static RP’s are desired. This configuration, however, requires the use of MSDP and is not a native function of the PIM Protocol.)



Tuning RP-Failover

Cisco.com

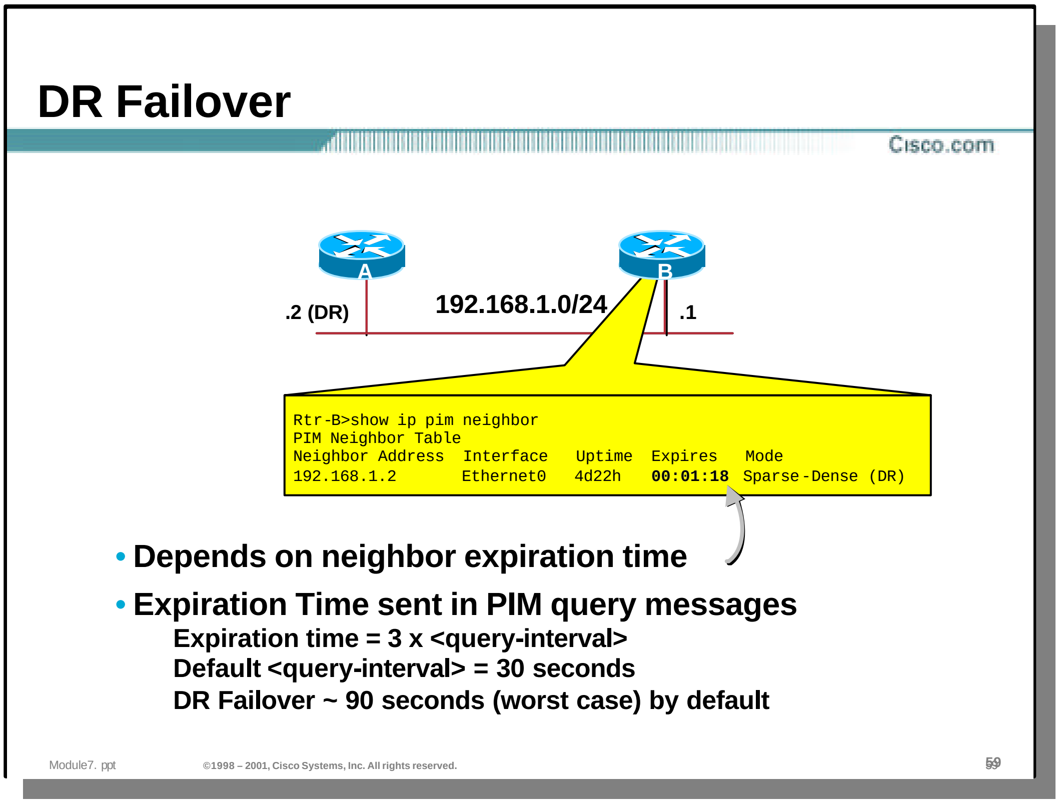
- Tune Candidate RPs
- New ‘interval’ clause added for C-RPs

```
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

Module7. ppt 58

- **Tuning RP Failover**
  - Prior to IOS release 12.0, the “rp-announce-interval” was fixed at 60 seconds. Beginning with IOS 12.0, the ‘`interval` *<seconds>*’ clause was added to the ‘`ip send-rp-announce`’ command. This new clause allows this interval to be tuned and hence tunes the “Holdtime” advertised in the RP Announcements.
  - By reducing the “rp-announce-interval”, RP failure is detected sooner and therefore failover to the backup RP occurs sooner. However, the reduced intervals between announcements results in an increase in RP Announcement traffic in the network. This is generally insignificant and worth the reduced RP failover times.
  - The minimum “rp-announce-interval” that may be set is 1 second. This corresponds to a worst case failover of 3 seconds.

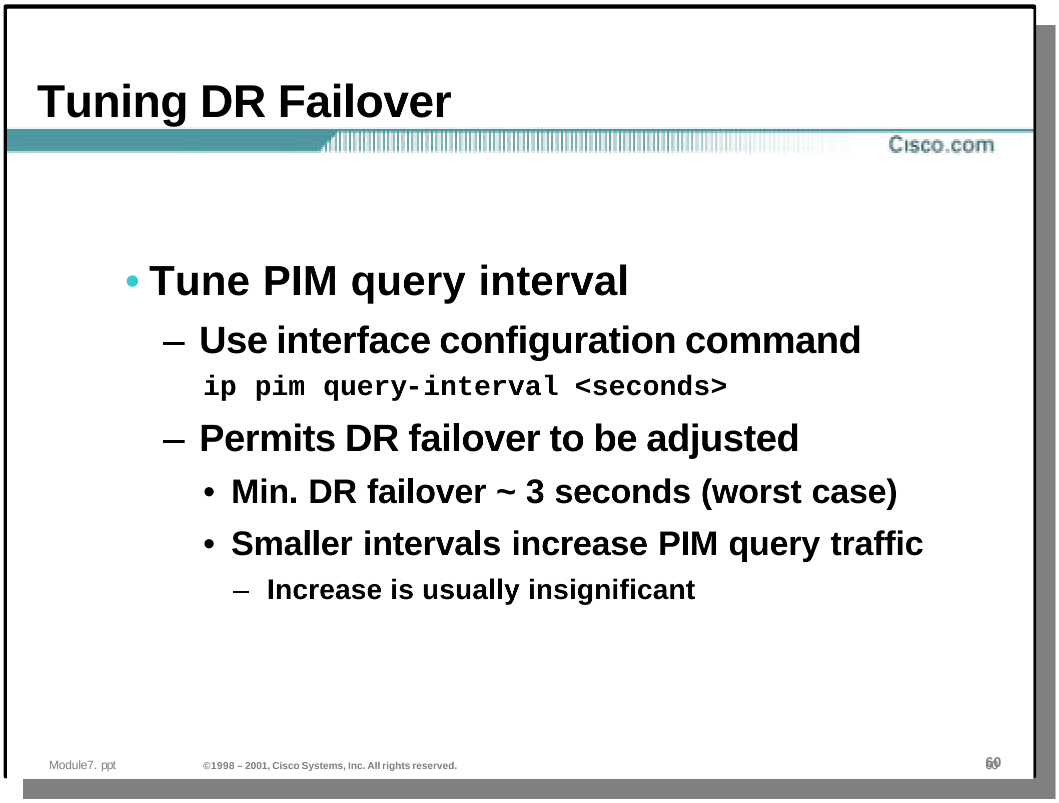


DR Failover

```
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**

- **Designated Router Failover**
  - When more than one multicast routers are connected to a LAN, one is elected as the DR and is responsible for sending Registers and Joins on behalf of sources and receivers that are active on the LAN segment. If this router fails, the other router(s) on the LAN segment will detect this and a new DR will be elected.
  - The length of time that it takes for the other routers on the LAN segment to detect that the DR has failed is dependant on the “Expire” time value advertised by the DR in its PIM Hello messages. This value is fixed at 3 times the PIM Query interval which governs how often the router sends PIM Hello messages on the local LAN interface. By default, the query interval is set to 30 seconds which results in an “expiration” time of 90 seconds. Therefore, the worst case scenario is that it will take the other routers on the LAN segment 90 seconds to detect the failure of the DR and elect a new one.



Tuning DR Failover

Cisco.com

- Tune PIM query interval
  - Use interface configuration command

    ```
    ip pim query-interval <seconds>
    ```

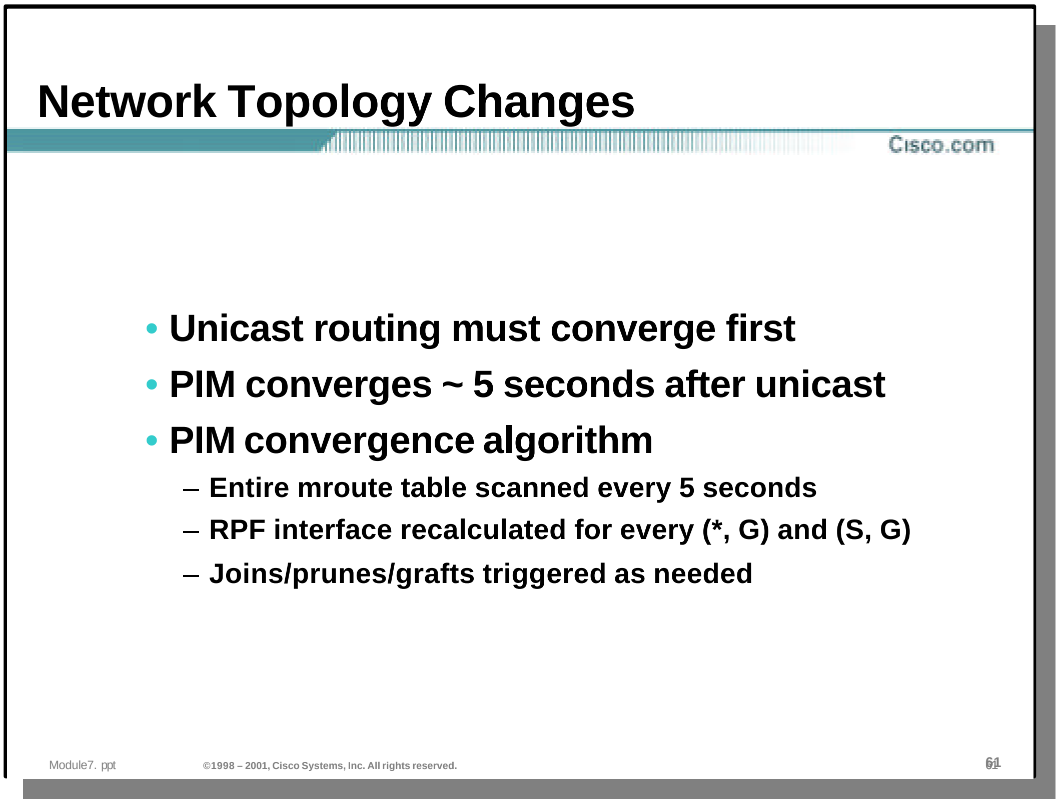
  - Permits DR failover to be adjusted
    - Min. DR failover ~ 3 seconds (worst case)
    - Smaller intervals increase PIM query traffic
      - Increase is usually insignificant

Module7. ppt

- **Tuning DR Failover**
  - The DR Failover process can be indirectly tuned by varying the PIM "query interval" on the interface. This is accomplished using the following IOS interface command:

    ```
    ip pim query-interval <seconds>
    ```

    By reducing this value from its default of 30 seconds, the period between PIM Hello messages is reduced as well as the "expiration" time advertised in the PIM Hello message.
  - The minimum query interval that can be configured is 1 second. This results in an "expiration" time of 3 seconds which is the worst case scenario for DR failover. However, reducing the query interval from its default of 30 seconds increases the amount of PIM Hello traffic on the local LAN. In most cases, this is an acceptable trade-off when faster DR Failover is desired.



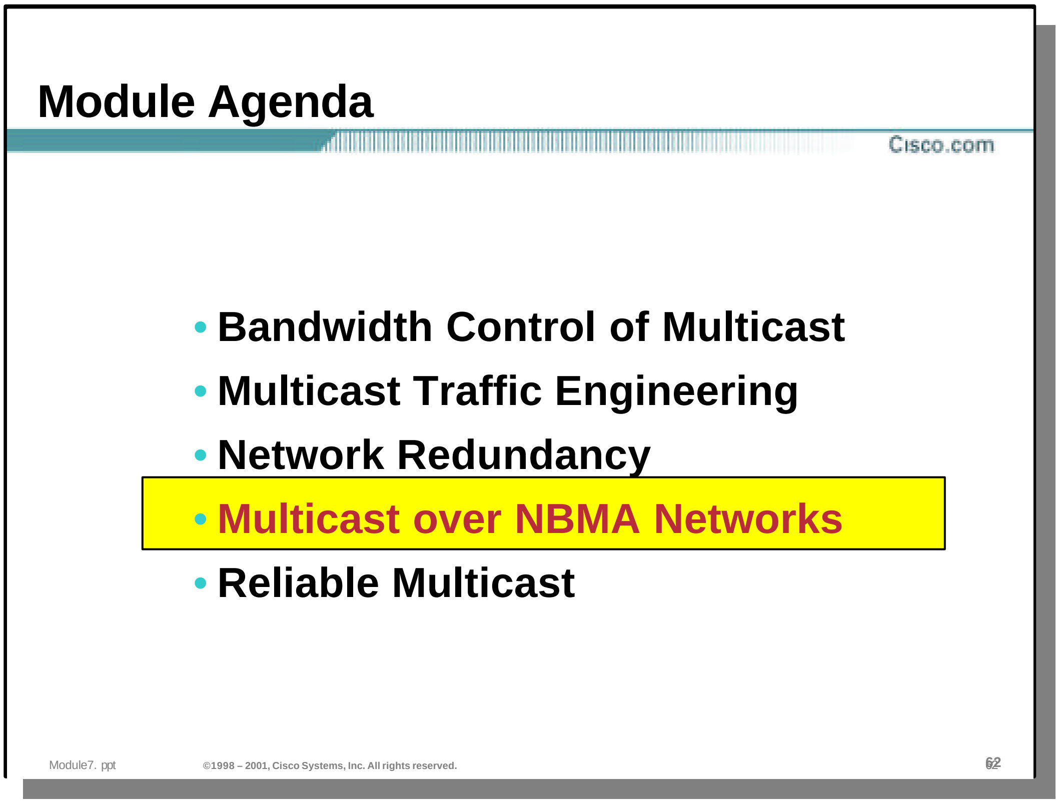
Network Topology Changes

Cisco.com

- 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

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 61

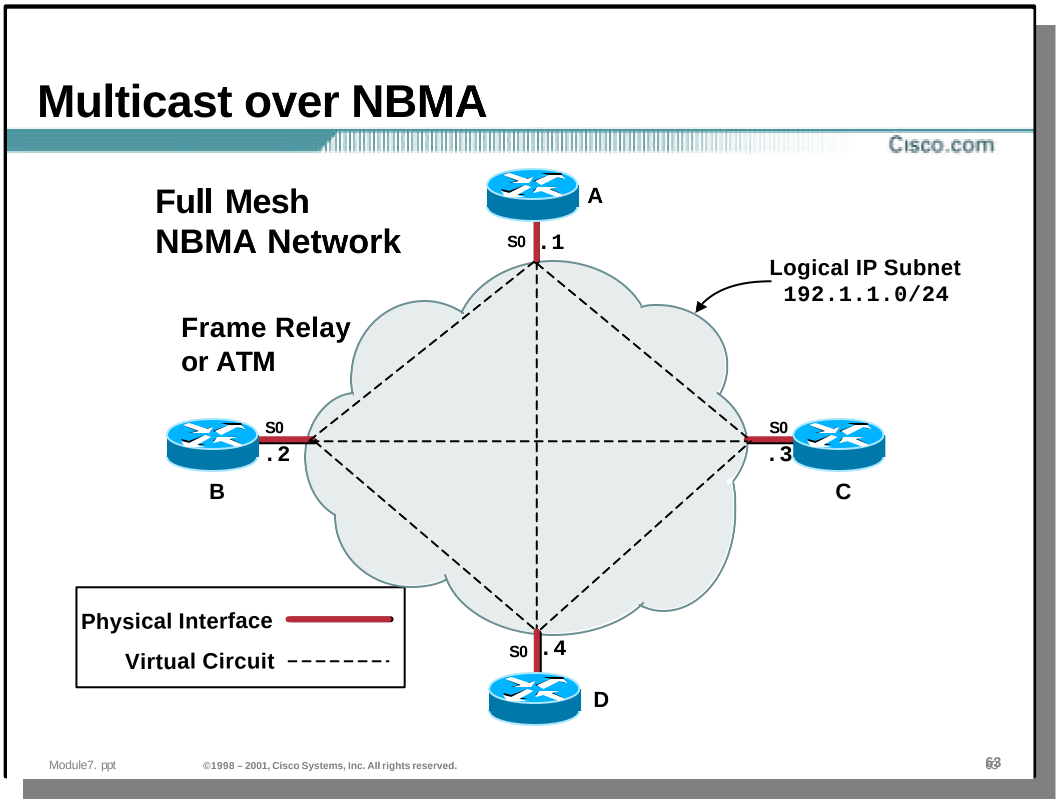
- **Network Topology Changes**
  – Convergence of the PIM distribution tree topology is dependent on the convergence of the unicast routing topology. This is because PIM normally makes us of the unicast routing table to calculate the correct RPF interface for each mroute table entry.
  – In order to synchronize changes in the unicast routing table with the PIM topology, the Cisco IOS implementation of PIM recalculates the RPF interfaces for every entry in the mroute table every 5 seconds. If an RPF interface changes, then the appropriate joins, prunes and/or grafts are sent by the router to rebuild the multicast distribution tree around the network failure.
  – The end result of the above is that the worst case convergence of PIM is approximately 5 seconds after unicast routing converges.



Module Agenda

- Bandwidth Control of Multicast
- Multicast Traffic Engineering
- Network Redundancy
- Multicast over NBMA Networks
- Reliable Multicast

Module7. ppt



Multicast over NBMA

Full Mesh
NBMA Network

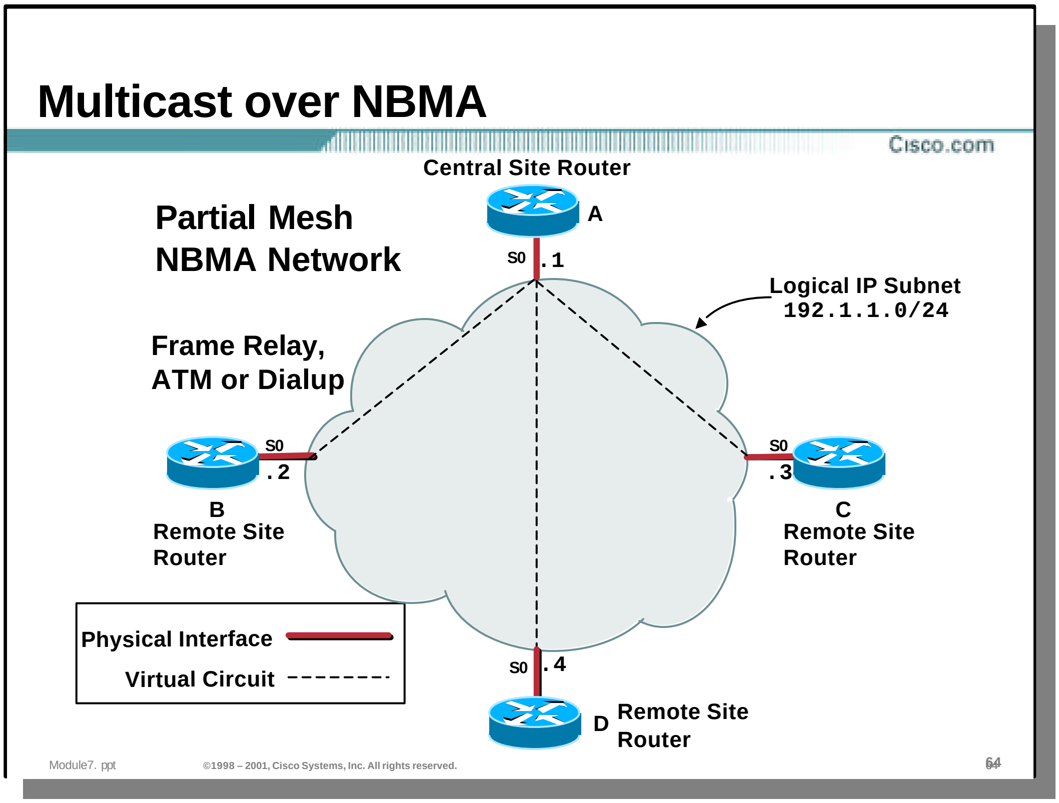
Frame Relay
or ATM

Logical IP Subnet
192.1.1.0/24

Physical Interface
Virtual Circuit

- **Multicast over NBMA**
  - Non-Broadcast, Multi-Access (NBMA) networks such as ATM and Frame relay are implemented using a Virtual Circuit (VC) concept. When these VC's are implemented using point-to-multipoint interfaces, the NBMA cloud is configured as a Logical IP Subnet (LIS). When this form of NBMA network connectivity is used, the special nature of how broadcast and multicast traffic is forwarded must be considered in order for IP Multicast in general and PIM in specific, to operate correctly.

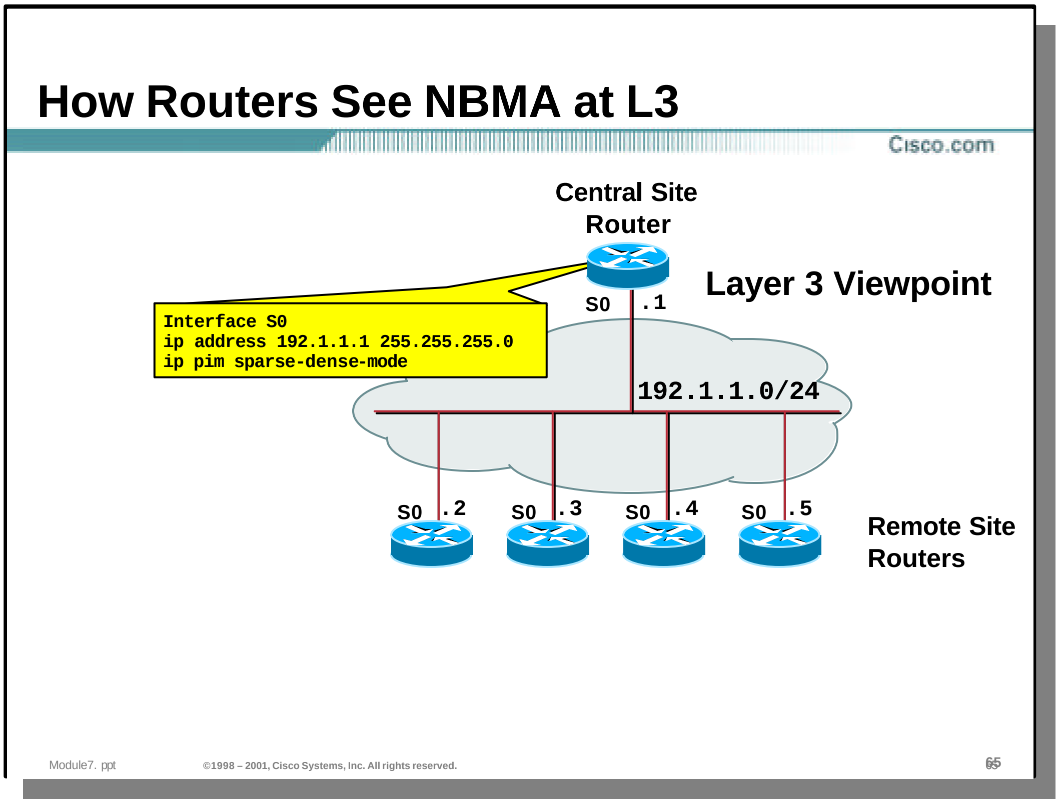
    Note: It is completely different situation when point-to-point sub-interfaces are used. In that case, the network is *not* a LIS and each point-to-point VC/sub-interface has its own subnet. Furthermore, the router sees the network as a collection of point-to-point links in with the same characteristics of serial links. *This section is **not** applicable to this point-to-point sub-interface model.*
  - The interconnectivity of the nodes in the NBMA cloud generally fall into two categories, Full Mesh and Partial Mesh. The example network shown in the drawing above is of a Full Mesh NBMA network. This has the following characteristics:
    - Each router has a single point-to-multipoint physical interface to the LIS.
    - Each router has a separate VC to every other router in the network.



Multicast over NBMA

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 64

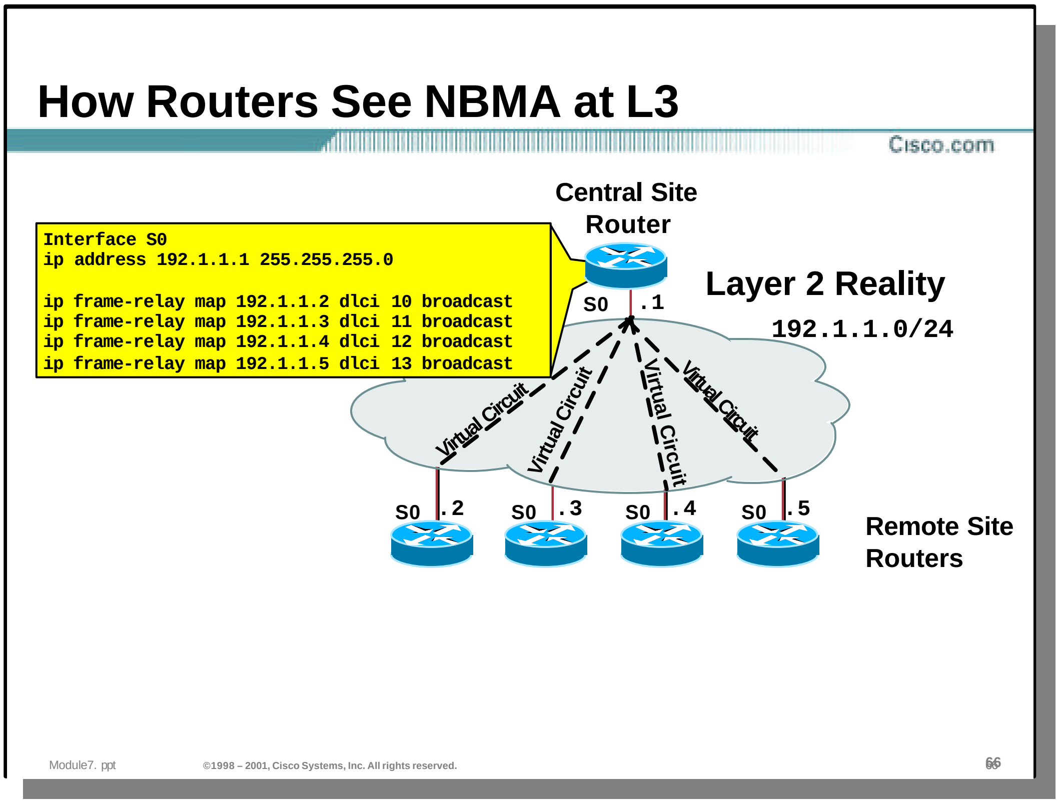
- **Multicast over NBMA**
  - The example network shown in the drawing above is of a Partial Mesh NBMA network. This has the following characteristics:
    - Each router has a single point-to-multipoint physical interface to the LIS.
    - Each router has does not have a separate VC to every other router in the network.
  - Instead of having a complete mesh that interconnects every router in the network, only a "paritial" set of VC's are configured. The typical configuration is for only the central site router to have a full set of VC's to every other remote site router in the network. It is this particular configuration that poses some restrictions on the design of IP Multicast.
  - Frame Relay and ATM are the scenarios that most people immediately think of when this type of network is mentioned. However, another very common network scenario also fits in this category: Dialup networks.
    - Most Dialup networks (Modem or ISDN) are configured such that all incoming dial connections are given an IP Address within a Logical IP Subnet. (Assigning each dialup connect its own subnet would use 4 IP addresses; 2 host addresses, 1 subnet address and 1 broadcast address. The LIS approach saves precious IP address as it only requires 2 addresses per dialup connection.)
    - ***As a result, the same issues that apply to the Partial Mesh, Frame Relay and ATM scenarios apply to Dialup networks.***



How Routers See NBMA at L3

```
Interface S0
ip address 192.1.1.1 255.255.255.0
ip pim sparse-dense-mode
```

- **Layer 3 Viewpoint**
  - When a point-to-multipoint NBMA network is configured using a LIS (as described in the previous slides), the router sees only a single physical interface from a Layer 3 perspective.
  - In the above example, the Central Site router sees **Serial0** as a single interface that is connected to the Remote site routes. Furthermore, this single interface appears (from a Layer 3 point of view) as having the same characteristics as an Ethernet. That is to say, any broadcast or multicast packet sent on **Serial0** will reach all remote site routers.



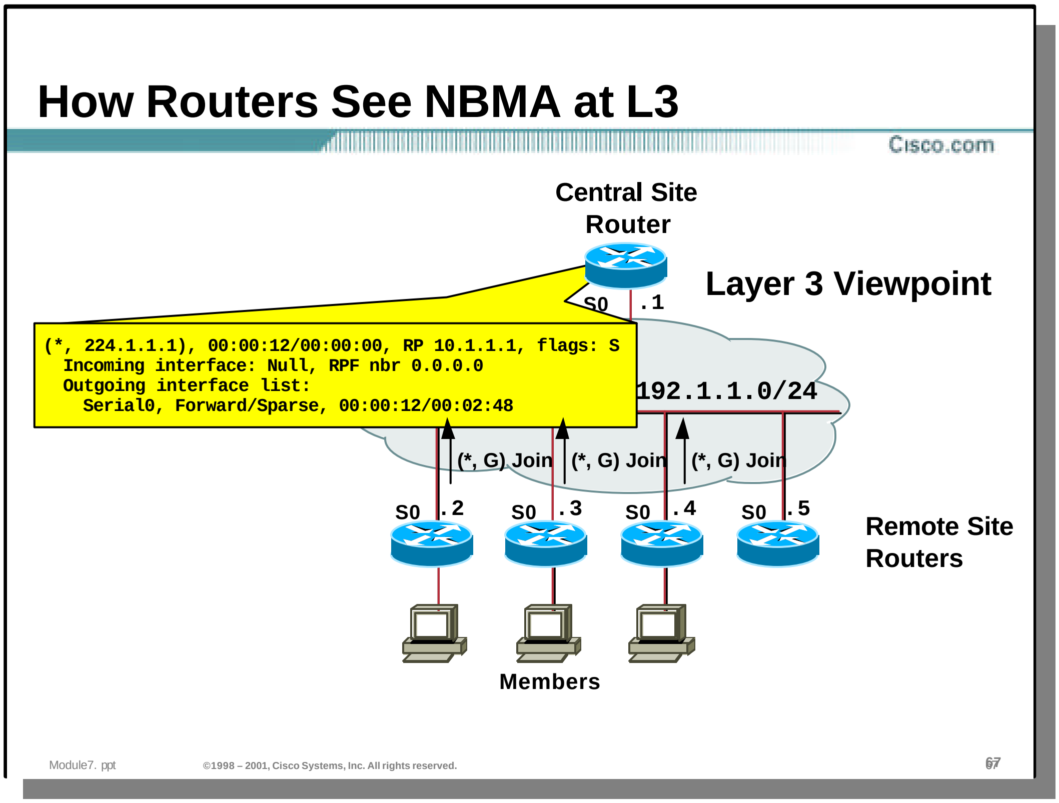
How Routers See NBMA at L3

```
Interface S0
ip address 192.1.1.1 255.255.255.0

ip frame-relay map 192.1.1.2 dlci 10 broadcast
ip frame-relay map 192.1.1.3 dlci 11 broadcast
ip frame-relay map 192.1.1.4 dlci 12 broadcast
ip frame-relay map 192.1.1.5 dlci 13 broadcast
```

Module7. ppt

66

- **Layer 2 Reality**
  - The Layer 2 reality is shown In the above example. The Central Site actually has separate VC’s configured on **Serial0** that connect it to all of the other remote site routers.

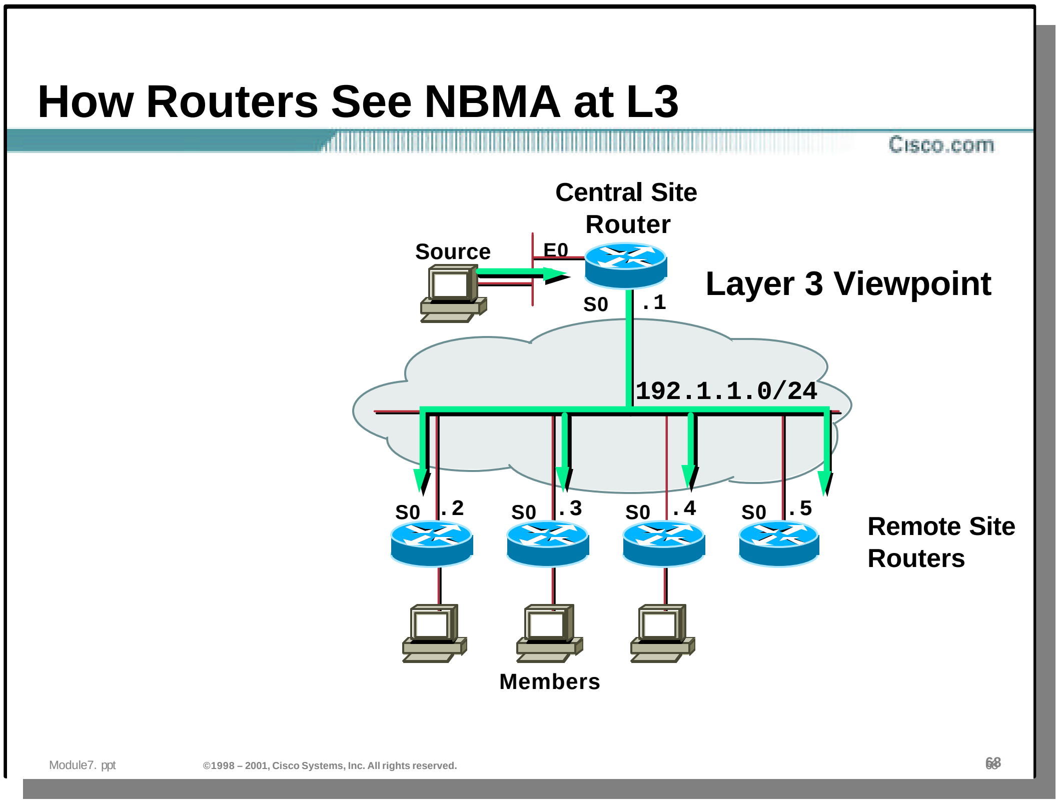


How Routers See NBMA at L3

```
(*, 224.1.1.1), 00:00:12/00:00:00, RP 10.1.1.1, flags: S
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial0, Forward/Sparse, 00:00:12/00:02:48
```

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 67

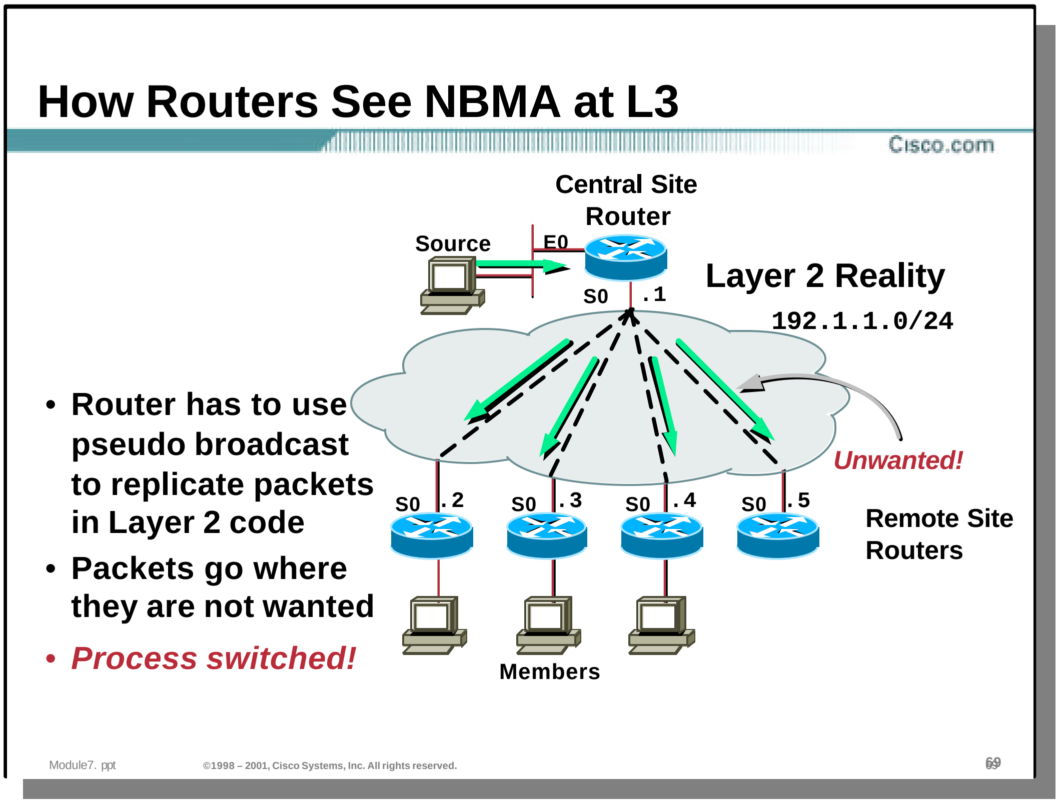
- **Layer 3 Viewpoint — (*, G) Joins**
  - – When several remote site routers join a group (as shown in the above drawing), the router’s Layer 3 viewpoint treats **Serial0** like an Ethernet interface capable of broadcast.
  - – The arriving (*, G) Joins result in only a single interface (**Serial0**) being put on the Outgoing Interface List of the (*, G) entry as shown in the above example.



How Routers See NBMA at L3

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 68

- **Layer 3 Viewpoint — Multicast Flow**
  - When a source for group “G” goes active, the router’s Layer 3 vi ewpoint causes it to treat **Serial0** like an Ethernet interface capable of broadcast. Therefore, the router queues a single copy of the packet to the **Serial0** output queue with the expectation that the packet will reach all remote site routers on the LIS.

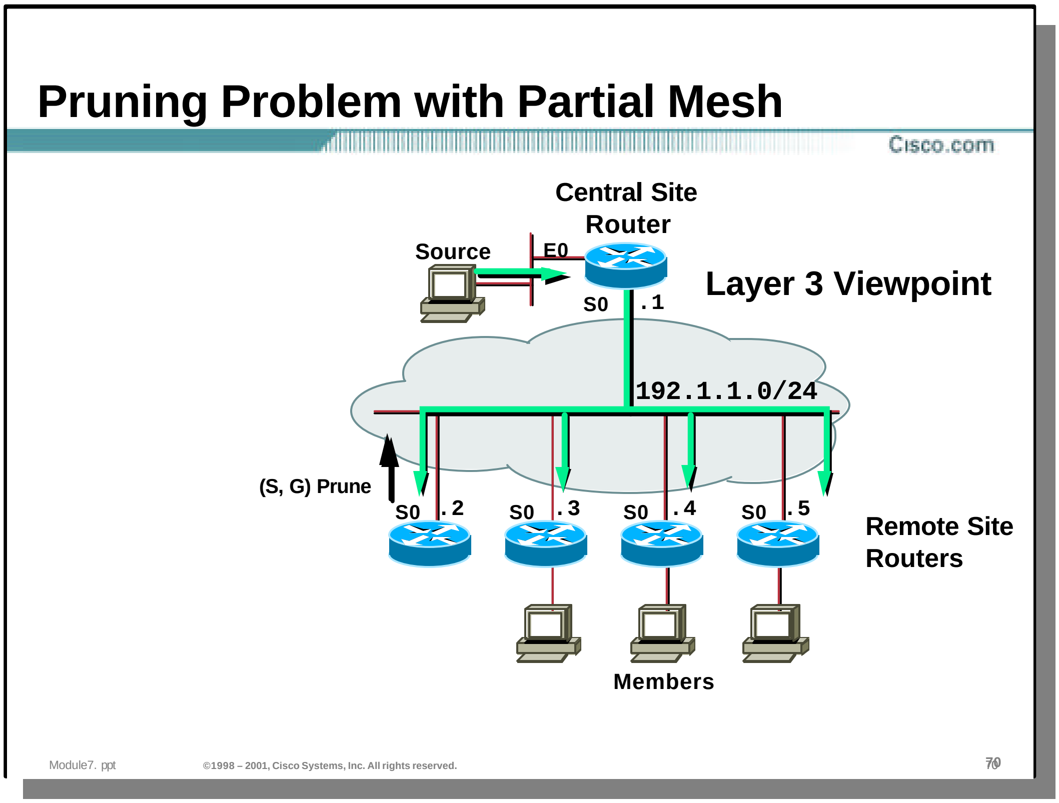


How Routers See NBMA at L3

- Router has to use pseudo broadcast to replicate packets in Layer 2 code
- Packets go where they are not wanted
- *Process switched!*

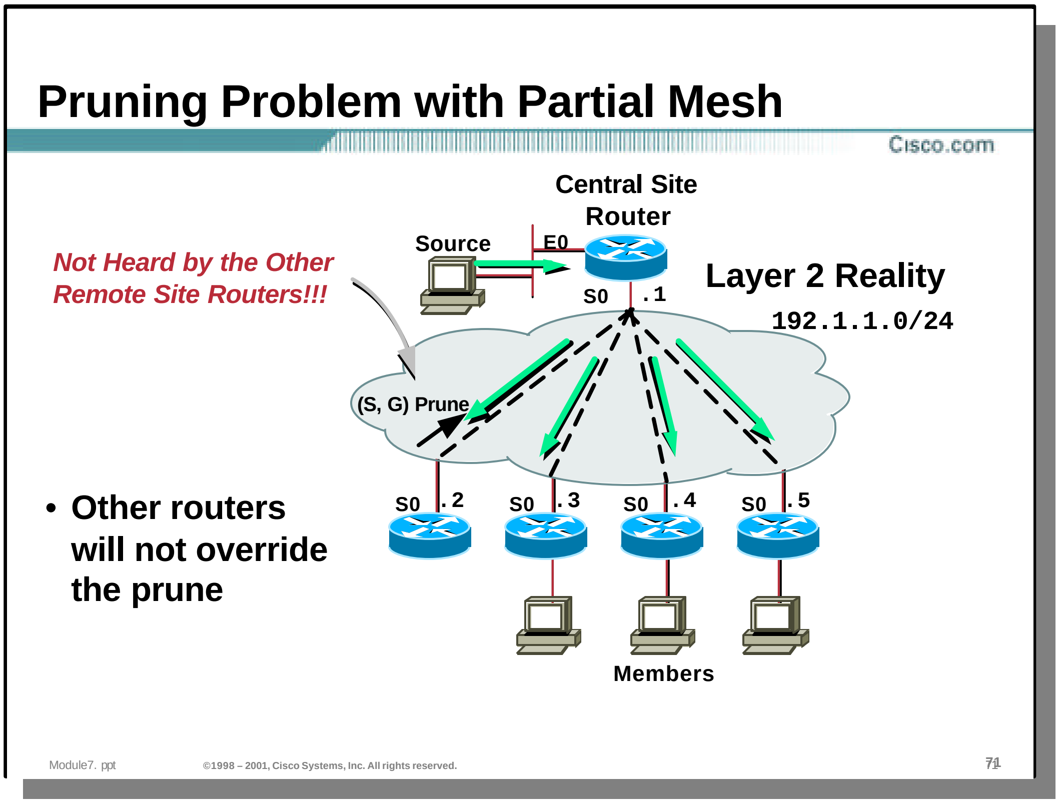
- **Layer 2 Reality — Multicast Flow**
  - The Layer 2 reality of this situation is that there is a separate VC configured on **Serial0** for each remote site router. Therefore, the Frame Relay (or ATM or Dial) interface driver is forced to replicate and send a separate copy of the multicast packet out each VC. (Often with different MAC headers.) This is referred to as “Pseudo Broadcast” and has the following implications:
    - Copies of the multicast packet are sent out all VC’s regardless of whether the remote router at the other end needs the multicast packet or not.
    - The packet replication process must be handled at the Process switching level. ***This has an enormous impact on router performance and causes throughput to suffer!***
    - These replicated packets are placed in a separate “Broadcast” output queue on the interface. This is a limited resource and can easily fill up at Process switching speeds resulting in dropped multicast packets. Drops of multicast data is bad enough but consider what happens if the dropped multicast packet is a PIM control message such as a Join or a Prune. (More on that later.)

    Remember: This scenario is only applicable for NBMA networks implemented using point-to-multipoint interfaces and a Logical IP subnet configuration. This is ***not*** applicable when point-to-point sub-interfaces are used.



Pruning Problem with Partial Mesh

- **Pruning problems with Partial Mesh NBMA Networks**
  - The misguided Layer 3 viewpoint of the routers in partial mesh networks also has an impact on normal PIM control mechanisms.
  - Consider the network example shown above where only a subset of the remote sites have members for a particular group.
    - Traffic is being sent (via the pseudo broadcast mechanism) from the source to all the remote site routers.
    - One remote site router does not have a member for this group yet it is receiving unwanted (S, G) traffic. It therefore responds by multicasting an (S,G) Prune PIM control message to the LIS for the Central site router to process.

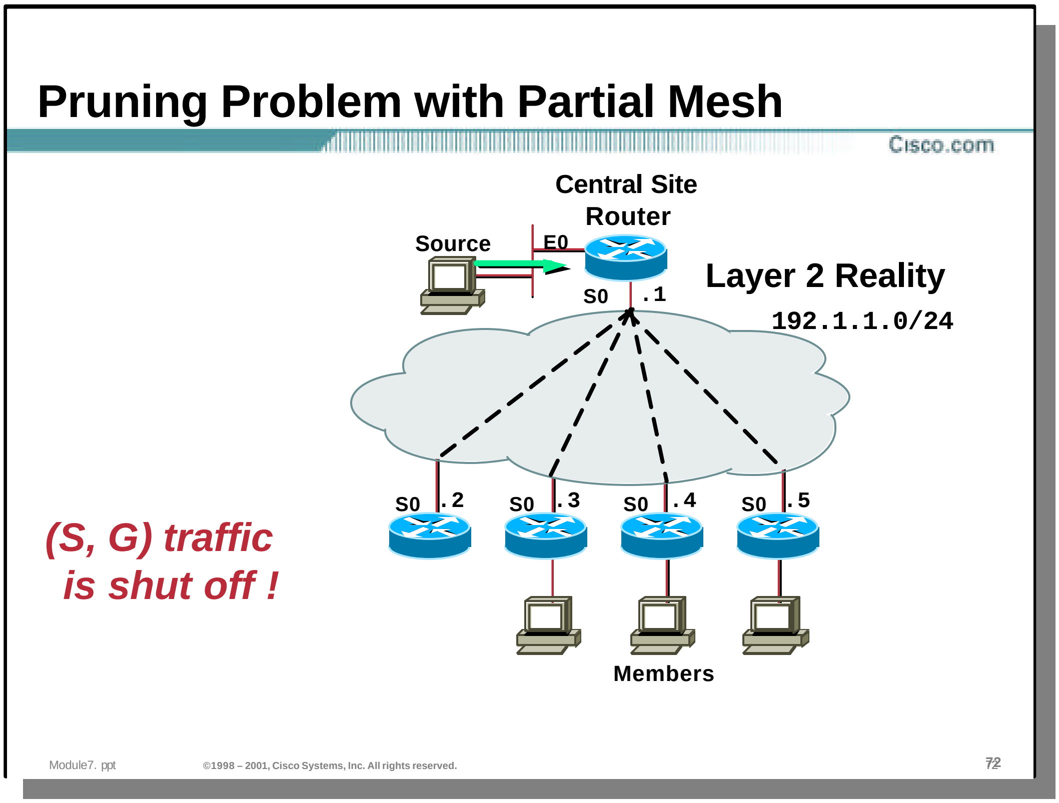


Pruning Problem with Partial Mesh

Not Heard by the Other Remote Site Routers!!!

- Other routers will not override the prune

- **Pruning problems with Partial Mesh NBMA Networks**
  - The Layer 2 reality of the situation is that the (S, G) Prune message is only sent up the VC to the Central site router. The other remote site routers in the partial mesh networks do not hear the (S, G) Prune.
  - As a result of not hearing the (S, G) Prune, the remote site routers do not know to send an (S, G) Join message to the Central site router to override the Prune.

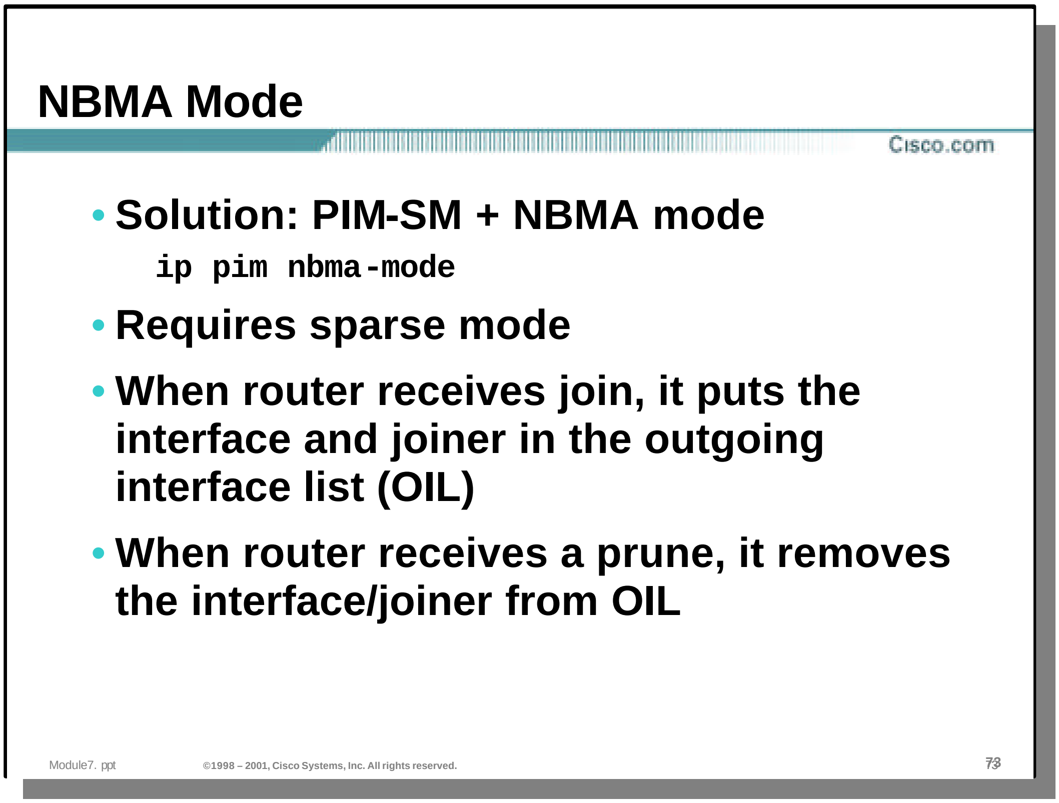


Pruning Problem with Partial Mesh

Cisco.com

(S, G) traffic is shut off !

- **Pruning problems with Partial Mesh NBMA Networks**
  - As a result of not hearing any (S, G) Joins to to override the Prune, the Central site router dutifully prunes interface **Serial0** after the 3 second prune delay.
    - ***This result in (S, G) traffic being shutoff to the other sites!***



NBMA Mode

- Solution: PIM-SM + NBMA mode

```
ip pim nbma-mode
```

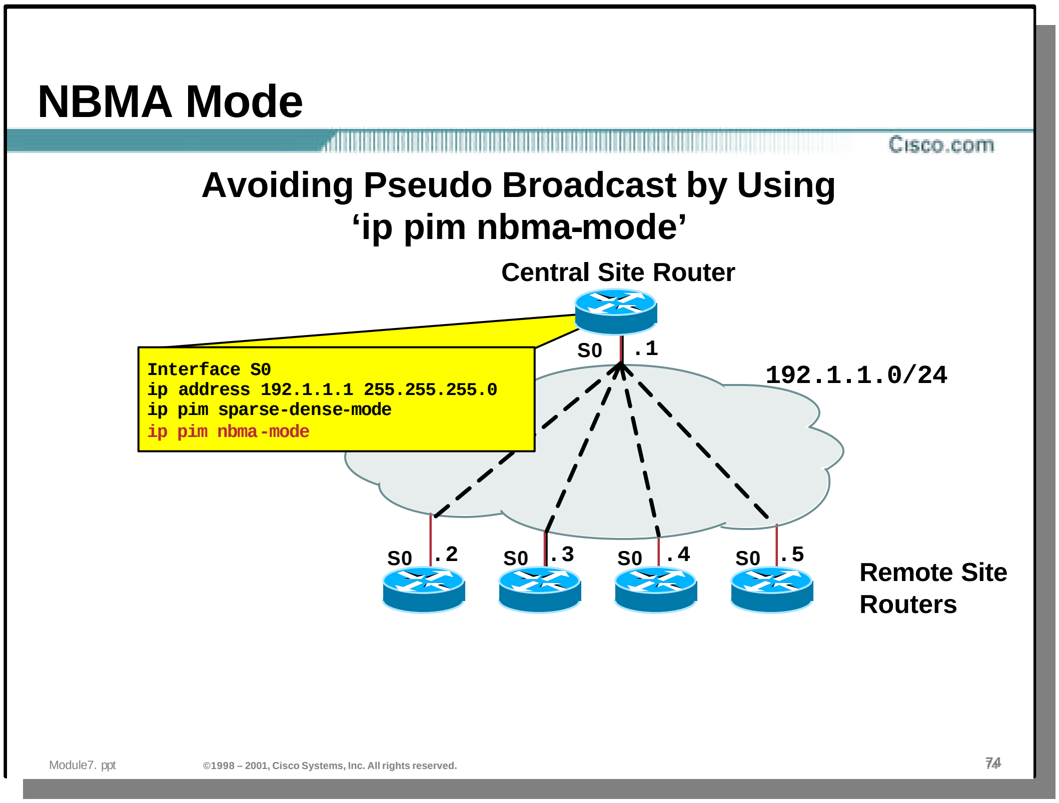
- Requires sparse mode
- When router receives join, it puts the interface and joiner in the outgoing interface list (OIL)
- When router receives a prune, it removes the interface/joiner from OIL

- **NBMA Mode Solution**
  - In order to deal with these problems introduced by NBMA networks, it is necessary to provide the router with information about the underlying Layer 2 topology of the NBMA network. Cisco IOS accomplishes this with the following interface command:

    ```
    ip pim nbma-mode
    ```

    In order for this command to function correctly, sparse mode must be in use. The reason that this is necessary should be clear from the following description of how `nbma-mode` operates.
  - When the ‘`ip pim nbma-mode`’ command is configured on an interface, the normal PIM control message processing is modified as follows:
    - When a Join message is received on the interface, the router puts both the **interface** and the **joiner** (usually in the form of the joiners IP address) in the Outgoing Interface List (OIL).
    - When a Prune message is received on the interface, the router removes the associated **interface/joiner** from the OIL.

    The method effectively maintains a picture of the active underlying Layer 2 topology in the OIL which allows the router to make the appropriate fowarding decisions at Layer 3.

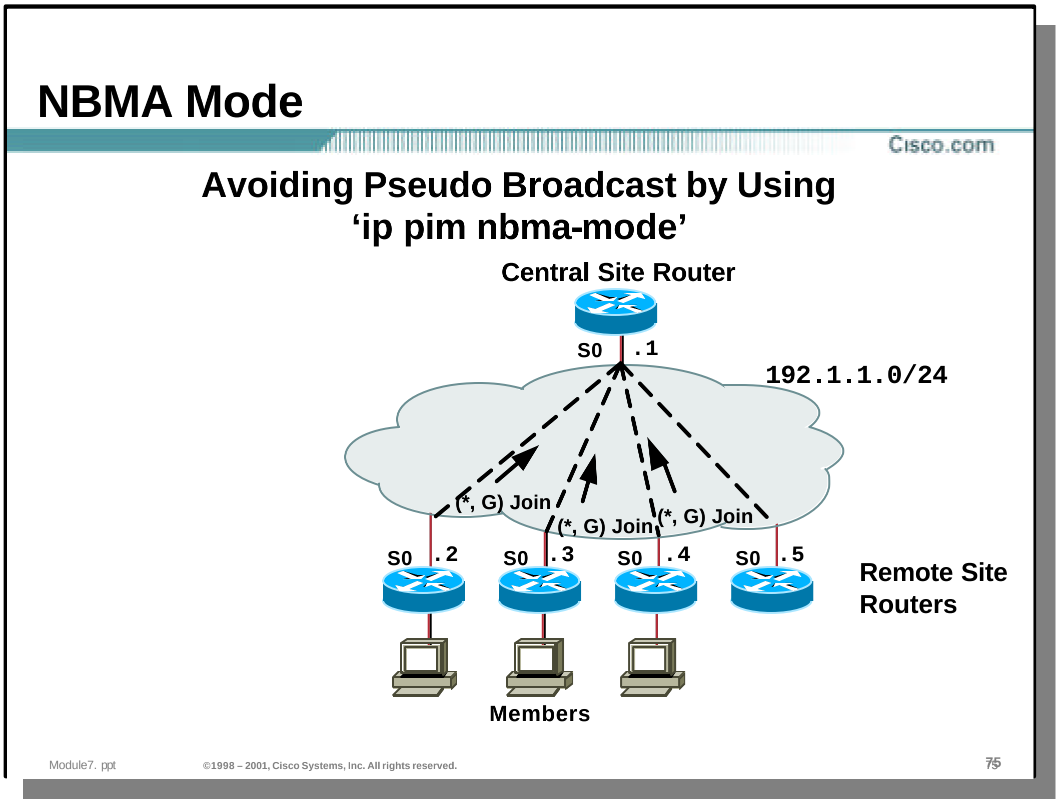


NBMA Mode

Avoiding Pseudo Broadcast by Using ‘ip pim nbma-mode’

```
Interface S0
ip address 192.1.1.1 255.255.255.0
ip pim sparse-dense-mode
ip pim nbma -mode
```

- **NBMA Mode Example**
  - Returning to our original example, we now configure ‘`ip pim nbma-mode`’ on **Serial0**.



NBMA Mode

Cisco.com

Avoiding Pseudo Broadcast by Using ‘ip pim nbma-mode’

Module7. ppt

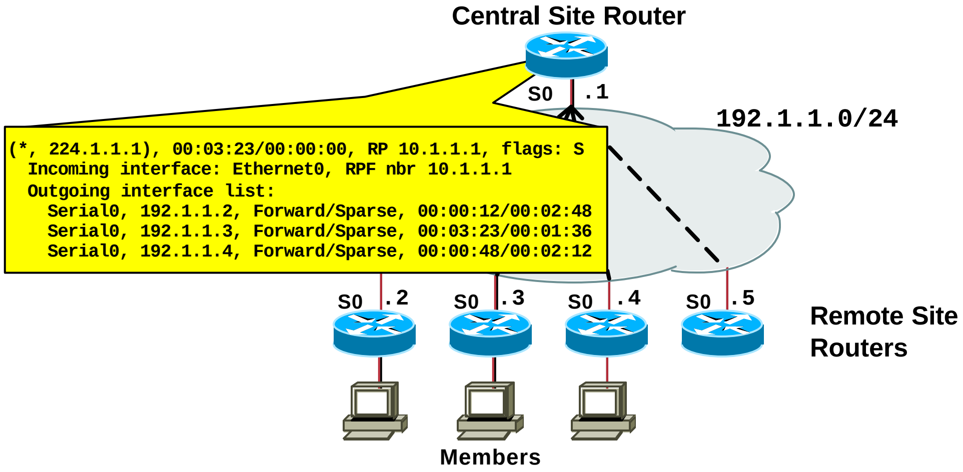
75

- **NBMA Mode Example**
  – Assume that some subset of the remote site routers have members for group “G” and therefore send (*, G) Joins toward the RP. (We are assuming the RP is at the Central site.)

NBMA Mode

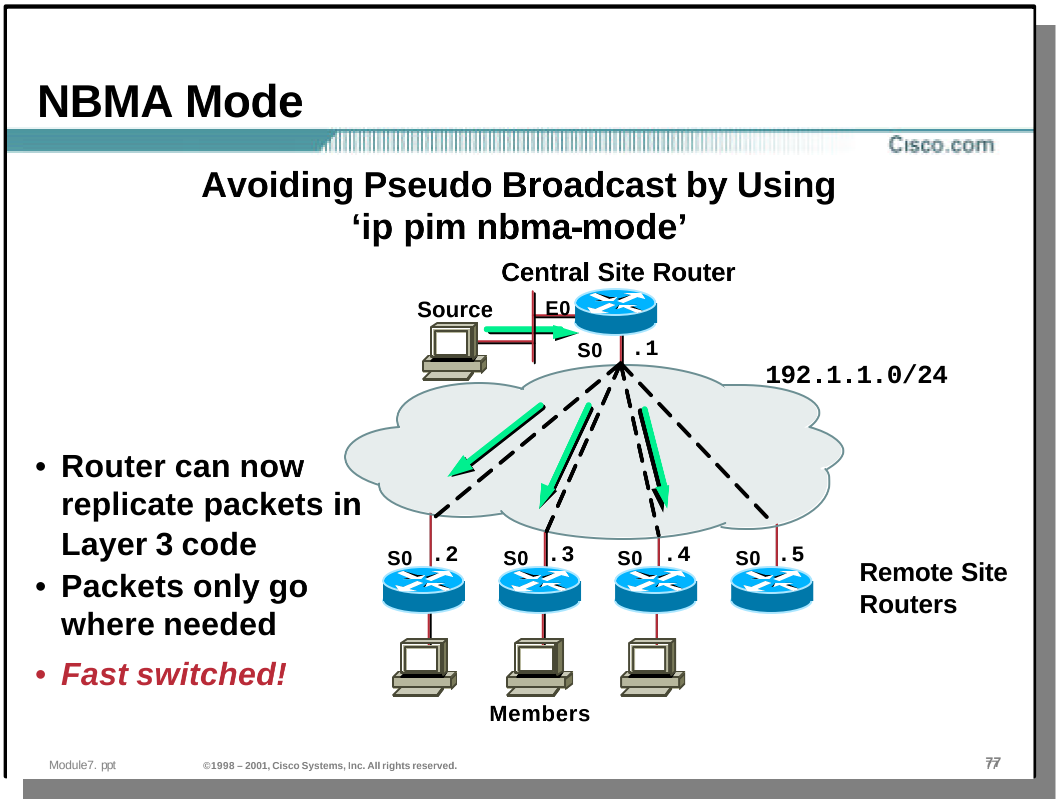
Cisco.com

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



```
(*, 224.1.1.1), 00:03:23/00:00:00, RP 10.1.1.1, flags: S
  Incoming interface: Ethernet0, RPF nbr 10.1.1.1
  Outgoing interface list:
    Serial0, 192.1.1.2, Forward/Sparse, 00:00:12/00:02:48
    Serial0, 192.1.1.3, Forward/Sparse, 00:03:23/00:01:36
    Serial0, 192.1.1.4, Forward/Sparse, 00:00:48/00:02:12
```

- **NBMA Mode Example**
  - When the Central site router receives these (*, G) Joins from three of the remote site routers, it adds a separate **interface/joiner** entry in the OIL for each of the received (*, G) Joins. This results in the OIL as shown in the example above.



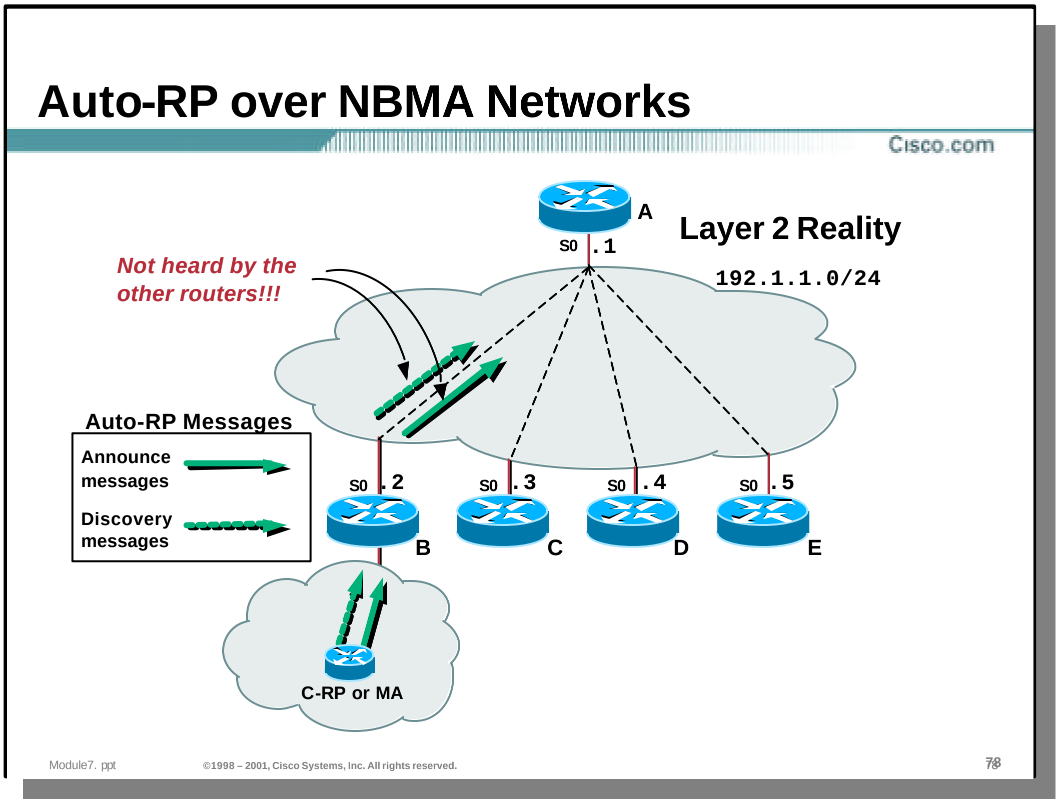
NBMA Mode

Cisco.com

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!***

- **NBMA Mode Example**
  - Because the router now has detailed information about the underlying Layer 2 topology (in the form of separate **interface/joiner** entries in the OIL) it can now replicate the multicast packets in the Layer 3 PIM code. This has the following advantages:
    - The multicast packets are only sent to those remote site routers that have joined the group.
    - Because the multicast fast-switching cache headers are a part of the OIL data structure, the router has the necessary MAC header information to perform fast-switching of the multicast traffic. ***This improves throughput considerably over the pseudo broadcast method which is process switched!***



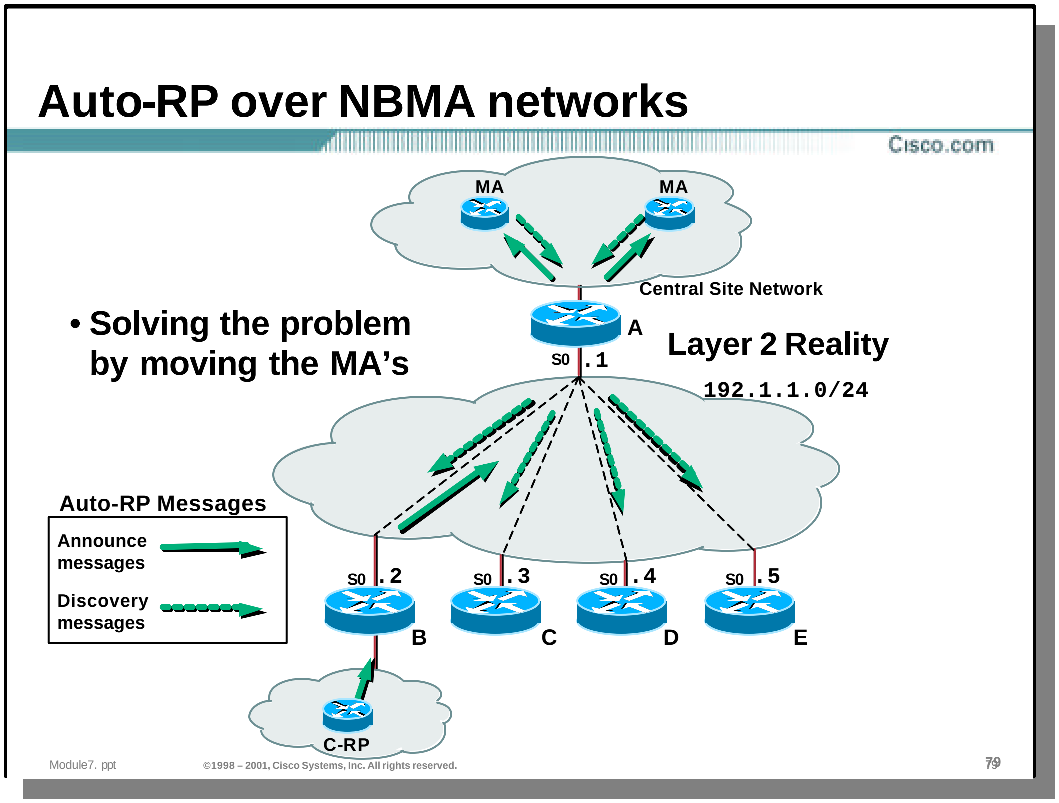
Auto-RP over NBMA Networks

- **Auto-RP over NBMA Networks**

  Because Auto-RP relies on Dense mode flooding of the two Auto-RP groups to function properly, care must be taken when using Auto-RP in partial-mesh NBMA networks.

  – The example above shows a Candidate-RP or Mapping Agent that has been configured at a remote site that is connected to a Hub-and-Spoke, partial mesh NBMA cloud.

  – Because the network is not fully meshed, Auto-RP Announcement and Discovery messages do not reach remote site routers C, D and E.

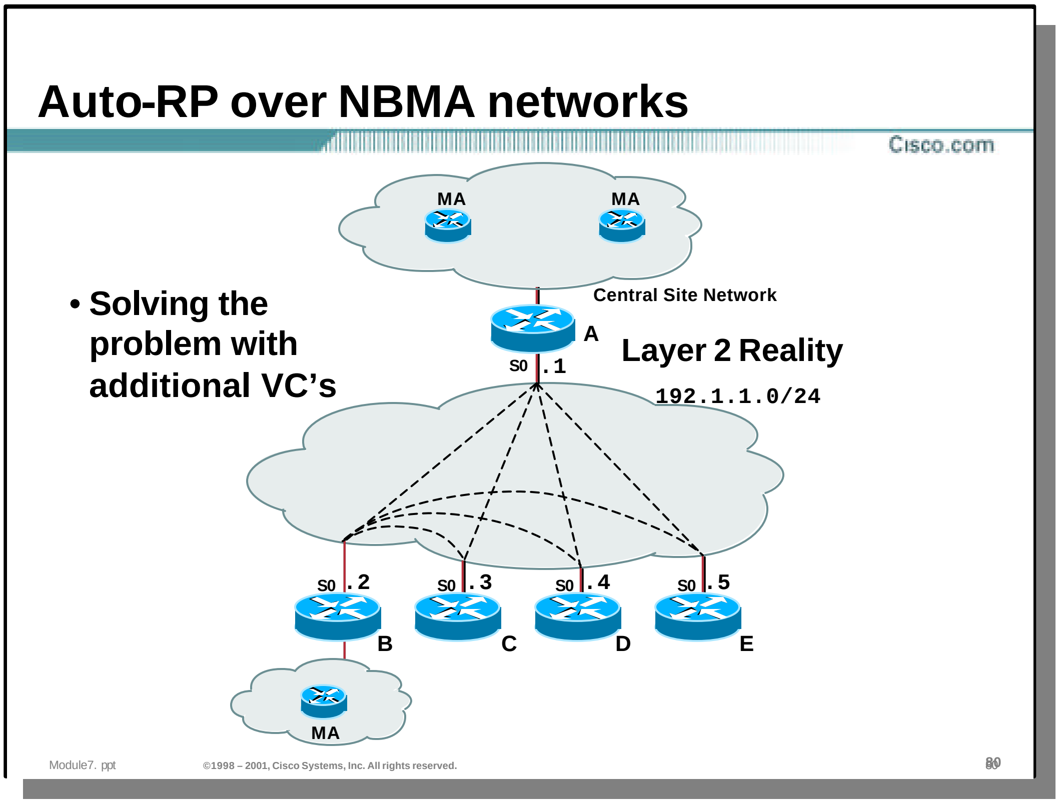


Auto-RP over NBMA networks

- Solving the problem by moving the MA's

Layer 2 Reality

- **Auto-RP over NBMA Networks**
  - One solution is to move the Mapping Agent(s) to the Central Site Network as shown in the example above.
    - Announcement messages from the Candidate-RP travel from the remote site to the Mapping Agent(s) in the Central Site.
    - The Mapping Agent(s) select the RP and send out Discovery messages which are flooded via Dense mode to all of the remote sites.

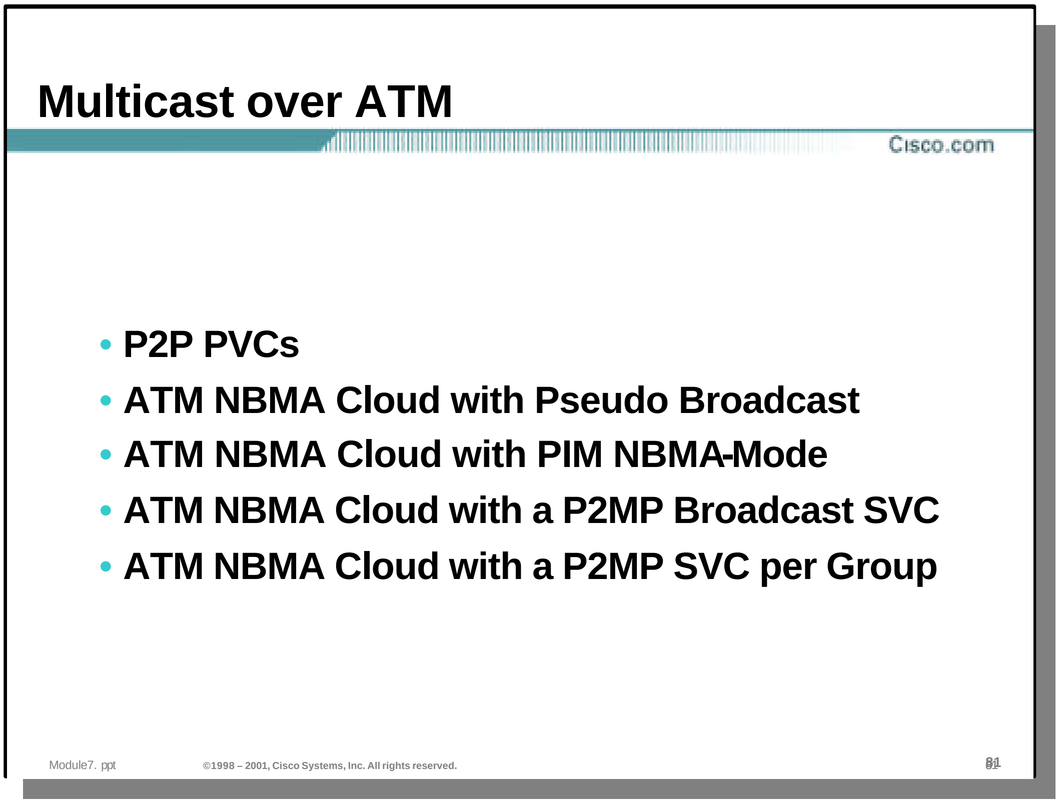


Auto-RP over NBMA networks

Cisco.com

- **Solving the problem with additional VC’s**

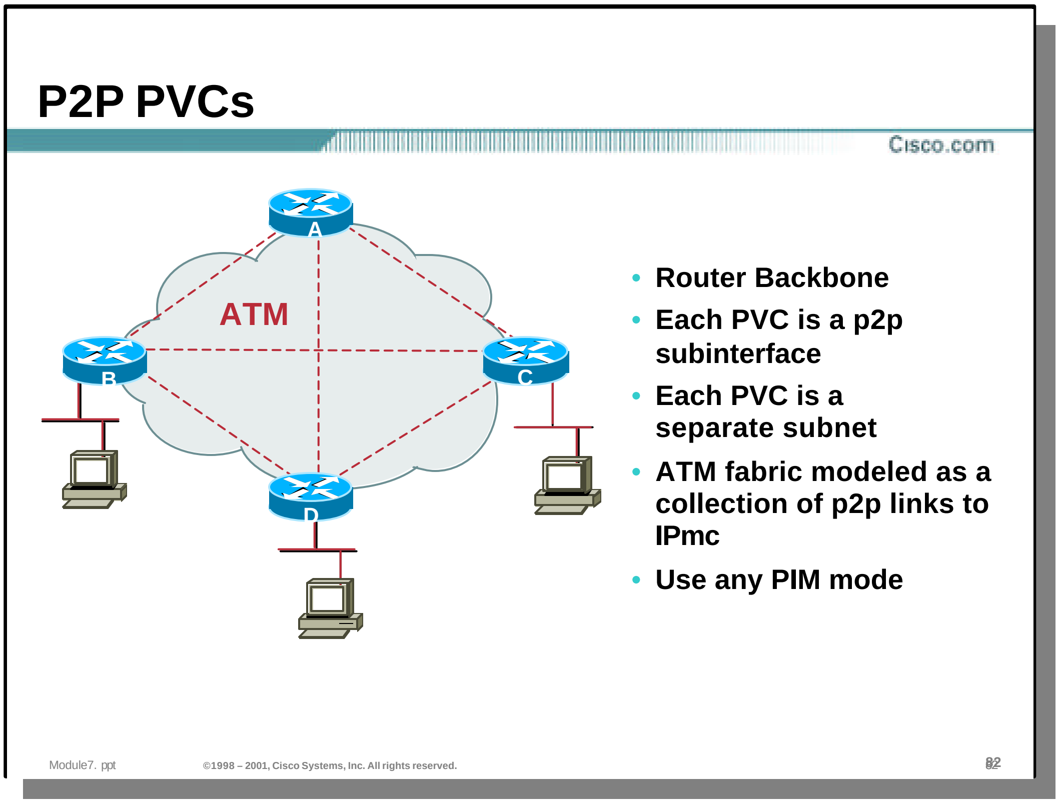
- **Auto-RP over NBMA Networks**
  - Another solution is to configure additional Virtual Circuits as shown in the example above.
    - Discovery messages can now be flooded via Dense mode to all of the other remote sites.
  - The obvious disadvantage is that more Virtual Circuits must be used which often increases the overal operational cost of the network since most Network Service Providers charge for each Virtual Circuit.



Multicast over ATM

- P2P PVCs
- ATM NBMA Cloud with Pseudo Broadcast
- ATM NBMA Cloud with PIM NBMA-Mode
- ATM NBMA Cloud with a P2MP Broadcast SVC
- ATM NBMA Cloud with a P2MP SVC per Group

- **Multicast over ATM**
  - There are several methods that can be employed to run Multicast over a core ATM network. Each of the above methods are addressed in the following section.

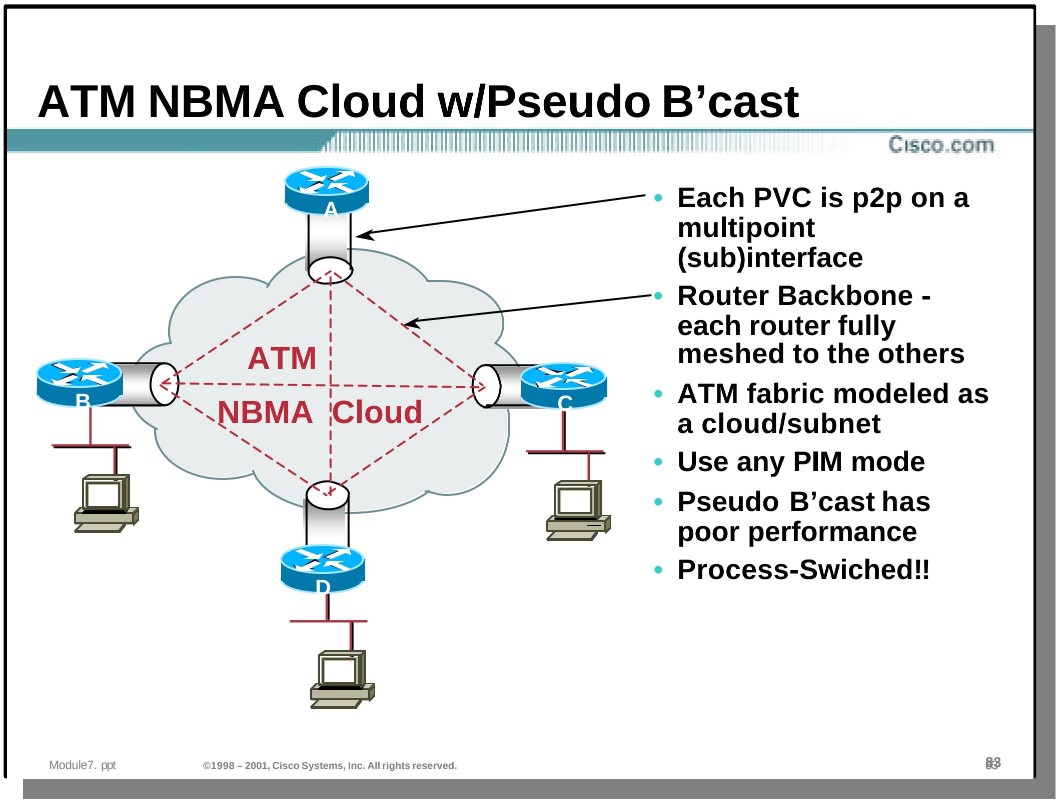


P2P PVCs

Cisco.com

- Router Backbone
- Each PVC is a p2p subinterface
- Each PVC is a separate subnet
- ATM fabric modeled as a collection of p2p links to IPmc
- Use any PIM mode

- **Comments**
  - Could use static PVCs or soft PVCs on the ATM switches
  - Could use SVCs with mapping and broadcast (PVCs more stable)
- **Advantages**
  - Works well when p2mp VCs are not available
  - Effective pruning and excellent configuration control since each VC looks like a separate interface
  - Fast switching supported
- **Disadvantages**
  - Router does replication - watch out for high bandwidth flows and/or high fanout
  - More configuration - more links, more subinterfaces, etc
  - Multiple copies of packet on the ATM fabric (unlike p2mp VCs)
  - Scalability - configuration gets larger as # of routers in mesh gets large



ATM NBMA Cloud w/Pseudo B’cast

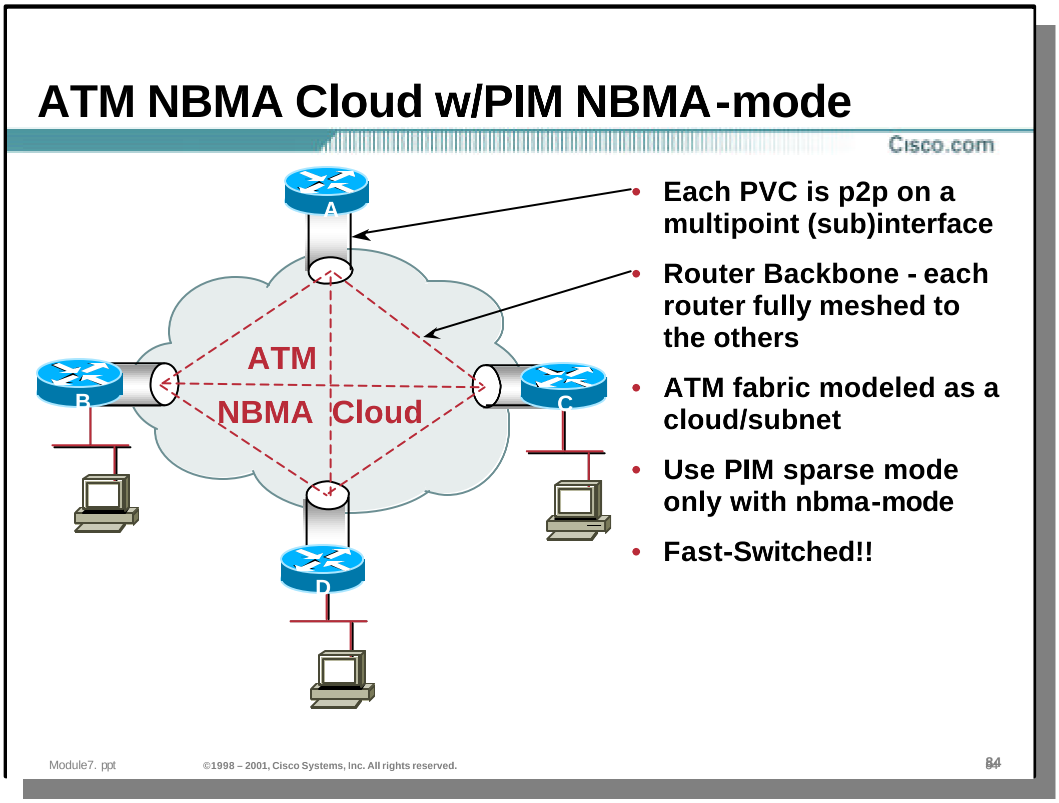
Cisco.com

- Each PVC is p2p on a multipoint (sub)interface
- Router Backbone - each router fully meshed to the others
- ATM fabric modeled as a cloud/subnet
- Use any PIM mode
- Pseudo B’cast has poor performance
- Process-Swiched!!

Module7. ppt

83

- **Comments**
  - – Identical to Frame Relay cloud but over ATM
  - – Could use static PVCs
  - – Could use SVCs with mapping and broadcast (PVCs more stable)
- **Advantages**
  - – Easy to configure.
  - – Works with any PIM mode.
- **Disadvantages**
  - – Router does replication - watch out for high bandwidth flows and/or high fanout
  - – Fast switching not supported
  - – Multiple copies of packet on the ATM fabric (unlike p2mp VCs)
  - – Scalability - configuration gets larger as # of routers in mesh gets large

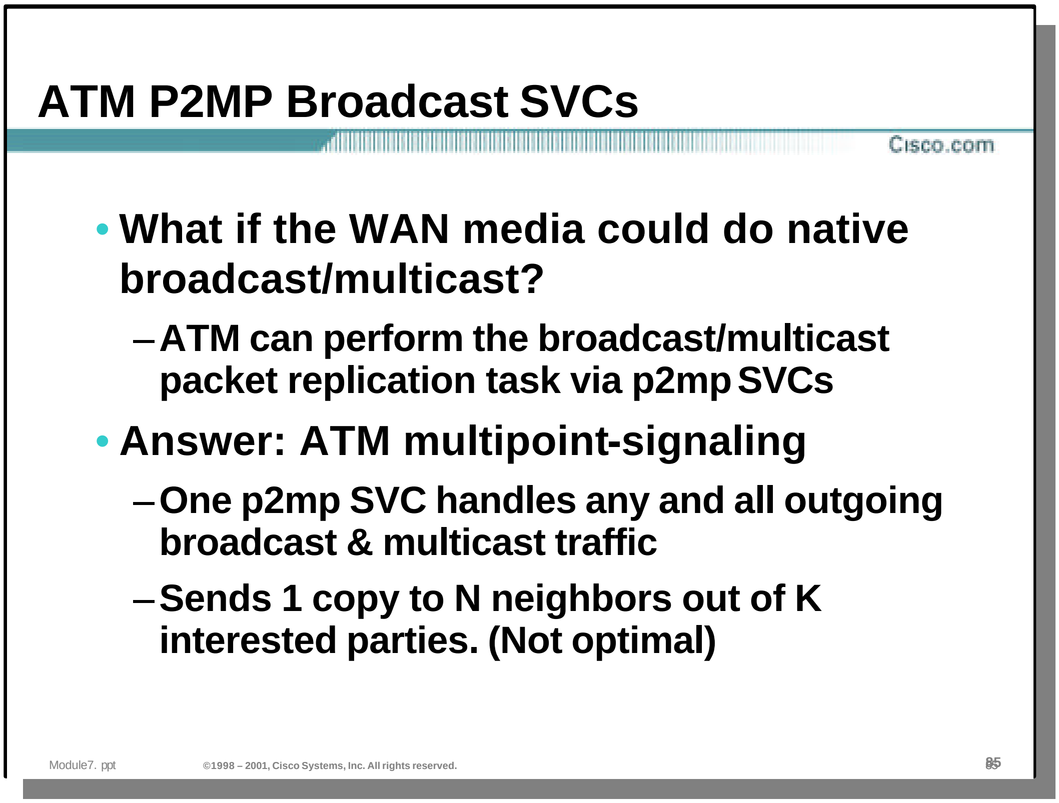


ATM NBMA Cloud w/PIM NBMA-mode

Cisco.com

- Each PVC is p2p on a multipoint (sub)interface
- Router Backbone - each router fully meshed to the others
- ATM fabric modeled as a cloud/subnet
- Use PIM sparse mode only with nbma-mode
- Fast-Switched!!

- **Comments**
  - Identical solution to example in the NBMA mode section but over ATM
  - Could use static PVCs or soft PVCs on the ATM switches
  - Could use SVCs with mapping and broadcast (PVCs more stable)
- **Advantages**
  - Traffic only goes to members who have joined the group via PIM SM even though it's a multipoint interface (I.e. we just don't replicate to each neighbor who has a broadcast map)
  - Works well when p2mp VCs are not available
  - Effective pruning since we can control replication to each neighbor/VC
  - Fast switching supported
- **Disadvantages**
  - Router does replication - watch out for high bandwidth flows and/or high fanout
  - More configuration - more links, more subinterfaces, must configure map list commands, etc
  - Multiple copies of packet on the ATM fabric (unlike p2mp VCs)
  - Scalability - configuration gets larger as # of routers in mesh gets large



ATM P2MP Broadcast SVCs

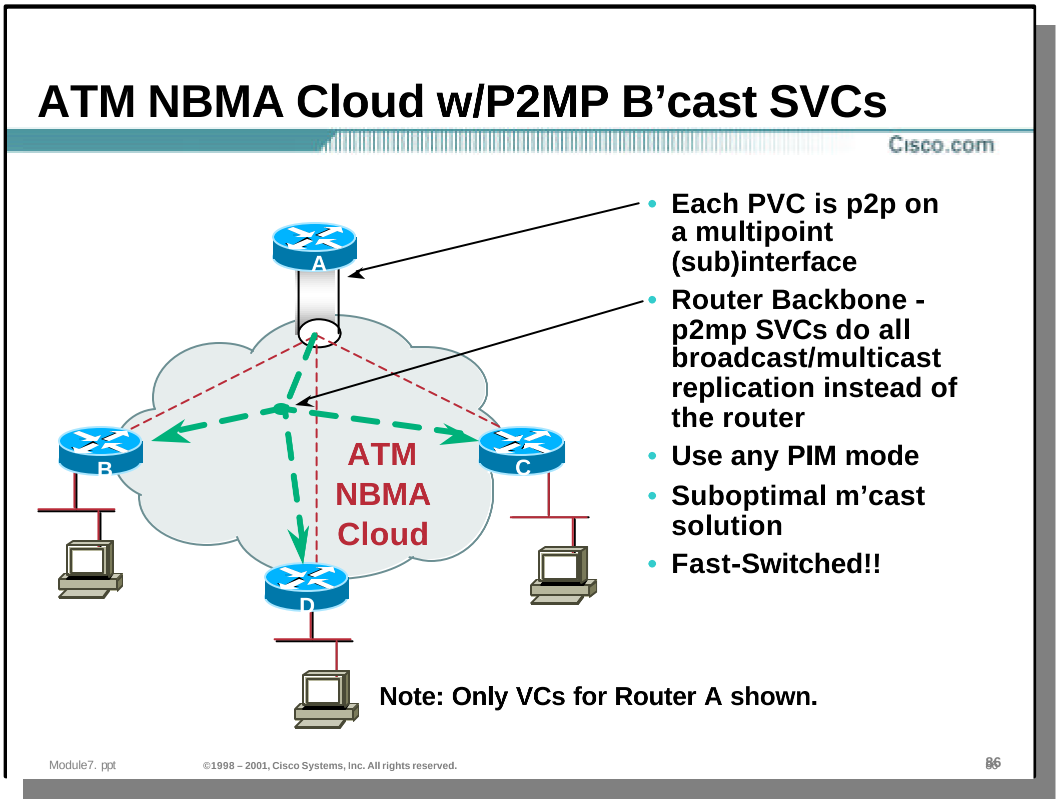
- What if the WAN media could do native broadcast/multicast?
  - ATM can perform the broadcast/multicast packet replication task via p2mp SVCs
- Answer: ATM multipoint-signaling
  - One p2mp SVC handles any and all outgoing broadcast & multicast traffic
  - Sends 1 copy to N neighbors out of K interested parties. (Not optimal)

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 85

- **ATM Point-to-Multipoint (p2mp) Broadcast SVC’s**

  Many ATM networks have the ability to build p2mp Switched Virtual Circuits. These p2mp SVC’s can be used to connect a router to all of the other routers in the ATM LIS. By configuring p2mp SVC’s (one originating from each router in the ATM LIS), the ATM network can be made responsible for performing the broadcast function.

  - ATM Multipoint Signaling is a feature in Cisco IOS that permits a router to be configured so that it will create a p2mp SVC to the other designated routers in the ATM LIS.
    - All outgoing broadcast and multicast traffic is sent on the p2mp SVC.
    - The router sends only a single packet to the ATM network which then reaches *N* neighbor routers out of *K* interested routers. (If the number of *K* interested routers is small compared to *N* number of routers in the LIS, the efficiency of this solution decreases substantially.)



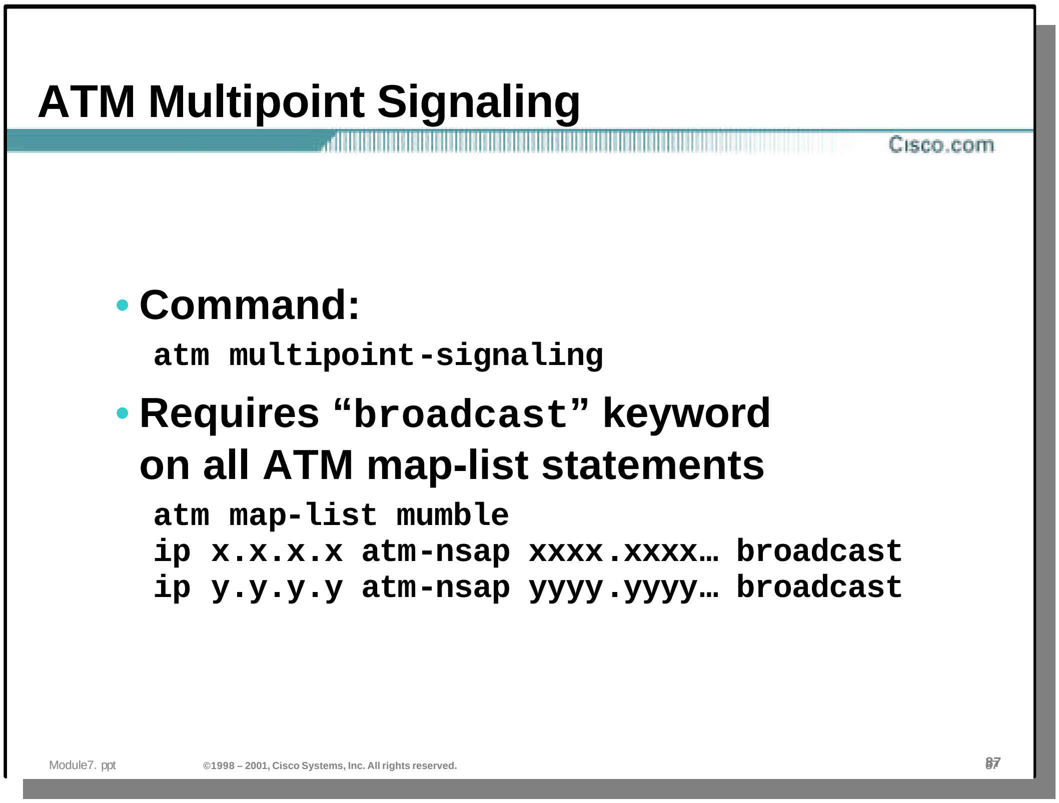
ATM NBMA Cloud w/P2MP B’cast SVCs

Cisco.com

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

Note: Only VCs for Router A shown.

- **Comments**
  - Leverage ATM fabric to do the broadcast/multicast replication vi a a single p2mp VC
  - Could use static P2P PVCs or soft PVCs on the ATM switches
- **Advantages**
  - ATM fabric does the replication instead of the router
  - Off-loads router CPU
  - Only one packet sent to the ATM fabric
  - Fast switching supported
- **Disadvantages**
  - All ATM routers get all the multicast groups even if they don’t care about some of them
  - ATM fabric must support p2mp VCs and have good replication performance
  - Leafs of the p2mp VC are configured with map-list commands with the broadcast keyword
  - Only useful for router-to-router backbones



ATM Multipoint Signaling

- **Command:**

```
atm multipoint-signaling
```

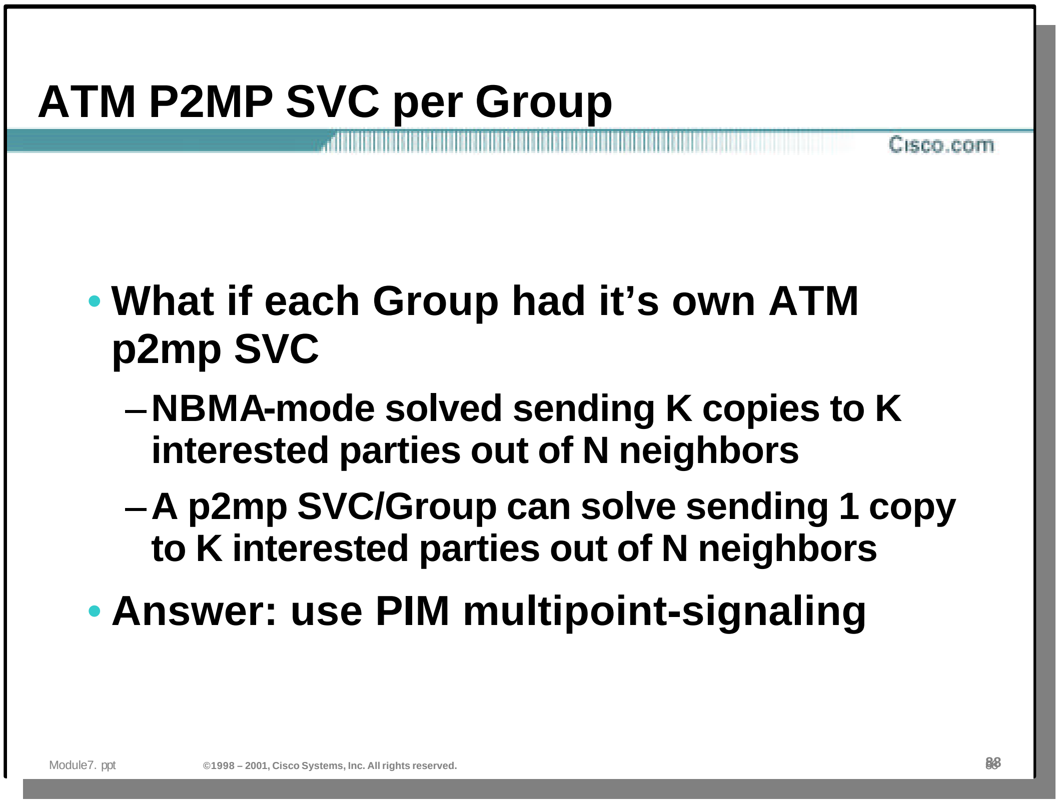
- **Requires “`broadcast`” keyword on all ATM map-list statements**

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

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 87

- **ATM Multipoint Signaling**
  - May be enabled via the following interface command

    `atm multipoint-signaling`
  - An ATM “map-list” must be configured to specify the IP to ATM NSAP address mapping of all the routers in the ATM LIS.
    - The ‘`broadcast`’ keyword must be configured on each entry of the ATM map list. This triggers the router to signal the ATM UNI layer to build a p2mp SVC to these routers in the LIS.

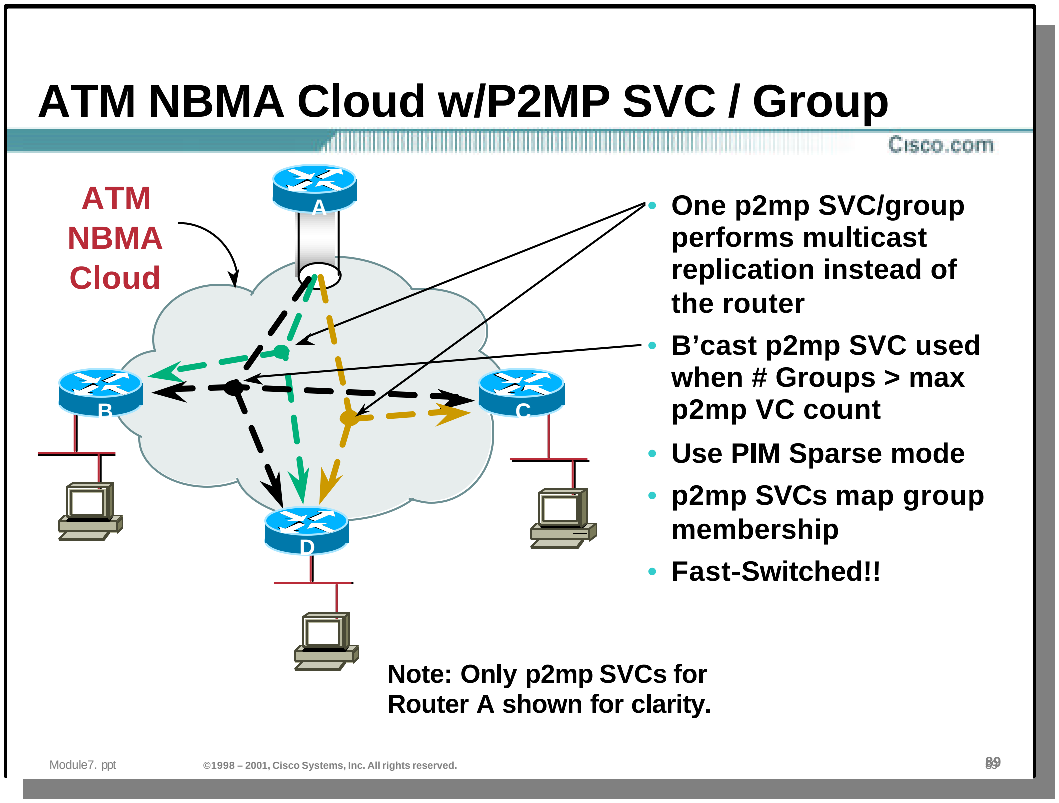


ATM P2MP SVC per Group

Cisco.com

- What if each Group had it’s own ATM p2mp SVC
  - –NBMA-mode solved sending K copies to K interested parties out of N neighbors
  - –A p2mp SVC/Group can solve sending 1 copy to K interested parties out of N neighbors
- Answer: use PIM multipoint-signaling

- **ATM Point-to-Multipoint SVC per Group**
  - – Using a single p2mp SVC for all multicast traffic is sub-optimal as precious bandwidth is consumed sending unwanted copies of the multicast packets to routers that do not have downstream members for the group.
  - – If each group had a dedicated p2mp SVC that connected to *only* those downstream routers that had joined the group (i.e have downstream members), then efficiency is maximized.
  - – ATM Multipoint Signaling is a feature in Cisco IOS that permits a router to be configured so that it will create per group p2mp SVC’s to the other routers that *have joined the group*.
    - A single copy of the packet is sent to exactly *K* interested routers out of a total of *N* neighbor routers.



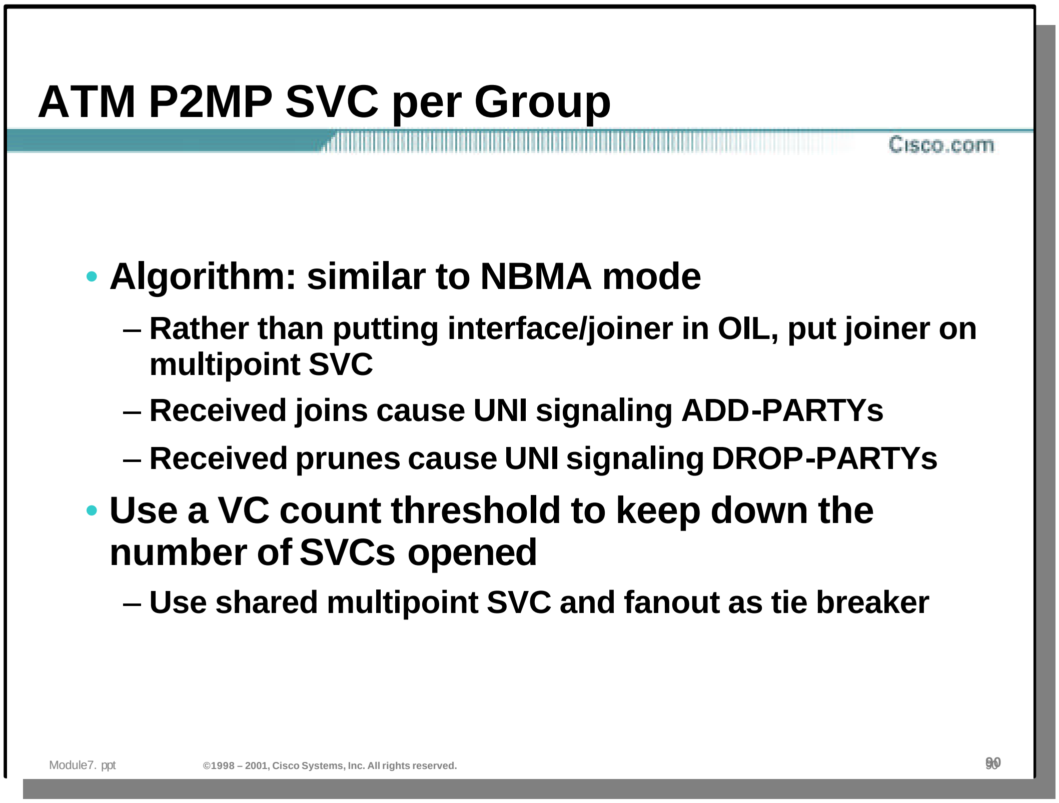
ATM NBMA Cloud w/P2MP SVC / Group

Cisco.com

- One p2mp SVC/group performs multicast replication instead of the router
- B'cast p2mp SVC used when # Groups > max p2mp VC count
- Use PIM Sparse mode
- p2mp SVCs map group membership
- Fast-Switched!!

Note: Only p2mp SVCs for Router A shown for clarity.

- **Comments**
  - Leverage ATM fabric to do the multicast replication via a p2mp VC per Group
  - Could use static P2P PVCs or soft PVCs on the ATM switches
  - Router cannot support unlimited number of p2mp VCs. Therefore Group to p2mp VC mapping is limited to a configured number of Groups.
  - A rate threshold can be set to cause low traffic Groups to drop back to the shared broadcast p2mp VC.
- **Advantages**
  - ATM fabric does the replication instead of the router
  - Off-loads router CPU
  - Only one packet sent to the ATM fabric
  - Fast switching supported
- **Disadvantages**
  - ATM fabric must support p2mp VCs and have good replication performance
  - Leafs of the p2mp VC are configured with map-list commands with the broadcast keyword
  - Only useful for router-to-router backbones

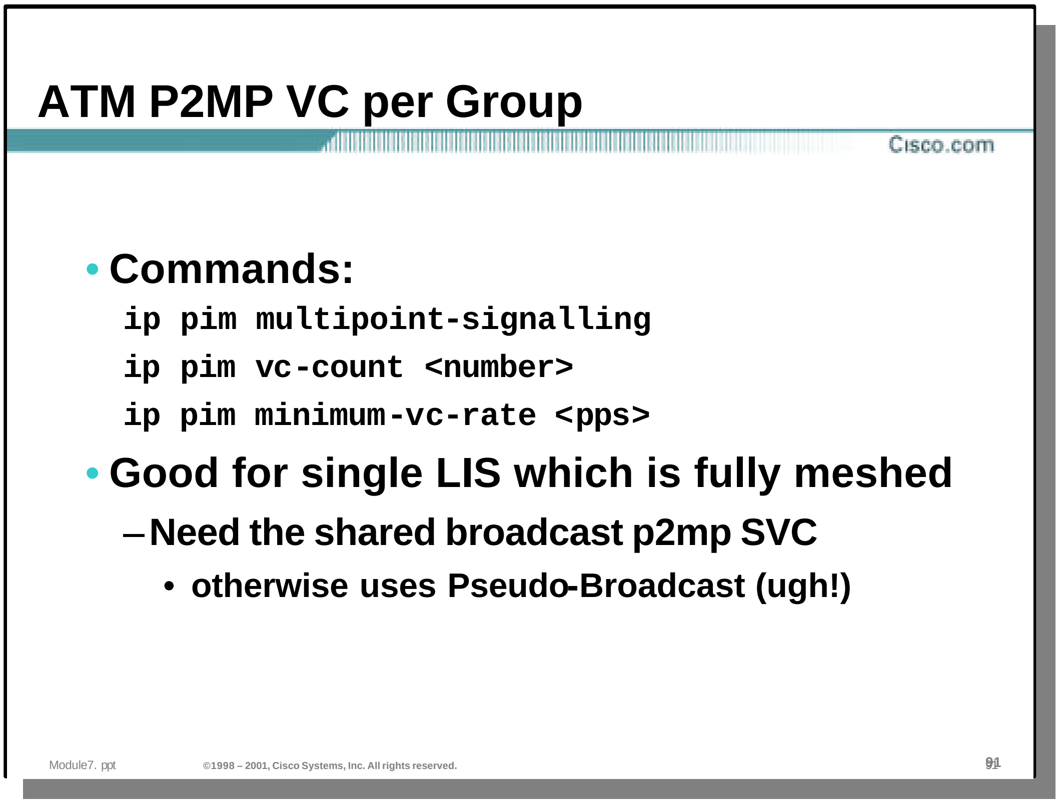


ATM P2MP SVC per Group

Cisco.com

- Algorithm: similar to NBMA mode
  - Rather than putting interface/joiner in OIL, put joiner on multipoint SVC
  - 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 SVCs opened
  - Use shared multipoint SVC and fanout as tie breaker

- **ATM P2MP SVC per Group**
  - This feature is implemented similar to the '`ip pim nbma-mode`' feature.
    - Each individual downstream router is not added to the outgoing interface list as a separate 'interface/joiner ID' as is done in '`ip pim nbma-mode`'.
    - Instead, the ATM interface and VC ID of the p2mp SVC is put in the outgoing interface list.
    - Received PIM Joins trigger the router to send an ADD-PARTY signal (with the NSAP address of the Joiner) to the ATM UNI layer.
    - Received PIM Leaves trigger the router to send an DROP-PARTY signal (with the NSAP address of the Joiner) to the ATM UNI layer.
  - ATM SVC's are limited resources in both the routers and switches in the ATM network. Therefore an upper limit must be placed on the total number of p2mp group SVC's that can be created. In order to accomplish this, a "VC Count" threshold is used to limit the number of SVC's opened.



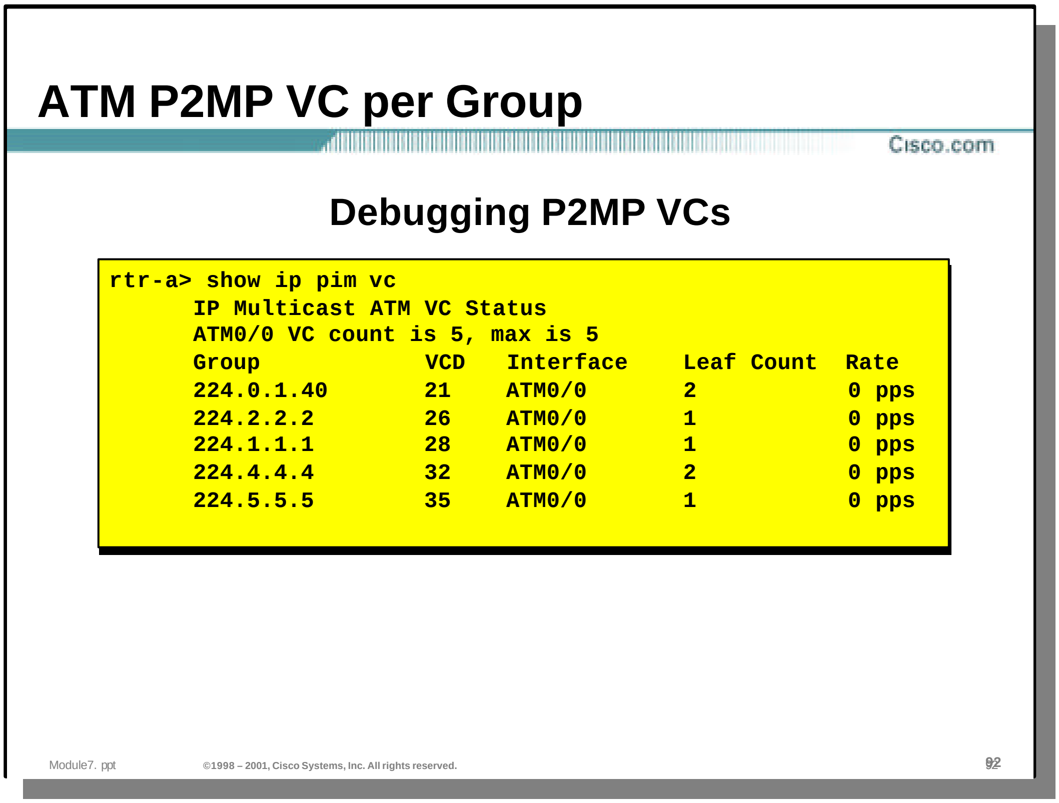
ATM P2MP VC per Group

- **Commands:**

```
ip pim multipoint-signalling
ip pim vc-count <number>
ip pim minimum-vc-rate <pps>
```

- **Good for single LIS which is fully meshed**
  - **Need the shared broadcast p2mp SVC**
    - **otherwise uses Pseudo-Broadcast (ugh!)**

- **[no] ip pim vc-count <number>**
  - Configures the maximum number of p2mp SVCs PIM opens. The default value is 200. When the router hits this maximum limit it will delete inactive p2mp SVCs so it may open other p2mp SVCs for new groups that might have activity.
- **[no] ip pim minimum-vc-rate <pps>**
  - Configures the minimum traffic rate to keep p2mp SVCs active. When the maximum number of p2mp SVCs are opened and a new p2mp SVC needs to be opened, the router will scan existing p2mp SVCs. SVCs that have a current 1 second rate less than or equal to <pps> are eligible for deletion. Ties are broken by group fanout. Higher fanout groups lose and are deleted. (The idea is that high fanout groups can be moved to the broadcast p2mp SVC with a minimum loss of efficiency since these p2mp SVC come closest to mapping to all routers in the LIS.)
  - If a p2mp SVC is deleted, it means that packets for its respective group do not have its own multipoint SVC. However, packets will flow over the shared broadcast/multicast p2mp SVC which delivers packets to all PIM neighbors. If all p2mp SVCs have a 1 minute rate more than <pps>, the new group will use the shared broadcast/multicast p2mp SVC.
  - The default value of '`minimum-vc-rate`' is 0 packets per second.



ATM P2MP VC per Group

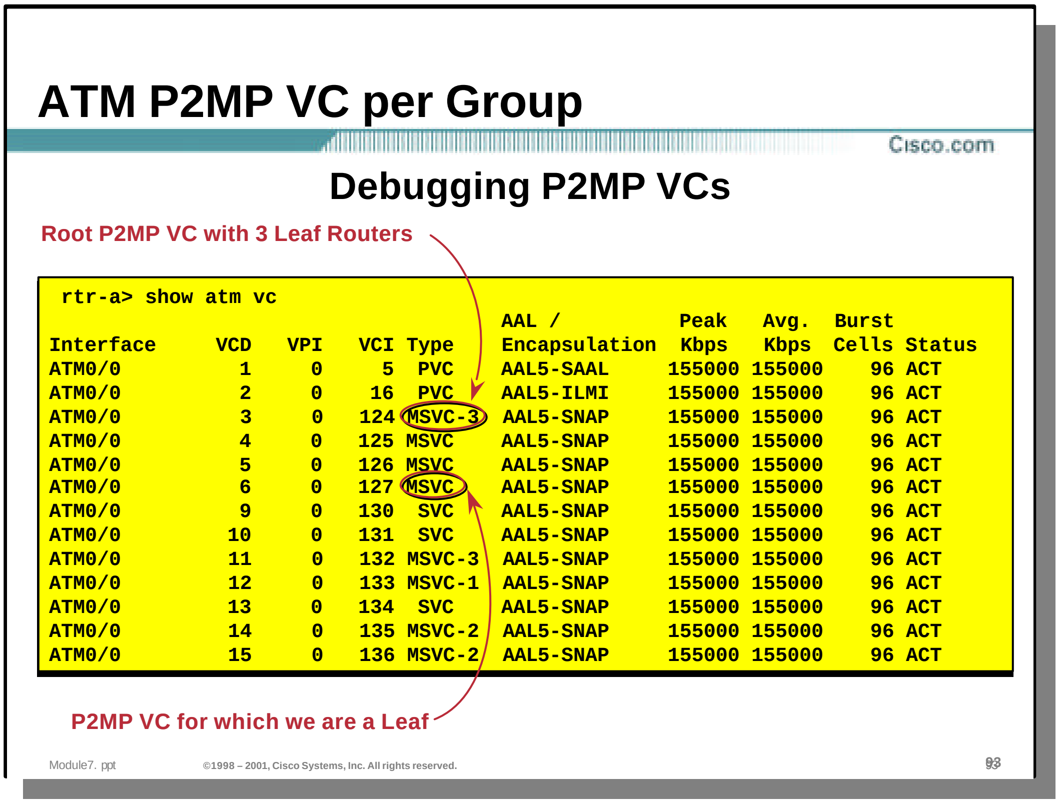
Debugging P2MP VCs

```
rtr-a> show ip pim vc
      IP Multicast ATM VC Status
      ATM0/0 VC count is 5, max is 5
      Group            VCD   Interface    Leaf Count  Rate
      224.0.1.40       21    ATM0/0       2           0 pps
      224.2.2.2        26    ATM0/0       1           0 pps
      224.1.1.1        28    ATM0/0       1           0 pps
      224.4.4.4        32    ATM0/0       2           0 pps
      224.5.5.5        35    ATM0/0       1           0 pps
```

- **Debugging P2MP SVC’s**

  `show ip pim vc`

  - Displays ATM VC status information for multipoint VCs opened by PIM. When <group-or-name> is specified, only the single group is displayed. When <interface> is specified, only the single ATM interface is displayed. [11.3]



ATM P2MP VC per Group

Debugging P2MP VCs

Root P2MP VC with 3 Leaf Routers

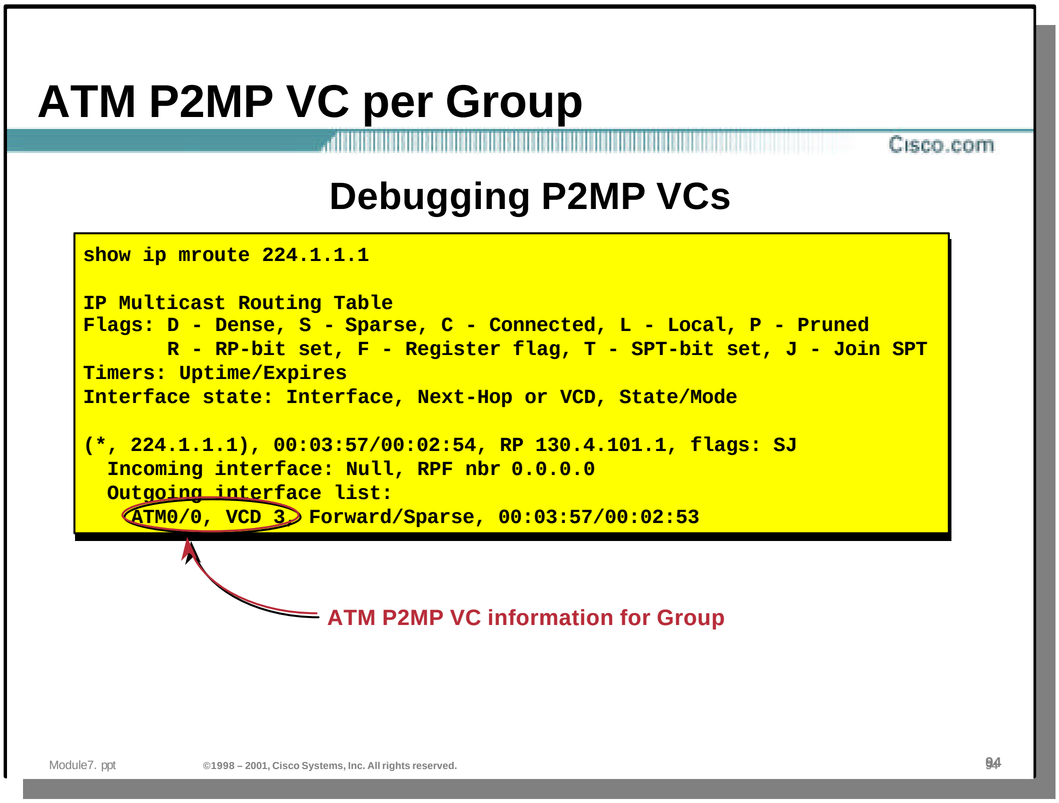
```
 rtr-a> show atm vc
                                      AAL /          Peak   Avg.  Burst
Interface     VCD   VPI   VCI Type    Encapsulation  Kbps   Kbps  Cells Status
ATM0/0          1     0     5  PVC    AAL5-SAAL     155000 155000    96 ACT
ATM0/0          2     0    16  PVC    AAL5-ILMI     155000 155000    96 ACT
ATM0/0          3     0   124 MSVC-3  AAL5-SNAP     155000 155000    96 ACT
ATM0/0          4     0   125 MSVC    AAL5-SNAP     155000 155000    96 ACT
ATM0/0          5     0   126 MSVC    AAL5-SNAP     155000 155000    96 ACT
ATM0/0          6     0   127 MSVC    AAL5-SNAP     155000 155000    96 ACT
ATM0/0          9     0   130  SVC    AAL5-SNAP     155000 155000    96 ACT
ATM0/0         10     0   131  SVC    AAL5-SNAP     155000 155000    96 ACT
ATM0/0         11     0   132 MSVC-3  AAL5-SNAP     155000 155000    96 ACT
ATM0/0         12     0   133 MSVC-1  AAL5-SNAP     155000 155000    96 ACT
ATM0/0         13     0   134  SVC    AAL5-SNAP     155000 155000    96 ACT
ATM0/0         14     0   135 MSVC-2  AAL5-SNAP     155000 155000    96 ACT
ATM0/0         15     0   136 MSVC-2  AAL5-SNAP     155000 155000    96 ACT
```

P2MP VC for which we are a Leaf

- **Debugging P2MP SVC’s**

  `show atm vc`

  - Displays ATM VC status information for all open VC’s. Notice the “MSVC-*n”* entries in the “Type” field. This indicates that the VC is a p2mp SVC and for which this router is the root. The number following the dash indicates the number of nodes on the p2mp SVC. Entries of “MSVC” without the dash indicate p2mp SVC’s for which this router is a member (i.e. some other router is the root of the p2mp SVC.)



ATM P2MP VC per Group

Cisco.com

Debugging P2MP VCs

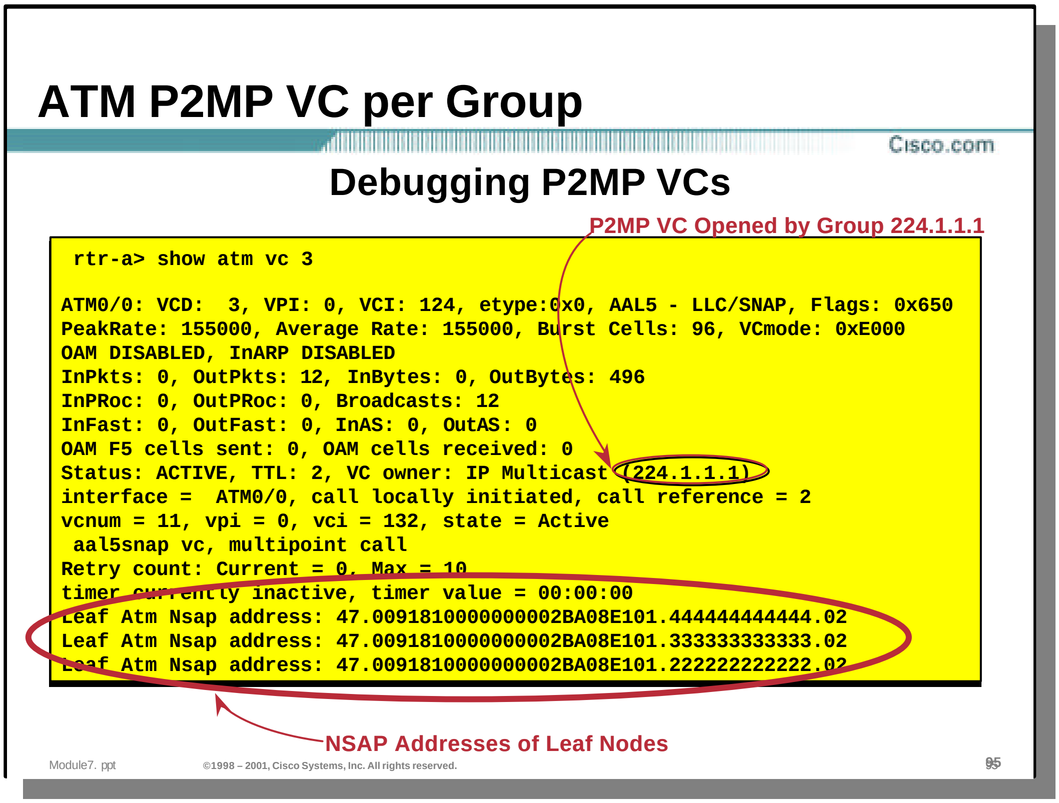
```
show ip mroute 224.1.1.1

IP Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 00:03:57/00:02:54, RP 130.4.101.1, flags: SJ
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    ATM0/0, VCD 3, Forward/Sparse, 00:03:57/00:02:53
```

ATM P2MP VC information for Group

- **Debugging P2MP SVC’s**

  `show ip mroute`

  - The information displayed in the outgoing interface list when p2mp SVC’s are in use is modified to reflect not only the interface but the Virtual Circuit Descriptor (VCD) number of the p2mp SVC as well.



ATM P2MP VC per Group

Debugging P2MP VCs

P2MP VC Opened by Group 224.1.1.1

```
 rtr-a> show atm vc 3

ATM0/0: VCD:  3, VPI: 0, VCI: 124, etype:0x0, AAL5 - LLC/SNAP, Flags: 0x650
PeakRate: 155000, Average Rate: 155000, Burst Cells: 96, VCmode: 0xE000
OAM DISABLED, InARP DISABLED
InPkts: 0, OutPkts: 12, InBytes: 0, OutBytes: 496
InPRoc: 0, OutPRoc: 0, Broadcasts: 12
InFast: 0, OutFast: 0, InAS: 0, OutAS: 0
OAM F5 cells sent: 0, OAM cells received: 0
Status: ACTIVE, TTL: 2, VC owner: IP Multicast (224.1.1.1)
interface =  ATM0/0, call locally initiated, call reference = 2
vcnum = 11, vpi = 0, vci = 132, state = Active
 aal5snap vc, multipoint call
Retry count: Current = 0, Max = 10
timer currently inactive, timer value = 00:00:00
Leaf Atm Nsap address: 47.0091810000000002BA08E101.444444444444.02
Leaf Atm Nsap address: 47.0091810000000002BA08E101.333333333333.02
Leaf Atm Nsap address: 47.0091810000000002BA08E101.222222222222.02
```

NSAP Addresses of Leaf Nodes

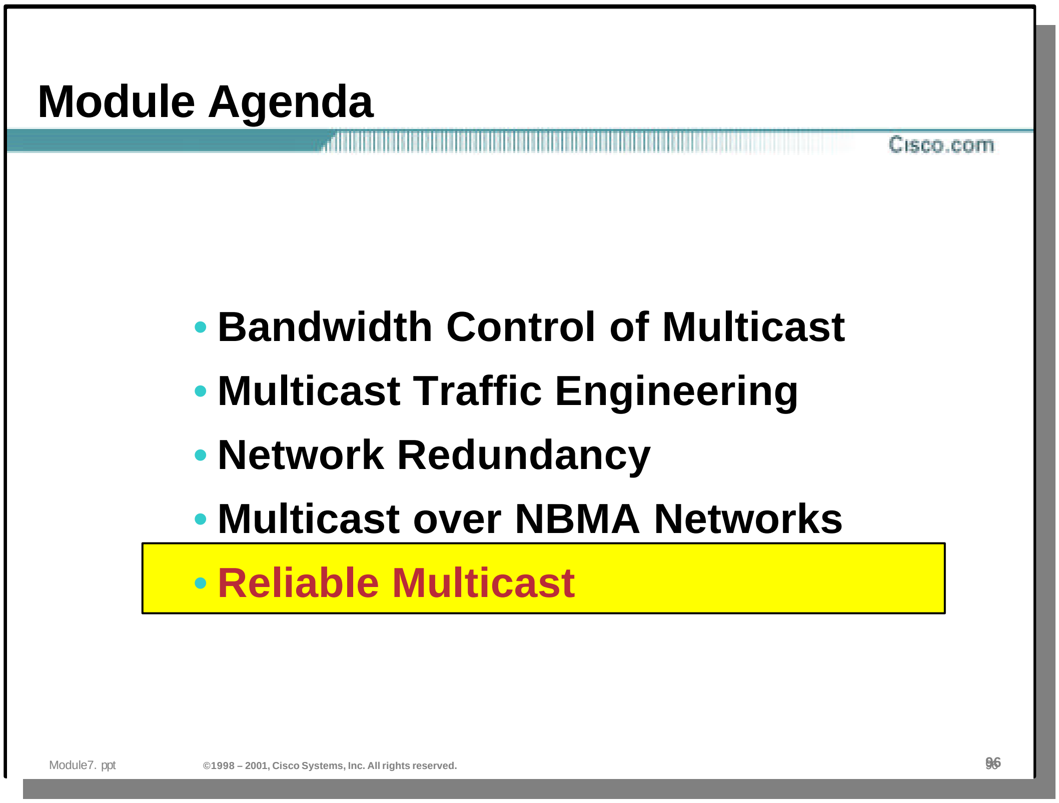
Module7. ppt

95

- **Debugging P2MP SVC's**

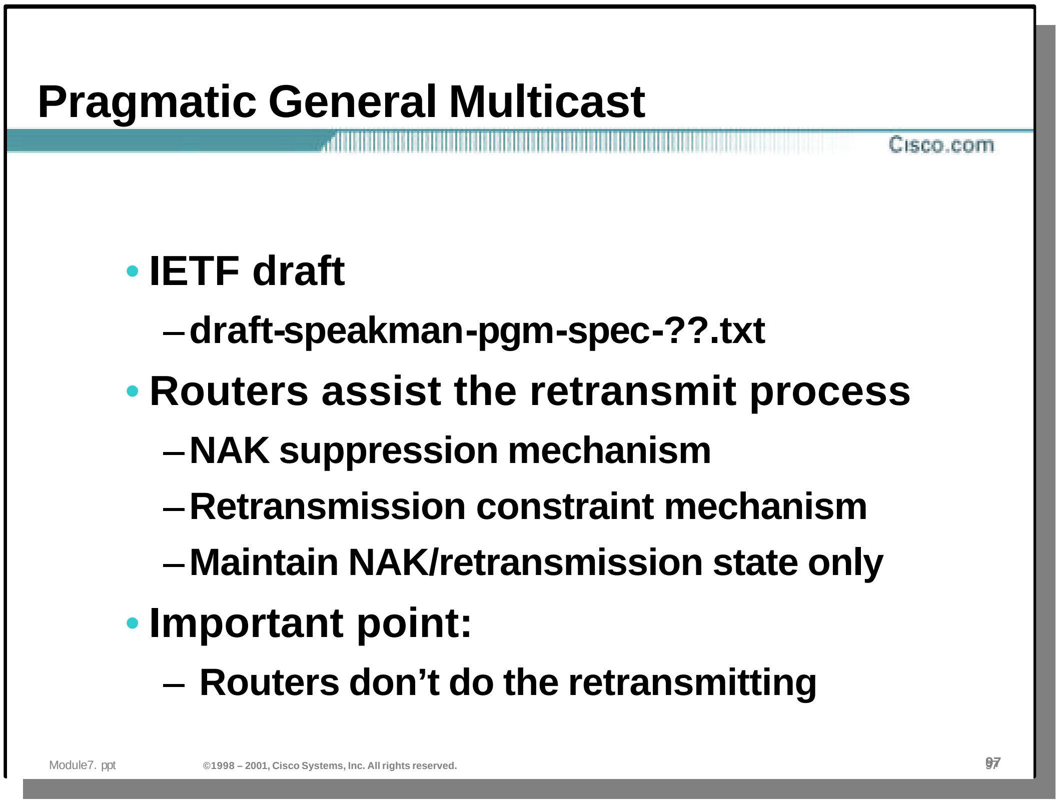
  `show atm vc <vcd>`

  - The output of this command reflects the IP Multicast group responsible for the p2mp SVC being created. In addition, a list of all of the other nodes on the p2mp SVC (listed by NSAP address) is also displayed.



Module Agenda

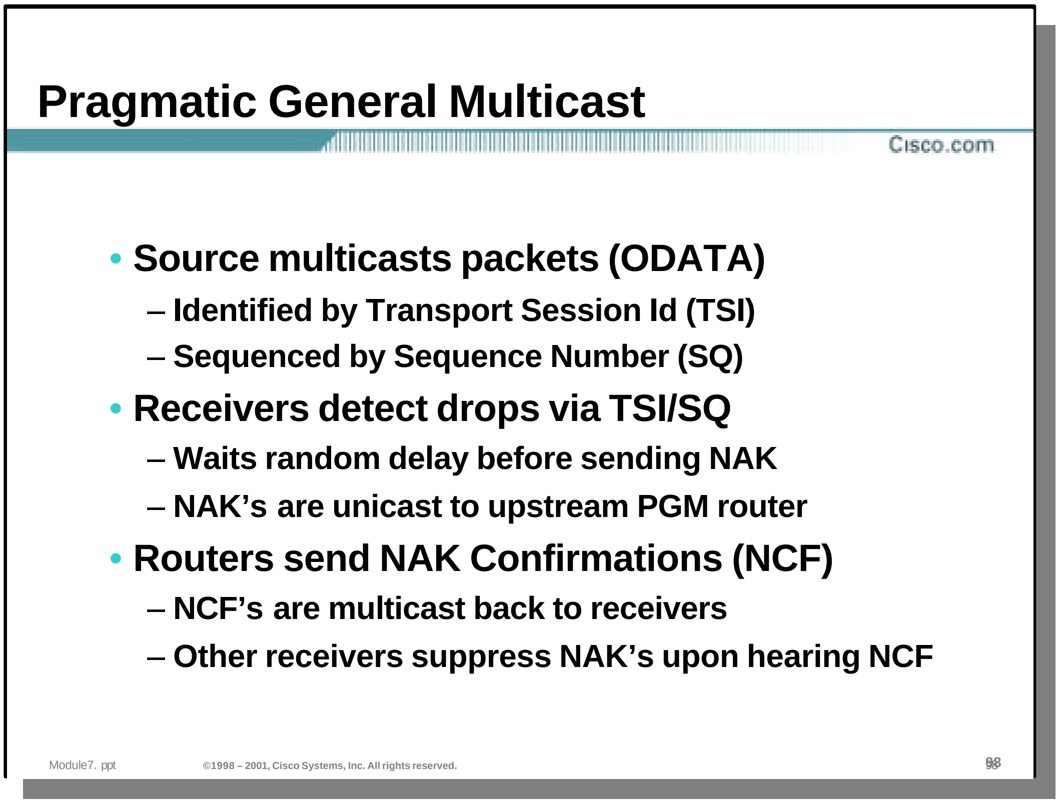
- Bandwidth Control of Multicast
- Multicast Traffic Engineering
- Network Redundancy
- Multicast over NBMA Networks
- **Reliable Multicast**



Pragmatic General Multicast

- IETF draft
  - draft-speakman-pgm-spec-??.txt
- Routers assist the retransmit process
  - NAK suppression mechanism
  - Retransmission constraint mechanism
  - Maintain NAK/retransmission state only
- Important point:
  - Routers don’t do the retransmitting

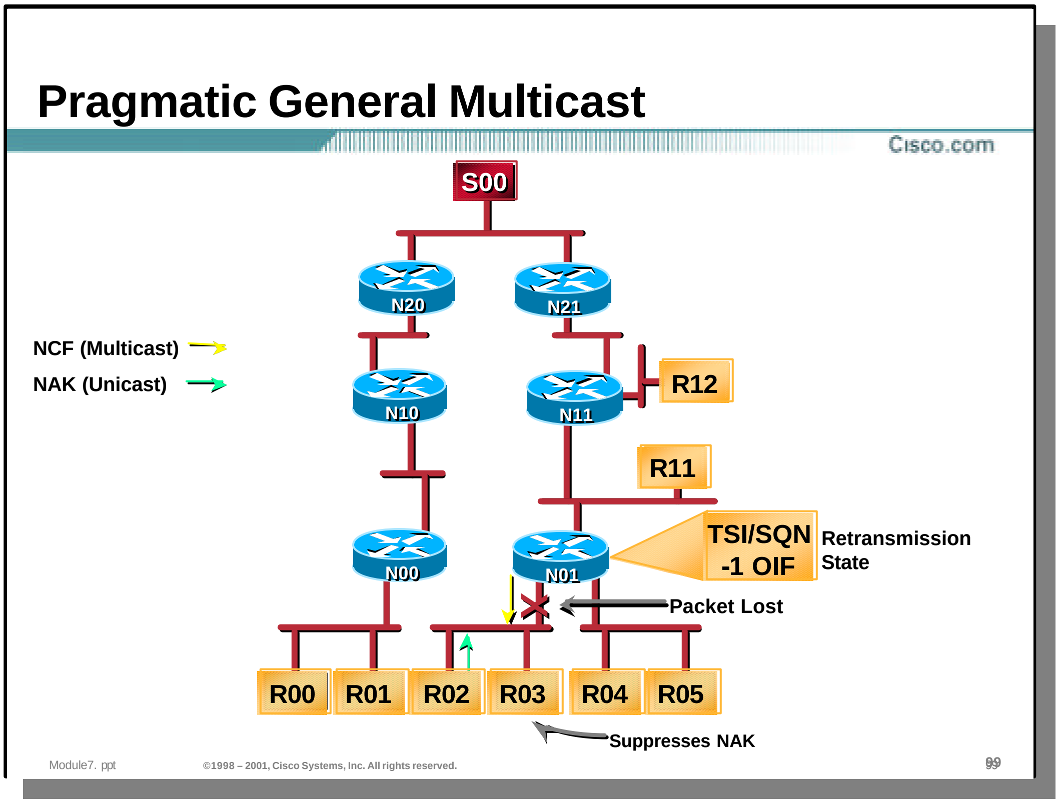
- **Pragmatic General Multicast**
  - PGM is a reliable multicast transport protocol for applications that require ordered duplicate-free, multicast data delivery from multiple sources to multiple receivers. PGM guarantees that a receiver in a multicast group either receives all data packets from transmissions and retransmissions, or can detect unrecoverable data packet loss. PGM is intended as a solution for multicast applications with basic reliability requirements. It is network layer-independent; the Cisco implementation of PGM Router Assist supports PGM over IP.
  - PGM Router Assist feature allows Cisco routers to support optimal operation of Pragmatic General Multicast (PGM). The PGM Reliable Transport Protocol itself is implemented on the source and receiver hosts.
    - PGM uses a NAK suppresion mechanism so that typically only one host sends a NAK for a particular lost packet.
    - Routers acknowledge the NAK by multicasting a NAK-Conf (NCF) message and by instantiating retransmission state. This state is later used to forward the retransmitted packet to only those portions of the network where the packet was reported as lost.
  - Important Point:
    - Routers do NOT do the retransmitting. Retransmission is normally done by the source.
  - The benefits of the PGM Router Assist are:
    - It saves bandwidth: The PGM Router Assist feature saves bandwidth by substantially reducing the number of negative acknowledgments (NAKs) to the source and by constraining the retransmissions to only those receivers that experience data loss.
    - It improves PGM Efficiency: The PGM Router Assist feature is not absolutely required for hosts that implement PGM, but PGM operates optimally in conjunction with routers that have this feature enabled.



Pragmatic General Multicast

- **Source multicasts packets (ODATA)**
  - **Identified by Transport Session Id (TSI)**
  - **Sequenced by Sequence Number (SQ)**
- **Receivers detect drops via TSI/SQ**
  - **Waits random delay before sending NAK**
  - **NAK's are unicast to upstream PGM router**
- **Routers send NAK Confirmations (NCF)**
  - **NCF's are multicast back to receivers**
  - **Other receivers suppress NAK's upon hearing NCF**

- **Pragmatic General Multicast**
  - PGM supports any number of sources within a multicast group, each fully identified by a globally unique Transport Session Identifier (TSI). Sequence numbers (SQN's) identifies packets. Thus the combination of TSI and SQN uniquely identifies a packet that must be retransmitted.
  - In the normal course of data transfer, a source multicasts sequenced data packets (ODATA), and receivers unicast selective negative acknowledgements (NAKs) for data packets detected to be missing from the expected sequence. Network elements (Cisco routers) forward NAKs PGM-hop-by-PGM-hop to the source, and confirm each hop by multicasting a NAK confirmation (NCF) in response on the interface on which the NAK was received.
  - Retransmissions (RDATA) may be provided either by the source itself or by a Designated Local Repairer (DLR) in response to a NAK, or by another receiver in response to an NCF.
  - NAKs provide the sole mechanism for reliability.
  - NAKs are sent continuously until the receiver doesn't get NCF. When router receives NAK, NAK confirmation (NCF) is sent on an interface from which NAK was received. NCF is sent using multicast so that other receivers can see it and suppress their NAKs.
  - NCFs are not propagated by PGM enabled routers.



Pragmatic General Multicast

NCF (Multicast)

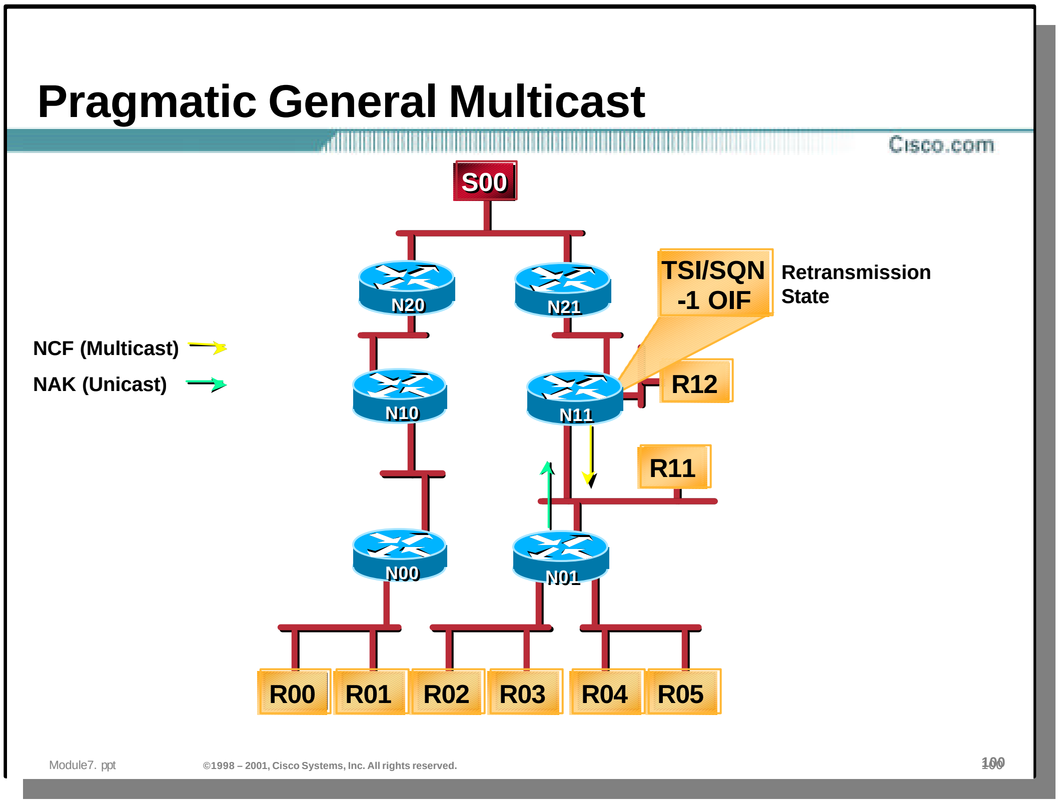
NAK (Unicast)

Retransmission State

Packet Lost

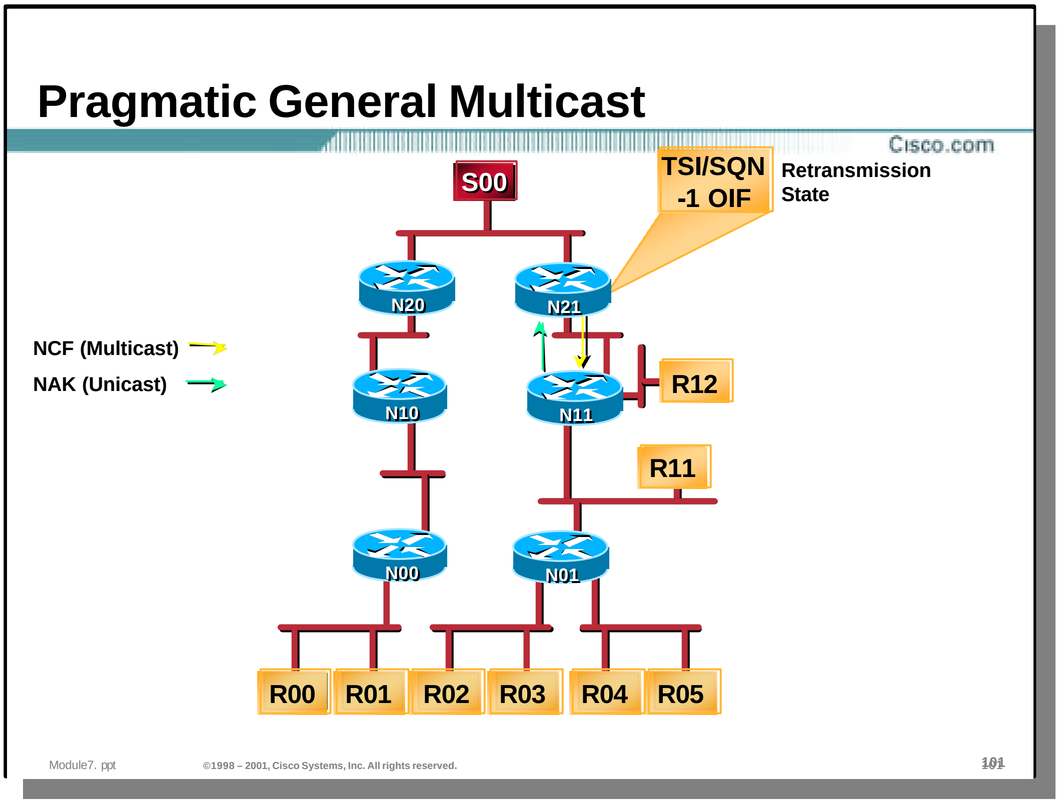
Suppresses NAK

- **PGM Example**
  - When a retransmission is needed to make up for a lost packet, a sequence of events occurs. This sequence is collectively depicted by this slide and the following three slides.
  - Upon detection of a missing data packet (error), a receiver repeatedly unicasts a NAK to the last-hop PGM router on the distribution tree from the source. A receiver repeats this NAK until it receives a NAK confirmation (NCF) multicast to the group from that PGM router. That router responds with an NCF to the first occurrence of the NAK and any further retransmissions of that same NAK from any receiver.



Pragmatic General Multicast

- **PGM Example**
  - In turn, the router repeatedly forwards the NAK to the upstream PGM router on the reverse of the distribution path from the source of the original data packet until it also receives an NCF from that router.



Pragmatic General Multicast

Cisco.com

S00

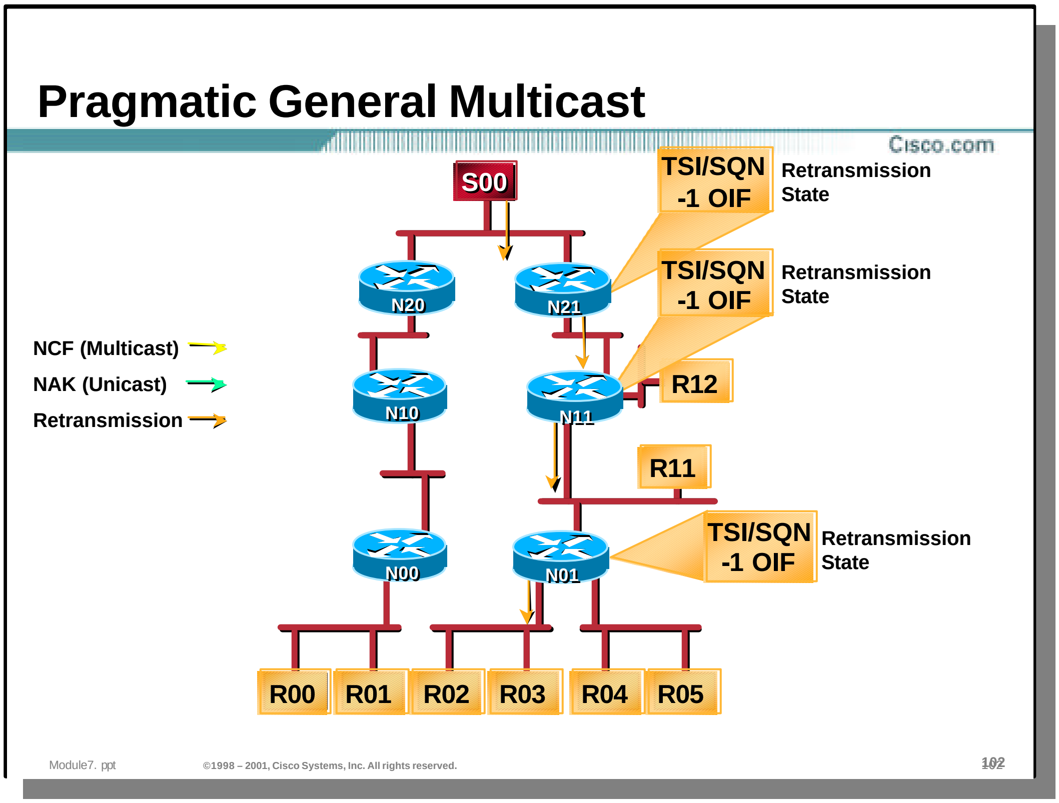
TSI/SQN -1 OIF

Retransmission State

NCF (Multicast)

NAK (Unicast)

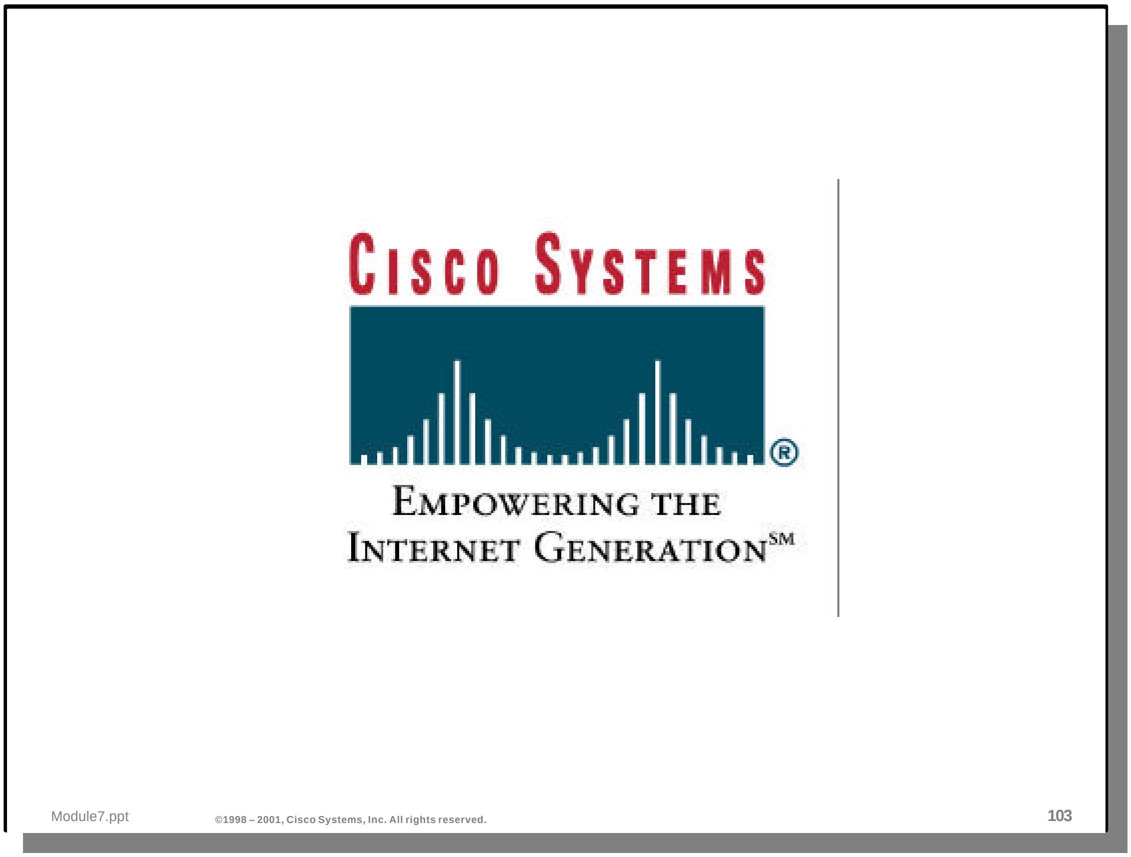
- **PGM Example**
  - – In turn, the router repeatedly forwards the NAK to the upstream PGM router on the reverse of the distribution path from the source of the original data packet until it also receives an NCF from that router. This occurs repeatedly as needed.



Pragmatic General Multicast

Module7. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 102

- **PGM Example**
  - Finally, the source itself receives and confirms the NAK by multicasting a NCF to the group. The source then retransmits the missing data packet to the group address.
  - PGM routers on the way forward the retransmitted packet according to the retransmission state in them - if the retransmission state exists the packet is forwarded to the interfaces via which NAKs were received.



CISCO SYSTEMS
EMPOWERING THE INTERNET GENERATION℠