BGP Multi-homing Design and Troubleshooting

Manigandan Ganesan

Cisco Support Community - Ask the Expert

- Today's featured expert is Manigandan Ganesan
- Ask him questions now about BGP – Multi-homing

Manigandan Ganesan

Routing Protocols Engineer, Cisco

Thank You for Joining Us Today

Today’s presentation will include audience polling questions

We encourage you to participate!

Thank You for Joining Us Today

Downloading the Presentation

- If you would like a copy of the presentation slides, click the “Download Presentation” button below the slide window

Poll Question

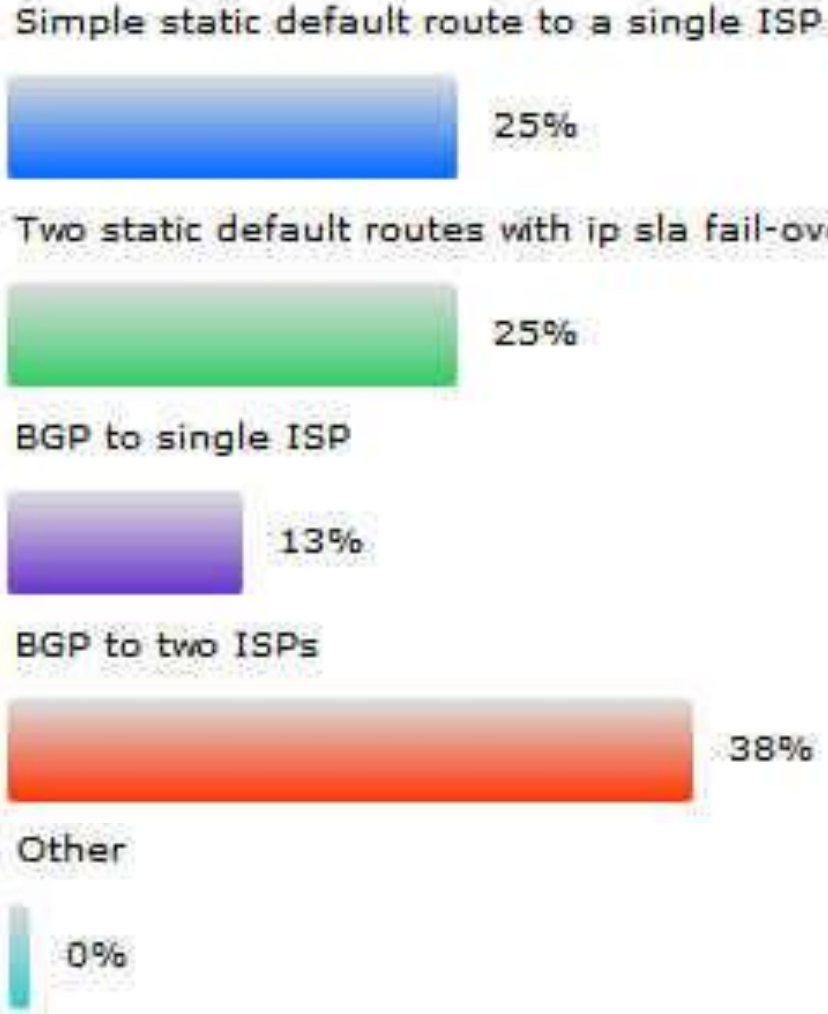
What type of internet connectivity set-up does your network currently have?

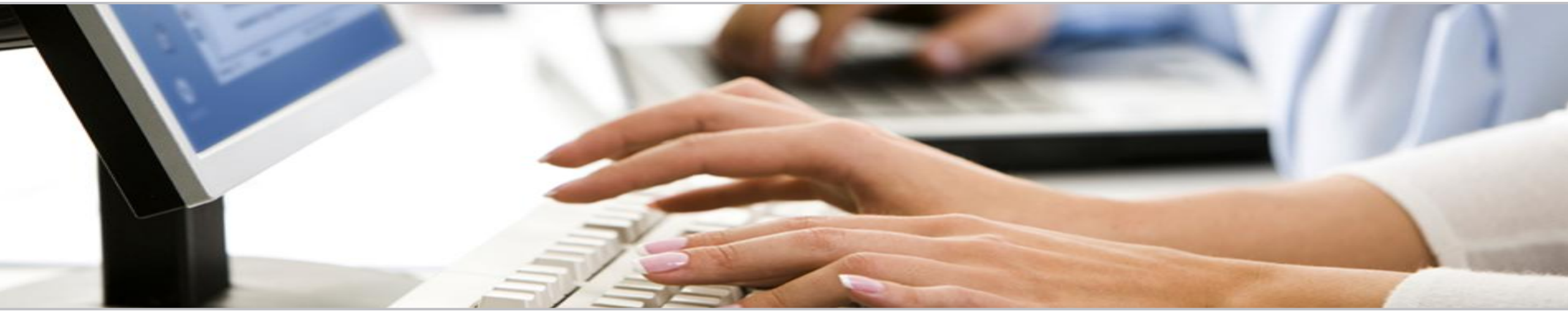
- ◯ Simple static default route to a single ISP gateway
- ◯ Two static default routes with ip sla fail-over
- ◯ BGP to single ISP
- ◯ BGP to two ISPs
- ◯ Other

Submit

Poll Response

What type of internet connectivity set-up does your network currently have?





Submit Your Questions Now

Use the Submit Text box Below the Slide Window. View Answers by Clicking on the “Refresh” Button

Agenda

- Multi-homing - Overview
- Various Options in Multi-homing
- Configuration of various multi-homing setups
- Troubleshooting typical BGP issues

Why Multi-home?

- **Redundancy**

One connection to internet means the network is dependent on:

- Local router (configuration, software, hardware)

- WAN media (physical failure, carrier failure)

- Upstream Service Provider (configuration, software, hardware)

Why Multi-home?

- **Reliability**

  - Business critical applications demand continuous availability

  - Lack of redundancy implies lack of reliability implies loss of revenue

Multi-homing: Definitions & Options

- **More than one link external to the local network**

- **Usually two external facing routers**

  (one router gives link and provider redundancy only)

Multihoming – A Deeper look at the Options

- **Will look at two cases:**

    1. **Multihoming to the same ISP**
    2. **Multihoming to different ISPs**

- **Will keep the examples easy**

  - **Understanding easy concepts will make the more complex scenarios easier to comprehend**

  - **All assume that the site multihoming has a /19 address block**

CASE 1 - Two links to the same ISP

In this case, we can run BGP with the same provider in two modes :

1. Fail-over mode

One link primary, the other link backup only

2. Load-sharing mode

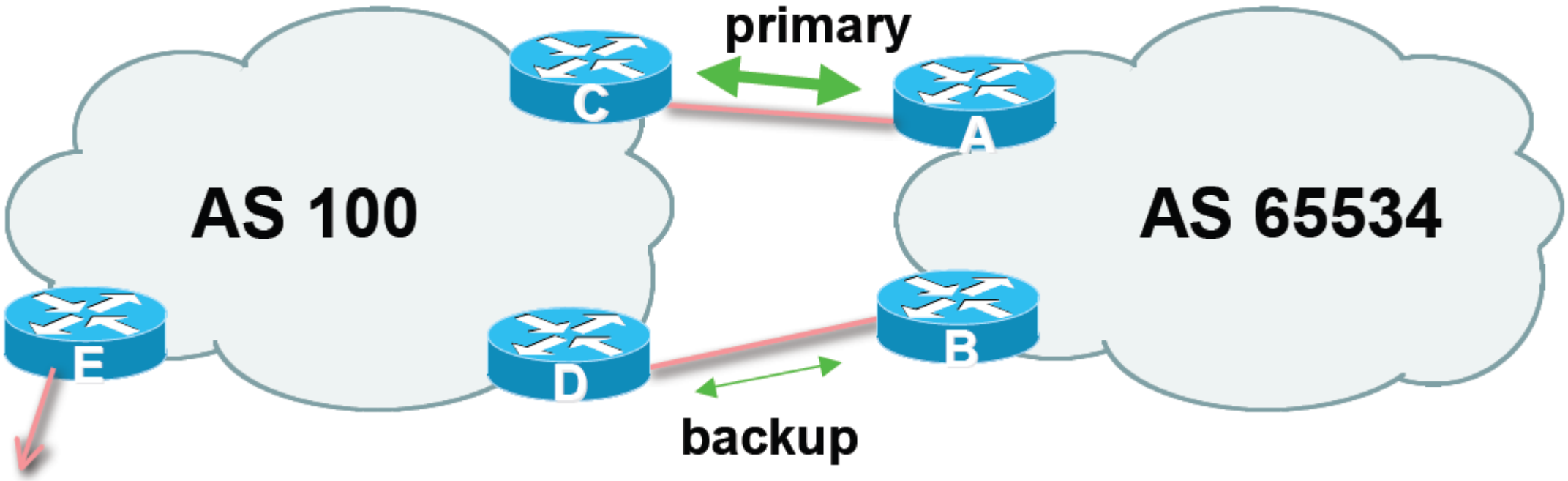
Two circuits active at the same time

1.a) Two links to the same ISP – Failover mode

- **Applies when end-site has bought a large primary WAN link to their upstream a small secondary WAN link as the backup**

  For example, primary path might be an E1, backup might be 64kbps

1.a) Two links to the same ISP – Failover mode (contd)



1.a) Two links to the same ISP – Failover mode (contd)

- **Announce /19 aggregate on each link**

  primary link:
  Outbound – announce /19 unaltered
  Inbound – receive default route

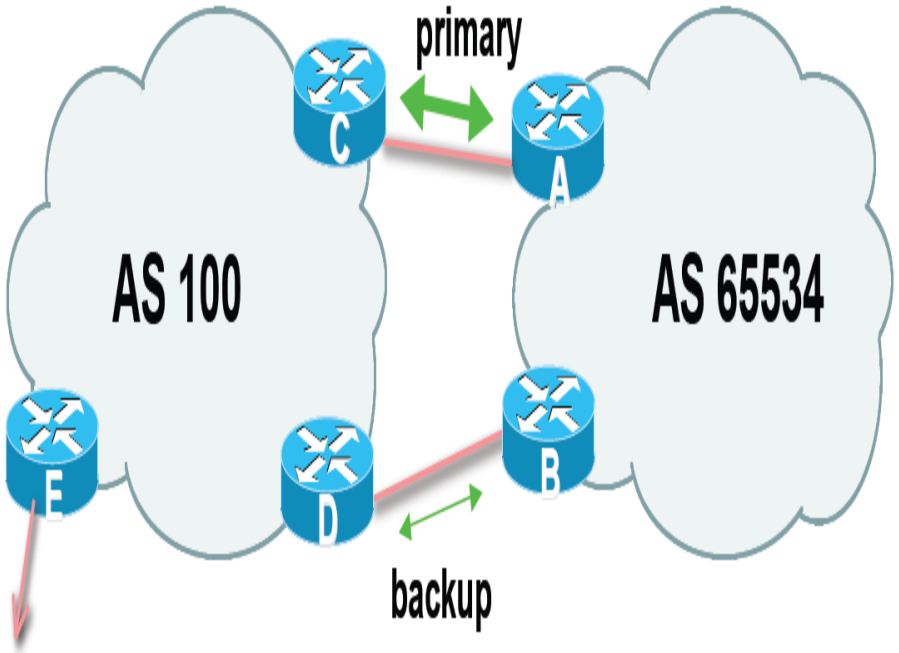
  backup link:
  Outbound – announce /19 with increased metric
  Inbound – received default, and reduce local preference

- **When one link fails, the announcement of the /19 aggregate via the other link ensures continued connectivity**

1.a) Two links to the same ISP – Failover mode (contd)

Router A Configuration

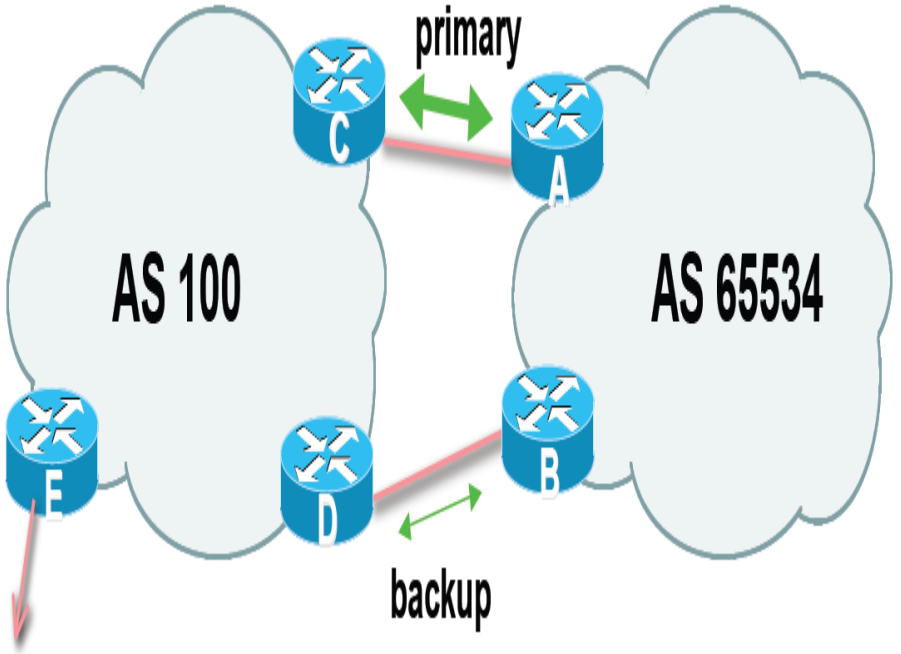
```
router bgp 65534
network 121.10.0.0 mask 255.255.224.0
neighbor 122.102.10.2 remote-as 100
neighbor 122.102.10.2 description RouterC
neighbor 122.102.10.2 prefix-list aggregate out
neighbor 122.102.10.2 prefix-list default in
!
ip prefix-list aggregate permit 121.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
```



1.a) Two links to the same ISP – Failover mode(contd)

Router B Configuration

```
router bgp 65534
network 121.10.0.0 mask 255.255.224.0
neighbor 122.102.10.6 remote-as 100
neighbor 122.102.10.6 description RouterD
neighbor 122.102.10.6 prefix-list aggregate out
neighbor 122.102.10.6 route-map routerD-out out
neighbor 122.102.10.6 prefix-list default in
neighbor 122.102.10.6 route-map routerD-in in
!
ip prefix-list aggregate permit 121.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
route-map routerD-out permit 10
match ip address prefix-list aggregate
set metric 10
route-map routerD-out permit 20
!
route-map routerD-in permit 10
set local-preference 90
```



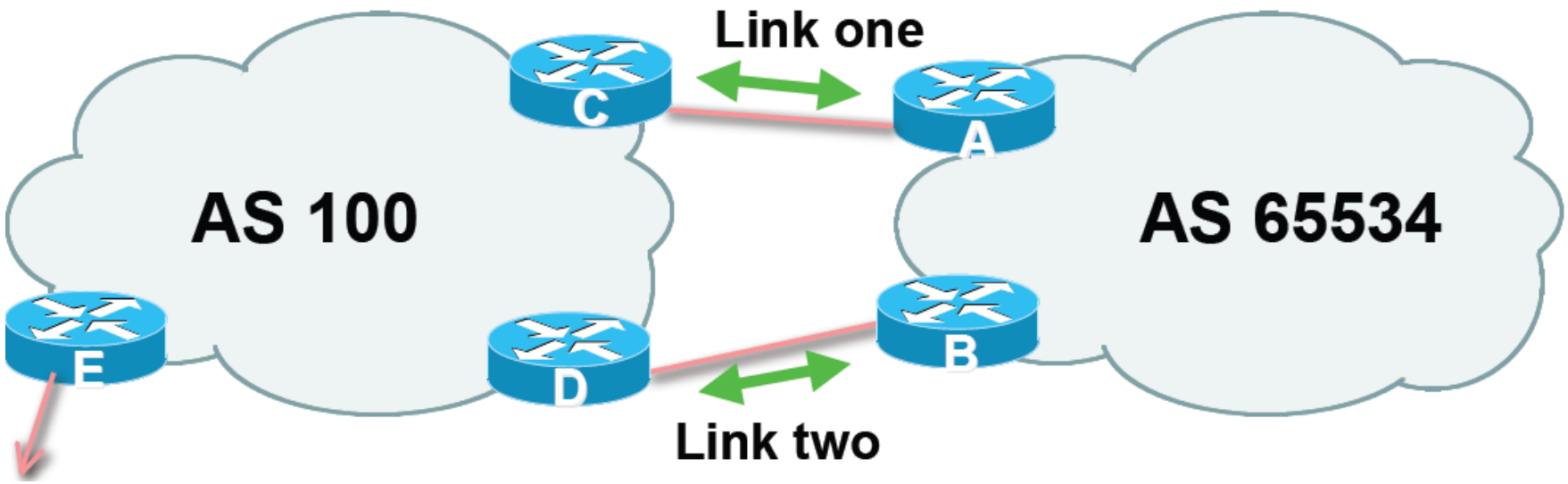
1.b) Two links to the same ISP – Load Balancing mode

- **More common case**

- **End sites tend not to buy circuits and leave them idle, only used for backup as in previous Example**

- **This example assumes equal capacity circuits - Unequal capacity circuits requires more refinement**

1.b) Two links to the same ISP – Load sharing mode (contd)



1.b) Two links to the same ISP – Load sharing mode (contd)

- **Announce /19 aggregate on each link**

- **Split /19 and announce as two /20s, one on each link**
  - **basic inbound loadsharing**
  - **assumes equal circuit capacity and even spread of traffic across address block**

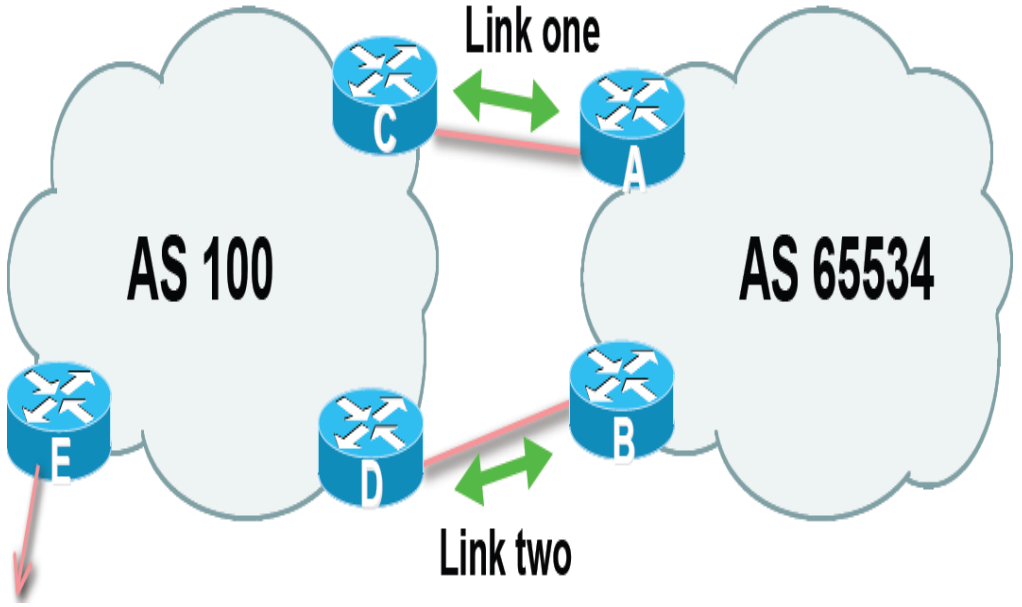
- **Vary the split until “perfect” loadsharing achieved**

- **Accept the default from upstream**
  - **basic outbound loadsharing by nearest exit**
  - **okay in first approx as most ISP and end-site traffic is inbound**

1.b) Two links to the same ISP – Load sharing mode(contd)

- **Router A Configuration**

```
router bgp 65534
network 121.10.0.0 mask 255.255.224.0
network 121.10.0.0 mask 255.255.240.0
neighbor 122.102.10.2 remote-as 100
neighbor 122.102.10.2 prefix-list routerC out
neighbor 122.102.10.2 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
ip prefix-list routerC permit 121.10.0.0/20
ip prefix-list routerC permit 121.10.0.0/19
!
ip route 121.10.0.0 255.255.240.0 null0
ip route 121.10.0.0 255.255.224.0 null0
```



Router B configuration is similar but with the other /20

CASE 2 - Two links to different ISPs

In this case, we run BGP with the different provider in two modes :

1. Fail-over mode

One link primary, the other link backup only

2. Load-sharing mode

Two circuits active at the same time

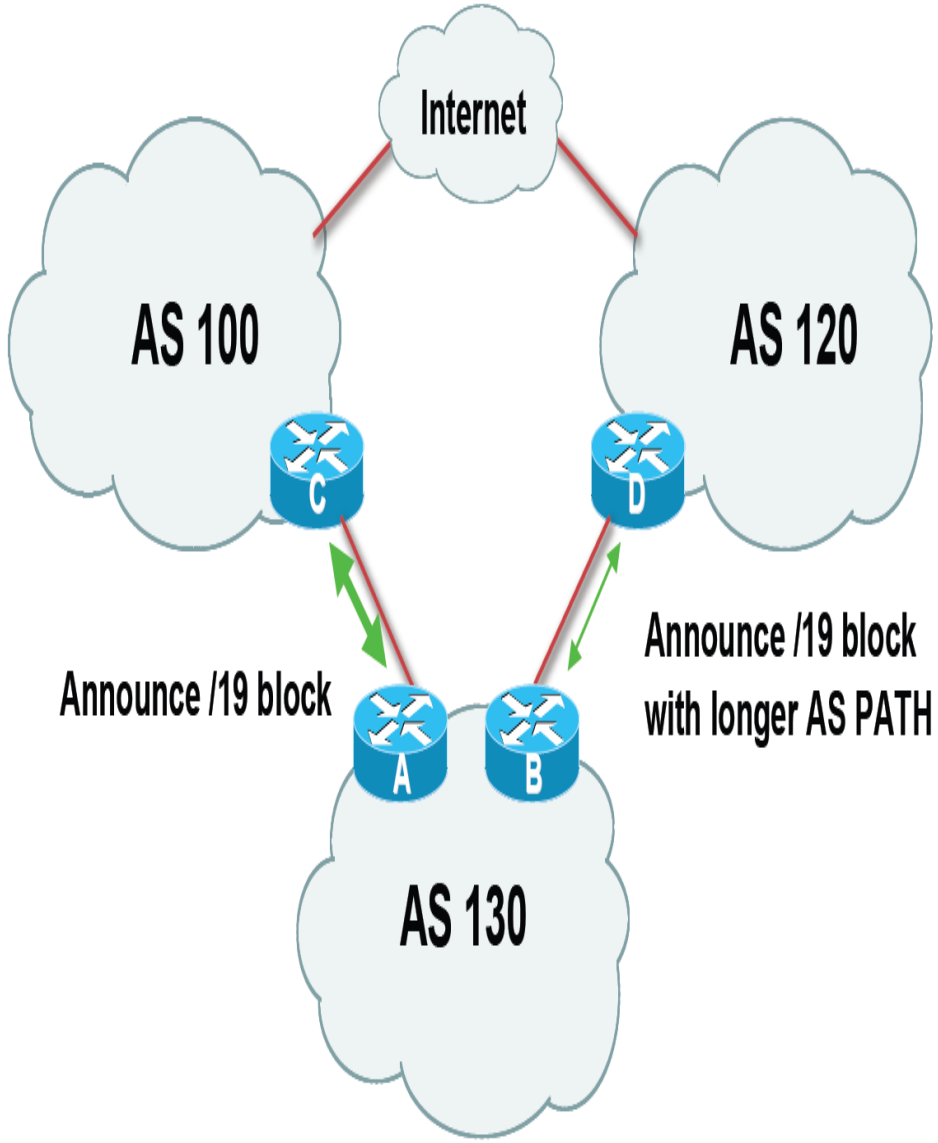
2.a) Two links to Different ISP – Fail-over mode

- **Announce /19 aggregate on each link**

  - primary link makes standard announcement

  - backup link lengthens the AS PATH by using AS PATH prepend

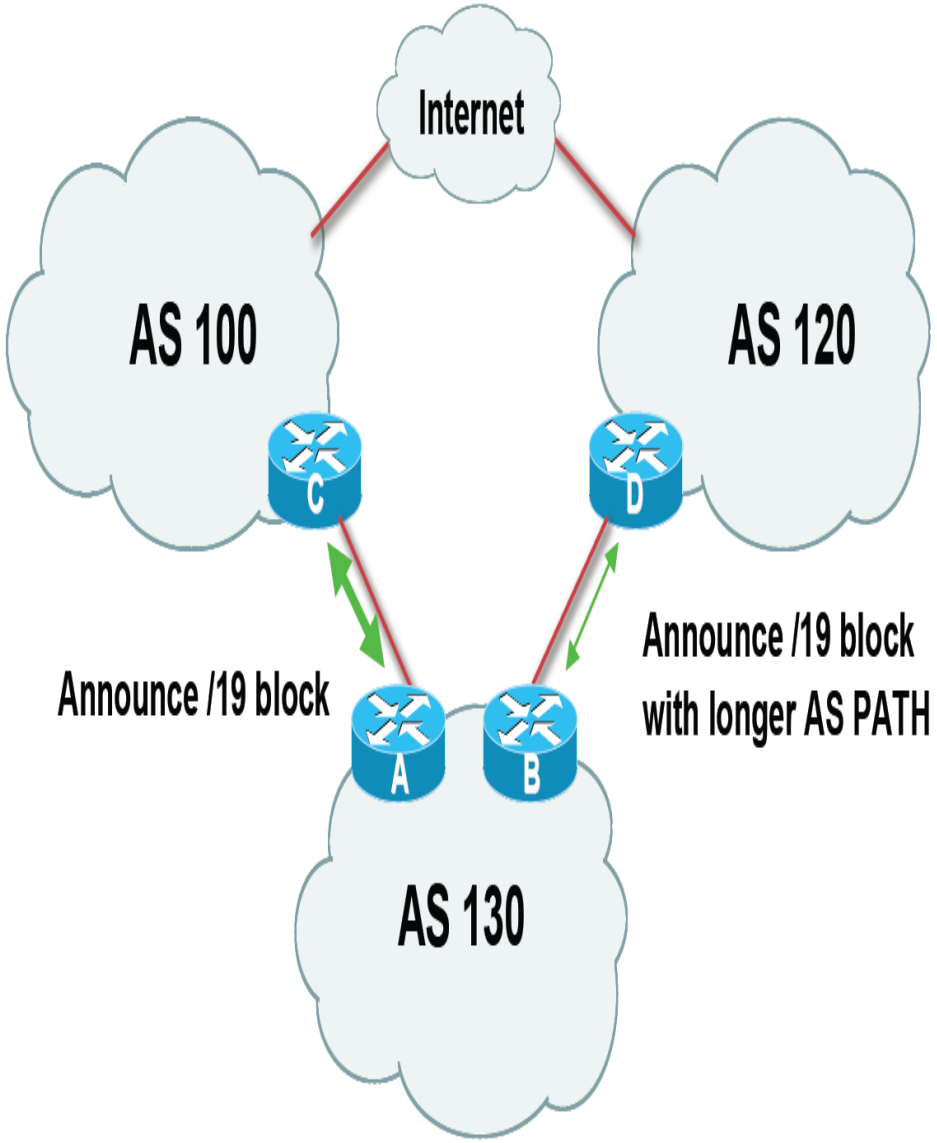
- **When one link fails, the announcement of the /19 aggregate via the other link ensures continued connectivity**



2.a) Two links to Different ISP – Fail-over mode

- **Router A Configuration**

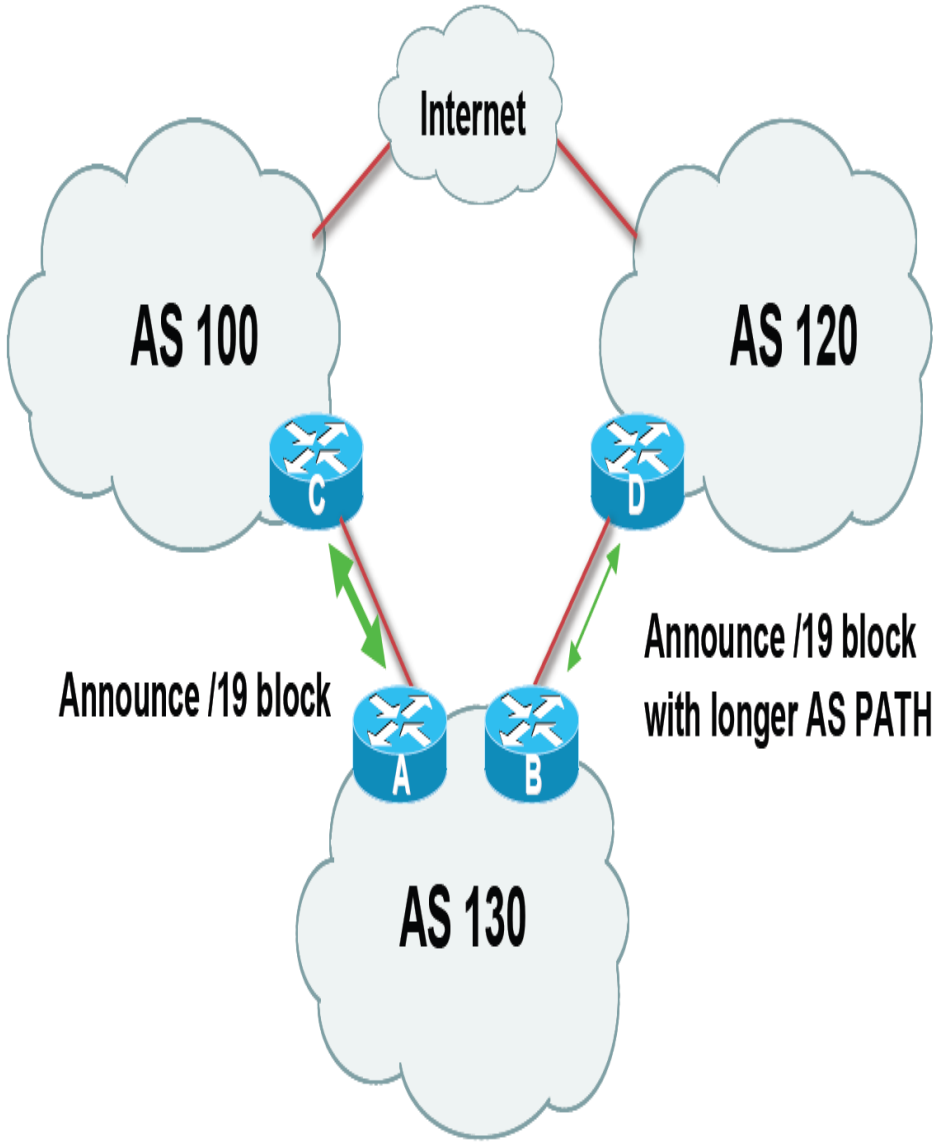
```
router bgp 130
network 121.10.0.0 mask 255.255.224.0
neighbor 122.102.10.1 remote-as 100
neighbor 122.102.10.1 prefix-list aggregate out
neighbor 122.102.10.1 prefix-list default in
!
ip prefix-list aggregate permit 121.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```



2.a) Two links to Different ISP – Fail-over mode

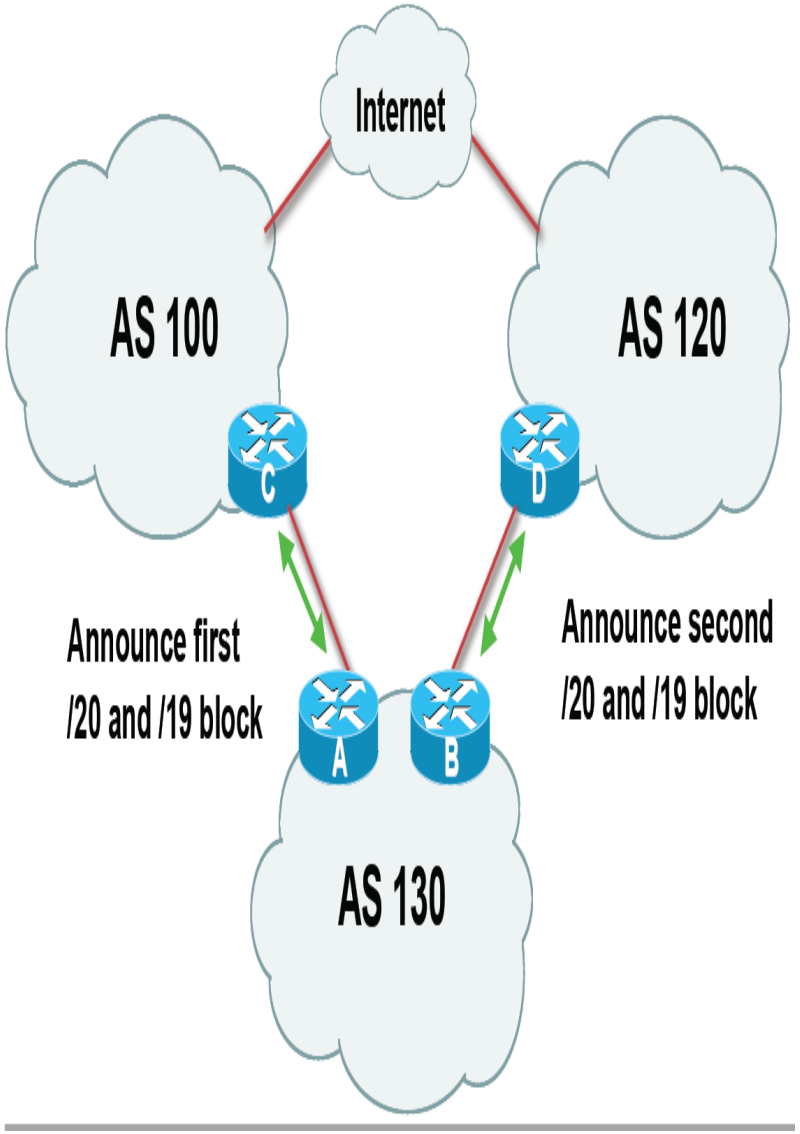
Router B Configuration

```
router bgp 130
network 121.10.0.0 mask 255.255.224.0
neighbor 120.1.5.1 remote-as 120
neighbor 120.1.5.1 prefix-list aggregate out
neighbor 120.1.5.1 route-map routerD-out out
neighbor 120.1.5.1 prefix-list default in
neighbor 120.1.5.1 route-map routerD-in in
!
ip prefix-list aggregate permit 121.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
route-map routerD-out permit 10
set as-path prepend 130 130 130
!
route-map routerD-in permit 10
set local-preference 80
```



2.b) Two links to Different ISP – Load-sharing mode

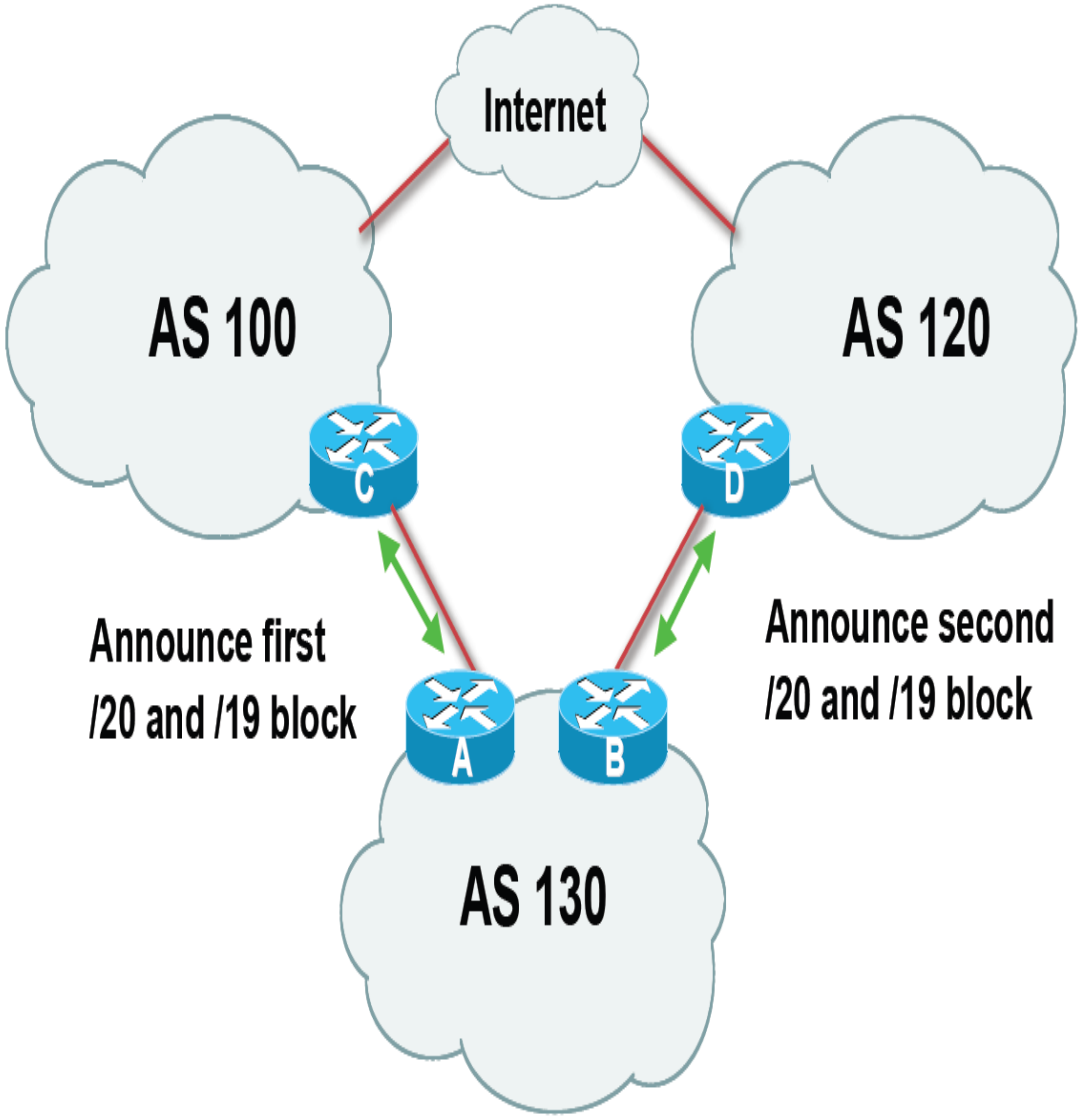
- **Announce /19 aggregate on each link**
- **Split /19 and announce as two /20s, one on each link**
  - **basic inbound loadsharing**
- **When one link fails, the announcement of the /19 aggregate via the other ISP ensures continued connectivity**



2.b) Two links to Different ISP – Load-sharing mode

- **Router A Configuration**

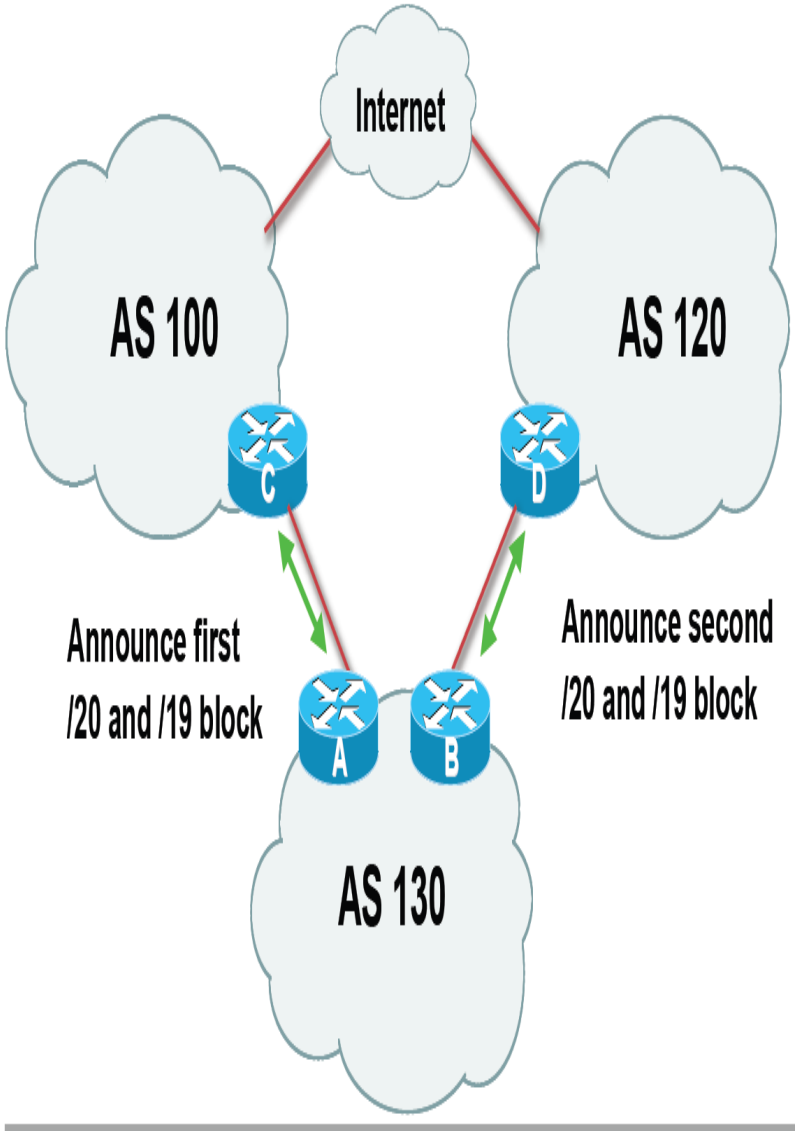
```
router bgp 130
network 121.10.0.0 mask 255.255.224.0
network 121.10.0.0 mask 255.255.240.0
neighbor 122.102.10.1 remote-as 100
neighbor 122.102.10.1 prefix-list firstblock out
neighbor 122.102.10.1 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
!
ip prefix-list firstblock permit 121.10.0.0/20
ip prefix-list firstblock permit 121.10.0.0/19
```



2.b) Two links to Different ISP – Load-sharing mode

Router B Configuration

```
router bgp 130
network 121.10.0.0 mask 255.255.224.0
network 121.10.16.0 mask 255.255.240.0
neighbor 120.1.5.1 remote-as 120
neighbor 120.1.5.1 prefix-list secondblock out
neighbor 120.1.5.1 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
!
ip prefix-list secondblock permit 121.10.16.0/20
ip prefix-list secondblock permit 121.10.0.0/19
```



Poll Question

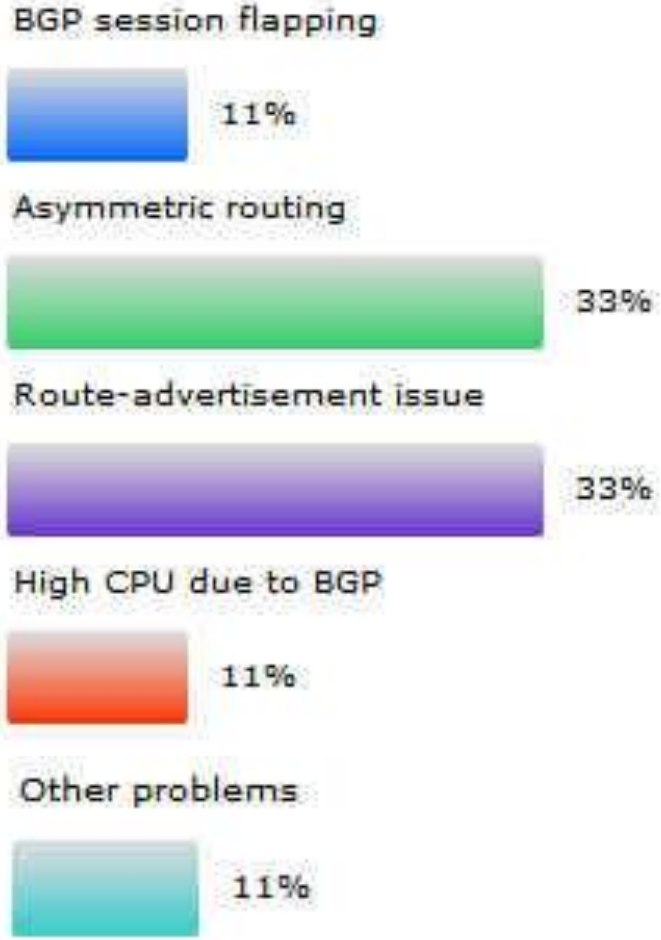
What is the most common problem that you face in BGP?

- ◯ BGP session flapping
- ◯ Asymmetric routing
- ◯ Route-advertisement issue
- ◯ High CPU due to BGP
- ◯ Other problems

Submit

Poll Response

What is the most common problem that you face in BGP?



TROUBLESHOOTING COMMON BGP PROBLEMS

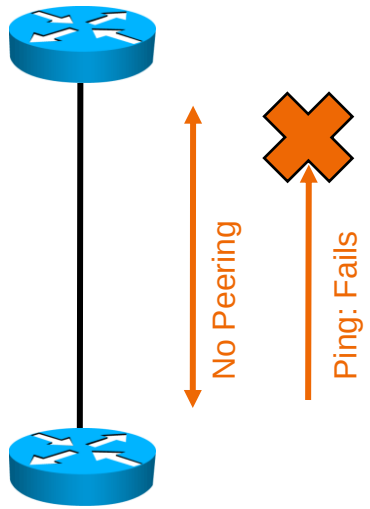
BGP Speakers Won't Peer

- This can be difficult to troubleshoot if you can only see one side of the connection
- Start with the simple things: check for common mistakes

  - Is it supposed to be configured for eBGP multihop?
  - Are the AS numbers right?

- Next, try pinging the peering address

  - If the ping fails, there's likely a connectivity problem



BGP Speakers Won't Peer

- Try some alternate ping options
- Is the local peering address the actual peering interface?

  If not, use extended ping to source from the loopback or actual peering address

  If this fails, there is an underlying routing problem

  The other router may not know how to reach your peering interface

```
Router>enable
Router#ping
Protocol [ip]:
Target IP address:
192.168.40.1
Repeat count [5]:
Datagram size [100]:
Timeout in seconds [2]:
Extended commands [n]: y
Source address or
interface: 172.16.23.2
```

BGP Speakers Won't Peer

- **Try extended ping to sweep a range of possible MTUs**
  - **Note the MTU at which the ping starts to fail**
  - **Make certain the interface is configured for that MTU size**
- **If these all fail**
  - **None of the pings work no matter how you try....**
  - **It's likely a transport problem**

```
Router>enable
Router#ping
Protocol [ip]:
Target IP address:
192.168.40.1
Repeat count [5]:
Datagram size [100]:
Timeout in seconds [2]:
Extended commands [n]: y
....
Sweep range of sizes [n]: y
Sweep min size [36]: 100
Sweep max size [18024]:
2500
Sweep interval [1]: 100
....
```

BGP Speakers Won't Peer

- **Remember that BGP runs on top of IP, and can be affected by:**

  - **Rate limiting**
  - **Traffic shaping**
  - **Tunneling problems**
  - **IP reachability problems (the underlying routing isn't working)**
  - **ACL / FW blocking port 179**
  - **TCP problems**

BGP Speakers Won't Peer

Useful Peer Troubleshooting Commands

show tcp brief all	**TCB Local Address Foreign Address (state)** **64316F14 1.1.1.1.12345 2.2.2.2.179 ESTAB** **6431BA8C *.179 2.2.2.2.* LISTEN**
show tcp statistics	Rcvd: 7005 Total, 10 no port **0 checksum error, 0 bad offset, 0 too short** .... **0 out-of-order packets (0 bytes)** .... 4186 ack packets (73521 bytes) .... Sent: 9150 Total, 0 urgent packets 4810 control packets (including 127 retransmitted) 2172 data packets (71504 bytes) ....

BGP Speakers Won't Peer

Useful Peer Troubleshooting Commands

debug ip tcp transactions

This can be very chatty, so be careful with this debug!

```
R1#sh log | i TCP0:

TCP0: state was ESTAB -> FINWAIT1 [12345 -> 2.2.2.2(179)]
TCP0: sending FIN
TCP0: state was FINWAIT1 -> FINWAIT2 [12345 -> 2.2.2.2(179)]
TCP0: FIN processed
TCP0: state was FINWAIT2 -> TIMEWAIT [12345 -> 2.2.2.2(179)]
TCP0: Connection to 2.2.2.2:179, advertising MSS 1460
TCP0: state was CLOSED -> SYNSENT [12346 -> 2.2.2.2(179)]
TCP0: state was SYNSENT -> ESTAB [12346 -> 2.2.2.2(179)]
TCP0: tcb 6430DCDC connection to 2.2.2.2:179, received MSS
1460, MSS is 1460
```

BGP Speakers Won't Peer

- **If the connectivity is good, the next step is to check BGP itself**
- **`debug ip bgp`**
  - **Use with caution**
  - **Configure so the output goes to the log, rather than the console**

```
logging buffered <size>
no logging console
```

  - **It's easier to find the problem points this way**

```
router#show log | i NOTIFICATION
```

BGP Speakers Won't Peer

- `show ip bgp neighbor 1.1.1.1 | include last reset`
  - This should give you the resets for a peer
  - The same information as is shown through `debug ip bgp`
- `bgp log-neighbor changes`
  - Provides much of the same information as `debug ip bgp`, as well

BGP Speakers Won't Peer

Source/Destination Address Matching

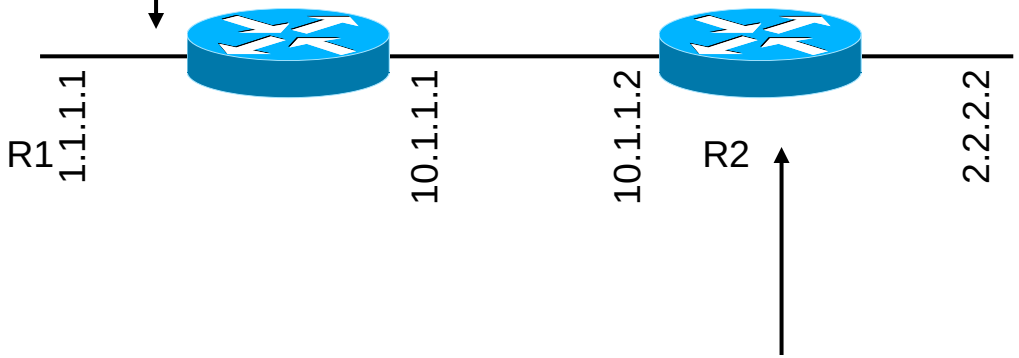
- **Both sides must agree on source and destination addresses**

  BGP will tear down the TCP session due to the conflict

- `debug ip bgp` on R1 shows

```
BGP: 2.2.2.2 passive open to
10.1.1.1
BGP: 2.2.2.2 passive open failed
- 10.1.1.1 is not update-source
Loopback0's address (1.1.1.1)
```

```
neighbor 2.2.2.2 remote-as 100
neighbor 2.2.2.2 update-source
loopback 0
```




```
neighbor 10.1.1.1 remote-as 100
neighbor 10.1.1.1 update-source
loopback 0
```

BGP Speakers Won't Peer

Bad Messages – What to look for ?

```
%BGP-3-NOTIFICATION: sent to neighbor 2.2.2.2 2/2
(peer in wrong AS) 2 bytes 00C8 FFFF FFFF FFFF FFFF
FFFF FFFF FFFF FFFF 002D 0104 00C8 00B4 0202 0202
1002 0601 0400 0100 0102 0280 0002 0202 00
```

BGP Speakers Won't Peer

Bad Messages – contd..

unknown subcode	The peer open notification subcode isn't known
incompatible BGP version	The version of BGP the peer is running isn't compatible with the local version of BGP
peer in wrong AS	The AS this peer is locally configured for doesn't match the AS the peer is advertising
BGP identifier wrong	The BGP router ID is the same as the local BGP router ID
unsupported optional parameter	There is an option in the packet which the local BGP speaker doesn't recognize
authentication failure	The MD5 hash on the received packet does not match the correct MD5 hash
unacceptable hold time	The remote BGP peer has requested a BGP hold time which is not allowed (too low)
unsupported/disjoint capability	The peer has asked for support for a feature which the local router does not support

High Processor Utilization

- **Why?**

  This could be for several reasons

  - High route churn is the most likely

```
router# show process cpu
  CPU utilization for five seconds: 100%/0%; one minute: 99%;
five minutes: 81%
....
139   6795740   1020252 6660 88.34% 91.63% 74.01%    0 BGP Router
```

High Processor Utilization

- **Check how busy the peers are - the Table Version**

  We have 150k routes and see the table version increase by 150k every minute … something is wrong

  We have 150k routes and see the table version increase by 300 every minute … sounds like normal network churn

  The InQ - Flood of incoming updates or build up of unprocessed updates

  The OutQ - Flood of outgoing updates or build up of untransmitted updates

```
router# show ip bgp summary

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ  OutQ Up/Down  /PfxRcd
10.1.1.1        4        64512  309453  157389    19981    0   253 22:06:44 111633
172.16.1.1      4        65101  188934    1047    40081   41     0 00:07:51 58430
```

High Processor Utilization

- **If the Table Version is Changing Quickly**
  - **Are we in initial convergence with this peer?**
  - **Is the peer flapping for some reason?**
  - **Examine the table entries from this peer: why are they changing?**
  - **If there is a group of routes which are constantly changing, consider route flap dampening**
- **If the InQ is high**
  - **You should see the table version changing quickly**
  - **If it's not, the peer isn't acting correctly**
  - **Consider shutting it down until the peer can be fixed**
- **If the OutQ is high**
  - **Lots of updates being generated**
  - **Check table versions of other peers**
  - **Check for underlying transport problems**

High Processor Utilization

- **Check on the BGP Scanner**

  - **Walks the table looking for changed next hops**
  - **Checks conditional advertisement**
  - **Imports from and exports to VPNv4 VRFs**

```
router# show processes | include BGP Scanner

 172 Lsi 407A1BFC 29144 29130 1000 8384/9000   0 BGP Scanner
```

High Processor Utilization

- **To relieve pressure on the BGP Scanner**

  Upgrade to newer code

  - Most of the work of the BGP Scanner has been moved to an event driven model
  - This has reduced the impact of BGP Scanner significantly

  Reduce route and view count

  Reduce or eliminate other processes which walk the RIB

  - SNMP routing table walks, for instance

  Deploy BGP Next Hop Tracking (NHT)

Poll Question

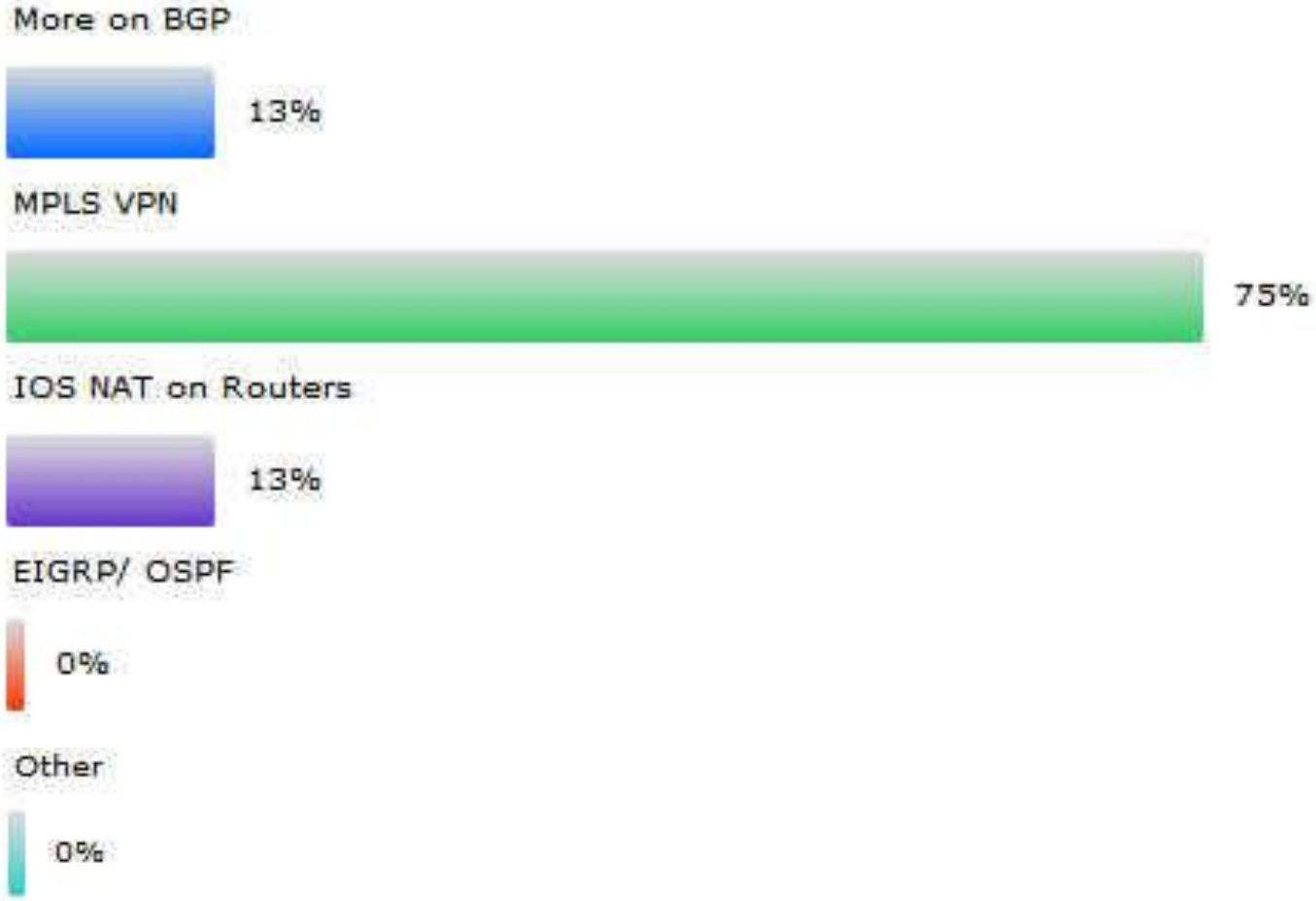
Which Routing topics would you like to see in a future Webcast?

- ◯ More on BGP
- ◯ MPLS VPN
- ◯ IOS NAT on Routers
- ◯ EIGRP/ OSPF
- ◯ Other

Submit

Poll Response

Which Routing topics would you like to see in a future Webcast?



Submit Your Questions Now

Use the Submit Text Box Below the Slide Window. View Answers by Clicking on the “Refresh” Button

Q&A

We Appreciate Your Feedback!

The first 10 listeners
who fill out an Evaluation
will receive a free:

$20 USD
Amazon Gift Certificate

To complete the evaluation, please click on Evaluation button under the slides.

Some Useful BGP Links

BGP Case studies:

http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a00800c95bb.shtml

BGP Command Reference:

http://www.cisco.com/en/US/docs/ios/iproute_bgp/command/reference/irg_book.html

BGP Configuration Guide:

http://www.cisco.com/en/US/docs/ios/iproute_bgp/configuration/guide/12_4/irg_12_4_book.html

Bug Tool-Kit:

http://www.cisco.com/pcgi-bin/Support/Bugtool/home.pl

Post Questions on Our Forum Here:

https://supportforums.cisco.com/community/netpro/network-infrastructure/routing

Thank You for Your Time

Please Take a Moment to Complete the Evaluation


CISCO