



Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

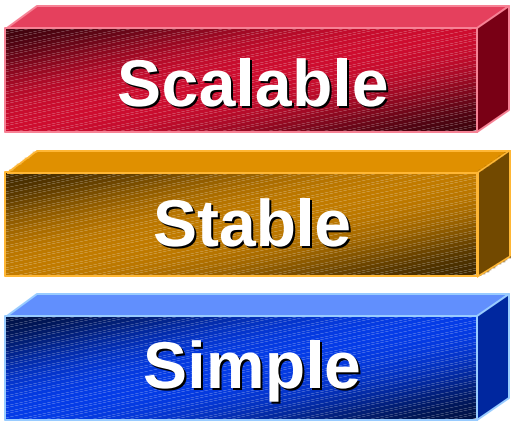
Complex Network Scalability

"BGP is the protocol **brains** that controls the router **brawn** between different Internet service providers…"

Boardwatch Magazine, April 1999,
Scaling Internet and Data Services...

© 1999, Cisco Systems, Inc.

Complex Network Scalability



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Agenda

- Scaling Your Network
- Case Studies
  - Troubleshooting
- BGP Extensions

© 1999, Cisco Systems, Inc.



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

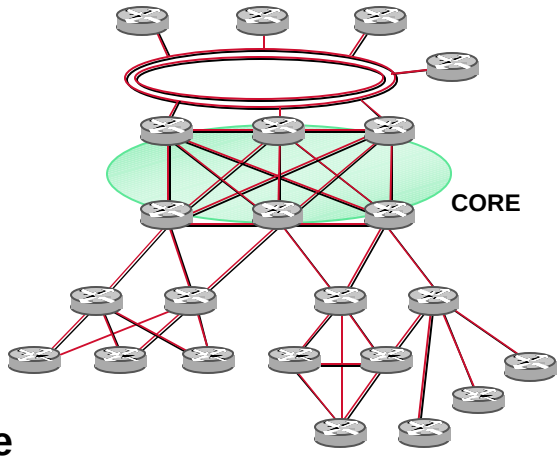
IGP Limitations

- **Amount of routing information in the network**
  - **Periodic updates/flooding**
  - **Long convergence times**
  - **Affects the core first**
- **Policy definition**
  - **Not easy to do**

© 1999, Cisco Systems, Inc.

BGP Cores—Sample Network

- **Geographically distributed**
- **Hierarchical**
- **Redundant**
- **Media independent**
- **A clearly identifiable core**



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

iBGP Core Migration Plan

- Configure BGP in all the core routers
  - Transit path
  - Turn synchronization off
- Route Generation
  - Use static routes to create summaries
  - Redistribution from the IGP is NOT recommended as it may cause instability

© 1999, Cisco Systems, Inc.

iBGP Core Migration Plan (Cont.)

- Route Generation—Example:

```
!
router bgp 109
network 200.200.200.0
network 201.201.0.0 mask 255.255.0.0
!
ip route 200.200.200.0 255.255.255.0 null0
ip route 201.201.0.0 255.255.0.0 null0
!
```


© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

iBGP Core Migration Plan (Cont.)

- Verify consistency of routing information
  - Compare the routing table against the BGP table—they must match!
- Change the distance parameters so that the BGP routes are preferred
  - distance bgp 20 20 20
  - All IGPs have a higher administrative distance

© 1999, Cisco Systems, Inc.

iBGP Core Migration Plan (Cont.)

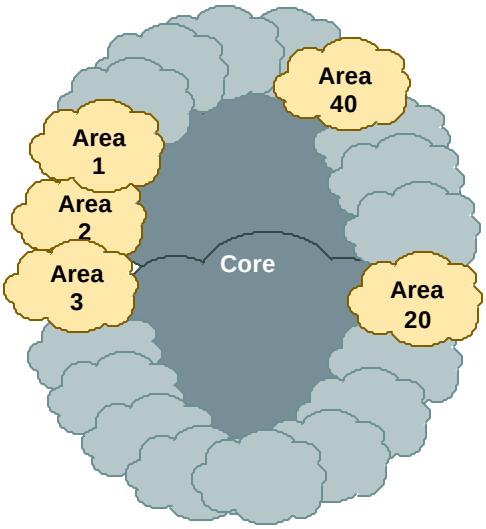
- Filter "non-core" IGP routes
  - Method will depend on the IGP used
  - May require the use of a different IGP process in the core if using a link state protocol
  - The routes to reach all the core links plus the BGP peering addresses must be carried by the IGP

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

iBGP Core Before...

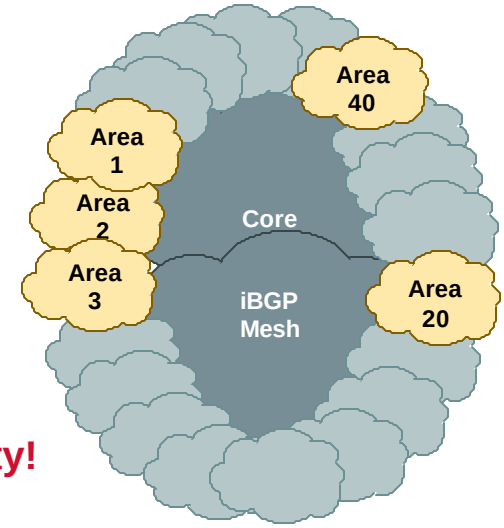
- IGP carries all the routes
- The core routers may be stressed due to the large number of routes



317
0901_04F9_c3 © 1999, Cisco Systems, Inc. www.cisco.com 13

iBGP Core After...

- Core:
  - IGP carries only core links plus peering address information
  - BGP carries all the routes
  - Increased Stability!



317
0901_04F9_c3 © 1999, Cisco Systems, Inc. www.cisco.com 14

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

iBGP Core Results

- The routes from the core cannot be redistributed back into the IGP
  - Non-core areas need a default route
  - Amount of routing information in non-core areas has been reduced!
- Full logical iBGP mesh
- External connections must be located in the core

© 1999, Cisco Systems, Inc.

Scaling Issues

- Full mesh core
  - High number of neighbors
  - Update generation
- Complex topologies
  - Not a “simple” hierarchical network
  - Multiple external and/or inter-region connections
  - Policy definition and enforcement

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Scaling Issues—Solutions

- **Reduce the number of updates**
  - **Peer groups**
- **Reduce the number of neighbors**
  - **Confederations**
  - **Route reflectors**
- **Use additional information to effectively apply policies**
  - **eBGP provides extra granularity**
  - **Confederations**

© 1999, Cisco Systems, Inc.



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

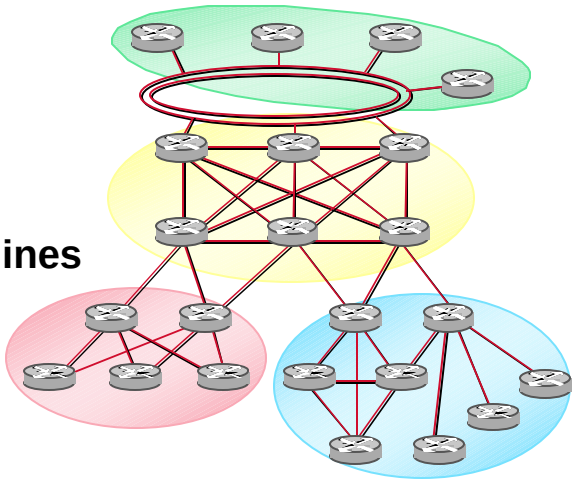
Implementation Strategy

- Divide the network into multiple regions/areas
- Connect each region using BGP
- Reconfigure the IGP in each region/area

© 1999, Cisco Systems, Inc.

Divide the Network into Pieces

- Where:
  - Geography
  - Department lines
  - Hierarchy
  - Etc.



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

eBGP Connections

- Assign an ASN to each region
  - Private ASNs maybe used and must be removed at the border of the network
    - neighbor x.x.x.x remove-private-AS
  - External connections only at the core
- Apply policy at inter-AS borders
  - May use AS_PATH filters to permit or deny route propagation to other regions

© 1999, Cisco Systems, Inc.

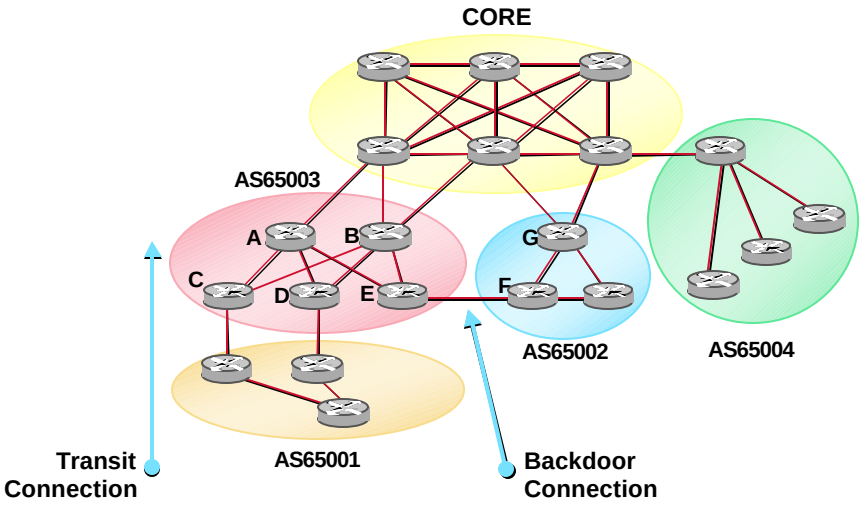
eBGP Connections (Cont.)

- Only the routers connected to the core need to run BGP
  - iBGP mesh in the core
- …Except if backdoor or transit connections exist
  - Routers in the transit path need to run BGP too

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

eBGP Connections (Cont.)



© 1999, Cisco Systems, Inc.

eBGP Connections—Routing

- Source the local routes for each AS at the border BGP routers
  - Use static routes and network statements
  - Verify consistency of routing information
- What about the IGP?
  - For each region/area it must carry routes to the infrastructure (all links), peering addresses and local destinations
  - Filter at the borders
  - May need to use an independent IGP process per AS

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Confederations

- Divide the AS into sub-AS
  - eBGP between sub-AS, but some iBGP information is kept
    - Preserve NEXT_HOP across the sub-AS (IGP carries this information)
    - Preserve LOCAL_PREF and MED

© 1999, Cisco Systems, Inc.

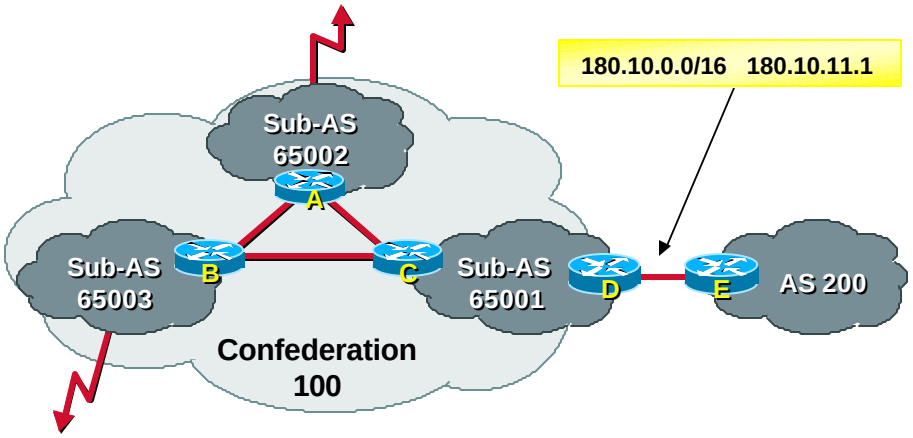
Confederations (Cont.)

- Visible to outside world as single AS
  - Each sub-AS uses a number from the private space
- iBGP speakers in sub-AS are fully meshed
  - The total number of neighbors is reduced by limiting the full mesh requirement to only the peers in the sub-AS

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Confederations—NEXT_HOP



© 1999, Cisco Systems, Inc.

Route Propagation Decisions

- **Same as with "normal" BGP:**
  - **From peer in same sub-AS → only to external peers**
  - **From external peers → to all neighbors**
- **"External peers" refers to**
  - **Peers outside the confederation**
  - **Peers in a different sub-AS**
    - **Preserve LOCAL_PREF, MED and NEXT_HOP**

© 1999, Cisco Systems, Inc.

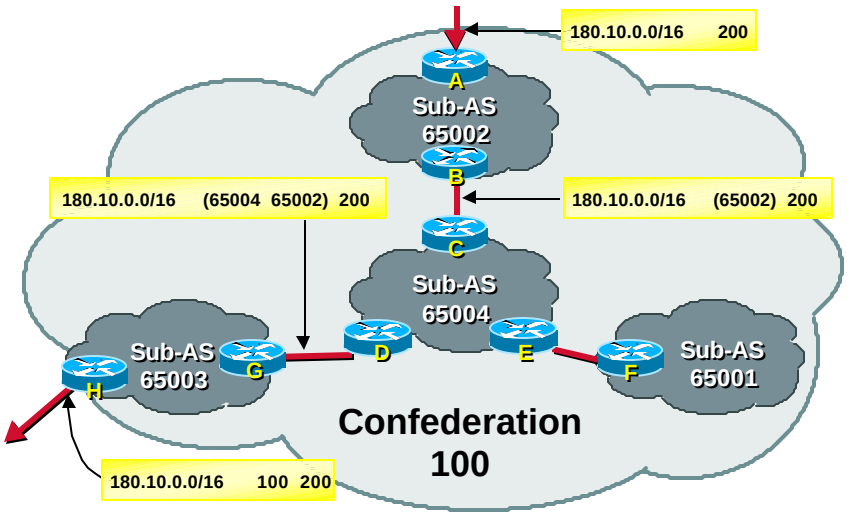
Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Confederations—AS_PATH

- Sub-AS traversed are carried as part of AS_PATH (AS_CONFED_SEQUENCE or AS_CONFED_SET) for loop avoidance
  - Not counted as regular AS when comparing AS_PATH
  - Paths with only confederation ASNs in the AS_PATH are skipped during MED comparison
    - bgp bestpath med confed

© 1999, Cisco Systems, Inc.

Confederation—AS_PATH (Cont.)



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Confederations—Migration I

- Same steps as when using eBGP connections, but external connections may be located anywhere in the network!
- What about the IGP?

  It must carry routes to the infrastructure (all links) and peering addresses (including external NEXT_HOP)

  One instance of the IGP for the whole AS

© 1999, Cisco Systems, Inc.

Confederations—Migration II

- Migration from a full iBGP mesh may be tricky as all the routers must be configured at one time

  bgp confederation identifier *realASN*

  bgp confederation peers *otherASNs*

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Confederations or Not?

	Internet Connectivity	Multi-Level Hierarchy	Policy Control	IGP	Migration Complexity
Confederations	Anywhere in the Network	Yes	Yes	One Instance Across the Network	Medium to High
eBGP Connections	Only in the Core	Yes	Yes	May Need Different Instances in Each Region	Low to Medium

Scalability and Stability Achieved by Both Methods!

© 1999, Cisco Systems, Inc.



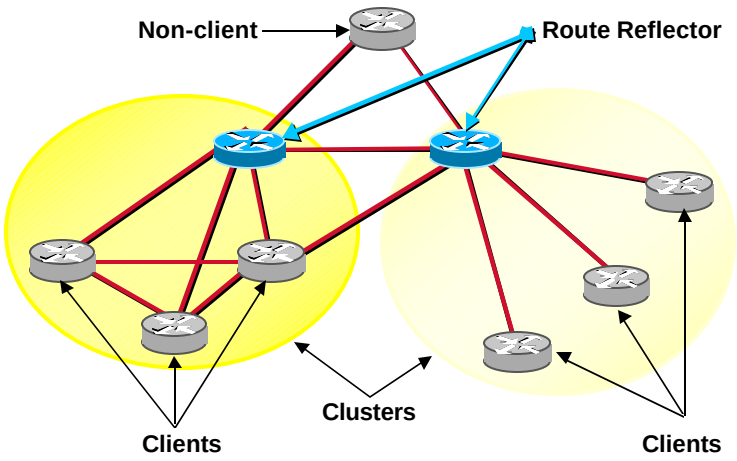
Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Route Reflectors

- **Provide additional control to allow router to advertise (reflect) iBGP learned routes to other iBGP peers**
  - **Method to reduce the size of the iBGP mesh**
- **Normal BGP speakers can coexist**
  - **Only the RR has to support this feature**
  - **neighbor x.x.x.x route-reflector-client**

© 1999, Cisco Systems, Inc.

Route Reflectors—Terminology



Lines Represent Both Physical Links and BGP Logical Connections

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Route Reflectors—Terminology (Cont.)

- **Route reflector**

  Router that reflects the iBGP information

- **Client**

  Routers between which the RR reflects updates (may be fully meshed among themselves)

- **Cluster**

  Set of one or more RRs and their clients (may overlap)

- **Non-client**

  iBGP neighbour outside the cluster

© 1999, Cisco Systems, Inc.

Route Reflectors—Loop Avoidance

- **Originator_ID attribute**

  Carries the RID of the originator of the route in the local AS (created by the RR)

- **Cluster_list attribute**

  The local cluster-id is added when the update is sent to (added by the RR)

  bgp cluster-id x.x.x.x

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

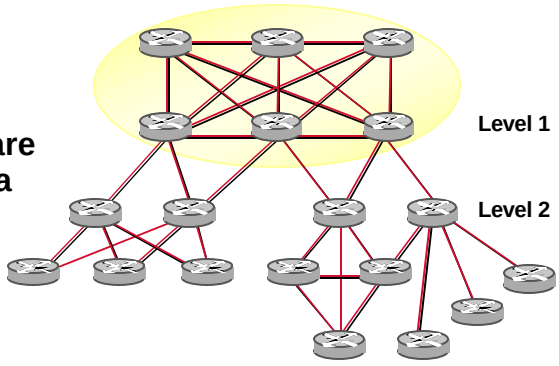
Reflection Decisions

- **Once the best path is selected:**
  - **From non-client reflect to all clients**
  - **From client → reflect to all non-clients AND other clients**
  - **From eBGP peer → reflect to all clients and non-clients**

© 1999, Cisco Systems, Inc.

Route Reflectors—Hierarchy

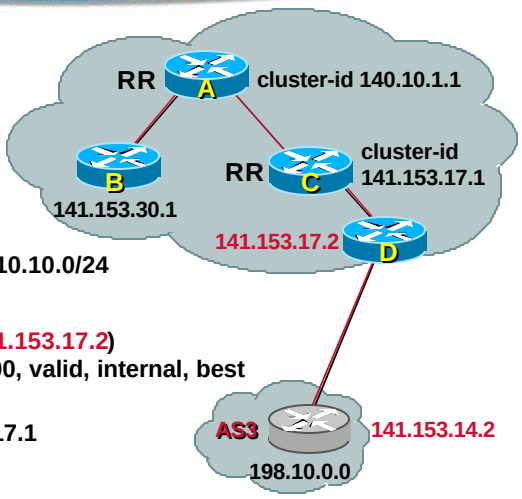
- **Clusters may be configured hierarchically**
  - **RRs in a cluster are clients of RRs in a higher level**
  - **Provides a "natural" method to limit routing information sent to lower levels**



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Hierarchical Route Reflectors



routerB>sh ip bgp 198.10.10.0

BGP routing table entry for 198.10.10.0/24

3

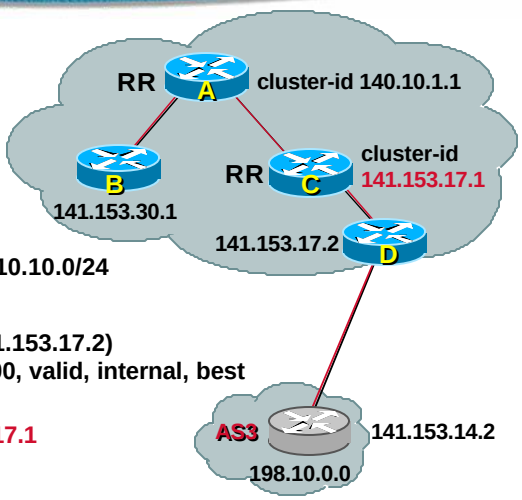
141.153.14.2 from 140.10.1.1 (141.153.17.2)
Origin IGP, metric 0, localpref 100, valid, internal, best

Originator : 141.153.17.2
Cluster list: 144.10.1.1, 141.153.17.1

Lines represent both physical links
and BGP logical connections

© 1999, Cisco Systems, Inc.

Hierarchical Route Reflectors



routerB>sh ip bgp 198.10.10.0

BGP routing table entry for 198.10.10.0/24

3

141.153.14.2 from 140.10.1.1 (141.153.17.2)
Origin IGP, metric 0, localpref 100, valid, internal, best

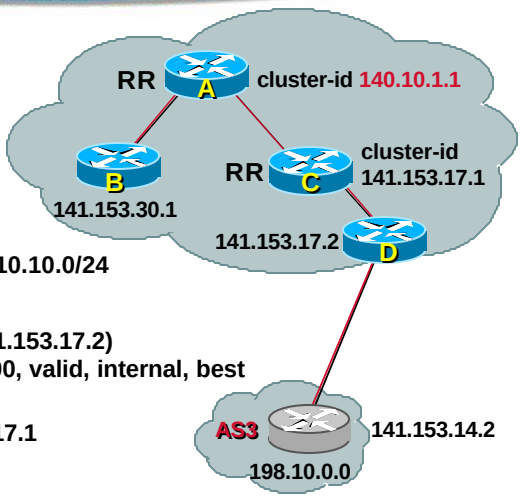
Originator : 141.153.17.2
Cluster list: 144.10.1.1, 141.153.17.1

Lines represent both physical links
and BGP logical connections

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Hierarchical Route Reflectors



routerB>sh ip bgp 198.10.10.0
BGP routing table entry for 198.10.10.0/24
3
141.153.14.2 from 140.10.1.1 (141.153.17.2)
Origin IGP, metric 0, localpref 100, valid, internal, best
Originator : 141.153.17.2
Cluster list: 144.10.1.1, 141.153.17.1

Lines represent both physical links
and BGP logical connections

© 1999, Cisco Systems, Inc.

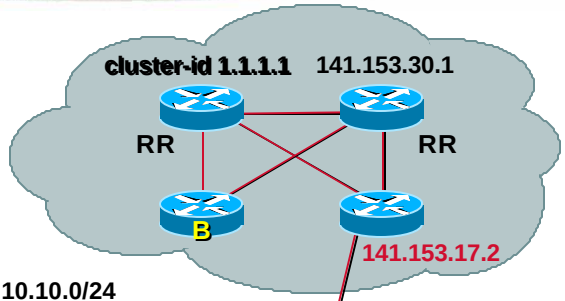
Route Reflectors—Redundancy

- Multiple RRs can be configured in the same cluster
  - Other RRs in the same cluster should be treated as iBGP peers (non-clients)
  - All RRs in the cluster must have the same cluster-id
- A router may be a client for RRs in different clusters

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Multiple Route Reflectors



routerB>sh ip bgp 198.10.10.0
BGP routing table entry for 198.10.10.0/24
3
141.153.14.2 from 141.153.30.1 (141.153.17.2)
Origin IGP, metric 0, localpref 100, valid, internal, best
Originator: 141.153.17.2
Cluster list: 1.1.1.1

eBGP

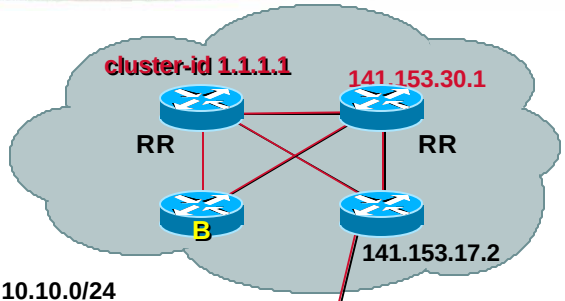


141.153.14.2
198.10.10.0/24s

Lines Represent Both Physical
Links and BGP Logical Connections

© 1999, Cisco Systems, Inc.

Multiple Route Reflectors



routerB>sh ip bgp 198.10.10.0
BGP routing table entry for 198.10.10.0/24
3
141.153.14.2 from 141.153.30.1 (141.153.17.2)
Origin IGP, metric 0, localpref 100, valid, internal, best
Originator: 141.153.17.2
Cluster list: 1.1.1.1

eBGP



141.153.14.2
198.10.10.0/24s

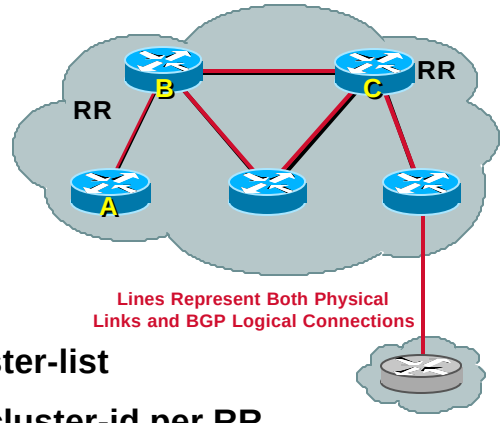
Lines Represent Both Physical
Links and BGP Logical Connections

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Multiple Route Reflectors

- The cluster-id must be different, otherwise B will not reflect any route to A if coming from C
  - B will detect its own cluster-id in the cluster-list
  - Tip: use a different cluster-id per RR



© 1999, Cisco Systems, Inc.

Route Reflectors—Migration

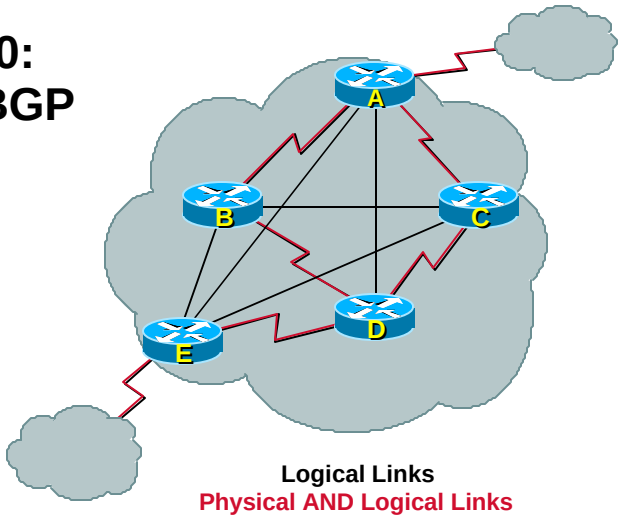
- Where to place the route reflectors?
  - Follow the physical topology!
  - This will guarantee that the packet forwarding won't be affected
- Configure one RR at a time
  - Eliminate redundant iBGP sessions
  - Place one RR per cluster

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Route Reflectors—Migration

- **Step 0: full iBGP mesh**

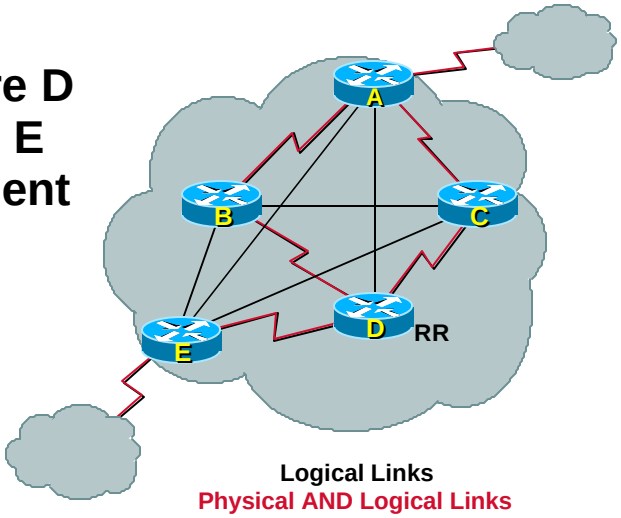


Logical Links
Physical AND Logical Links

© 1999, Cisco Systems, Inc.

Route Reflectors—Migration

- **Step 1: configure D as a RR; E is the client**



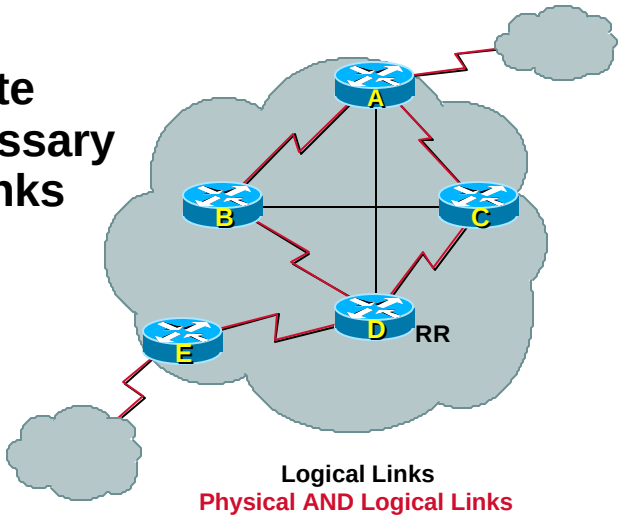
Logical Links
Physical AND Logical Links

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Route Reflectors—Migration

- **Step 2: eliminate unnecessary iBGP links**

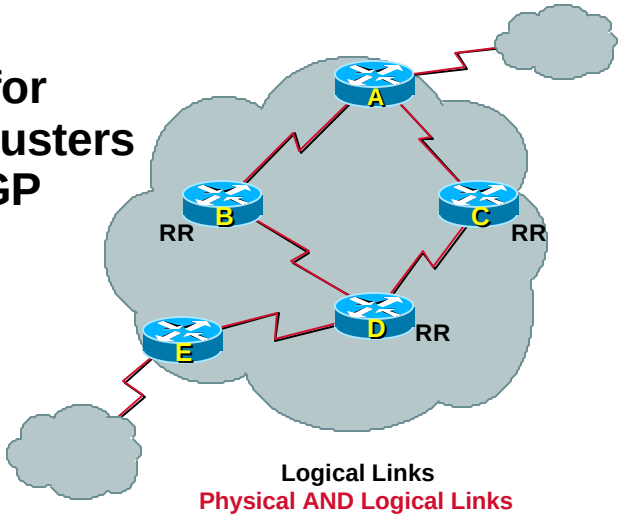


Logical Links
Physical AND Logical Links

© 1999, Cisco Systems, Inc.

Route Reflectors—Migration

- **Step 3: repeat for other clusters and iBGP links**



Logical Links
Physical AND Logical Links

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

RR: Other Issues

- The set clause for outbound route-maps does not affect routes reflected to iBGP peers
- The nexthop-self command will only affect the next-hop of eBGP learned routes (the next-hop of reflected routes should not be changed)

© 1999, Cisco Systems, Inc.

Route Reflectors—Results

- Number of neighbors is reduced
  - No need for full iBGP mesh
- Number of routes propagated is reduced
  - Each RR advertises only the best path to its clients
- Stability and Scalability are achieved!

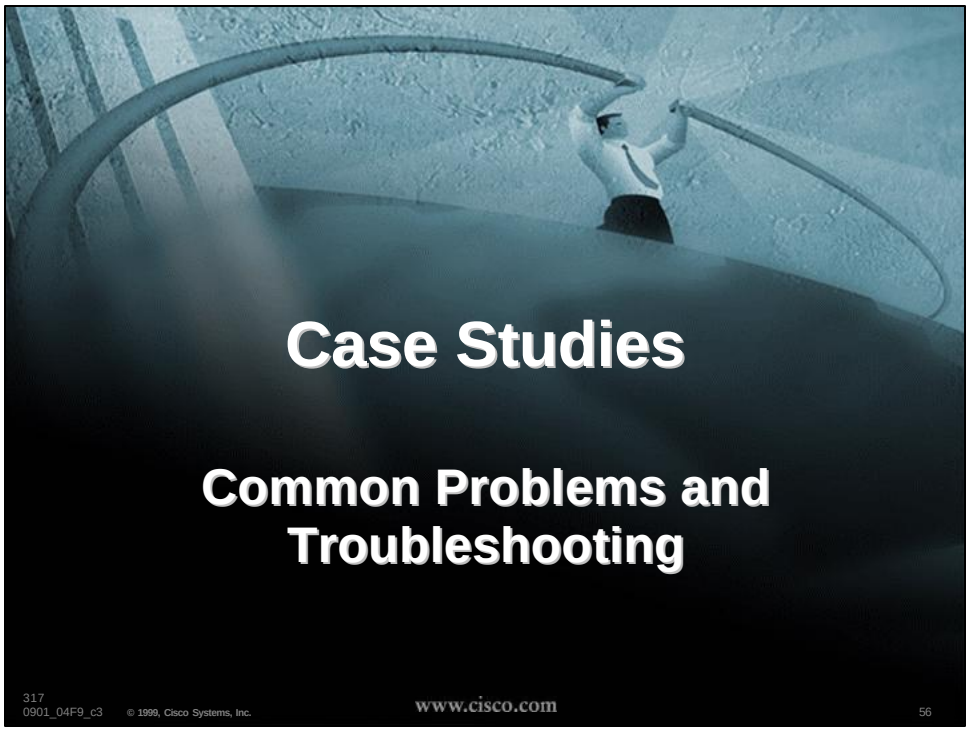
© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

To Reflect or Not to Reflect

	Internet Connectivity	Multi-Level Hierarchy	Policy Control	Scalability	Migration Complexity
Confederations	Anywhere in the Network	Yes	Yes	Medium	Medium to High
Route Reflectors	Anywhere in the Network	Yes	Yes	Very High	Very Low

317
0901_04F9_c3 © 1999, Cisco Systems, Inc. www.cisco.com 55



317
0901_04F9_c3 © 1999, Cisco Systems, Inc. www.cisco.com 56

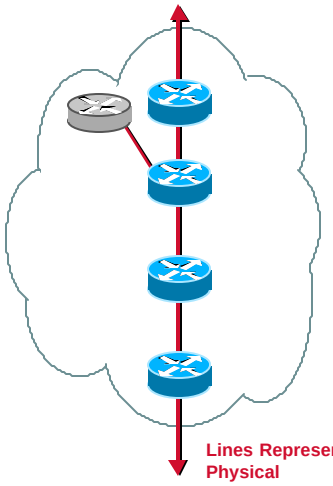
Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

RR—Physical Topology

- RRs relax the logical full-mesh requirements that iBGP has

  Some configurations… "may not yield the same route-selection result as that of the full iBGP mesh…"

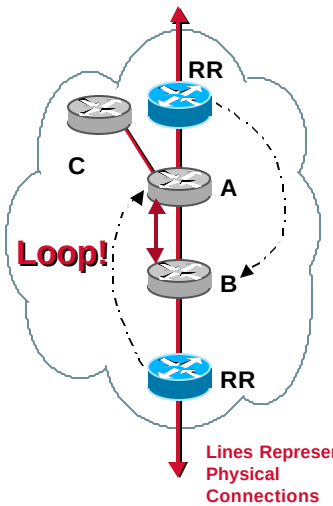
  draft-idr-route-reflect-v2, April 99



© 1999, Cisco Systems, Inc.

RR—Physical Topology

- Not following the physical topology may cause routing loops!



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

RR—Physical Topology

- **Symptom**

```
routerC#traceroute 7.7.7.7

Tracing the route to 7.7.7.7
  1 10.105.1.71 4 msec 4 msec 8 msec
rtrB 2 140.10.50.6 188 msec 4 msec 4 msec
rtrA 3 140.10.50.5 4 msec 4 msec 4 msec
  4 140.10.50.6 4 msec 8 msec 8 msec
  5 140.10.50.5 8 msec 8 msec 8 msec
  6 140.10.50.6 8 msec 4 msec 8 msec
```


© 1999, Cisco Systems, Inc.

RR—Physical Topology

```
routerA#show ip bgp 7.7.7.7
BGP routing table entry for 7.0.0.0/8
 1
   21.21.21.1 (metric 201) from 2.1.1.1 (2.1.1.1)
     Origin IGP,valid, internal, best
routerA#show ip route 21.21.21.1
Routing entry for 21.21.21.0/24
Routing Descriptor Blocks:
 * 140.10.50.6, from 140.10.50.6, via Serial0
```

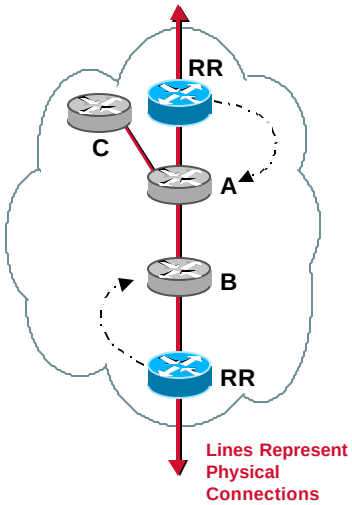
```
routerB#show ip bgp 7.7.7.7
BGP routing table entry for 7.0.0.0/8
 1
   22.22.22.1 (metric 201) from 3.3.3.1 (3.3.3.1)
     Origin IGP, valid, internal, best
routerB#show ip route 22.22.22.1
Routing entry for 22.22.22.0/24
Routing Descriptor Blocks:
 * 140.10.50.5, from 140.10.50.5, via Serial0
```


© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

RR—Physical Topology

- **Solution:**
  Follow the physical topology!

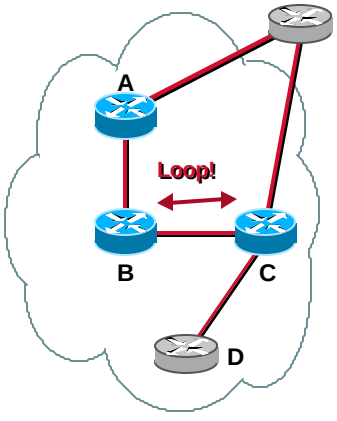


© 1999, Cisco Systems, Inc.

RR—Physical Topology II

- **Symptom**

```
routerD#traceroute 7.1.1.1
  1 1.1.1.2 24 msec 24 msec 40 msec
rtrB 2 156.1.1.1 28 msec 48 msec 24 msec
rtrC 3 156.1.1.2 24 msec 24 msec 24 msec
  4 156.1.1.1 28 msec 28 msec 24 msec
  5 156.1.1.2 28 msec 28 msec 28 msec
  6 156.1.1.1 28 msec 28 msec 32 msec
```



© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

RR—Physical Topology II

routerC#show ip bgp 7.0.0.0
BGP routing table entry for 7.0.0.0/8
1
150.10.10.1 (metric 115) from 150.10.10.1 (150.20.20.1)
Origin IGP, valid, external, best
routerC#show ip route 150.10.10.1
Routing entry for 150.10.10.1/32
Routing Descriptor Blocks:
* 156.1.1.1, from 150.20.20.1, via Ethernet2/1/1

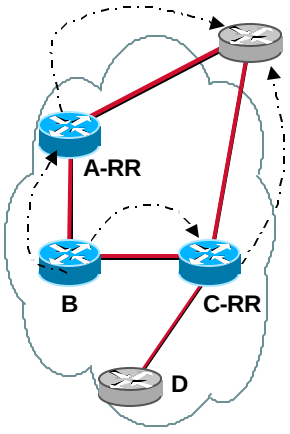
routerB#show ip bgp 7.0.0.0
BGP routing table entry for 7.0.0.0/8
1
156.1.1.2 from 156.1.1.2 (212.212.212.1)
Origin IGP, valid, internal, best
routerB#show ip route 156.1.1.2
Routing entry for 156.1.1.0/24
Routing Descriptor Blocks:
* directly connected, via Ethernet1

© 1999, Cisco Systems, Inc.

RR—Physical Topology II

- **Problem**

```
routerC#show running-config
router bgp 134
 neighbor 150.10.10.1 remote-as 1
 neighbor 150.10.10.1 ebgp-multihop 255
 neighbor 150.10.10.1 update-source Loopback0
 neighbor 156.1.1.1 remote-as 134
 neighbor 156.1.1.1 route-reflector-client
 neighbor 156.1.1.1 next-hop-self
!
```



Lines Represent
Physical Connections

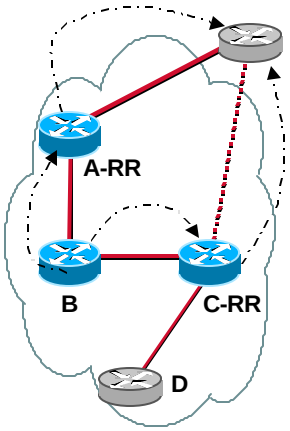
© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

RR—Physical Topology II

- **Problem**

```
routerC#show running-config
router bgp 134
 neighbor 150.10.10.1 remote-as 1
 neighbor 150.10.10.1 ebgp-multihop 255
 neighbor 150.10.10.1 update-source Loopback0
 neighbor 156.1.1.1 remote-as 134
 neighbor 156.1.1.1 route-reflector-client
 neighbor 156.1.1.1 next-hop-self
!
ip route 150.10.10.1 255.255.255.255 s0 250
```



Lines Represent
Physical Connections

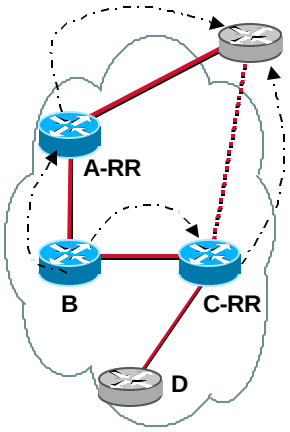
© 1999, Cisco Systems, Inc.

RR—Physical Topology II

- **Solution**

  Establish the eBGP peering permanently through the "backup" link

  Use LOCAL_PREF or MED to break any tie!



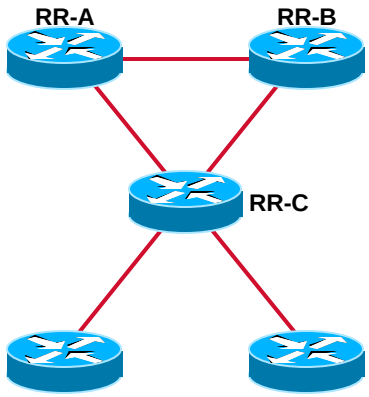
Lines Represent
Physical Connections

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Clusters with Multiple RRs

- It is possible to have multiple RRs in one cluster for redundancy
- Hierarchical clusters help scale your network

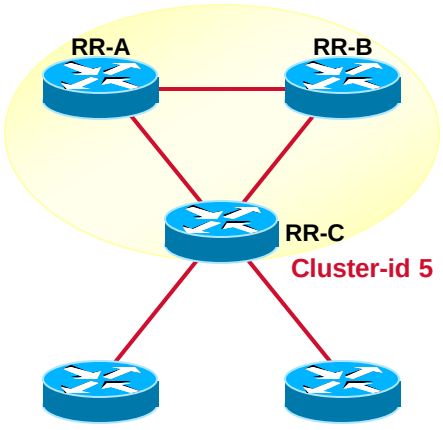


Lines Represent Physical and Logical Connections

© 1999, Cisco Systems, Inc.

Clusters with Multiple RRs

- A and B are core routers
  - Carry routes to the rest of the network
- Symptom
  - RR-C is not receiving any routes



Lines Represent Physical and Logical Connections

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

Clusters with Multiple RRs

- **Problem**

  After resetting the session and using debug ip bgp:

  BGP: 1.1.1.1 Route Reflector cluster loop received cluster-id 0.0.0.5
  BGP: 2.2.2.2 Route Reflector cluster loop received cluster-id 0.0.0.5

  C is configured with the same cluster-id as A and B!

  ```
  routerC:
  !
  router bgp 1
   bgp cluster-id 5
  ...
  !
  ```


© 1999, Cisco Systems, Inc.

Clusters with Multiple RRs

- **Solution**

  In hierarchical route reflector configurations, each level must have a different cluster-id

  Recommendation: use a different cluster-id per route reflector

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

eBGP Multihop

- **Symptom**

  The eBGP peering is established, but convergence is not complete even after several hours

```
routerA#show ip bgp summary

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
150.10.10.1     4     1    3550    3570      847    0  206 05:53:51        100
```


© 1999, Cisco Systems, Inc.

eBGP Multihop

```
routerA#show ip route 150.10.10.1
Routing entry for 150.10.10.1/32
  Routing Descriptor Blocks:
  10.105.1.71, from 150.20.20.1, 00:06:14 ago, via POS2/1/0
* 156.1.1.1, from 150.20.20.1, 00:06:14 ago, via POS2/1/1
routerA#ping 150.10.10.1
Sending 5, 100-byte ICMP Echos to 150.10.10.1:     !!!!!
Success is 100 percent, round-trip min/avg/max = 4/64/296 ms
```

```
Reply to request 0
 Record route:
   (156.1.1.2)
   (195.5.5.1)
   (10.105.1.134)
   (150.10.10.1)
   (10.105.1.76)
   (195.5.5.2)
   (156.1.1.1)
   (211.211.211.1) <*>

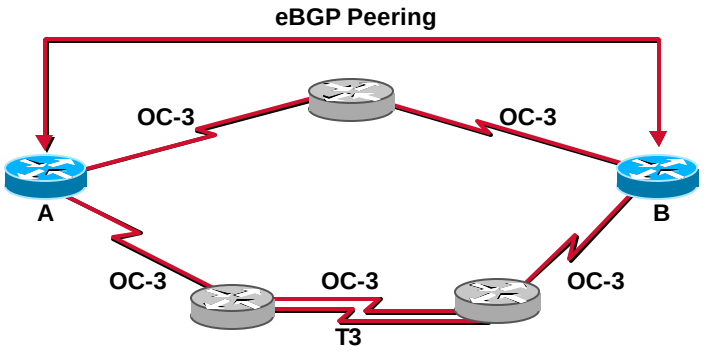
Reply to request 1
 Record route:
   (10.105.1.69)
   (140.10.50.5)
   (150.10.10.1)
   (140.10.50.6)
   (10.105.1.71)
   (211.211.211.1) <*>
```


© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

eBGP Multihop

- **Problem: peers configured with eBGP-multihop 2**



© 1999, Cisco Systems, Inc.

eBGP Multihop

- **Solution**

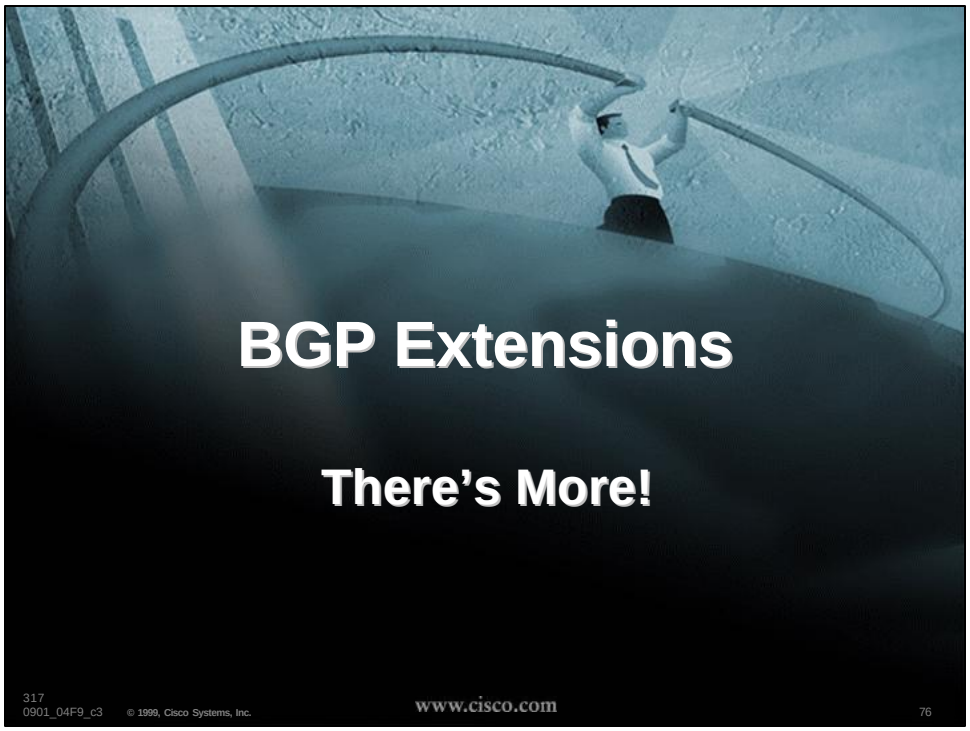
  The paths have different number of hops between them—make sure that the TTL is enough for the longest path

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

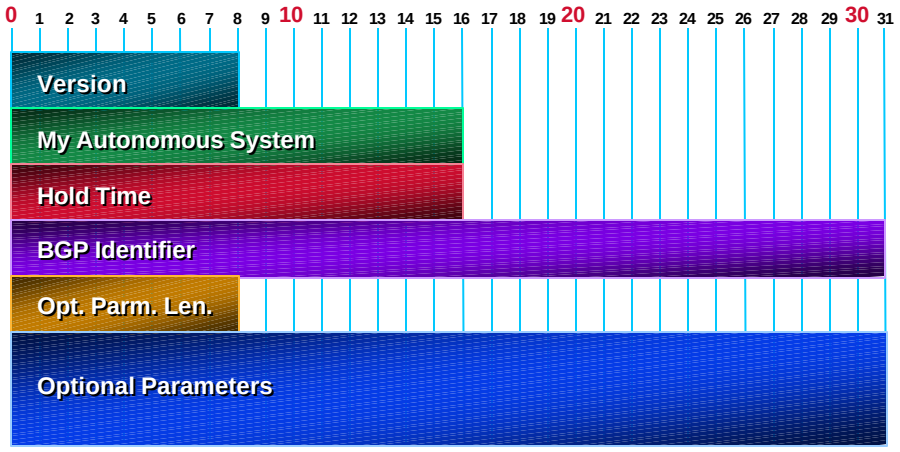
Common Problems—Conclusions

- BGP is a simple protocol
  - Straight forward state machine
  - Rides over TCP
  - Easy “basic” configuration
- BGP is also very flexible
  - Many options and knobs!



Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

OPEN Message



317
0901_04F9_c3 © 1999, Cisco Systems, Inc. www.cisco.com 77

Capabilities Negotiation

- Allows for the advertisement of capabilities (type 2)
- Backwards compatible
  - New error subcode introduced to indicate which capabilities are not supported—the session must be reset



draft-ietf-idr-bgp4-cap-neg, Feb. 1999

317
0901_04F9_c3 © 1999, Cisco Systems, Inc. www.cisco.com 78

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Capabilities Negotiation

- Current capabilities
  - 1 multiprotocol
  - 128 route refresh
  - 129 outbound route filter

© 1999, Cisco Systems, Inc.

Route Refresh Capability

- Facilitates non-disruptive policy changes
- No configuration is needed
- No additional memory is used
- clear ip bgp x.x.x.x [soft] in

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Managing Policy Changes

clear ip bgp <addr> [soft] [in|out]

- **<addr> may be any of the following**

x.x.x.x	IP address of a peer
*	all peers
ASN	all peers in an AS
external	all external peers
peer-group <name>	all peers in a peer-group

© 1999, Cisco Systems, Inc.

Outbound Route Filter Capability

- **Allows for the use of the neighbor's inbound prefix-list as part of the local outbound policy (Currently only for IPv4 unicast NLRI)**
  - **Reduces the number of updates**
  - **5 sec. delay after session is established, before updates are sent**

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.

PrefixList-ORF

- By default, this capability is not advertised to any neighbor

  neighbor x.x.x.x capability prefix-filter

  Can't be advertised to peer-group members

- To push out a prefix-list

  clear ip bgp x.x.x.x in prefix-list

  Also requests a route refresh

© 1999, Cisco Systems, Inc.

Multiprotocol Extensions—rfc2283

MP_REACH_NLRI Attribute

Address Family Identifier (2 Octets)
Subsequent Address Family Identifier (1 Octet)
Length of Next Hop Network Address (1 Octet)
Network Address of Next Hop (Variable)
Number of First SNP As (1 Octet)
Length of First SNP A (1 Octet)
Length of First SNP A (1 Octet)
First SNP A (Variable)
…
Length of Last SNP A (1 Octet)
Last SNP A (Variable)
Network layer Reachability Information (Variable)

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Address Family Identifiers

- Address family identifier—rfc1700
  - 1 IPv4
  - 2 IPv6
  - 8 E.164
- Sub-AFI (for IPv4)
  - 1 unicast
  - 2 multicast
  - 3 unicast + multicast

Multiprotocol Extensions I

- mBGP

  Used to propagate multicast source information
- The different NLRI types allow for diverging topologies

  The NEXT_HOP information is different

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Multiprotocol Extensions II

- MPLS VPN
  - Used to carry both intra- and inter-VPN routing information
- New AFI—VPN-IPv4
- NLRI format for VPN addresses
  - Tag
  - VPNID (32 bits)
  - Prefix (variable length, 0-32 bits)

© 1999, Cisco Systems, Inc.

Extended Community Attribute

- Extended range
  - 8 octets
- Structure
  - Type: value
  - Value may be of the form AS:xxx
- Same functionality as existing attribute

draft-ramachandra-bgp-ext-communities, March 1999

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.
0901_04F9_c3.scr

Complex Network Scalability

- **Scalable**

  Confederations, route reflectors, and multiprotocol support

- **Stable**

  Network isolation, capability to handle large amount of data

- **Simple**

  … But flexible and extendible

© 1999, Cisco Systems, Inc.

For Further Reference:

- **Advanced IP Network Design**
  White, et. All—Cisco Press 1999
- **BGP4**
  Stewart—Addison Wesley 1999
- **Internet Routing Architectures**
  Halabi—Cisco Press 1997
- **IETF IDR Working Group**
  (http://www.ietf.org)

© 1999, Cisco Systems, Inc.

Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.





Copyright © 1998, Cisco Systems, Inc. All rights reserved. Printed in USA.