Multiprotocol BGP (MBGP)

Module 10

 8/14/2001 3:35 PM 1

Copyright ? ?1998-2000, Cisco Systems, Inc. Module10.ppt 1

Module Objectives

- **Understand the basic concepts of BGP**
- **Explain the MBGP extensions to BGP**
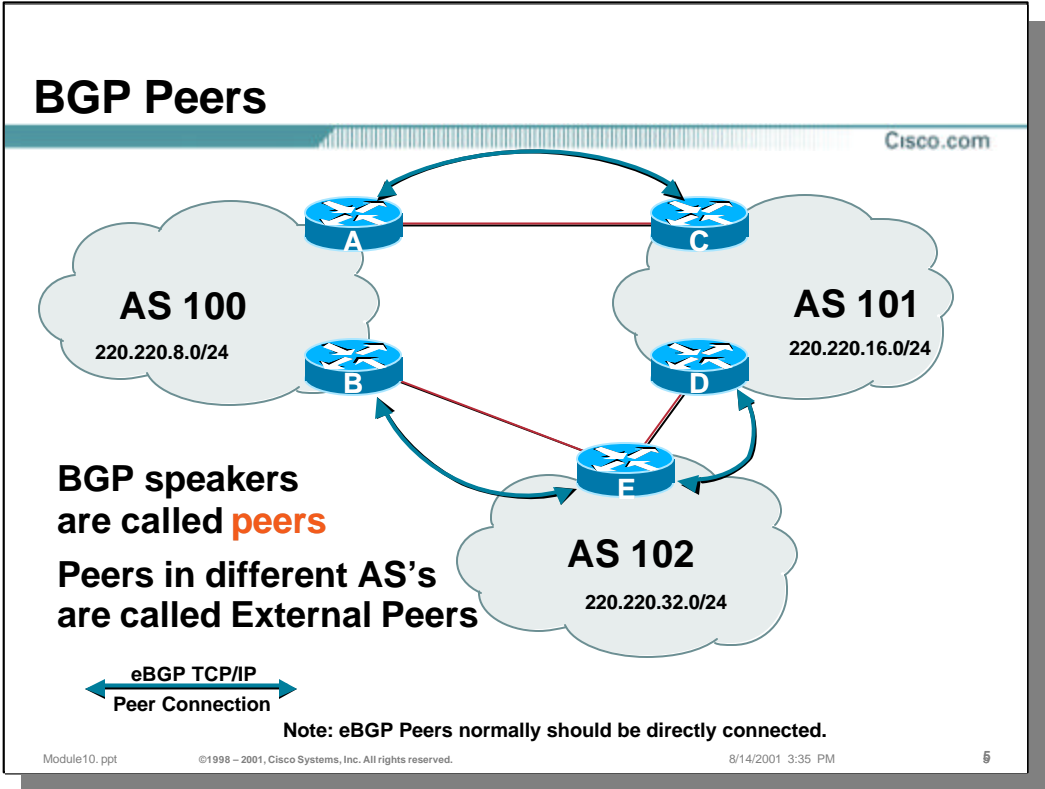- **Identify steps associated with configuring and debugging MBGP**

 8/14/2001 3:35 PM 2

Agenda

- **BGP Review**
- **MBGP Overview**
- **MBGP Update Messages**
- **MBGP Capability Negotiation**
- **MBGP NLRI Information**
- **Advanced MBGP Features**
- **New 12.1 MBGP Syntax**
- **Debugging MBGP**

3

BGP — Border Gateway Protocol

- **Routing Protocol used between AS's**
- **Currently Version 4**
- **Runs over TCP**
- **Path Vector Protocol**
- **Incremental Updates**

 8/14/2001 3:35 PM 4

BGP Peers

AS 100
220.220.8.0/24

AS 101
220.220.16.0/24

AS 102
220.220.32.0/24

BGP speakers
are called peers

Peers in different AS's
are called External Peers

eBGP TCP/IP
Peer Connection

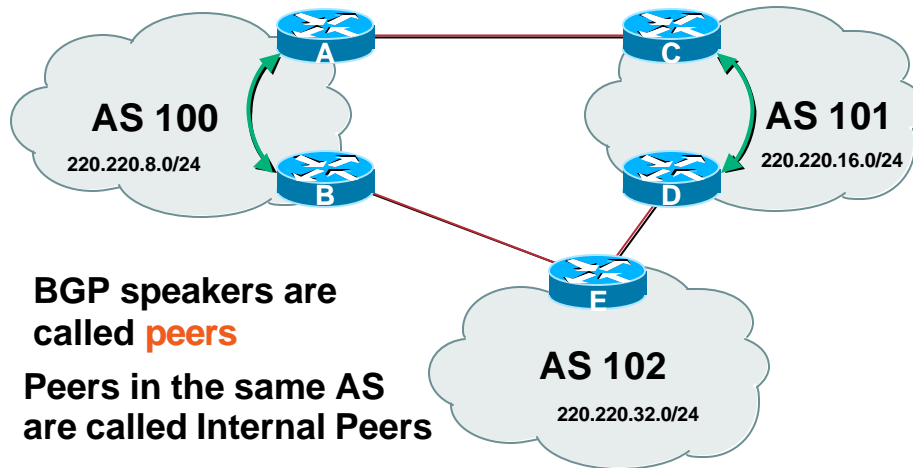Note: eBGP Peers normally should be directly connected.

- **External BGP Peers**
  - An external BGP peering is established when two BGP speakers in different Autonomous Systems (AS) are connected via a TCP/IP connection.
  - External BGP peers are normally directly connected to each other. This simplifies the establishment of the TCP/IP session since the two BGP peers do not have to rely on an IGP or static routing to establish the TCP/IP path.

BGP Peers

AS 100

220.220.8.0/24

AS 101

220.220.16.0/24

AS 102

220.220.32.0/24

BGP speakers are called peers

Peers in the same AS are called Internal Peers

iBGP TCP/IP Peer Connection

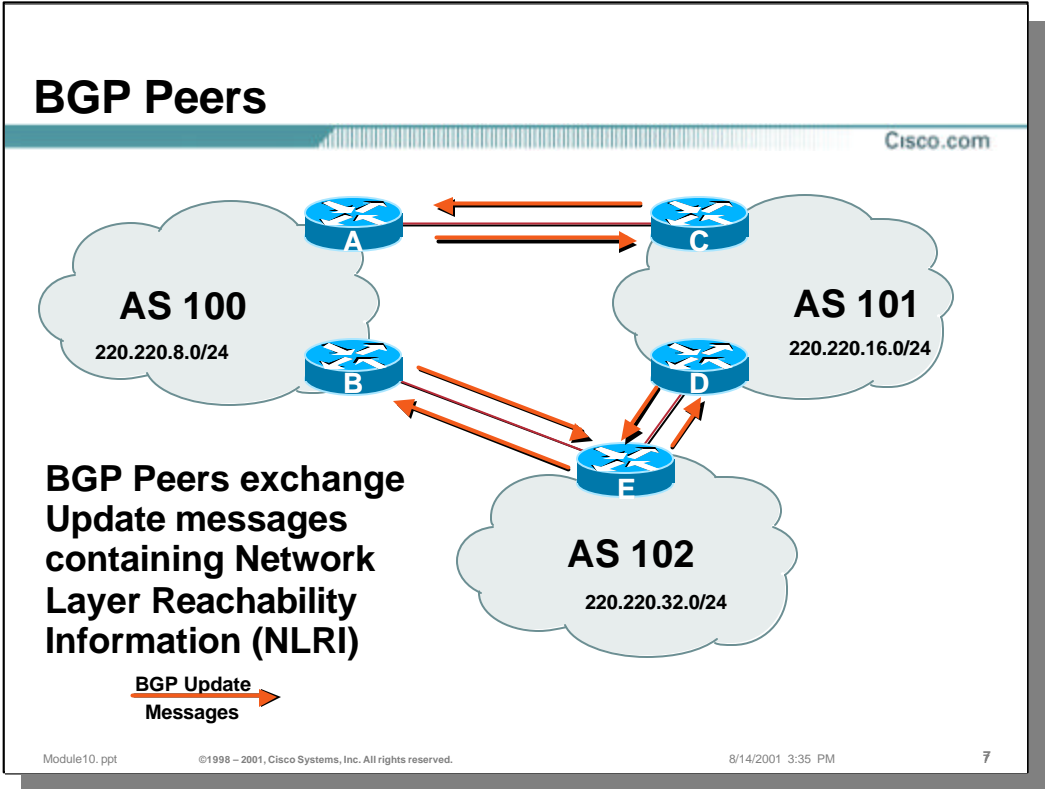**Note: iBGP Peers don't have to be directly connected.**

Module10. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/14/2001 3:35 PM 6
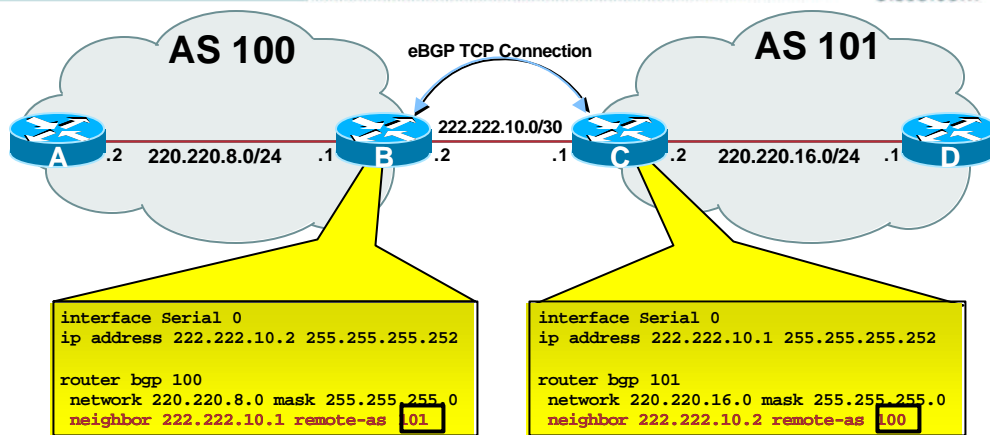
- **Internal BGP Peers**
  - An internal BGP peering is established when two BGP speakers in the same Autonomous Systems (AS) are connected via a TCP/IP connection.
  - Internal BGP peers are normally not directly connected to each other. For example, it is assumed that routers A & B in the above drawing are some distance apart and are not directly connected. This means that IGP routing information must be available in order to establish the TCP/IP session between the two iBGP peers.

BGP Peers

AS 100

220.220.8.0/24

AS 101

220.220.16.0/24

BGP Peers exchange Update messages containing Network Layer Reachability Information (NLRI)

AS 102

220.220.32.0/24

BGP Update Messages

8/14/2001 3:35 PM 7

- **BGP Update Messages**
  - Routing information is exchanged between BGP peers in the form of BGP Update messages which contain Network Layer Reachability Information (NLRI).

Configuring BGP Peers

AS 100

eBGP TCP Connection

AS 101

A   .2   220.220.8.0/24   .1   B   .2   222.222.10.0/30   .1   C   .2   220.220.16.0/24   .1   D

```
interface Serial 0
ip address 222.222.10.2 255.255.255.252

router bgp 100
 network 220.220.8.0 mask 255.255.255.0
 neighbor 222.222.10.1 remote-as 101
```

```
interface Serial 0
ip address 222.222.10.1 255.255.255.252

router bgp 101
 network 220.220.16.0 mask 255.255.255.0
 neighbor 222.222.10.2 remote-as 100
```

- **BGP Peering sessions are established using the BGP "neighbor" configuration command**
  - External (eBGP) is configured when AS numbers are different

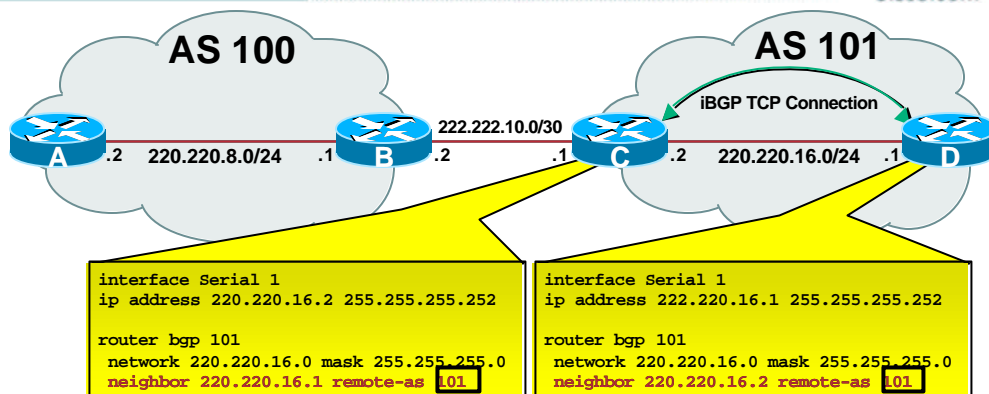 8/14/2001 3:35 PM 8

- **Configuring BGP Peers**
  - The BGP "neighbor" configuration command is used to establish the TCP/IP connection between two BGP peers.
  - For example, the following command instructs Router B to open a TCP/IP session to 222.222.10.1 (Router C):

    ```
    neighbor 222.222.10.1 remote-as 101
    ```

  - Likewise, the following command instructs Router C to open its end of the TCP/IP session to Router B via the 222.222.10.2 IP address

    ```
    neighbor 222.222.10.2 remote-as 100
    ```

  - The BGP peering established by these two commands will be an External BGP session because the AS numbers are different. (i.e. Routers B & C are in different AS.)

Configuring BGP Peers

AS 100

AS 101

iBGP TCP Connection

A | .2 220.220.8.0/24 .1 | B | .2   .1 | C | .2 220.220.16.0/24 .1 | D

222.222.10.0/30

```
interface Serial 1
ip address 220.220.16.2 255.255.255.252

router bgp 101
 network 220.220.16.0 mask 255.255.255.0
 neighbor 220.220.16.1 remote-as 101
```

```
interface Serial 1
ip address 222.220.16.1 255.255.255.252

router bgp 101
 network 220.220.16.0 mask 255.255.255.0
 neighbor 220.220.16.2 remote-as 101
```

- **BGP Peering sessions are established using the BGP "neighbor" configuration command**
  - External (eBGP) is configured when AS numbers are different
  - Internal (iBGP) is configured when AS numbers are same

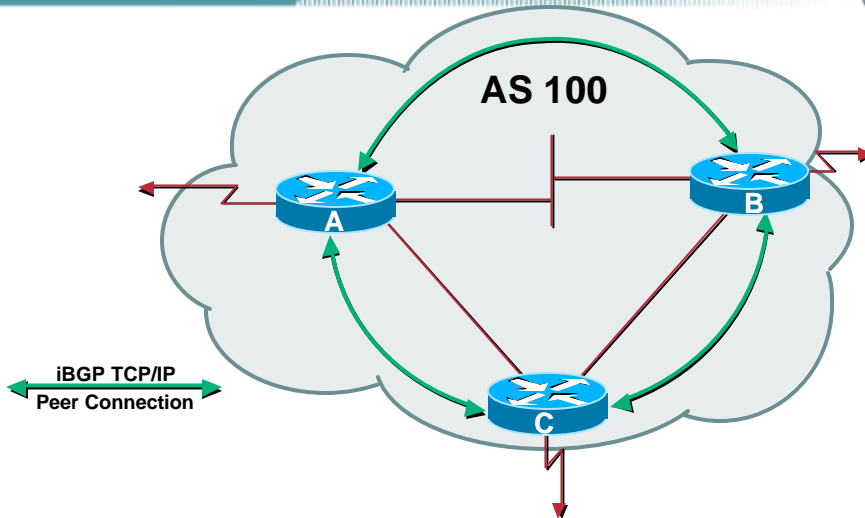 8/14/2001 3:35 PM 9

- **Configuring BGP Peers**
  - In this example, the following command instructs Router C to open a TCP/IP session to 222.222.16.1 (Router D):

    ```
    neighbor 222.222.16.1 remote-as 101
    ```

  - Likewise, the following command instructs Router D to open its end of the TCP/IP session to Router C via the 222.222.16.2 IP address

    ```
    neighbor 222.222.16.2 remote-as 101
    ```

  - The BGP peering established by these two commands will be an Internal BGP session because the AS numbers are different. (i.e. Routers B & C are in the same AS.)

Configuring BGP Peers

AS 100

A

B

C

iBGP TCP/IP
Peer Connection

• **Each iBGP speaker must peer with every other iBGP speaker in the AS**

• Configuring BGP Peers

– Normally, all Internal BGP (iBGP) speakers in an AS must peer with every other iBGP speaker in the AS. In other words, all iBGP speakers in the AS must be fully meshed.

• Note: Special BGP techniques such as Route Reflectors and BGP Confederations may be employed to avoid having to fully mesh all iBGP speakers in an AS. However, the details of these techniques are beyond the scope of this tutorial. Consult other documentation on Deploying BGP.

Configuring BGP Peers

AS 100

215.10.7.1

215.10.7.2

215.10.7.3

iBGP TCP/IP
Peer Connection
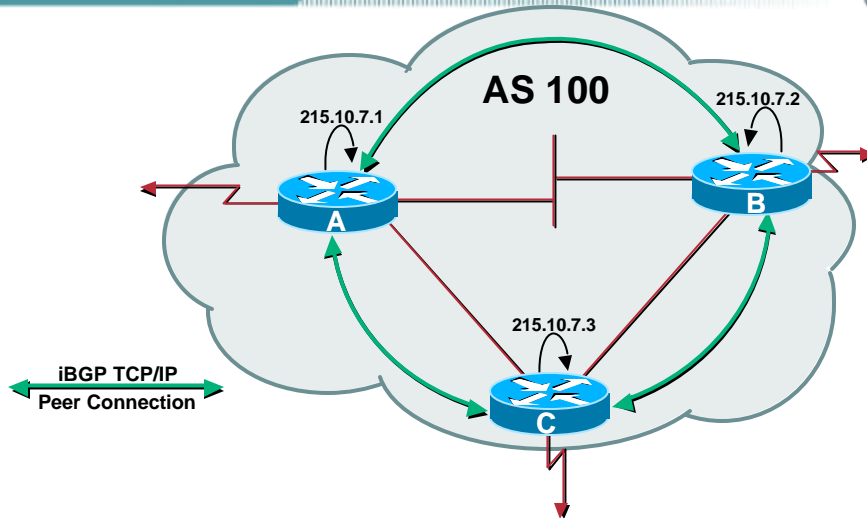
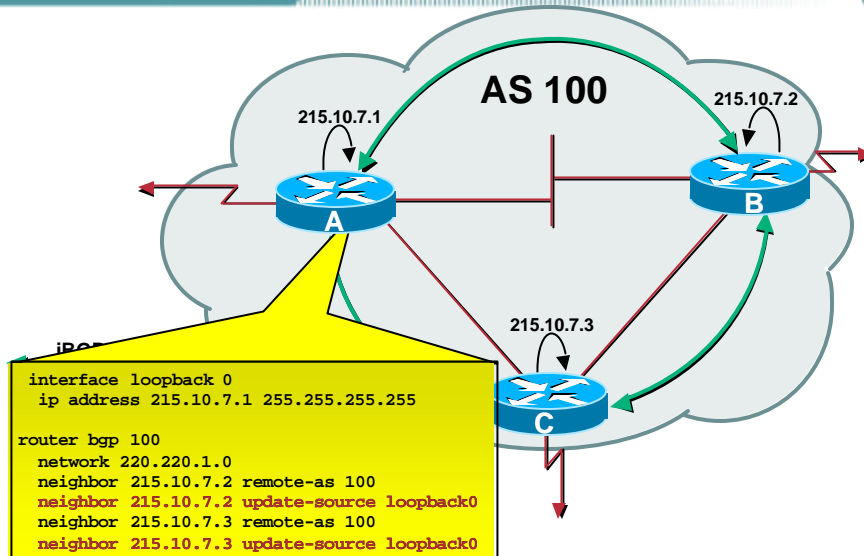• **Loopback interface are normally used as peer connection end-points**

• **Configuring BGP Peers**
  – Loopback interfaces are often used as the peering address in order to avoid BGP peering sessions from being lost when a single physical interface goes down.

Cisco.com

AS 100

215.10.7.1

215.10.7.2

215.10.7.3

A

B

C

iBGP

```
interface loopback 0
 ip address 215.10.7.1 255.255.255.255

router bgp 100
 network 220.220.1.0
 neighbor 215.10.7.2 remote-as 100
 neighbor 215.10.7.2 update-source loopback0
 neighbor 215.10.7.3 remote-as 100
 neighbor 215.10.7.3 update-source loopback0
```

Module10. ppt

8/14/2001 3:35 PM

12

• **Configuring BGP Peers**

– In this example, Router A has been configured to establish iBGP sessions to routers B and C using the address of their Loopback interfaces 215.10.7.2 and 215.10.7.3, respectively. This is accomplished by the following commands:

```
neighbor 215.10.7.2 remote-as 100
neighbor 215.10.7.2 update-source loopback0
neighbor 215.10.7.3 remote-as 100
neighbor 215.10.7.3 update-source loopback0
```

– Notice that the "update-source" commands are used in the above configuration so that router A will make the TCP/IP connections using the address of *its* Loopback interface, 215.10.7.1. If this is not done, the TCP/IP connection will not come up as there will be a mismatch in source/destination addresses.

Configuring BGP Peers

AS 100

215.10.7.1

215.10.7.2

215.10.7.3

iBGP TCP/IP Peer Connection

```
interface loopback 0
  ip address 215.10.7.2 255.255.255.255

router bgp 100
  network 220.220.5.0
  neighbor 215.10.7.1 remote-as 100
  neighbor 215.10.7.1 update-source loopback0
  neighbor 215.10.7.3 remote-as 100
  neighbor 215.10.7.3 update-source loopback0
```

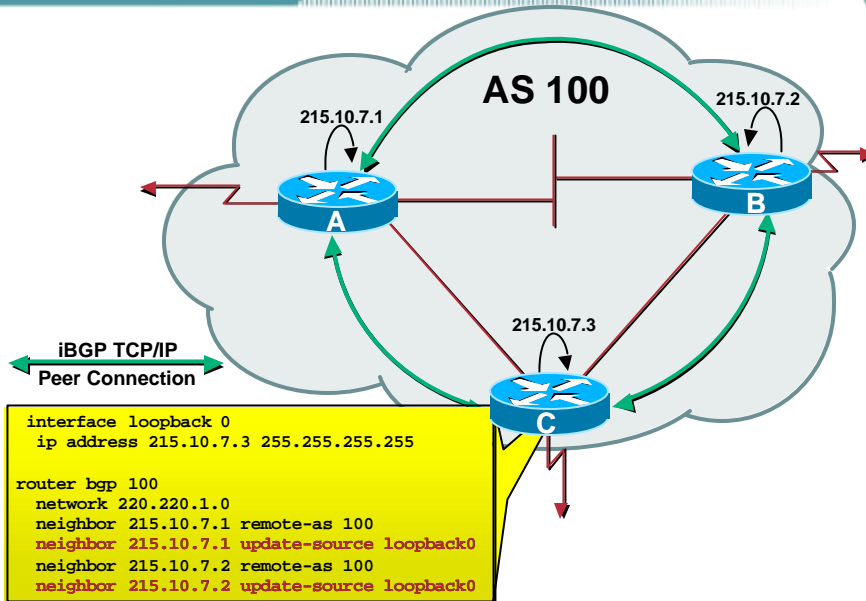 8/14/2001 3:35 PM 13

- **Configuring BGP Peers**
  - In this example, Router B has been configured to establish iBGP sessions to routers A and C using the address of their Loopback interfaces 215.10.7.1 and 215.10.7.3, respectively. This is accomplished by the following commands:

    ```
    neighbor 215.10.7.1 remote-as 100
    neighbor 215.10.7.1 update-source loopback0
    neighbor 215.10.7.3 remote-as 100
    neighbor 215.10.7.3 update-source loopback0
    ```

  - Notice that the "update-source" commands are used in the above configuration so that router B will make the TCP/IP connections using the address of *its* Loopback interface, 215.10.7.2. If this is not done, the TCP/IP connection will not come up as there will be a mismatch in source/destination addresses.

Configuring BGP Peers

AS 100

215.10.7.1

215.10.7.2

215.10.7.3

iBGP TCP/IP
Peer Connection

```
interface loopback 0
  ip address 215.10.7.3 255.255.255.255

router bgp 100
  network 220.220.1.0
  neighbor 215.10.7.1 remote-as 100
  neighbor 215.10.7.1 update-source loopback0
  neighbor 215.10.7.2 remote-as 100
  neighbor 215.10.7.2 update-source loopback0
```
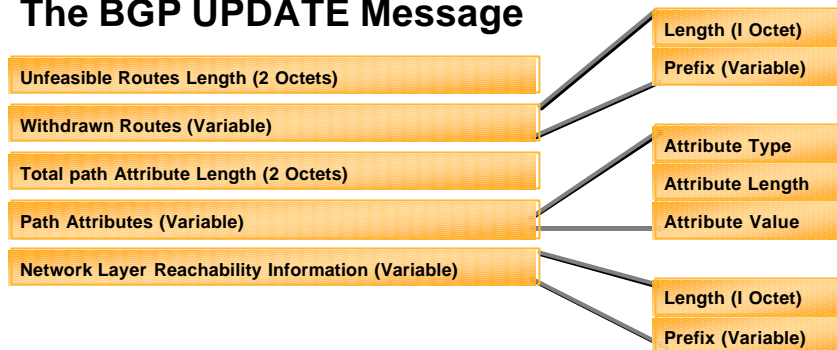
 8/14/2001 3:35 PM 14

- **Configuring BGP Peers**
  - In this final example, Router C has been configured to establish iBGP sessions to routers A and B using the address of their Loopback interfaces 215.10.7.1 and 215.10.7.2, respectively. This is accomplished by the following commands:

    ```
    neighbor 215.10.7.1 remote-as 100
    neighbor 215.10.7.1 update-source loopback0
    neighbor 215.10.7.2 remote-as 100
    neighbor 215.10.7.2 update-source loopback0
    ```

  - Notice that the "update-source" commands are used in the above configuration so that router C will make the TCP/IP connections using the address of *its* Loopback interface, 215.10.7.3. If this is not done, the TCP/IP connection will not come up as there will be a mismatch in source/destination addresses.

BGP Update Messages

The BGP UPDATE Message

| Unfeasible Routes Length (2 Octets) |
| Withdrawn Routes (Variable) |
| Total path Attribute Length (2 Octets) |
| Path Attributes (Variable) |
| Network Layer Reachability Information (Variable) |

- Length (I Octet)
- Prefix (Variable)

- Attribute Type
- Attribute Length
- Attribute Value

- Length (I Octet)
- Prefix (Variable)

- **A BGP update is used to advertise a single feasible route to a peer, or to withdraw multiple unfeasible routes**

- **Each update message contains attributes, like origin, AS-Path, Next-Hop, ……..**

- **BGP Update Messages**
    - The drawing above shows the basic format of the BGP Update Message. These messages are used to advertise a single feasible route to a peer as well as to optionally withdraw one or more unfeasible routes.
        - The first section of the Update Message is the Withdrawn Routes. This is a variable length field whose length is specified in the first two bytes. Each withdrawn route consists of a length and the prefix of the route being withdrawn.
        - The second section of the message is the Path Attributes section. This is also a variable length section whose length is specified in the first two bytes of the section. Each attribute is encoded in a TLV (Type, Length, Value) format that specifies certain attributes that are attributed to the Network Layer Reachability Information (NLRI) contained in the next section.
        - The last section is the Network Layer Reachability Information (NLRI) field which consists of a length and prefix of the advertised route.

BGP Updates — NLRI

- **Network Layer Reachability Information**
- **Used to advertise feasible routes**
- **Composed of:**
  - **Network Prefix**
  - **Mask Length**

 8/14/2001 3:35 PM 16

- **Network Layer Reachability Information (NLRI)**
  - This information is the heart and soul of the BGP routing protocol. NLRI are feasible routes that are exchanged between BGP speakers.
  - NLRI is composed of:
    - Mask Length (or the length of the Prefix)
    - Network Prefix

BGP Updates — Attributes

- **Used to convey information associated with NLRI**
  - **AS path**
  - **Next hop**
  - Local preference
  - Multi-Exit Discriminator (MED)
  - Community
  - Origin
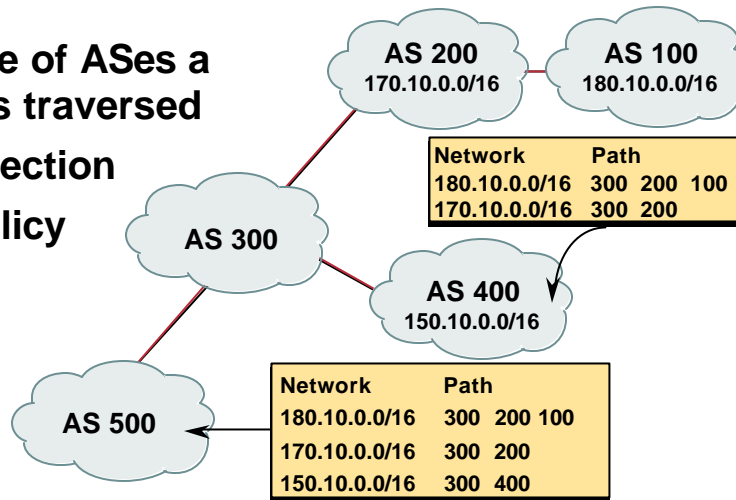 - Aggregator

 8/14/2001 3:35 PM 17

- **BGP Attributes**
  - Attributes are associated with the NLRI (i.e. the route) being advertised. These attributes include (but are not limited to) the following:
    - AS path
    - Next-Hop
    - Local Preference
    - Multi-Exit Discriminator (MED)
    - Community
    - Origin
    - Aggregator
  - The discussion of all of these attributes is beyond the scope of this tutorial. However, the "AS-Path" and "Next-Hop" attributes are discussed in the following sections since they are fundamental to the basic operation of BGP.

AS-Path Attribute

- **Sequence of ASes a route has traversed**
- **Loop detection**
- **Apply policy**

AS 200
170.10.0.0/16

AS 100
180.10.0.0/16

Network	Path
180.10.0.0/16	300 200 100
170.10.0.0/16	300 200

AS 300

AS 400
150.10.0.0/16

AS 500

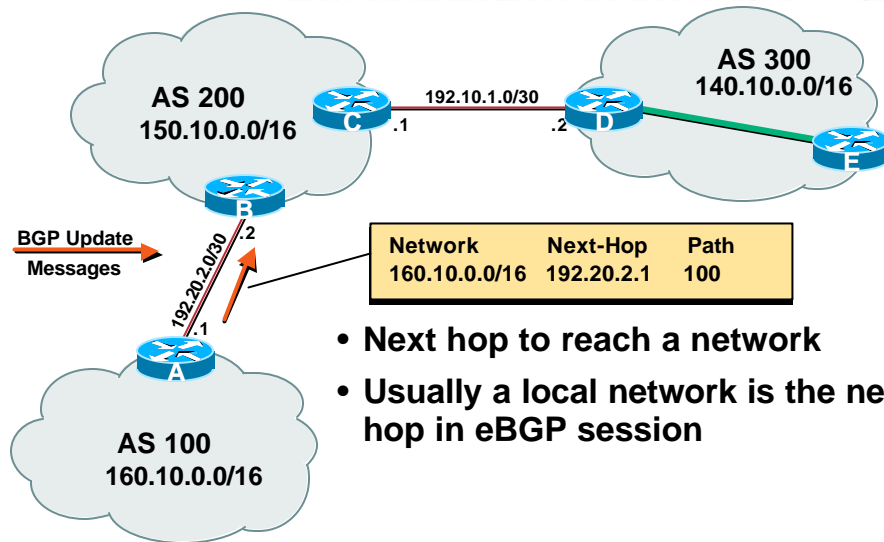| Network	Path
180.10.0.0/16	300 200 100
170.10.0.0/16	300 200
150.10.0.0/16	300 400

- **AS-Path Attribute**
  - This attribute describes the sequence of AS numbers that must be traversed to reach the network whose prefix is advertised in the NLRI field of the Update message. As each eBGP speaker in the network forwards this Update message on to its eBGP neighbors, the local AS number is prepended to this AS-Path list.
  - In the above example, network 180.10.0.0/16 resides inside of AS100. Notice that after this network prefix is reaches AS 500, the AS-Path for network 180.10.0.0/16 is "300 200 100". This means that traffic destined for this network must travel to AS 300, then on to AS 200 and finally AS 100 where network 180.10.0.0 resides.
  - The same thing occurs for networks 170.10.0.0/16 and 150.10.0.0/16. An Update messages are originated by AS 200 and AS 400, respectively. When these Update messages reach AS 500, a separate entry is maintained for each network along with its unique AS-Path.

Next Hop Attribute

AS 300
140.10.0.0/16

192.10.1.0/30

AS 200
150.10.0.0/16

C .1

.2 D

E

B
.2

BGP Update
Messages

192.20.2.0/30

Network	Next-Hop	Path
160.10.0.0/16	192.20.2.1	100

.1
A

AS 100
160.10.0.0/16

- **Next hop to reach a network**
- **Usually a local network is the next hop in eBGP session**
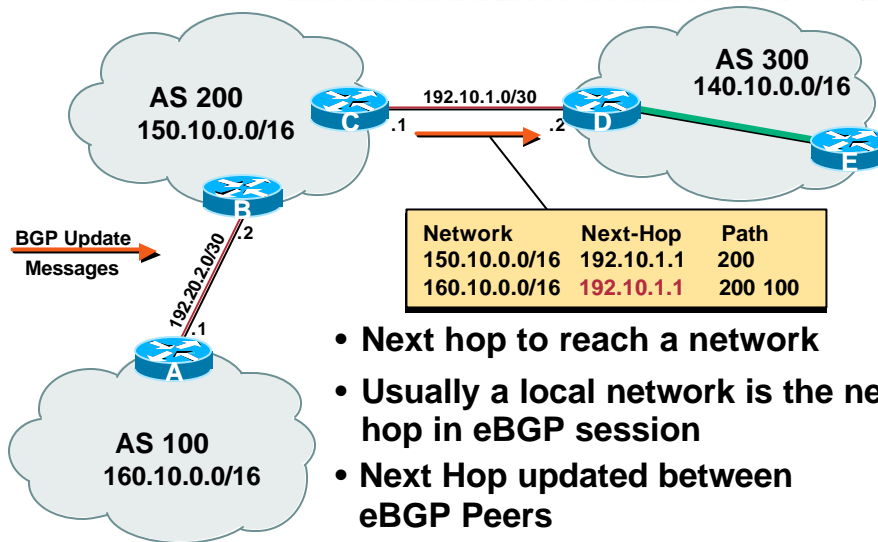
 8/14/2001 3:35 PM 19

- **Next-Hop Attribute**
    - The Next-Hop attribute contains the IP address of the next-hop router to which traffic destined for the network specified in the NLRI is to be sent. This is normally a directly connected network in the case of eBGP sessions.
    - In the above example, network 160.10.0.0/16 resides in AS 100. Router A originates an Update message containing this network in the NLRI and sends this to Router B as shown. The "Next-Hop" attribute in the Update message contains the the IP address of Router A's serial port, namely 192.20.2.1. This information instructs Router B that traffic for network 160.10.0.0/16 should be sent to 192.20.2.1 (Router A) for forwarding on to the destination.

Next Hop Attribute

AS 300
140.10.0.0/16

AS 200
150.10.0.0/16

192.10.1.0/30

C .1 → .2 D

E

B .2

BGP Update
Messages →

192.20.2.0/30 .1

A

AS 100
160.10.0.0/16

Network	Next-Hop	Path
150.10.0.0/16	192.10.1.1	200
160.10.0.0/16	192.10.1.1	200 100

- **Next hop to reach a network**
- **Usually a local network is the next hop in eBGP session**
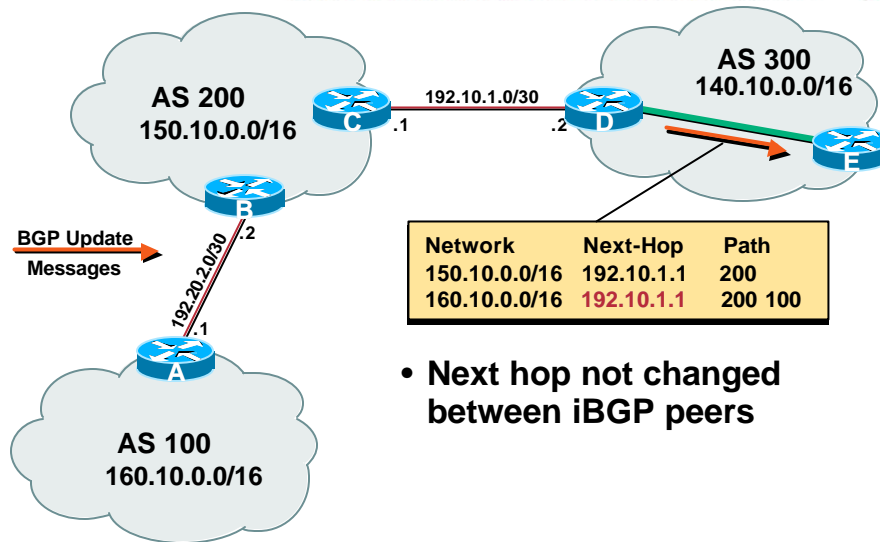- **Next Hop updated between eBGP Peers**

8/14/2001 3:35 PM

- **Next-Hop Attribute**
  – The Next-Hop attribute is normally updated with the local IP address of the eBGP router when an Update message is forwarded to another eBGP peer.
  – In the above example, the Update for network 160.10.0.0/16 is being forwarded by Router C to its eBGP peer Router D. Notice that the "Next-Hop" attribute in the Update message has been updated to contain the the IP address of Router C's serial port, namely 192.10.1.1. This information instructs Router D that traffic for network 160.10.0.0/16 should be sent to 192.10.1.1 (Router C) for forwarding on to the destination.

Cisco.com

AS 300
140.10.0.0/16

AS 200
150.10.0.0/16

192.10.1.0/30

.1   .2   D

C   E

B

.2

BGP Update
Messages

192.20.2.0/30

.1

A

Network	Next-Hop	Path
150.10.0.0/16	192.10.1.1	200
160.10.0.0/16	192.10.1.1	200 100

AS 100
160.10.0.0/16

- **Next hop not changed between iBGP peers**

- **Next-Hop Attribute**
    - The Next-Hop attribute is *not* updated when the Update message is being sent to an iBGP peer.
    - In the above example, the Update for network 160.10.0.0/16 is being forwarded by Router D to its iBGP peer Router E. Notice that the "Next-Hop" attribute for network 160.10.0.0/16 has *not* been updated and still contains the the IP address of Router C's serial port, namely 192.10.1.1. This means that the IGP running in AS 300 must contain routing information for 192.10.1.1 so that Router E can resolve how to forward the traffic for network 160.10.0.0/16 across AS 300 to Router D.
        - Note: The requirement of carrying this Next-Hop information through the IGP (in this case a route to 192.10.1.1) can be eliminated by the use of the "next-hop-self" command at Router D. This forces Router D to update the Next-Hop attribute with its own IP address when sending the Update to its iBGP neighbor, Router E.

Next Hop Attribute (more)

- **IGP should carry route to next hops**
- **Recursive route look-up**
- **Unlinks BGP from actual physical topology**
- **Allows IGP to make intelligent forwarding decision**

- **Next-Hop Attribute**
  - In general, the IGP should carry a route to the Next-Hop address as these addresses are often outside the address space in the IGP.
    - iBGP speakers must perform a recursive route lookup to resolve the BGP Next-Hop information to a local IGP next-hop. (In other words, the iBGP router must determine the internal network next hop in the direction of iBGP speaker on the other side of the AS that advertised the network.
    - This uncouples BGP from the actual physical topology of the network inside of the AS. As long as the IGP can find a path through the network to reach the Next-Hop address, then transient traffic can be routed across the AS to the exit-point iBGP router.
    - This also permits the IGP to make intelligent forwarding decisions based on the internal metrics set in the local network.

BGP Updates — Withdrawn Routes

- **Used to "withdraw" network reachability**
- **Each Withdrawn Route is composed of:**
  - **Network Prefix**
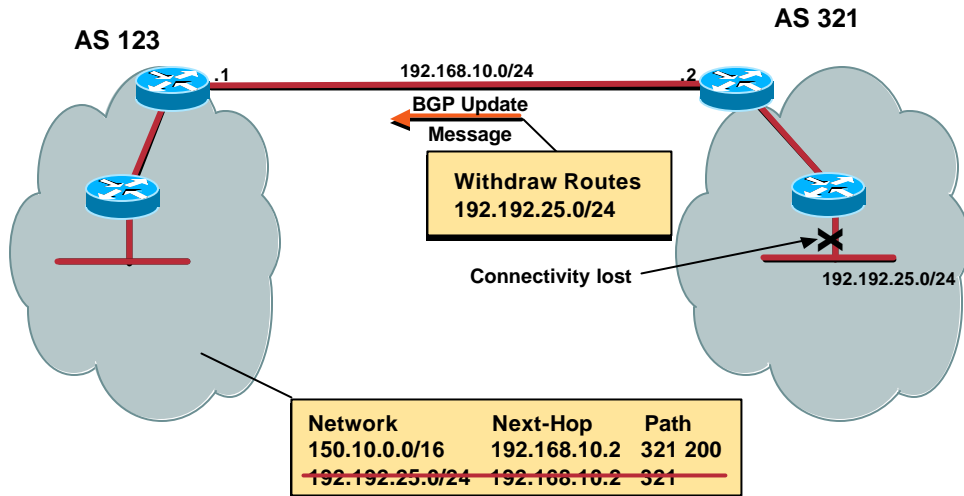 - **Mask Length**

8/14/2001 3:35 PM 23

- **Withdrawn Routes**
  - This section of the Update message contains zero or more routes (prefix) that are to be "withdrawn". The message is used to inform a BGP neighbor that the specified routes are no longer reachable.

BGP Updates — Withdrawn Routes

AS 321

AS 123

.1   192.168.10.0/24   .2

BGP Update
Message

Withdraw Routes
192.192.25.0/24

Connectivity lost

192.192.25.0/24

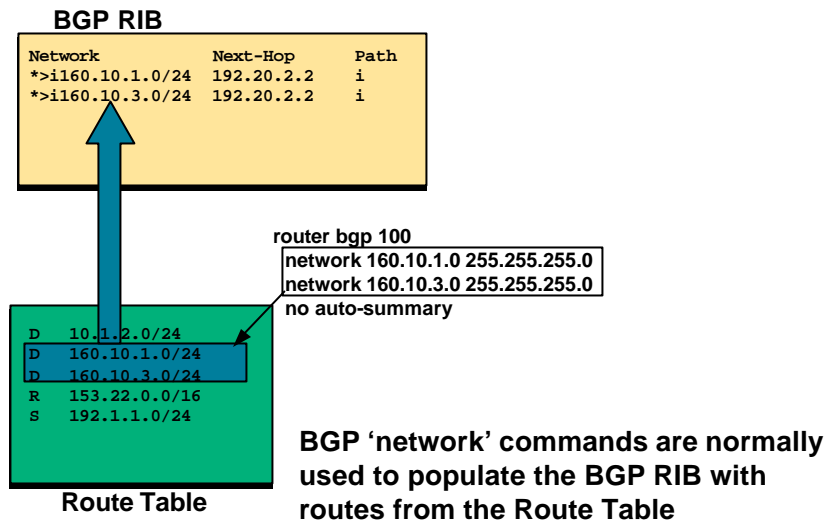| Network	Next-Hop	Path
150.10.0.0/16	192.168.10.2	321 200
192.192.25.0/24	192.168.10.2	321

 8/14/2001 3:35 PM   24

- **Withdrawn Routes**
  - In this example, network 192.192.25.0/24 was previously advertised to AS 123. However, the only interface to this network has failed. As a result, an Update message is sent to AS 123 with the prefix of network 192.192.25.0/24 in the Withdrawn Routes section of the message. The eBGP peer in AS 123 will update the information in its BGP Routing Information Base (RIB) to mark this route as withdrawn.

BGP Routing Information Base

BGP RIB

```
Network          Next-Hop      Path
*>i160.10.1.0/24  192.20.2.2    i
*>i160.10.3.0/24  192.20.2.2    i
```

router bgp 100
network 160.10.1.0 255.255.255.0
network 160.10.3.0 255.255.255.0
no auto-summary

```
D    10.1.2.0/24
D   160.10.1.0/24
D   160.10.3.0/24
R   153.22.0.0/16
S   192.1.1.0/24
```

**BGP 'network' commands are normally
used to populate the BGP RIB with
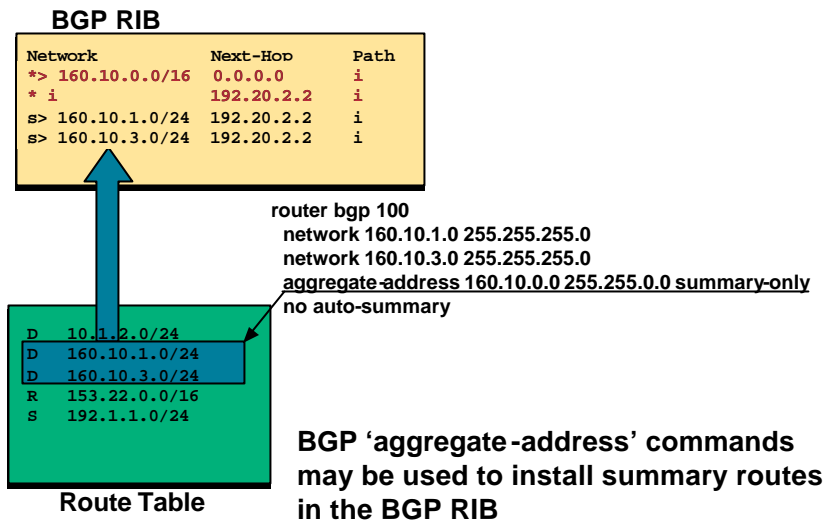routes from the Route Table**

Route Table

 8/14/2001 3:35 PM

- **BGP Routing Information Base**
    - The BGP Routing Information Base contains all of the prefixes along with their associated attributes that have been received from other BGP peers. In addition, the router can be configured to inject information from its local Route Table into its local BGP RIB. This information is, in turn, advertised to the router's BGP peers.
    - The Cisco IOS "network" command is the most common method of injecting local Route Table information into the BGP RIB. In the example shown above, prefixes 160.10.1.0/24 and 160.10.3.0/24 are being injected via the use of the network command.
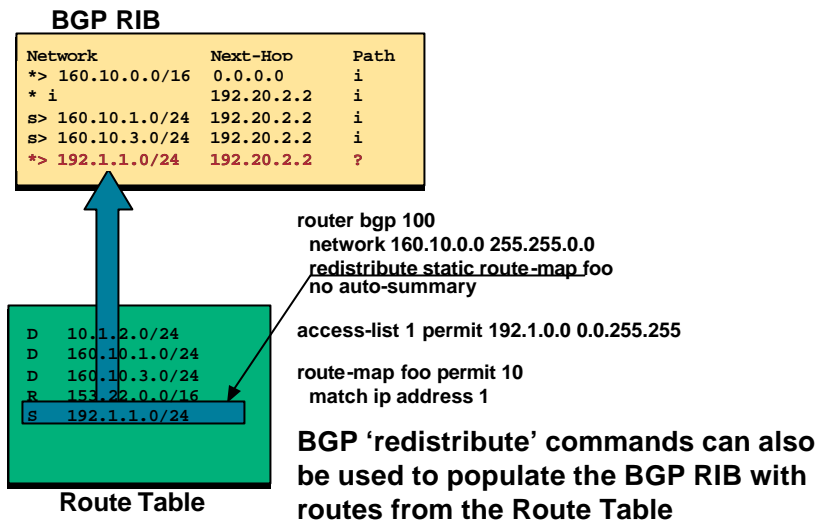
BGP RIB

```
Network            Next-Hop       Path
*> 160.10.0.0/16   0.0.0.0        i
*  i               192.20.2.2     i
s> 160.10.1.0/24   192.20.2.2     i
s> 160.10.3.0/24   192.20.2.2     i
```

router bgp 100
  network 160.10.1.0 255.255.255.0
  network 160.10.3.0 255.255.255.0
  aggregate-address 160.10.0.0 255.255.0.0 summary-only
  no auto-summary

```
D   10.1.2.0/24
D   160.10.1.0/24
D   160.10.3.0/24
R   153.22.0.0/16
S   192.1.1.0/24
```

Route Table

BGP 'aggregate-address' commands may be used to install summary routes in the BGP RIB

- **BGP Routing Information Base**
  - Prefix aggregation may be accomplished by the use of the Cisco IOS "aggregate-address" command as shown in the example above.
    - In this example, the use of the "summary-only" clause results in only the aggregate address being advertised to the router's BGP peers.
    - The small "s" to the left of the other 160.10.x.x entries in the BGP RIB indicate that these prefixes are "suppressed" and are not being advertised to the router's BGP peers.

BGP Routing Information Base

BGP RIB

```
Network           Next-Hop      Path
*> 160.10.0.0/16  0.0.0.0       i
*  i              192.20.2.2    i
s> 160.10.1.0/24  192.20.2.2    i
s> 160.10.3.0/24  192.20.2.2    i
*> 192.1.1.0/24   192.20.2.2    ?
```

```
router bgp 100
  network 160.10.0.0 255.255.0.0
  redistribute static route-map foo
  no auto-summary

access-list 1 permit 192.1.0.0 0.0.255.255

route-map foo permit 10
  match ip address 1
```

```
D    10.1.2.0/24
D    160.10.1.0/24
D    160.10.3.0/24
R    153.22.0.0/16
S    192.1.1.0/24
```

Route Table

BGP 'redistribute' commands can also be used to populate the BGP RIB with routes from the Route Table

- **BGP Routing Information Base**
  - The "redistribute" command may also be used to inject specific routes from the Route Table into the BGP RIB as shown in the example above.
    - In this example, all static routes specified by the route-map "foo" are being injected into the BGP RIB.
    - In this case, the static route for network 192.1.1.0/24 is a match and is being injected into the BGP RIB

BGP Routing Information Base

IN Process

BGP RIB

OUT Process

```
Network              Next-Hop       Path
*>i160.10.1.0/24  192.20.2.2     i
*>i160.10.3.0/24  192.20.2.2     i
*> 173.21.0.0/16  192.20.2.1     100
```

Update → Update →

Network	Next-Hop	Path
173.21.0.0/16	192.20.2.1	100
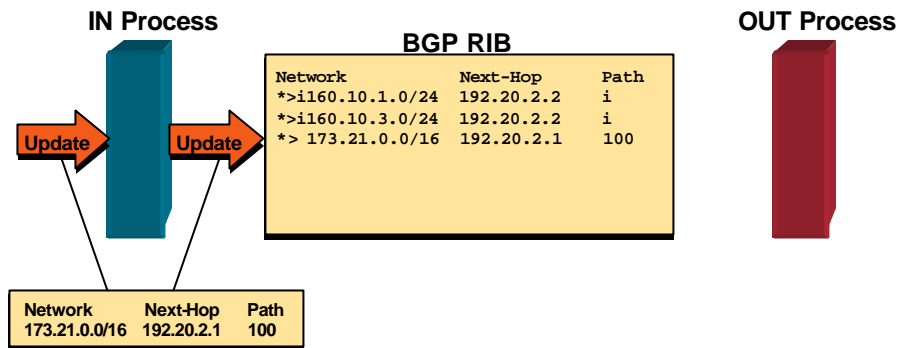
- **BGP "in" process**
  - **receives path information from peers**
  - **results of BGP path selection placed in the BGP table**
  - **"best path" flagged (denoted by ">")**

 8/14/2001 3:35 PM

- **BGP Input Process**
  - The BGP Input Process receives all incoming BGP Update messages from the other BGP peers. Although not shown in the above example, multiple paths to a particular network (prefix) may exist. The Input Process applies the various selection criteria to the incoming information and selects the "best path" for a given prefix. The selected "best path" is indicated by the ">" character as shown in the above example.

BGP Routing Information Base

IN Process　　　　　　　　**OUT Process**

BGP RIB

Network	Next-Hop	Path
*>i160.10.1.0/24	192.20.2.2	i
*>i160.10.3.0/24	192.20.2.2	i
*> 173.21.0.0/16	192.20.2.1	100

Update　　**Update**

Network	Next-Hop	Path
160.10.1.0/24	192.20.2.2	200
160.10.3.0/24	192.20.2.2	200
173.21.0.0/16	192.20.2.1	200 100

Next-Hop changed

- **BGP "out" process**
  - **builds update using info from RIB**
  - **may modify update based on config**
  - **Sends update to peers**

　　8/14/2001 3:35 PM　29

- **BGP Output Process**
  - The BGP Output Process is responsible for constructing and sending BGP Update messages to the router's other BGP peers. The contents of these Update messages may be modified using certain IOS configuration commands so that the desired routing policy is established.
    - In the example above, the next hop address being used in the Update message has been modified (possibly using the "next-hop-self" configuration command.

BGP Routing Information Base

BGP RIB

```
Network            Next-Hop      Path
*>i160.10.1.0/24   192.20.2.2    i
*>i160.10.3.0/24   192.20.2.2    i
*> 173.21.0.0/16   192.20.2.1    100
```

```
D   10.1.2.0/24
D   160.10.1.0/24
D   160.10.3.0/24
R   153.27.0.0/16
S   192.1.1.0/24
B   173.21.0.0/16
```

Route Table

- **Best paths installed in routing table if:**
  - **prefix and prefix length are unique**
  - **lowest "protocol distance"**

- **Injection of BGP prefixes into the local Route Table**
  - The "best-path" BGP prefixes are installed in the local routing table if:
    - The prefix/length are unique AND
    - the BGP prefix is the lowest Administrative Distance.

MBGP—Multiprotocol BGP

- **MBGP Overview**
- **MBGP Update Messages**
- **MBGP Capability Negotiation**
- **MBGP NLRI exchange**
- **MBGP-DVMRP redistribution**
- **BGP-to-MBGP redistribution**

 8/14/2001 3:35 PM 31

MBGP Overview

- **MBGP: Multiprotocol BGP**
  - **Defined in RFC 2283 (extensions to BGP)**
  - **Can carry different types of routes**
    - **IPv4 Unicast**
    - **IPv4 Multicast**
    - **IPv6 Unicast**
  - **May be carried in same BGP session**
  - **Does not propagate multicast state info**
    - **Still need PIM to build Distribution Trees**
  - **Same path selection and validation rules**
    - **AS-Path, LocalPref, MED, …**

 8/14/2001 3:35 PM 32

- **MBGP Overview**
  - Multiprotocol BGP (MBGP) is defined in RFC 2283. This RFC defines extensions to the existing BGP protocol to allow it to carry more than just IPv4 route prefixes. Examples of some of the new types of routing information include (but are not limited to):
    - IPv4 prefixes for Unicast routing
    - IPv4 prefixes for Multicast RPF checking
    - IPv6 prefixes for Unicast routing
  - A common misconception is that MBGP is a replacement for PIM. This is incorrect. MBGP does not propagate *any* multicast state information nor does it build any sort of multicast distribution trees. MBGP *can* distribute unicast prefixes that can be used for the multicast RPF check.
  - Because MBGP is an extension to the existing BGP protocol, the same basic rules apply to path selection, path validation, etc.

MBGP Overview

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

 8/14/2001 3:35 PM 33

- **Routing Information Bases**
  - Previously, BGP only maintained a single Routing Information Base (RIB) for IPv4 unicast prefixes. In the case of MBGP, separate RIB's must be maintained for each type of routing information being exchanged. This implies that a separate Unicast RIB (U-RIB) and a separate Multicast RIB (M-RIB) can be maintained by MBGP.
    - A new "nlri" keyword was added to the Cisco IOS command structure to differentiate between the U-RIB and the M-RIB. (Note: This keyword will soon be depreciated in order to generalize MBGP for other protocols such as IPv6. Consult you latest IOS Documentation for the correct syntax.)
  - Unicast RIB (U-RIB)
    - This RIB contains the unicast prefixes that was previously used by BGP for IPv4 unicast traffic forwarding.
  - Multicast RIB (M-RIB)
    - This new RIB contains the same *type* of unicast prefixes contained in the U-RIB except that the prefixes stored in the M-RIB are used to RPF check arriving multicast traffic.

MBGP Overview

- **Cisco IOS® MBGP implementation**
  - **All the familiar BGP configuration knobs**
  - **Carries multicast routes in MP_REACH_NLRI**
    - **Still carries unicast routes in old style NLRI**
  - **NLRI capability negotiation**
  - **Redistribution between MBGP and DVMRP**
  - **Redistribution of BGP stubs into MBGP**

 8/14/2001 3:35 PM 34

- **Cisco IOS Implementation of MBGP**
  - All of the familiar BGP configuration capabilities are maintained and may be applied separately to the U-RIB and/or the M-RIB.
  - The initial implementation of MBGP carries the multicast routes in the new MBPG attribute MP_REACH_NLRI while the unicast routes are still carried in the old style NLRI format.
  - Capability negotiation has been added to permit MBGP peers to negotiate which route information will be exchanged over a session.
  - The ability to redistribute MBGP multicast NLRI into DVMRP and vice versa has been added to permit a transition from legacy DVMRP based networks to native multicast networks.
  - Translation of BGP unicast NLRI into MBGP multicast NLRI is supported to permit easy transition of non MBGP capable networks.

MBGP Update Message

The MBGP UPDATE Message

| Length (I Octet) |
| Prefix (Variable) |

Unfeasible Routes Length (2 Octets)

Withdrawn Routes (Variable)

| Attribute Type |
| Attribute Length |

Total path Attribute Length (2 Octets)

Path Attributes (Variable)

Attribute Value

Network Layer Reachability Information (Variable)

| Length (I Octet) |
| Prefix (Variable) |
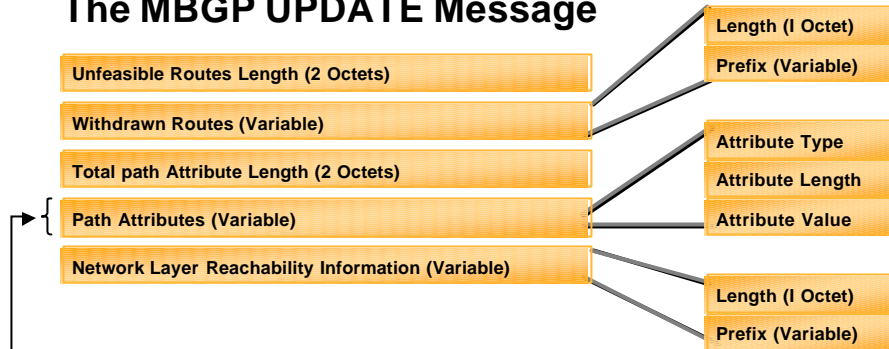
- **New Multiprotocol Attributes added to Path Attributes:**
  - **MP_REACH_NLRI**
  - **MP_UNREACH_NLRI**

 8/14/2001 3:35 PM 35

- **MBGP Update Message**
  - The MBGP Update message is identical to the old BGP Update message with the exception that two new attribute types have been added. These two attributes are:
    - MP_REACH_NLRI
    - MP_UNREACH_NLRI

MBGP Update Message
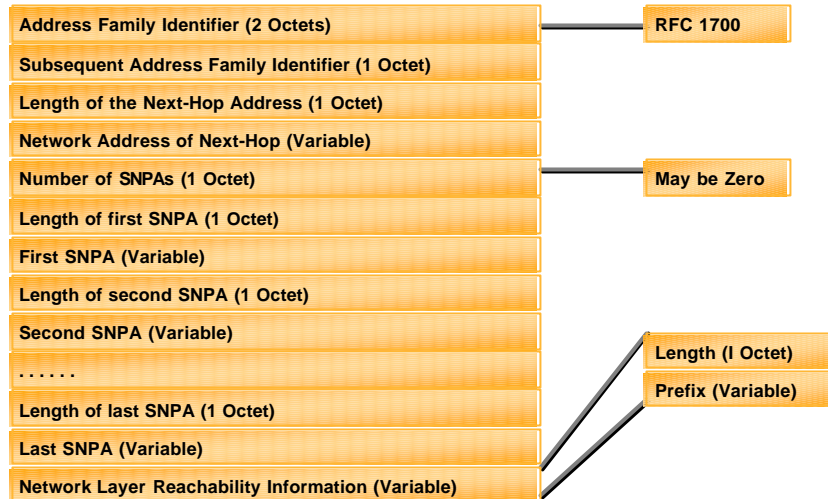
MP_REACH_NLRI Attribute

Address Family Identifier (2 Octets)	RFC 1700
Subsequent Address Family Identifier (1 Octet)	
Length of the Next-Hop Address (1 Octet)	
Network Address of Next-Hop (Variable)	
Number of SNPAs (1 Octet)	May be Zero
Length of first SNPA (1 Octet)	
First SNPA (Variable)	
Length of second SNPA (1 Octet)	
Second SNPA (Variable)	Length (I Octet)
. . . . . .	Prefix (Variable)
Length of last SNPA (1 Octet)	
Last SNPA (Variable)	
Network Layer Reachability Information (Variable)	

- **MP_REACH_NLRI Attribute**
    - The key characteristics of this new attribute is the Address Family Identifier and Sub-Address Family Identifier fields. These two fields define the type of routing information that is carried in the NLRI field of this attribute.
    - The Next-Hop Address information is contained in the field following the AFI and Sub-AFI fields.
    - Following the Next-Hop Address fields are zero or more SNPA fields. These field contain the attributes associated with the NLRI field. (For IPv4 AFI's, these attributes are the same as the old style BGP attributes.)
    - Finally, the NLRI field contains the Length and Prefix information of the route that is being advertised as reachable.

MBGP Update Message

- **Address Family Information (AFI)**
    - **Identifies Address Type (see RFC1700)**
        - **AFI = 1 (IPv4)**
        - **AFI = 2 (IPv6)**

- **Sub-Address Family Information (Sub-AFI)**
    - **Sub category for AFI Field**
    - **Address Family Information (AFI) = 1 (IPv4)**
        - **Sub-AFI = 1 (NLRI is used for unicast)**
        - **Sub-AFI = 2 (NLRI is used for multicast RPF check)**
        - **Sub-AFI = 3 (NLRI is used for both unicast and multicast RPF check)**

 8/14/2001 3:35 PM 37

- **Address Family Information (AFI)**
    - This field is based on the address families defined in RFC1700.
        - AFI = 1 (IPv4)
        - AFI = 2 (IPv6)

- **Sub-Address Family Information (Sub-AFI)**
    - This field contains further information regarding the type of routing information being exchanged in the NLRI field. The following are the current definitions for Sub-AFI codes associated with the IPv4 Address Family:
        - Sub-AFI = 1 (NLRI is used for unicast routing)
        - Sub-AFI = 2 (NLRI is used for multicast RPF check)
        - Sub-AFI = 3 (NLRI is used for both unicast routing and multicast RPF checking

MBGP Update Message

MP_UNREACH_NLRI Attribute

Address Family Identifier (2 Octets)

Subsequent Address Family Identifier (1 Octet)

Withdrawn Routes (Variable) ──── **Length (I Octet)**

Prefix (Variable)

- **MP_UNREACH_NLRI Attribute**
  - This new attribute permits unfeasible routes of the new protocol types to be withdrawn in the same fashion as the Withdrawn Routes field is used in BGP.
  - Notice that this attribute also carries the AFI and Sub-AFI fields along the associated Length and Prefix of the withdrawn route.

MBGP—Capability Negotiation

- **BGP routers establish BGP sessions through the OPEN message**
- **OPEN message contains optional parameters**
- **BGP session is terminated if OPEN parameters are not recognised**
- **New parameter: CAPABILITIES**
  - **Multiprotocol extension**
  - **Multiple routes for same destination**

8/14/2001 3:35 PM
39

- **MBGP Capability Negotiation**
  - MBGP has extended the Open Message format to include a new optional Capability negotiation parameters.
  - MBGP routes negotiation the lowest common set of capabilities using these Capability option fields.
  - If two MBGP peers are unable to agree on the Capabilities supported, the MBGP session is terminated and an error message written to the console.

Module10.ppt   39

MBGP—Capability Negotiation

- **New "nlri" keyword on "neighbor" command**
  ```
  neighbor <foo> remote-as <asn> nlri multicast unicast
  ```
- **Configures router to negotiate either or both types of NLRI**
- **If neighbor configures both or subset, common NRLI is used in both directions**
- **If there is no match, notification is sent and peering doesn't come up**

 8/14/2001 3:35 PM 40

- **MBGP Capability Negotiation**
  - A new "nlri" keyword is used on the "neighbor" command to control which capabilities are to be enabled with which peer. For example the command:
    - neighbor foo remote-as 100 nlri multicast unicast
  - results in both multicast and unicast capabilities being enabled to neighbor "foo".
  - If "foo" has configured the same set of abilities, then both unicast and multicast NLRI can be exchanged via the session. If the two peers do not match, the lowest common subset is used.
  - If there is no match between the capabilities, the peering will not come up.

MBGP — Capability Negotiation

AS 321

AS 123

.1   192.168.100.0/24   .2

```
router bgp 123
  neighbor 192.168.100.2 remote-as 321 nlri unicast multicast
. . .
```

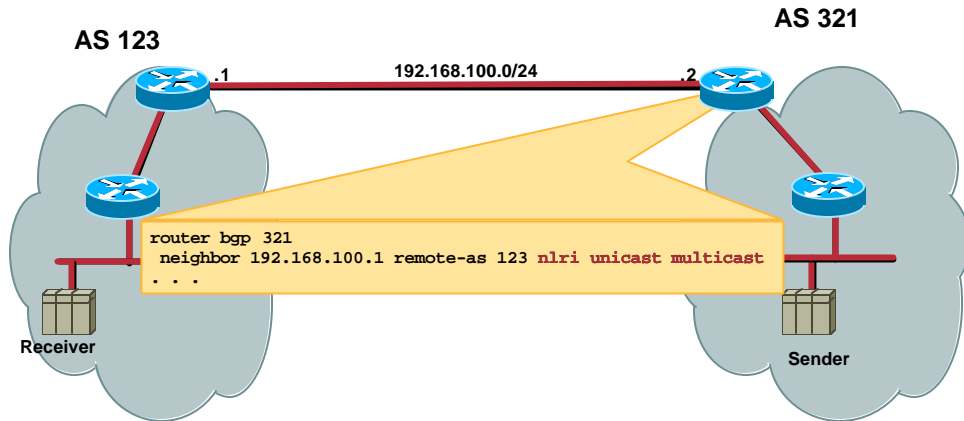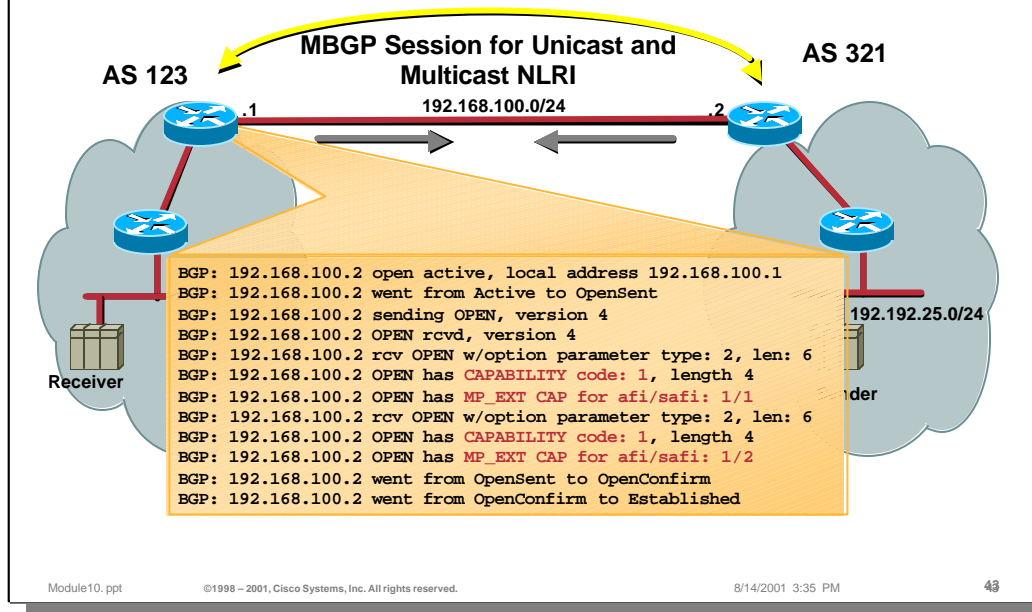**Receiver**

Sender

8/14/2001 3:35 PM

41
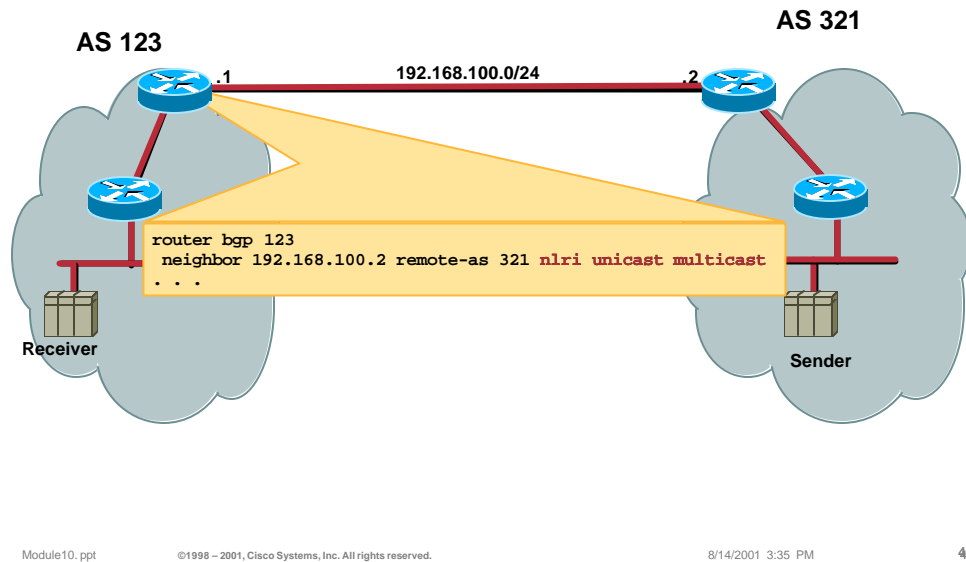
- **MBGP Capability Negotiation**
  - In this example, the router on the left is configured to peer with the router on the right. The command:
    - neighbor 192.168.100.2 remote-as 321 nlri unicast multicast
  -         instructs the router on the left to attempt to negotiate both unicast and multicast NLRI exchange.

MBGP — Capability Negotiation

AS 123

AS 321

.1   192.168.100.0/24   .2

```
router bgp 321
  neighbor 192.168.100.1 remote-as 123 nlri unicast multicast
. . .
```

Receiver

Sender

 8/14/2001 3:35 PM 42

• **MBGP Capability Negotiation**

– In this example, the router on the right is configured to peer with the router on the left. The command:

```
neighbor 192.168.100.1 remote-as 123 nlri unicast multicast
```

instructs the router on the right to attempt to negotiate *both* unicast and multicast NLRI exchange.

MBGP — Capability Negotiation

MBGP Session for Unicast and Multicast NLRI

AS 123

AS 321

.1   192.168.100.0/24   .2

192.192.25.0/24

Receiver

```
BGP: 192.168.100.2 open active, local address 192.168.100.1
BGP: 192.168.100.2 went from Active to OpenSent
BGP: 192.168.100.2 sending OPEN, version 4
BGP: 192.168.100.2 OPEN rcvd, version 4
BGP: 192.168.100.2 rcv OPEN w/option parameter type: 2, len: 6
BGP: 192.168.100.2 OPEN has CAPABILITY code: 1, length 4
BGP: 192.168.100.2 OPEN has MP_EXT CAP for afi/safi: 1/1
BGP: 192.168.100.2 rcv OPEN w/option parameter type: 2, len: 6
BGP: 192.168.100.2 OPEN has CAPABILITY code: 1, length 4
BGP: 192.168.100.2 OPEN has MP_EXT CAP for afi/safi: 1/2
BGP: 192.168.100.2 went from OpenSent to OpenConfirm
BGP: 192.168.100.2 went from OpenConfirm to Established
```

 8/14/2001 3:35 PM 43

- **MBGP Capability Negotiation**
  - In this example, the two routers can be seen exchanging Capabilities in the router debug message output. In this case, both unicast and multicast NLRI has been successfully negotiated. Therefore, both unicast and multicast NLRI will be exchanged in this single MBGP peer session

MBGP — Capability Negotiation

- **If neighbor doesn't include the CAPABILITY parameters in open, Cisco backs off and reopens with no capability parameters**

- **Peering comes up in unicast-only mode**

- **Hidden command**

  ```
  neighbor <foo> dont-capability-negotiate
  ```

 8/14/2001 3:35 PM 44

- **MBGP Capability Negotiation**
  - In cases where the far end of the attempted peering is not running MBGP, the router at that end will not send an OPEN message with the Capability option. Therefore, the Cisco router will retry the OPEN handshake without any Capability options.
    - This means the peering will come up and only IPv4 unicast NLRI will be exchanged.
  - In some cases, non MBGP routers have been known to crash or get stuck in a loop during the OPEN handshake. If this occurs, the Cisco MBGP router can be configured with the following hidden command to not attempt the Capability negotiation:
    - neighbor <foo> don't-capability-negotiate

MBGP — NLRI Information

RIB's may be populated by:

- **Network commands**

    ```
    network <foo> <foo-mask> [nlri multicast unicast]
    ```

    - **New "nlri" keyword controls in which RIB the matching route(s) is(are) stored**

        - **M-RIB if "multicast" keyword specified**
        - **U-RIB if "unicast" keyword specified (or if nlri clause omitted)**
        - **Both RIB's if both keywords specified**

 8/14/2001 3:35 PM 45

- **Controlling RIB population**
    - Like the "neighbor" command, the "network" command also accepts the new "nlri" keyword. This provides control as to which RIB (U-RIB, M-RIB, or both) that the network information is injected.
    - If the "nlri" clause is omitted, the Unicast RIB (U-RIB) is assumed.

MBGP — NLRI Information

Unicast RIB

```
Network          Next-Hop     Path
*>i160.10.1.0/24  192.20.2.2   i
*>i160.10.3.0/24  192.20.2.2   i
```

Multicast RIB

```
Network          Next-Hop     Path
```

```
D   10.1.2.0/24
D   160.10.1.0/24
D   160.10.3.0/24
R   153.22.0.0/16
S   192.1.1.0/24
```

Route Table

router bgp 100
  network 160.10.1.0 255.255.255.0 nlri unicast
  network 160.10.3.0 255.255.255.0 nlri unicast
  no auto-summary

New 'nlri' keyword used to control RIB population. (e.g. 'network' command)
  • **Unicast RIB only**

   8/14/2001 3:35 PM   46

- **RIB Population Example**
  – In this example, only the "unicast" keyword is used in the "nlri" clause. This causes the matching networks in the local Route Table to be injected into only the MBGP Unicast RIB.

MBGP — NLRI Information

Unicast RIB

Network	Next-Hop	Path

Multicast RIB

Network	Next-Hop	Path
*>i160.10.1.0/24	192.20.2.2	i
*>i160.10.3.0/24	192.20.2.2	i

```
D   10.1.2.0/24
D   160.10.1.0/24
D   160.10.3.0/24
R   153.22.0.0/16
S   192.1.1.0/24
```

Route Table

```
router bgp 100
   network 160.10.1.0 255.255.255.0 nlri multicast
   network 160.10.3.0 255.255.255.0 nlri multicast
   no auto-summary
```

New 'nlri' keyword used to control RIB population. (e.g. 'network' command)
- **Unicast RIB only**
- **Multicast RIB only**

- **RIB Population Example**
  - In this example, only the "multicast" keyword is used in the "nlri" clause. This causes the matching networks in the local Route Table to be injected into only the MBGP Multicast RIB.

MBGP — NLRI Information

Unicast RIB

```
Network          Next-Hop      Path
*>i160.10.1.0/24  192.20.2.2    i
*>i160.10.3.0/24  192.20.2.2    i
```

Multicast RIB

```
Network          Next-Hop      Path
*>i160.10.1.0/24  192.20.2.2    i
*>i160.10.3.0/24  192.20.2.2    i
```

router bgp 100
network 160.10.1.0 255.255.255.0 nlri unicast multicast
network 160.10.3.0 255.255.255.0 nlri unicast multicast
no auto-summary

```
D    10.1.2.0/24
D    160.10.1.0/24
D    160.10.3.0/24
R    153.22.0.0/16
S    192.1.1.0/24
```

Route Table

**New 'nlri' keyword used to control RIB
population. (e.g. 'network' command)**
- **Unicast RIB only**
- **Multicast RIB only**
- **Both RIB's**

8/14/2001 3:35 PM   48

- **RIB Population Example**
  – In this example, both the "unicast" and "multicast" keywords are used in the "nlri"
    clause. This causes the matching networks in the local Route Table to be
    injected into both the MBGP Unicast and Multicast RIBs.

RIB's may be populated by:

- **Redistribution commands**

```
router bgp 100
 redistribute <unicast> [<process>] route-map foo

access-list 1 permit 192.10.0.0 0.0.255.255

route-map foo permit 10
 match ip address 1
 set nlri multicast
```

  – **Route map 'set nlri' clause controls which RIB the matching route(s) is(are) stored**

 8/14/2001 3:35 PM 49

- **RIB Population Example**
  – This example demonstrates that the "nlri" clause may be used in a route-map that is in turn, used in a redistribute command. This may be useful in some scenarios where tighter control is desired as to what is injected into the desired RIB.
  – In this example, any network in the 192.10.0.0/16 range will be injected into only the MBGP Multicast RIB.

MBGP — NLRI Information

Unicast RIB

Network	Next-Hop	Path

Multicast RIB

Network	Next-Hop	Path
*>i192.1.1.0/24	192.20.2.2	?

```
router bgp 100
  redistribute static route-map foo
access-list 1 permit 192.1.1.0 0.0.0.255
route-map foo permit 10
  match ip address 1
  set nlri multicast
```

```
D   10.1.2.0/24
D   160.10.1.0/24
D   160.10.3.0/24
R   153.22.0.0/16
S   192.1.1.0/24
```
Route Table

New 'nlri' keyword used to control RIB population in a 'route-map'
(Multicast RIB only example)

• **RIB Population Example**
 – In this example, network 192.1.1.0/24 is being injected into only the MBGP Multicast RIB as a result of route-map "foo".

MBGP — NLRI Information

Cisco.com

- **Other "nlri" keyword commands**

  - **Aggregation**

    `aggregate-address <foo> <foo-mask> [nlri multicast unicast]`

    - **Generates an aggregate route for network <foo>**

  - **Default Origination**

    `neighbor <foo> default-originate [nlri multicast unicast]`

    - **Originates a default route to the neighbor**

  - **In Route Maps**

    `match nlri multicast unicast`

    - **Matches on the NLRI type**

    `set nlri multicast unicast`

    - **Injects the matched route into the specified unicast or multicast RIB**

Module10. ppt   ©1998 – 2001, Cisco Systems, Inc. All rights reserved.   8/14/2001 3:35 PM   51

- **Controlling RIB population**
  - The new "nlri" keyword can also be applied in other MBGP commands. This includes:
    - aggregate-address
    - default-originate
    - route-maps

Module10.ppt   51

MBGP — NLRI Information

RIB's are also populated by:

- **Receiving MP_REACH_NLRI from Peers**
  - **Storage controlled by AFI/SAFI value:**
    - **AFI/SAFI = 1/1 (IPv4 / Unicast) : U-RIB only**
    - **AFI/SAFI = 1/2 (IPv4 / Multicast) : M-RIB only**
    - **AFI/SAFI = 1/3 (IPv4 / Unicast-Multicast) : Both RIB's**
- **Receiving NLRI (old style) from Peers**
  - **Stored in RIB only**

- **RIB Population from Update Messages**
  - Information received in the form of Update messages are written into either the Unicast RIB, the Multicast RIB or both RIB's depending on the value of the AFI/Sub-AFI value.
    - AFI/Sub-AFI = 1/1 (IPv4 / Unicast): U-RIB only
    - AFI/Sub-AFI = 1/2 (IPv4 / Multicast): M-RIB only
    - AFI/Sub-AFI = 1/3 (IPv4 / Unicast-Multicast): Both RIB's
    - Old style NLRI (i.e. not in a MP_REACH_NLRI Attribute): U-RIB only

MBGP — NLRI Information

Unicast RIB

```
Network            Next-Hop       Path
*>i160.10.1.0/24   192.20.2.2     i
*>i160.10.3.0/24   192.20.2.2     i
*>i192.192.2.0/24  192.168.200.2 300 200 i
```

BGP Update from Peer

```
MP_REACH_NLRI: 192.192.2/24
AFI: 1, Sub-AFI: 1 (unicast)
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2
```

Multicast RIB

```
Network            Next-Hop       Path
*>i160.10.1.0/24   192.20.2.2     i
*>i160.10.3.0/24   192.20.2.2     i
```

- **Storage of arriving NLRI information depends on AFI/SAFI fields in the Update message**
  - **Unicast RIB only (AFI=1/SAFI=1 or old style NLRI)**

 8/14/2001 3:35 PM 53

- **Receiving Update Messages**
  - Storage of arriving NLRI information depends on the AFI/Sub-AFI fields in the MP_REACH_NLRI attribute.
  - In the above example, an Update message is received that contains an MP_REACH_NLRI with an AFI/Sub-AFI = 1/1 which indicates that the information is for the Unicast RIB. As a result, the information is processed by the MBGP input process and written into the U-RIB as shown.

MBGP — NLRI Information

Unicast RIB

```
Network            Next-Hop      Path
*>i160.10.1.0/24   192.20.2.2    i
*>i160.10.3.0/24   192.20.2.2    i
```

BGP Update from Peer

```
MP_REACH_NLRI: 192.192.2/24
AFI: 1, Sub-AFI: 2 (multicast)
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2
```

Multicast RIB

```
Network             Next-Hop        Path
*>i160.10.1.0/24    192.20.2.2      i
*>i160.10.3.0/24    192.20.2.2      i
*>i192.192.2.0/24   192.168.200.2   300 200 i
```

- **Storage of arriving NLRI information depends on AFI/SAFI fields in the Update message**
  - **Unicast RIB only (AFI=1/SAFI=1 or old style NLRI)**
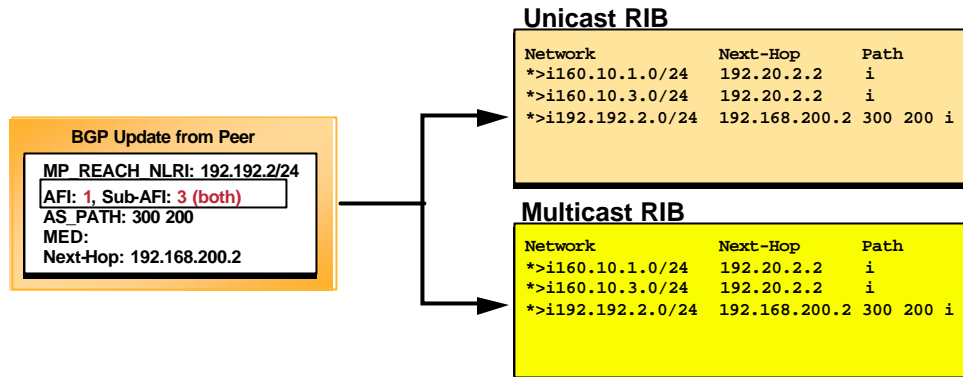 - **Multicast RIB only (AFI=1/SAFI=2)**

 8/14/2001 3:35 PM 54

- **Receiving Update Messages**
  - In the above example, an Update message is received that contains an MP_REACH_NLRI with an AFI/Sub-AFI = 1/2 which indicates that the information is for the Multicast RIB. As a result, the information is processed by the MBGP input process and written into the M-RIB as shown.

MBGP — NLRI Information

Unicast RIB

```
Network            Next-Hop      Path
*>i160.10.1.0/24   192.20.2.2    i
*>i160.10.3.0/24   192.20.2.2    i
*>i192.192.2.0/24  192.168.200.2 300 200 i
```

BGP Update from Peer

```
MP_REACH_NLRI: 192.192.2/24
AFI: 1, Sub-AFI: 3 (both)
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2
```

Multicast RIB

```
Network            Next-Hop      Path
*>i160.10.1.0/24   192.20.2.2    i
*>i160.10.3.0/24   192.20.2.2    i
*>i192.192.2.0/24  192.168.200.2 300 200 i
```

- **Storage of arriving NLRI information depends on AFI/SAFI fields in the Update message**
  - **Unicast RIB only (AFI=1/SAFI=1 or old style NLRI)**
  - **Multicast RIB only (AFI=1/SAFI=2)**
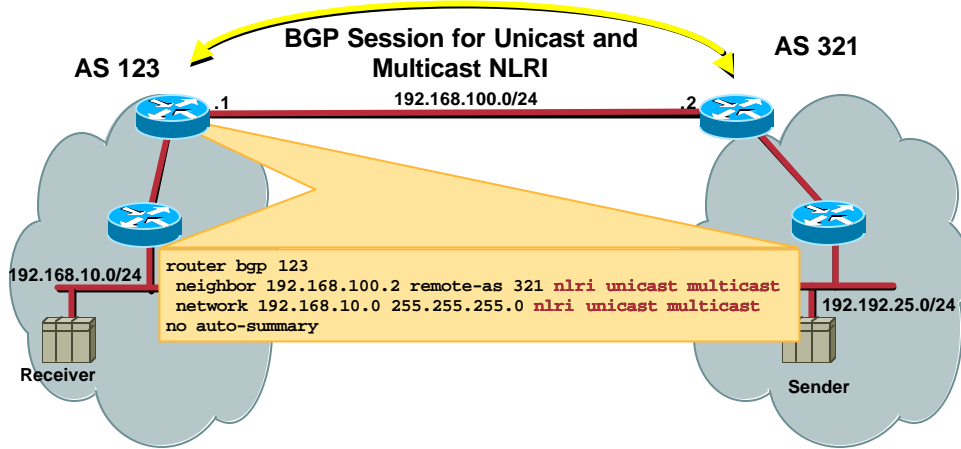 - **Both RIB's (AFI=1/SAFI=3)**

 8/14/2001 3:35 PM 55

- **Receiving Update Messages**
  - In the above example, an Update message is received that contains an MP_REACH_NLRI with an AFI/Sub-AFI = 1/3which indicates that the information is for both the Unicast and Multicast RIBs. As a result, the information is processed by the MBGP input process and written into both the U-RIB and the M-RIB as shown.

Cisco.com

Congruent Topologies

BGP Session for Unicast and Multicast NLRI

AS 123

AS 321

.1   **192.168.100.0/24**   .2

192.168.10.0/24

```
router bgp 123
  neighbor 192.168.100.2 remote-as 321 nlri unicast multicast
  network 192.168.10.0 255.255.255.0 nlri unicast multicast
no auto-summary
```

192.192.25.0/24

Receiver

Sender

8/14/2001 3:35 PM

56

- **Congruent Unicast-Multicast Topologies**
  - When the flows of both unicast and multicast traffic follow the same path, it is said that the two topologies are "congruent".
  - In the example above, both the "neighbor" and the "network" statements in the router on the left, contain the "nlri unicast multicast" clause. This causes this router to negotiate a single BGP session over which both unicast and multicast NLRI are exchanged.
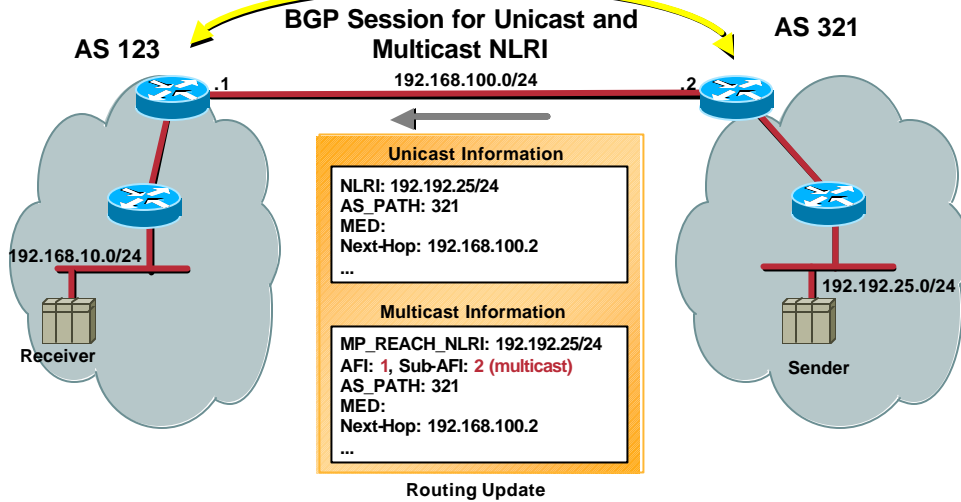
- **Congruent Unicast-Multicast Topologies**
  - Continuing with the example from the previous page, both the "neighbor" and the "network" statements in the router on the right also contain the "nlri unicast multicast" clause. This causes this router to also negotiate a single BGP session over which both unicast and multicast NLRI are exchanged. In this simple case, the Unicast and Multicast topologies are congruent.
  - Once the two routers have opened the session, the result is that the Unicast and Multicast topologies will be congruent.

MBGP — NLRI Information

Congruent Topologies

BGP Session for Unicast and Multicast NLRI

AS 123

AS 321

.1   192.168.100.0/24   .2

Unicast Information

NLRI: 192.192.25/24
AS_PATH: 321
MED:
Next-Hop: 192.168.100.2
...

192.168.10.0/24

Multicast Information

MP_REACH_NLRI: 192.192.25/24
AFI: 1, Sub-AFI: 2 (multicast)
AS_PATH: 321
MED:
Next-Hop: 192.168.100.2
...

192.192.25.0/24

Receiver

Sender

Routing Update
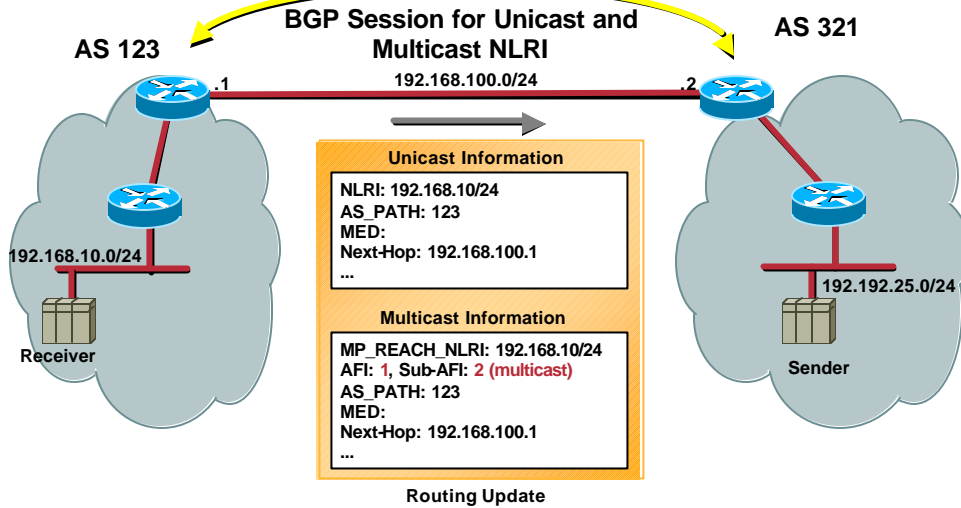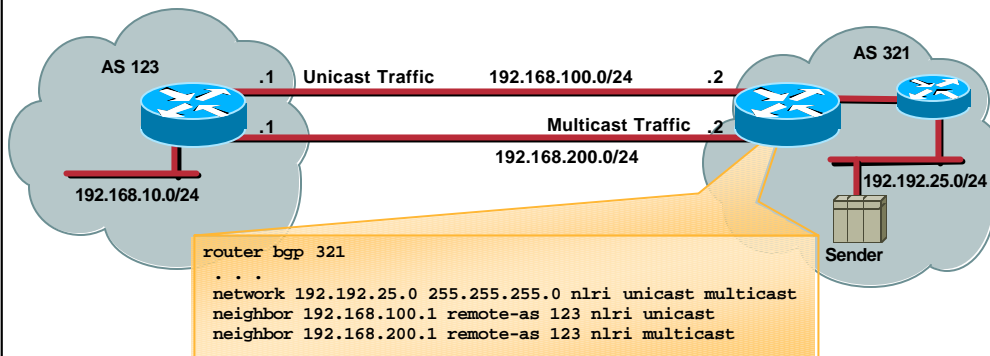
 8/14/2001 3:35 PM 58

- **Congruent Unicast-Multicast Topologies**
  - Continuing with the example from the previous pages we see the router on the right sending an Update message containing both an old style NLRI for the unicast traffic and an MP_REACH_NLRI attribute containing the multicast NLRI information.
    - Notice that the "Next-Hop" address is identical for both NLRI. This results in the Unicast and Multicast topologies being congruent.

MBGP — NLRI Information

Congruent Topologies

BGP Session for Unicast and Multicast NLRI

AS 123

AS 321

.1   **192.168.100.0/24**   .2

192.168.10.0/24

Receiver

192.192.25.0/24

Sender

Unicast Information

NLRI: 192.168.10/24
AS_PATH: 123
MED:
Next-Hop: 192.168.100.1
...

Multicast Information

MP_REACH_NLRI: 192.168.10/24
AFI: 1, Sub-AFI: 2 (multicast)
AS_PATH: 123
MED:
Next-Hop: 192.168.100.1
...

Routing Update

8/14/2001 3:35 PM
59

- **Congruent Unicast-Multicast Topologies**
  - Continuing with the example from the previous pages we see the router on the left sending an Update message. This message also containing both an old style NLRI for the unicast traffic and an MP_REACH_NLRI attribute containing the multicast NLRI information.
    - Notice that the "Next-Hop" address is identical for both NLRI. This results in the Unicast and Multicast topologies being congruent.

Cisco.com

Incongruent Topologies

AS 123

.1 Unicast Traffic 192.168.100.0/24 .2

AS 321

.1 Multicast Traffic .2

192.168.200.0/24

192.168.10.0/24

192.192.25.0/24

Sender

```
router bgp 321
. . .
 network 192.192.25.0 255.255.255.0 nlri unicast multicast
 neighbor 192.168.100.1 remote-as 123 nlri unicast
 neighbor 192.168.200.1 remote-as 123 nlri multicast
```

8/14/2001 3:35 PM 60

- **Incongruent Unicast-Multicast Topologies**
  - In some cases it is desirable to have the multicast traffic follow one path and the unicast another. When this occurs, the topologies are said to be "incongruent".
  - In the example above, two separate "neighbor" statements are used so that the router on the right will negotiate and establish two separate BGP sessions with its neighbor router on the left. This is accomplished with the following two commands

    ```
    neighbor 192.168.100.1 remote-as 123 nlri unicast
    neighbor 192.168.200.1 remote-as 123 nlri multicast
    ```

    - The first command instructs the router on the right to negotiate a unicast-only BGP session over the top serial link while the second command instructs it to negotiate a multicast-only session over the bottom serial link.
  - The router on the right is also instructed to inject network 192.192.25.0/24 into both its local U-RIB and M-RIB using the following command:

    ```
    network 192.192.25.0 255.255.255.0 nlri unicast multicast
    ```

    - Once this network is injected into both RIBs, its network prefix will be advertised to the router on the right. *However, the unicast nlri for this network will be advertised over the top BGP session and the multicast nlri over the bottom BGP session.*

Cisco.com

Incongruent Topologies

AS 123

AS 321

.1 **Unicast Traffic** 192.168.100.0/24 .2

.1 **Multicast Traffic** .2

192.168.200.0/24

192.168.10.0/24

192.192.25.0/24

Sender

```
router bgp 123
. . .
network 192.168.10.0 255.255.255.0 nlri unicast multicast
neighbor 192.168.100.2 remote-as 321 nlri unicast
neighbor 192.168.200.2 remote-as 321 nlri multicast
```

8/14/2001 3:35 PM
61

- **Incongruent Unicast-Multicast Topologies**
  - Continuing with the example, two separate "neighbor" statements are also used on the router on the left so it will negotiate and establish two separate BGP sessions with its neighbor router on the right. This is accomplished with the following two commands

    ```
    neighbor 192.168.100.2 remote-as 321 nlri unicast
    neighbor 192.168.200.2 remote-as 321 nlri multicast
    ```

    - The first command instructs the router on the right to negotiate a unicast-only BGP session over the top serial link while the second command instructs it to negotiate a multicast-only session over the bottom serial link.
  - The router on the right is also instructed to inject network 192.168.10.0/24 into both its local U-RIB and M-RIB using the following command:
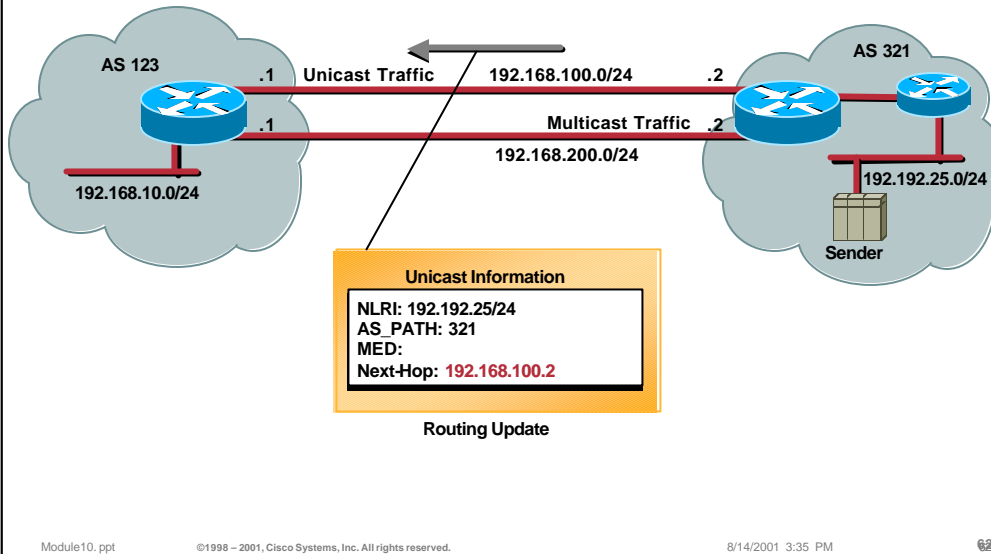
    ```
    network 192.168.10.0 255.255.255.0 nlri unicast multicast
    ```

    - Once this network is injected into both RIBs, its network prefix will be advertised to the router on the right. *However, the unicast nlri for this network will be advertised over the top BGP session and the multicast nlri over the bottom BGP session.*

MBGP — NLRI Information

Incongruent Topologies

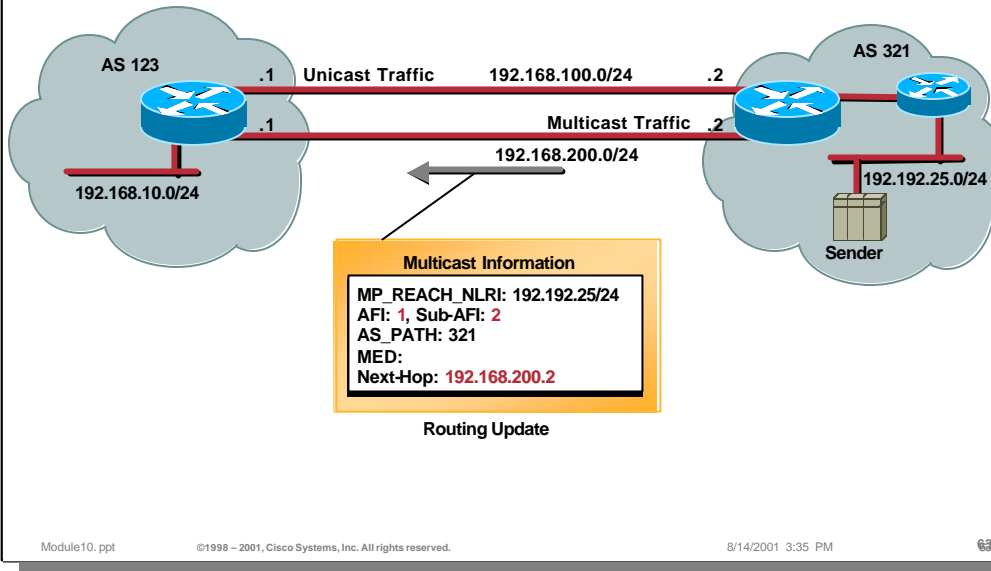AS 123

.1 Unicast Traffic 192.168.100.0/24 .2 AS 321

.1 Multicast Traffic .2

192.168.200.0/24

192.168.10.0/24

192.192.25.0/24

Sender

Unicast Information

NLRI: 192.192.25/24
AS_PATH: 321
MED:
Next-Hop: 192.168.100.2

Routing Update

   8/14/2001 3:35 PM 62

- **Incongruent Unicast-Multicast Topologies**
  - Once the two separate BGP sessions have been established, unicast and multicast NLRI will be exchanged over their respective sessions as shown in the example above.
  - In the drawing above, the router on the right is sending a BGP Update message over the top serial link (the unicast session) containing only a unicast NLRI for network 192.192.25.0/24.
    - Notice that the "Next-Hop" information indicates that unicast traffic for this network should be sent to 192.168.100.2, which corresponds to the top serial link. This will result in unicast traffic destined for the 192.168.100.2 network flowing over the top serial link.
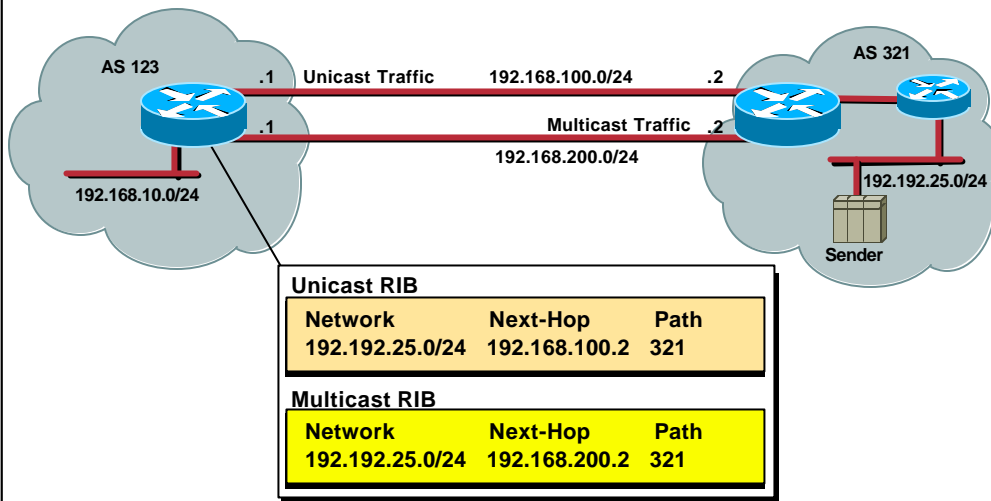
MBGP — NLRI Information

Incongruent Topologies

AS 123

AS 321

.1 Unicast Traffic   192.168.100.0/24   .2

.1   Multicast Traffic   .2

192.168.200.0/24

192.168.10.0/24

192.192.25.0/24

Sender

Multicast Information

MP_REACH_NLRI: 192.192.25/24
AFI: 1, Sub-AFI: 2
AS_PATH: 321
MED:
Next-Hop: 192.168.200.2

Routing Update

8/14/2001 3:35 PM

63

- **Incongruent Unicast-Multicast Topologies**
  - Continuing with NLRI exchange example shown In the drawing above, the router on the right is now sending a BGP Update message over the bottom serial link (the multicast session) containing only a multicast NLRI for network 192.192.25.0/24.
    - Notice that the "Next-Hop" information indicates that multicast traffic from any source in the network 192.192.25.0/24 should RPF to the 192.168.200.2 address, which corresponds to the bottom serial link. This will result in multicast traffic from any source in network 192.168.100.2 flowing over the bottom serial link.

MBGP — NLRI Information

Cisco.com

Incongruent Topologies

AS 123

AS 321

.1 Unicast Traffic 192.168.100.0/24 .2

.1 Multicast Traffic .2
192.168.200.0/24

192.192.25.0/24

192.168.10.0/24

Sender

Unicast RIB

Network	Next-Hop	Path
192.192.25.0/24	192.168.100.2	321

Multicast RIB

Network	Next-Hop	Path
192.192.25.0/24	192.168.200.2	321

8/14/2001 3:35 PM

64

- **Incongruent Unicast-Multicast Topologies**
  - In the final drawing of our example we see that the router on the left has processed the two Update messages and inserted an entry for network 192.192.25/24 in both the U-RIB and the M-RIB.
  - Notice that the "Next-Hop" information in the entry in the M-RIB indicates that the Next-Hop is 192.168.200.2 while the "Next-Hop" information in the entry in the U-RIB indicates that the Next-Hop is 192.168.100.2. Because each RIB entry for this network has a different Next-Hop address, the unicast and multicast traffic will flow over different paths.

Unicast-Multicast NLRI Translation

- **BGP stubs that don't have MBGP support need to get their prefixes into the Multicast backbone**

- **They get external routes via MBGP default or static default**

8/14/2001 3:35 PM
65

- **Unicast to Multicast NLRI Translation**
  - In some transition cases, it may be desirable to convert unicast NLRI into Multicast NLRI on behalf of an old style BGP stub AS that would like to receive multicast traffic but is unable to upgrade to MBGP.
    - Because this is a BGP stub AS, it is possible for this AS to simple RPF using either a static route or the unicast NLRI prefixes received from the upstream AS.

Unicast-Multicast NLRI Translation

- **Use command**

  `neighbor <foo> translate-update [nlri multicast]`

- **Arriving BGP Updates are translated into an MP_REACH_NLRI attribute**
  - **As if the neighbor sent AFI 1/SAFI 2 routes**
  - **Results written into the Multicast RIB**

- **Original BGP Update processed as normal**
  - **Results written into the Unicast RIB**

- **Unicast to Multicast NLRI Translation**
  - Unicast-Multicast NLRI Translation may be enabled by the following command:
    - neighbor <foo> translate-update nlri multicast
    - This instructs the router to clone another copy of the received BGP Update message and to convert the unicast NLRI into a multicast NLRI using the MBGP MP_REACH_NLRI attribute. This update is processed as if it were received by the BGP Stub AS and the results are written into the M-RIB.
    - The original BGP Update containing the unicast NLRI is processed as normal and the results written into the U-RIB

Unicast-Multicast NLRI Translation

'translate-update' Front-end　　　　　　　　　**BGP IN Process**

```
NLRI: 192.192.2/24
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2
```

BGP Update from Peer

```
NLRI: 192.192.2/24
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2
```

```
MP_REACH_NLRI: 192.192.2/24
AFI: 1, Sub-AFI: 2 (multicast)
AS_PATH: 300 200
MED:
Next-Hop: 192.168.200.2
```

- **Arriving Unicast update intercepted by 'translate-update' Front-end**
- **A translated Multicast update is created & passed to the IN Process**
- **Original Unicast update is passed on to the IN Process**
- **Both updates processed normally by the IN Process**

- **Unicast to Multicast NLRI Translation**
  - The above drawing graphically depicts the sequence of events when "translate-update" has been for a BGP neighbor.
    - At the left of the drawing a old style BGP Update arrives containing unicast NLRI.
    - The Update message is intercepted by the "Translate-Update" front-end process and a Multicast Update message is created and passed on to the MBGP Input process.
    - The original Unicast Update message is also passed on unchanged to the MBGP Input process.
    - Both updates are processed normally and the results written into the appropriate RIB.

Unicast-Multicast NLRI Translation

Unicast NLRI Only Stub AS
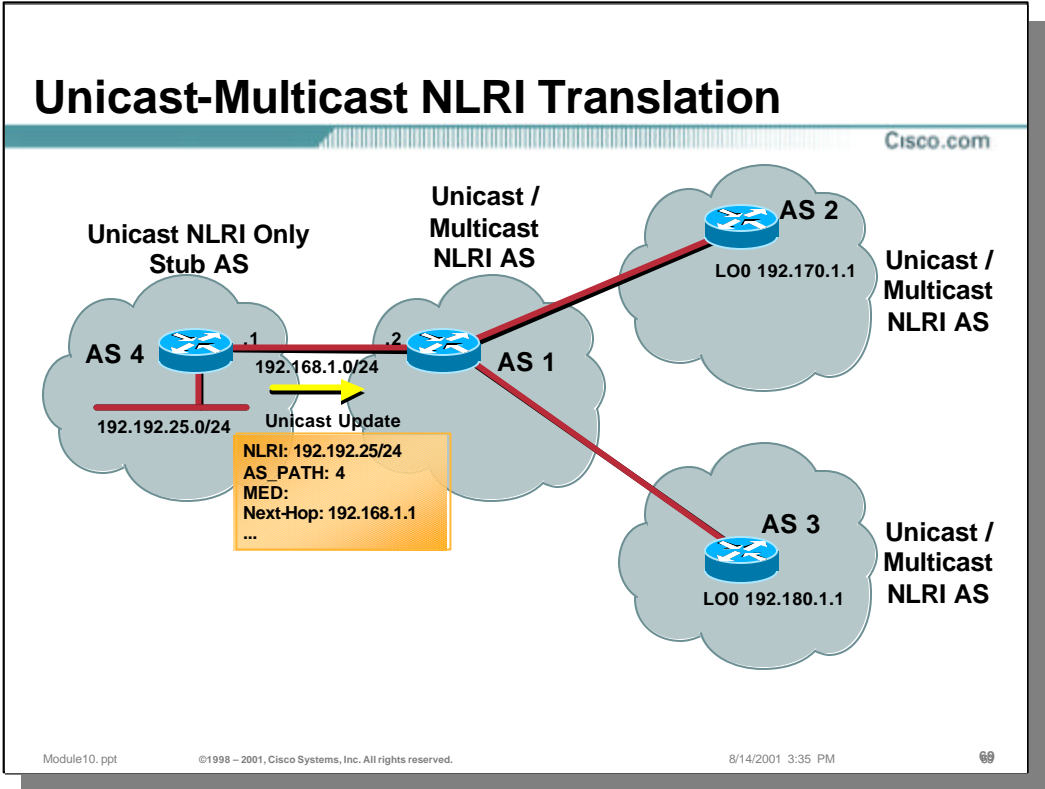
Unicast / Multicast NLRI AS

AS 2

LO0 192.170.1.1

Unicast / Multicast NLRI AS

AS 4

.1   .2

192.168.1.0/24   **AS 1**

192.192.25.0/24

AS 3

LO0 192.180.1.1

Unicast / Multicast NLRI AS

```
router bgp 1
. . .
 neighbor 192.168.1.1 remote-as 4
 neighbor 192.168.1.1 translate-update nlri multicast
 neighbor 192.170.1.1 remote-as 2 nlri unicast multicast
 neighbor 192.180.1.1 remote-as 3 nlri unicast multicast
. . .
```

8/14/2001 3:35 PM   68

- **Unicast to Multicast NLRI Translation**
  - In this drawing we see a BGP-Only stub AS (AS 4) using AS 1 as their transit for both unicast and multicast traffic. However, in order for other AS's to be able to correctly RPF towards AS 4 for network 192.192.25.0/24, the router in AS 1 has been configured to do unicast-multicast NLRI translation as shown in the configuration above.
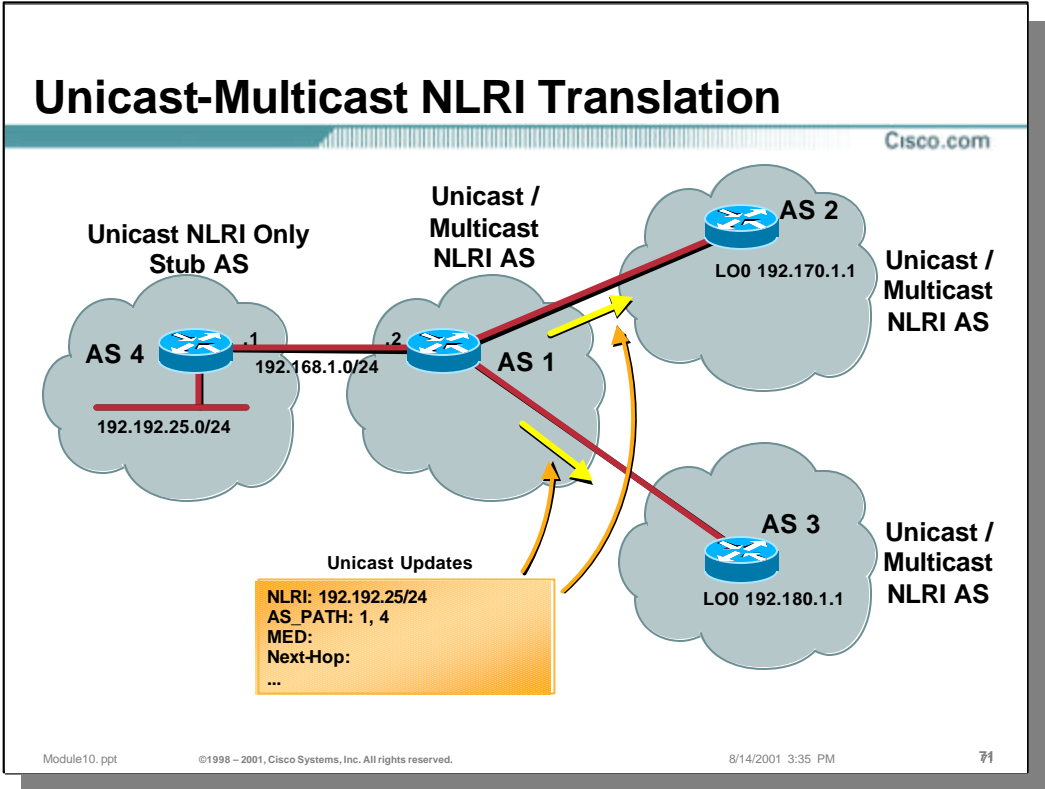
Unicast-Multicast NLRI Translation

Unicast NLRI Only Stub AS

Unicast / Multicast NLRI AS

AS 4

AS 2

LO0 192.170.1.1

Unicast / Multicast NLRI AS

.1

192.168.1.0/24

.2

AS 1

192.192.25.0/24

Unicast Update

NLRI: 192.192.25/24
AS_PATH: 4
MED:
Next-Hop: 192.168.1.1
...

AS 3

LO0 192.180.1.1

Unicast / Multicast NLRI AS

8/14/2001 3:35 PM   69

- **Unicast to Multicast NLRI Translation**
  - Continuing with the example, we see prefix 192.192.25.0/24 being sent in an old-style BGP Update message to the router in AS 1.

Unicast-Multicast NLRI Translation

Cisco.com

Unicast / Multicast NLRI AS

Unicast NLRI Only Stub AS

AS 2
LO0 192.170.1.1

Unicast / Multicast NLRI AS

AS 4

.1
192.168.1.0/24
.2

AS 1

192.192.25.0/24

AS 3
LO0 192.180.1.1

Unicast / Multicast NLRI AS

Multicast Updates

MP_REACH_NLRI: 192.192.25/24
AFI: 1, Sub-AFI: 2 (multicast)
AS_PATH: 1, 4
MED:
Next-Hop:
...

8/14/2001 3:35 PM
70
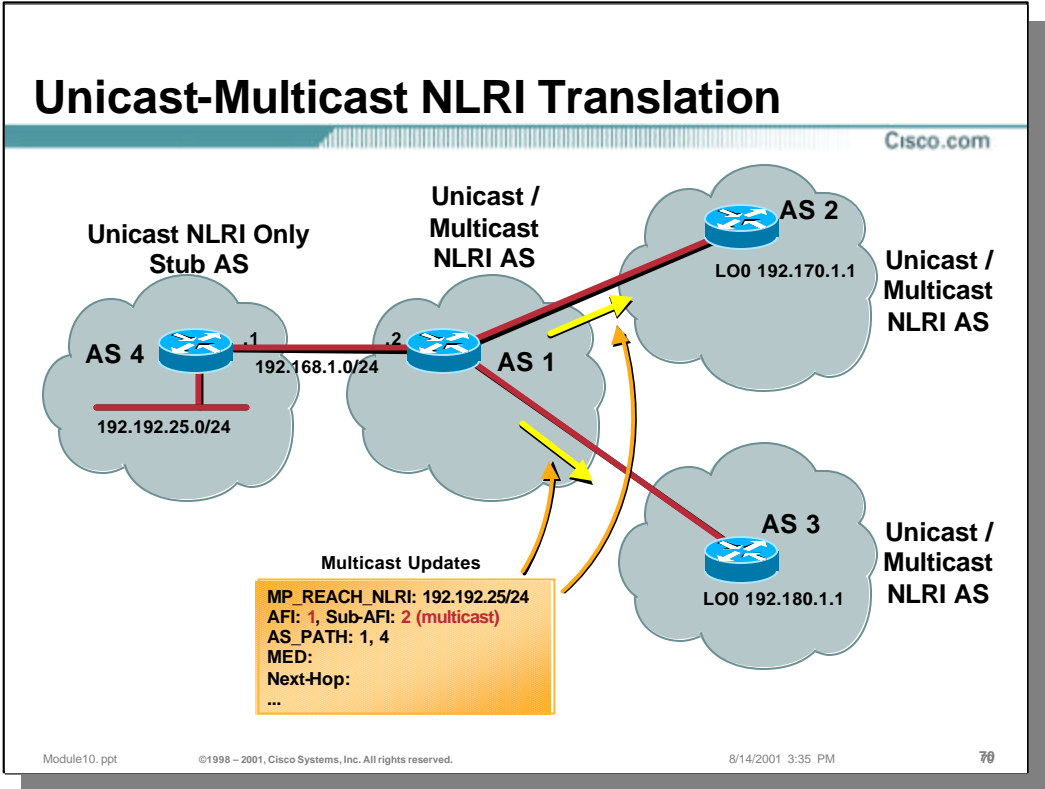
- **Unicast to Multicast NLRI Translation**
  – The router in AS 1 has been configured to perform unicast-multicast NLRI translation and therefore converts the incoming unicast (old-style) NLRI into a Multicast NLRI in an MBGP Update message using the MP_REACH_NLRI attribute. This translated NLRI is processed normally by the router in AS1 which results in it forwarding Multicast NLRI updates to its neighbors for network 192.192.25.0/24.

Unicast-Multicast NLRI Translation

- **Unicast to Multicast NLRI Translation**
  - Finally, the original unicast NLRI update is processed normally by the router in AS1 which results in it forwarding unicast NLRI updates to its neighbors for network 192.192.25.0/24.

Unicast-Multicast NLRI Translation

- **Route Maps can be used for finer control over which prefixes are translated.**
- **Omit 'nlri' clause in translate-update cmd**

  ```
  neighbor <foo> translate-update
  ```

- **Use 'set nlri' in Route map**

  ```
  neighbor <foo> remote-as 4 bgp-to-mbgp in
  access-list 1 permit 192.192.25.0 0.0.0.255
  route-map bgp-to-mbgp permit 10
  match ip address 1
  set nlri multicast
  ```

- **Unicast to Multicast NLRI Translation**
  - Route Maps may be used to provide finer control over which prefixes are translated into Multicast NLRI. When this is desired, the "nlri" clause is omitted from the "translate-update" command. (The "translate-update" command is still necessary for the neighbor.)
  - The example above shows a configuration that will result in only prefixes in the 192.192.25.0/24 range being translated into Multicast NLRI.
    - Because the "translate-update" function has been enabled for this neighbor, all arriving unicast NLRI are translated (by the "translate-update" front-end) into Multicast NLRI.
    - The Route map in the above example will only match on prefixes in the 192.192.25.0/24 range that are received from neighbor "foo".
    - The "set nlri multicast" clause in the body of the Route Map causes only these matching prefixes to be injected into the local M-RIB. As a result, only prefixes in the above range will be translated.

DVMRP to MBGP Redistribution

- **You can also put routes in the MRIB that are currently in the DVMRP routing table**

  ```
  router bgp <asn>
  redistribute dvmrp route-map <map>
  ```

- **You can do your typical set operations**

- **Used when connecting DVMRP access points into the MBGP backbone**

- **Used at strategic interconnect points with the old DVMRP MBONE**

- **DVMRP to MBGP Redistribution**
  - Routes in the router's local DVMRP routing table may be "redistributed" into the MBGP Multicast RIB using standard Route Maps. This is typically done where legacy DVMRP networks still exist and the received DVMRP routes need to be injected into the M-RIB.
    - Careful redistributing of these DVMRP routes into the M-RIB will allow other MBGP routers to RPF correctly for sources in the DVMRP network.
    - This is considered a "transitional" function that permits DVMRP networks to be phased out gracefully and should be used only a strategic interconnect points with old DVMRP networks. (Otherwise, multicast route loops and multicast instability can occur.)

DVMRP to MBGP Redistribution
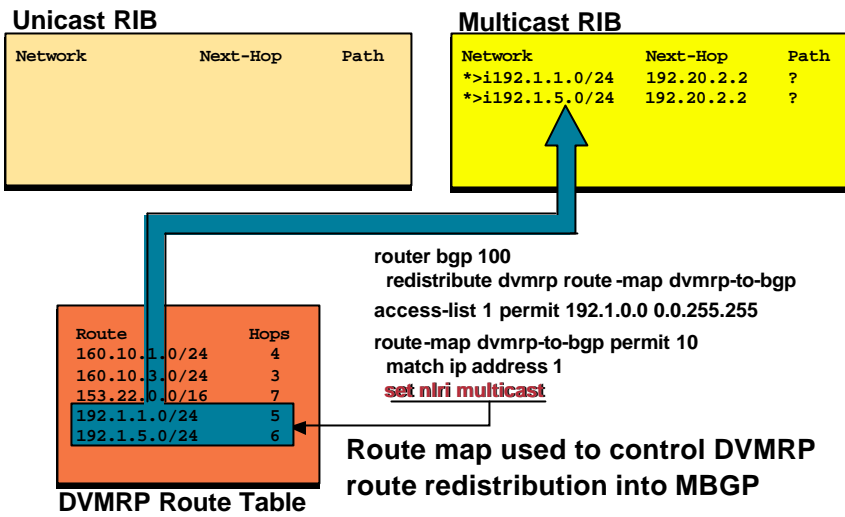
Unicast RIB

Network	Next-Hop	Path

Multicast RIB

Network	Next-Hop	Path
*>i192.1.1.0/24	192.20.2.2	?
*>i192.1.5.0/24	192.20.2.2	?

```
router bgp 100
   redistribute dvmrp route -map dvmrp-to-bgp
access-list 1 permit 192.1.0.0 0.0.255.255
route-map dvmrp-to-bgp permit 10
   match ip address 1
   set nlri multicast
```

Route	Hops
160.10.1.0/24	4
160.10.3.0/24	3
153.22.0.0/16	7
192.1.1.0/24	5
192.1.5.0/24	6

Route map used to control DVMRP route redistribution into MBGP

DVMRP Route Table

- **DVMRP to MBGP Redistribution**
  - In the above configuration example, a Route Map is being used to "redistribute" DVMRP Routes from the router's local DVMRP routing table into the MBGP Multicast RIB. Notice that the "set nlri multicast" clause in the body of the Route Map is what causes this to occur.

DVMRP to MBGP Redistribution

AS 2
LO0 192.170.1.1

DVMRP Tunnel

AS 1

**DVMRP Stub
Network**

AS 3
LO0 192.180.1.1

```
router bgp 1
 neighbor 192.168.1.1 remote-as 4
 neighbor 192.170.1.1 remote-as 2 nlri unicast multicast
 neighbor 192.180.1.1 remote-as 3 nlri unicast multicast
 redistribute dvmrp route-map dvmrp-to-mbgp
 access-list 1 permit 192.168.0.0 0.0.255.255
 route-map dvmrp-to-mbgp permit 10
 match ip address 1
 set nlri multicast
```
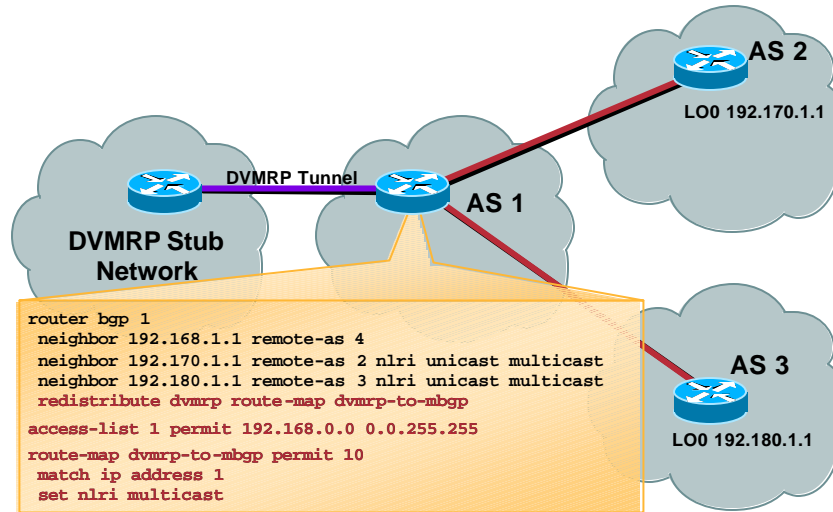
 8/14/2001 3:35 PM 75

- **DVMRP to MBGP Redistribution**
  - The drawing above is another example of DVMRP to MBGP redistribution.
    - Notice that the configuration in the router in AS1 uses a Route Map that matches on any DVMRP prefixes in the 192.168.0.0/16 range via "access-list 1".
    - The "set nlri multicast" clause in the body of the Route Map causes the DVMRP matching routes to be injected into the MBGP M-RIB.

DVMRP to MBGP Redistribution
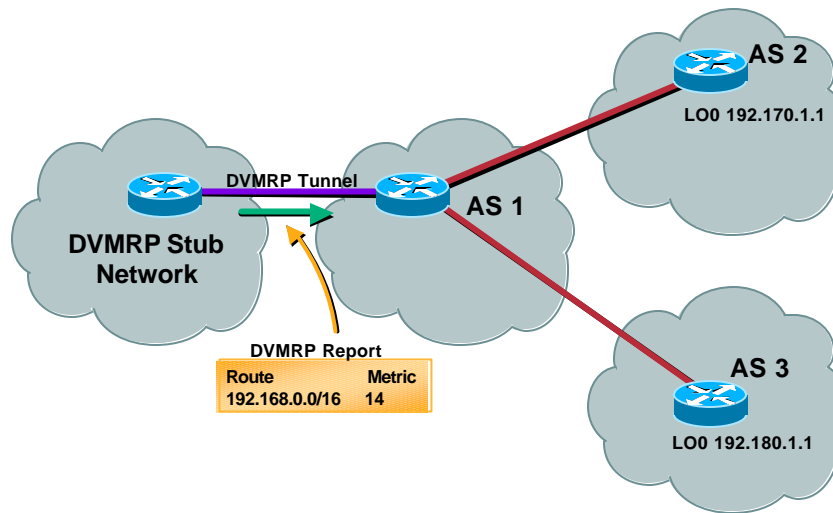
- **DVMRP to MBGP Redistribution**
  – When the router in AS1 receives a DVMRP Route Report for network 192.168.0.0/16, it is installed in the DVMRP Route table.
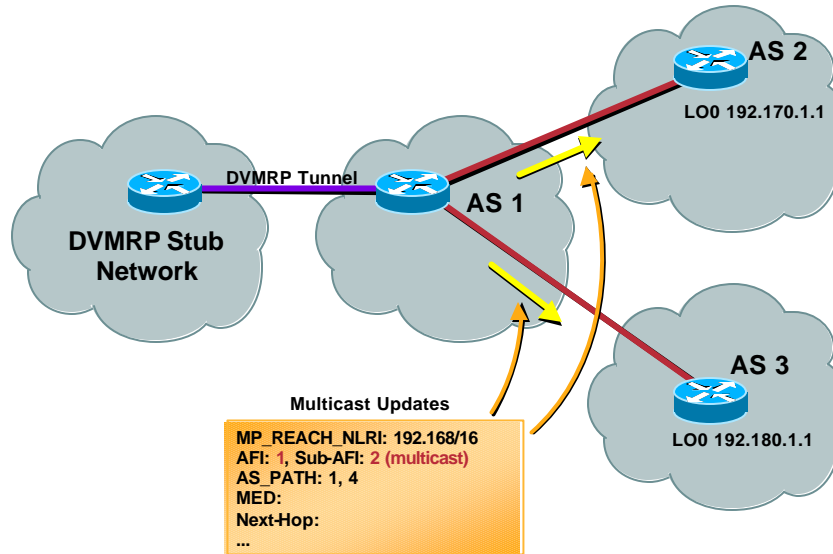
DVMRP to MBGP Redistribution

• DVMRP to MBGP Redistribution

- – As a result of the Route Map associated with the "redistribute dvmrp" command that was configured on the router in AS1, the DVMRP route, 192.168.0.0/16 is a match and is injected into the MBGP M-RIB. Once this has been done, the 192.168.0.0/16 prefix is advertised to MBGP peer routers in AS2 and AS3 as a Multicast NLRI.

MBGP to DVMRP Redistribution

- **MBGP routes can be sent into DVMRP**

  ```
  interface tunnel0
  ip dvmrp metric 1 route-map <map> mbgp
  ```

- **Can use typical match operations**

- **However, we recommend tail sites using DVMRP access to accept DVMRP default route**

 8/14/2001 3:35 PM 78

- **MBGP to DVMRP Redistribution**
  - Prefixes in the MBGP M-RIB may also be advertised as DVMRP routes to the DVMRP neighbor at the other end of Tunnel 0 by the use of the following interface commands:
    - interface tunnel 0
    - ip dvmrp metric 1 route-map mrib-to-dvmrp mbgp
    - The above command causes any prefixes that are matched by the "mrib-to-dvmrp" Route Map to be sent in DVMRP Route Reports out Tunnel 0.
    - Note: The "mbgp" keyword on the above command is a little confusing as it really implies the M-RIB of the local MBGP process. This syntax may change in the near future.
  - Typically, ISP's reconfigure their connectivity to legacy DVMRP networks so that they are stub networks as they phase out support for DVMRP over time. This is done so that the ISP may simply their support for any remaining DVMRP networks by only injecting the default route into the DVMRP network.

12.1 MBGP Syntax Changes

- **MBGP History**
  - **Introduced in IOS 12.0S**
  - **'nlri' clause added to BGP syntax**
    - **Covered IPv4 Unicast and Multicast NLRI only**
    - **Didn't cover other protocols or address-families**
  - **Extending 'nlri' syntax considered**
    - **Rejected as being too inflexible**
- **New syntax as of 12.1/12.0(7)T**
  - **Exception: 12.0S train retains old syntax**

 8/14/2001 3:35 PM 79

- **MBGP History**
  - Support for Multiprotocol BGP was first introduced in IOS release 12.0S. In order to support different types of NLRI exchange, the "nlir" clause was added to many of the existing BGP configuration commands. However, this initial release only supported "ipv4 unicast" and "ipv4 multicast" NLRI.
  - In order to support other types of NLRI, (such as IPv6) it was decided that the nlri syntax was not suitable and a new change in the syntax was introduced beginning in IOS releases 12.1 and 12.0(7)T.
    - The 12.0S train still retains the old syntax.

12.1 MBGP Syntax Changes

- **New 'address-family' structure**

```
router bgp <asn>
 address-family <afi> [<sub-afi>]
              .
              .
 exit-address-family
```

  - **Replaces most 'nlri' clauses**
  - **Separates configurations by address family**
    - **Note: 'ipv4-unicast' is default address-family**
      - **Implied 'address-family ipv4 unicast' block**
      - **To override this default behavior, use:**
        ```
        no bgp default ipv4-unicast
        ```

   8/14/2001 3:35 PM 80

- **Address-family structure**
  - In order to support different types of address-family and sub-address family NLRI, the "address-family" block was added to the BGP configuration command syntax.
    - The address-family block replaces many of the old 'nlri' clauses. However, there are still some commands that retain the use of the 'nlri' keyword.
  - The address-family block is used to separate groups of configuration commands by address-family/sub-address family.
  - In order to remain as backwards compatible as possible, the default address family is 'ipv4 unicast'. This results in an "implied" 'address-family ipv4 unicast' block.
    - This default behavior is sometimes confusing because it merges common bgp configuration commands with ipv4 unicast specific commands. In addition, neighbor definition "implies" ipv4 unicast capability negotiation by default. This makes specifying "ipv4 multicast" only neighbors a bit confusing.
    - The 'no bgp default ipv4-unicast' command may be used to override this rather confusing behavior. When this command has been configured, a separate 'address-family ipv4 unicast' block can be configured. This allows the configuration to clearly separate ipv4 unicast and multicast into separate address-family blocks.

12.1 MBGP Syntax Changes

- **New 2-step neighbor configuration**
  - **Neighbors are first identified**

    ```
    neighbor <foo> remote-as <asn>
    ```
  - **Then they are 'activated' by address family**

    ```
    neighbor <foo> activate
    ```
    - **Controls MBGP capability negotiation**
  - **Exception:**
    - **Unicast neighbors are automatically activated in the implied 'address-family ipv4 unicast' block**
    - **This default behavior can be overridden with:**

      ```
      no bgp default ipv4-unicast
      ```

 8/14/2001 3:35 PM 81

- **2-Step Neighbor Configuration**
  - In the new bgp syntax, BGP neighbors (peers) are first identified and *then* activated by address family.
    - The well known '**neighbor <foo> remote-as <asn>**' command simply identifies the neighbor with which a peering connection is to be established.
    - The '**neighbor <foo> activate**' command is used inside of an address family block to specify that this type of NLRI is to be exchanged with neighbor <foo>. For example, the '**neighbor <foo> activate**' command in the **"address-family ipv4 multicast"** block, "activates" the exchange of ipv4 multicast NLRI exchange. The '**no neighbor <foo> activate**' form will explicitly disable the exchange of this type of NLRI with neighbor <foo>.
  - EXCEPTION:

    In order to provide backward compatibility, ipv4 unicast NLRI exchange is automatically activated in the "implied" ipv4 unicast address family block. *No explicit "activate" command is necessary to activate ipv4 unicast NLRI exchange.*
    - This default behavior is often confusing to someone trying to configure multiple types of NLRI exchange since the "**activate**" command must be configured for all other NLRI types *EXCEPT* ipv4 unicast.
    - To disable this behavior and permit explicit ipv4 unicast configuration using an ipv4 unicast address family block, the '**no bgp default ipv4-unicast**' command may be configured.

12.1 MBGP Syntax Changes

Address-Family Example:

Implied ipv4 unicast address family block with implied neighbor 'activate' commands

```
router bgp 101
 no synchronization
 bgp log-neighbor-changes
 network 172.16.21.0 mask 255.255.255.0
 network 192.168.1.0
 neighbor 172.16.1.2 remote-as 301
 neighbor 172.16.11.2 remote-as 201
 no neighbor 172.16.11.2 activate
 no auto-summary
 !
 address-family ipv4 multicast
 neighbor 172.16.11.2 activate
 network 172.16.21.0 mask 255.255.255.0
 network 192.168.1.0
 exit-address-family
```

Explicit ipv4 multicast address family block

- **Address Family Example**
  - In the above example, the "implied" ipv4 unicast address family block consists of the following lines:
    ```
    network 172.16.21.0 mask 255.255.255.0
    network 192.168.1.0
    neighbor 172.16.1.2 remote-as 301
    neighbor 172.16.11.2 remote-as 201
    no neighbor 172.16.11.2 activate
    ```
    - Notice that the '**neighbor 172.16.1.2 remote-as 301**' command also *implies* a '**neighbor 172.16.1.2 activate**' for ipv4 unicast NLRI exchange to neighbor 172.16.1.2.
    - In addition, the '**no neighbor 172.16.11.2 activate**' overrides the *implied* '**neighbor 172.16.11.2 activate**' for ipv4 unicast NLRI exchange to neighbor 172.16.11.2.
  - The bottom section of the example is the ipv4 multicast address family block. Notice that only neighbor 172.16.11.2 is activated for ipv4 multicast NLRI exchange. In this case, the '**no neighbor 172.16.1.2 activate**' is implied.

12.1 MBGP Syntax Conversion

- **12.0S – 12.1 Syntax Conversion**
  - **Occurs when upgrading to 12.1 or later**
  - **Can result in some confusing changes**

 8/14/2001 3:35 PM 83

- **12.1 Syntax Conversion**
  - Conversion of the old 12.0S syntax to the new 12.1 syntax occurs automatically when the IOS version running in the router is upgraded from 12.0S to 12.1.
  - Unfortunately, the conversion can result in some configurations that are not immediately obvious as to what they do.

12.0S – 12.1 Syntax Conversion

- **Before Conversion**

```
router bgp 5
  network 171.69.214.0 mask 255.255.255.0 nlri unicast multicast
  neighbor 171.69.214.38 remote-as 2 nlri unicast
  neighbor 171.69.214.50 remote-as 2 nlri multicast
```

- **After Conversion**

```
router bgp 5
  network 171.69.214.0 mask 255.255.255.0
  neighbor 171.69.214.38 remote-as 2
  neighbor 171.69.214.50 remote-as 2
  no neighbor 171.69.214.50 activate      ◄── Overrides implied 'neighbor activate'
  !                                            for the ipv4 unicast address family
  address-family ipv4 multicast
  neighbor 171.69.214.50 activate
  network 171.69.214.0 mask 255.255.255.0
  exit-address-family
```

- **Syntax Conversion Example**

    Prior to conversion, the 12.0S configuration was:

```
router bgp 5
  network 171.69.214.0 mask 255.255.255.0 nlri unicast multicast
  neighbor 171.69.214.38 remote-as 2 nlri unicast
  neighbor 171.69.214.50 remote-as 2 nlri multicast
```

    After the conversion, the configuration has the *implied* ipv4 address block in the top lines as follows:

```
router bgp 5
  network 171.69.214.0 mask 255.255.255.0
  neighbor 171.69.214.38 remote-as 2
  neighbor 171.69.214.50 remote-as 2
  no neighbor 171.69.214.50 activate
```

- Again, the '**neighbor 172.69.224.38 remote-as 2**' command in this *implied* ipv4 unicast address family block has an *implied* '**neighbor 172.69.224.38 activate**'. This automatically activates ipv4 unicast NLRI exchange as was taking place in the original 12.0S configuration.

- In addition, the '**no neighbor 172.69.214.50 activate**' *overrides* the implied '**neighbor 172.69.214.50 activate**' that would normally occur in the implied ipv4 unicast block. This prevents ipv4 unicast NLRI exchange with this neighbor.

- The ipv4 multicast address family block also contains the '**neighbor 172.69.214.50** activate' command which explicitly activates ipv4 multicast NLRI exchange with this neighbor

– Finally, the '**network 171.69.224.0 mask 255.255.255.0**' command appears in both the *implied* ipv4 unicast address family block AND the *explicit* ipv4 multicast address family block. This causes this network to be injected into both the ipv4 Unicast and Multicast RIBs.

12.1 MBGP Syntax Tricks

- **Disabling default ipv4 unicast family**

  ```
  no bgp default ipv4-unicast
  ```

  - **Allows a clear separation of unicast and multicast configurations**
  - **All ipv4 unicast commands are placed in a separate (explicit) address-family block**

 8/14/2001 3:35 PM 85

- **Syntax Tricks**
  - By using the '**no bgp default ipv4-unicast**' command, we can disable the default, *implied* ipv4 unicast address family block and its *implied* neighbor activation commands. (Which can be quite confusing in a multiprotocol envirionment.)
    - When this command is configured, it allows a clear separation of ipv4 unicast and multicast configurations commands. In addition, 'activate' commands must be explicitly configured in each section.

12.1 MBGP Syntax Changes

Using 'no bgp default ipv4-unicast'

Common neighbor definition block

Explicit ipv4 unicast address family block

Explicit ipv4 multicast address family block

```
router bgp 101
 no synchronization
 bgp log-neighbor-changes
 no bgp default ipv4-unicast
 neighbor 172.16.1.2 remote-as 301
 neighbor 172.16.11.2 remote-as 201
 no auto-summary
 !
 address-family ipv4 unicast
 network 172.16.21.0 mask 255.255.255.0
 network 192.168.1.0
 neighbor 172.16.1.2 activate
 exit-address-family
 !
 address-family ipv4 multicast
 neighbor 172.16.11.2 activate
 network 172.16.21.0 mask 255.255.255.0
 network 192.168.1.0
 exit-address-family
```

- **Example Configuration with 'no bgp default ipv4-unicast'**
  - In this example, the '**no bgp default ipv4-unicast**' command has been added to the common bgp configuration block at the top of the bgp configuration.
    - In this section, we also identify the bgp neighbors using the normal neighbor command. However, no 'activate' commands are implied because the '**no bgp default ipv4-unicast**' command has been configured.
  - The next section is the explicit ipv4 unicast address family block. In this section we see that two network prefix are being injected into the ipv4 Unicast RIB and that neighbor 172.16.1.2 has been activated for ipv4 unicast NLRI exchange. (No others are implied.)
  - The last section is the ipv4 multicast address family block. Here too the same network prefix are being injected into the ipv4 Multicast RIB and that neighbor 172.16.11.2 has been activated for ipv4 multicast NLRI exchange.

Debugging MBGP

show ip bgp neighbor

```
asimov# show ip bgp neighbor
BGP neighbor is 10.0.10.3, remote AS 1, internal link
Index 2, Offset 0, Mask 0x4
BGP version 4, remote router ID 193.78.81.4
BGP state = Established, table version = 4, up for 22:32:50
Last read 00:00:49, hold time is 180, keepalive interval is 60 seconds
Neighbor NLRI negotiation:
Configured for unicast and multicast routes
Peer negotiated unicast and multicast routes
Exchanging unicast and multicast routes
Minimum time between advertisement runs is 5 seconds
Received 8916 messages, 0 notifications, 0 in queue
Sent 8923 messages, 0 notifications, 0 in queue
Connections established 4; dropped 3
Last reset 22:32:59, due to User reset
0 accepted unicast prefix consume 0 bytes of memory
0 history unicast paths consume 0 bytes of memory
Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Local host: 10.0.10.1, Local port: 11004
Foreign host: 10.0.10.3, Foreign port: 179
```

- **Debugging MBGP**
  - The above command may be used to debug the status of a (M)BGP peer connection with a neighbor. Notice that the highlighted text indicates exactly what capabilities and NLRI are being exchanged between the router and this peer.
    - Note: If the "Neighbor NLRI negotiation" field is missing, only unicast NLRI information is being exchanged.

Debugging MBGP

Old 12.0S Syntax

show ip bgp

```
asimov# show ip bgp
BGP table version is 4, local router ID is 10.0.100.1
Status codes: s suppressed, d damped, h history, * valid,
> best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

Network          Next Hop      Metric LocPrf Weight Path
*>10.0.100.0/24  0.0.0.0            0          32768 i
```

show ip mbgp

```
asimov# show ip mbgp
MBGP table version is 6, local router ID is 10.0.100.1
Status codes: s suppressed, d damped, h history, * valid, >
best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

Network          Next Hop       Metric LocPrf Weight Path
*>10.0.70.0/24   10.0.20.2      307200        32768 i
*>10.0.80.0/24   10.0.20.2       10000        32768 ?
```

- **Debugging MBGP**
  - Two different commands are currently used to display the contents of the Unicast RIB and the Multicast RIB. These are:
    - "show ip bgp"   Shows the contents of the U-RIB
    - "show ip mbgp"   Shows the contents of the M-RIB
  - The information displayed by the above commands is fundamentally the same. The only difference is on is the contents of the Unicast RIB and the other is the contents of the Multicast RIB.
  - Note:The syntax of the above commands will change in the near future to avoid the confusing practice of referring to "mbgp" as meaning Multicast NLRI or Multicast RIB instead of "Multiprotocol BGP".

Debugging MBGP

New 12.1 Syntax

show ip bgp ipv4 unicast

```
asimov# show ip bgp ipv4 unicast
BGP table version is 4, local router ID is 10.0.100.1
Status codes: s suppressed, d damped, h history, * valid,
> best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

Network           Next Hop      Metric LocPrf Weight Path
*>10.0.100.0/24   0.0.0.0            0         32768 i
```

show ip bgp ipv4 multicast

```
asimov# show ip bgp ipv4 multicast
BGP table version is 6, local router ID is 10.0.100.1
Status codes: s suppressed, d damped, h history, * valid, >
best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

Network           Next Hop      Metric LocPrf Weight Path
*>10.0.70.0/24    10.0.20.2     307200        32768 i
*>10.0.80.0/24    10.0.20.2      10000        32768 ?
```

- **Debugging MBGP**
  - Two different commands are currently used to display the contents of the Unicast RIB and the Multicast RIB. These are:
    - "show ip bgp"   Shows the contents of the U-RIB
    - "show ip mbgp"   Shows the contents of the M-RIB
  - The information displayed by the above commands is fundamentally the same. The only difference is on is the contents of the Unicast RIB and the other is the contents of the Multicast RIB.
  - Note:The syntax of the above commands will change in the near future to avoid the confusing practice of referring to "mbgp" as meaning Multicast NLRI or Multicast RIB instead of "Multiprotocol BGP".

Debugging MBGP

Same for both Old and New Syntax

- **MBGP "debug" commands**

  `debug ip mbgp updates`

  - **Logs Multicast related information passed in MBGP update messages.**

  `debug ip bgp updates`

  - **Logs Unicast related information passed in MBGP update messages.**

  `debug ip mbgp dampening [<acl>]`

  - **Logs multicast route flap dampening activity.**

  `debug ip bgp dampening [<acl>]`

  - **Logs multicast route flap dampening activity.**

 8/14/2001 3:35 PM 90

- **Debugging MBGP**
  - Use the following command to display Multicast NLRI passed in MBGP update messages.
    - debug ip mbgp updates
  - Use the following command to display Unicast NLRI passed in MBGP update messages.
    - debug ip mbgp updates
  - Use the following command to display Multicast route flap dampening activity.
    - debug ip mbgp dampening [<acl>]
  - Use the following command to display Unicast route flap dampening activity.
    - debug ip bgp dampening [<acl>]
  - Note: The syntax of the above commands will change in the near future.

MBGP—Summary

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

 8/14/2001 3:35 PM 91

- **MBGP — Summary**
  - MBGP solves part of the inter-domain multicast problem by allowing AS's to exchange Multicast RPF information in the for of MBGP Multicast NLRI.
    - Because this is accomplished using an extension to the BGP protocol to make it support multiple protocols (i.e. "Multiprotocol BGP"), the same BGP configuration knobs are available for both Unicast and Multicast information.
    - The separation of unicast and multicast prefixes into separate Unicast and Multicast RIB's permits unicast and multicast traffic to follow different paths if desired.
  - MBGP is only one piece of the overall Inter-domain Multicast solution and PIM must still be used to:
    - Build the multicast distribution trees. (Typically via PIM-SM.)
    - Actually RPF check and forward multicast traffic.
    - PIM-SM is recommended as it permits the use of MSDP which solves most of the remaining issues and is covered in another section.