

Introduction to the Border Gateway Protocol

Campus Network Design & Operations Workshop



These materials are licensed under the Creative Commons Attribution-NonCommercial 4.0 International license
(<http://creativecommons.org/licenses/by-nc/4.0/>)



UNIVERSITY OF OREGON

Last updated 10th October 2017



Border Gateway Protocol

- A Routing Protocol used to exchange routing information between different networks
 - Exterior gateway protocol
- Described in RFC4271
 - RFC4276 gives an implementation report on BGP
 - RFC4277 describes operational experiences using BGP
- The Autonomous System is the cornerstone of BGP
 - It is used to uniquely identify networks with a common routing policy

BGP

- Path Vector Protocol
- Incremental Updates
- Many options for policy enforcement
- Classless Inter Domain Routing (CIDR)
- Widely used for Internet backbone
- Autonomous systems

Path Vector Protocol

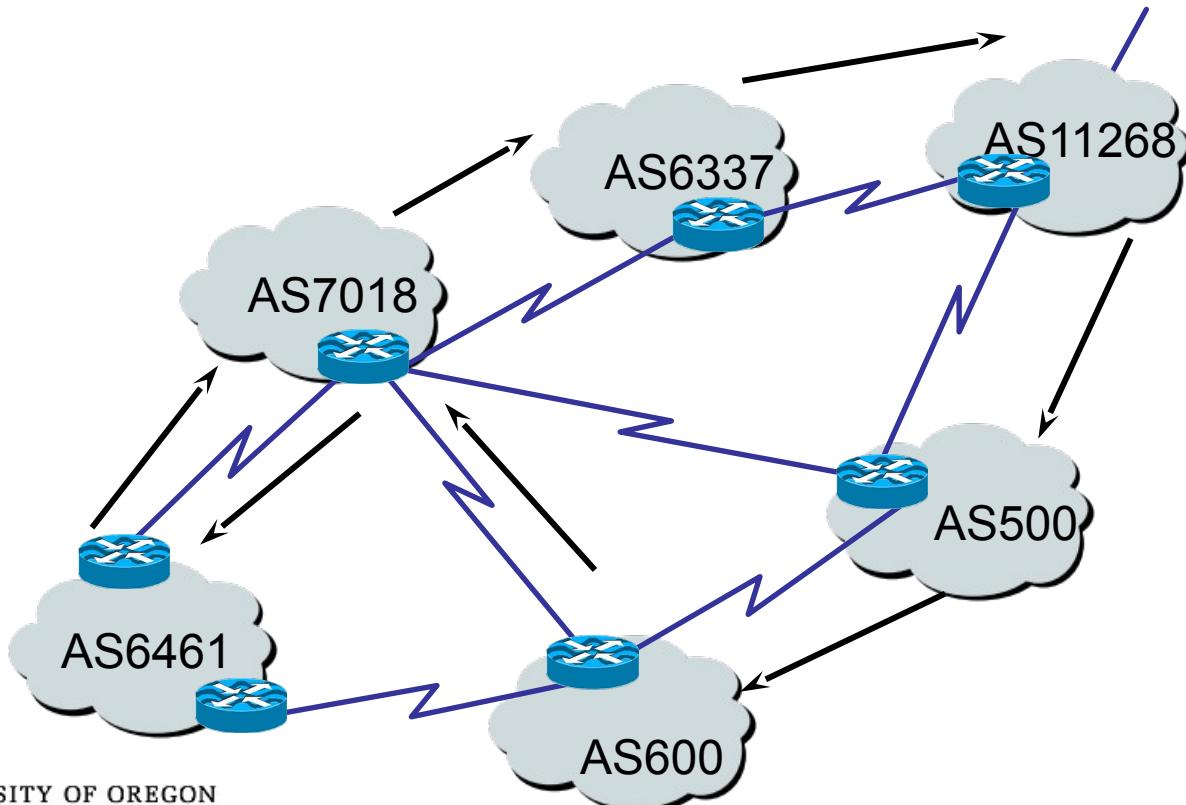
- BGP is classified as a path vector routing protocol (see RFC 1322)
 - A path vector protocol defines a route as a pairing between a destination and the attributes of the path to that destination.

```
12.6.126.0/24 207.126.96.43 1021 0 6461 7018 6337 11268 i
```



AS Path

Path Vector Protocol



Definitions

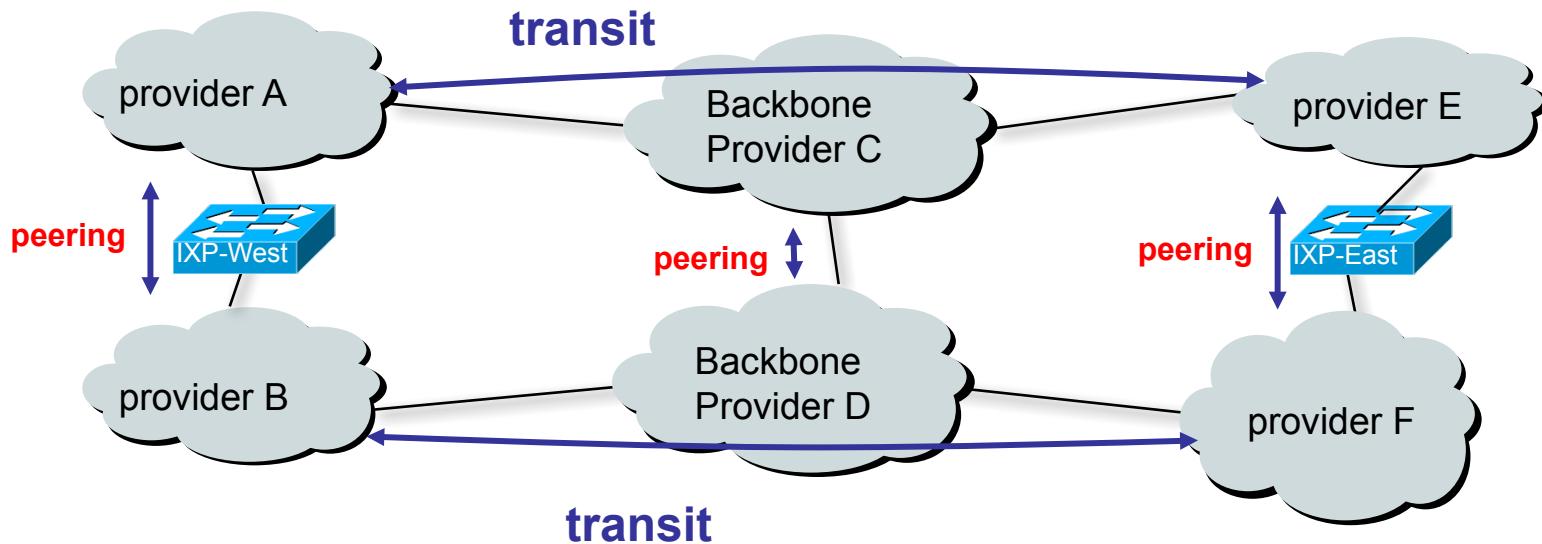
- Transit – carrying traffic across a network
 - (Commercially: for a fee)
- Peering – exchanging routing information and traffic
 - (Commercially: between similar sized networks, and for no fee)
- Default – where to send traffic when there is no explicit match in the routing table

Default Free Zone

The default free zone is made up of Internet routers which have routing information about the whole Internet, and therefore do not need to use a default route

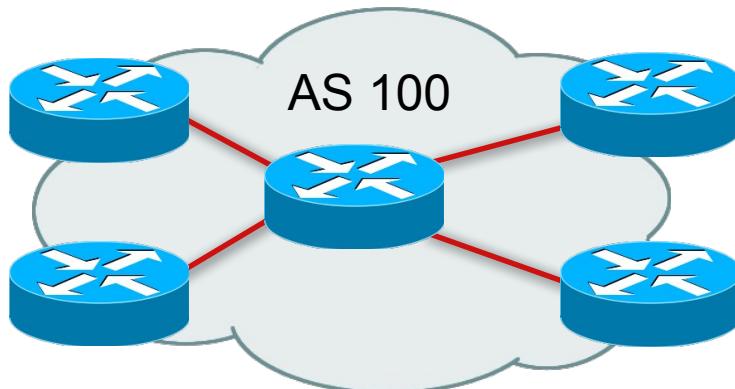
NB: is not related to where an ISP
is in the hierarchy

Peering and Transit example



A and B peer for free, but need transit arrangements
with C and D to get packets to/from E and F

Autonomous System (AS)



- Collection of networks with same routing policy
- Single routing protocol
- Usually under single ownership, trust and administrative control
- Identified by a unique 32-bit integer (ASN)

Autonomous System Number

Range:	
0-4294967295	(32-bit range – RFC6793)
	(0-65535 was original 16-bit range)

Usage:	
0 and 65535	(reserved)
1-64495	(public Internet)
64496-64511	(documentation – RFC5398)
64512-65534	(private use only)
23456	(represent 32-bit range in 16-bit world)
65536-65551	(documentation – RFC5398)
65552-4199999999	(public Internet)
4200000000-4294967295	(private use only)

- 32-bit range representation specified in RFC5396
 - Defines “asplain” (traditional format) as standard notation

Autonomous System Number

- ASNs are distributed by the Regional Internet Registries
 - They are also available from upstream ISPs who are members of one of the RIRs
- The entire 16-bit ASN pool has been assigned to the RIRs
 - Around 43000 16-bit ASNs are visible on the Internet
- Each RIR has also received a block of 32-bit ASNs
 - Out of 20100 assignments, around 16000 are visible on the Internet (October 2017)
- See www.iana.org/assignments/as-numbers

Configuring BGP in Cisco IOS

- This command enables BGP in Cisco IOS:

```
router bgp 100
```

- For ASNs > 65535, the AS number can be entered in either plain or dot notation:

```
router bgp 131076
```

- Or

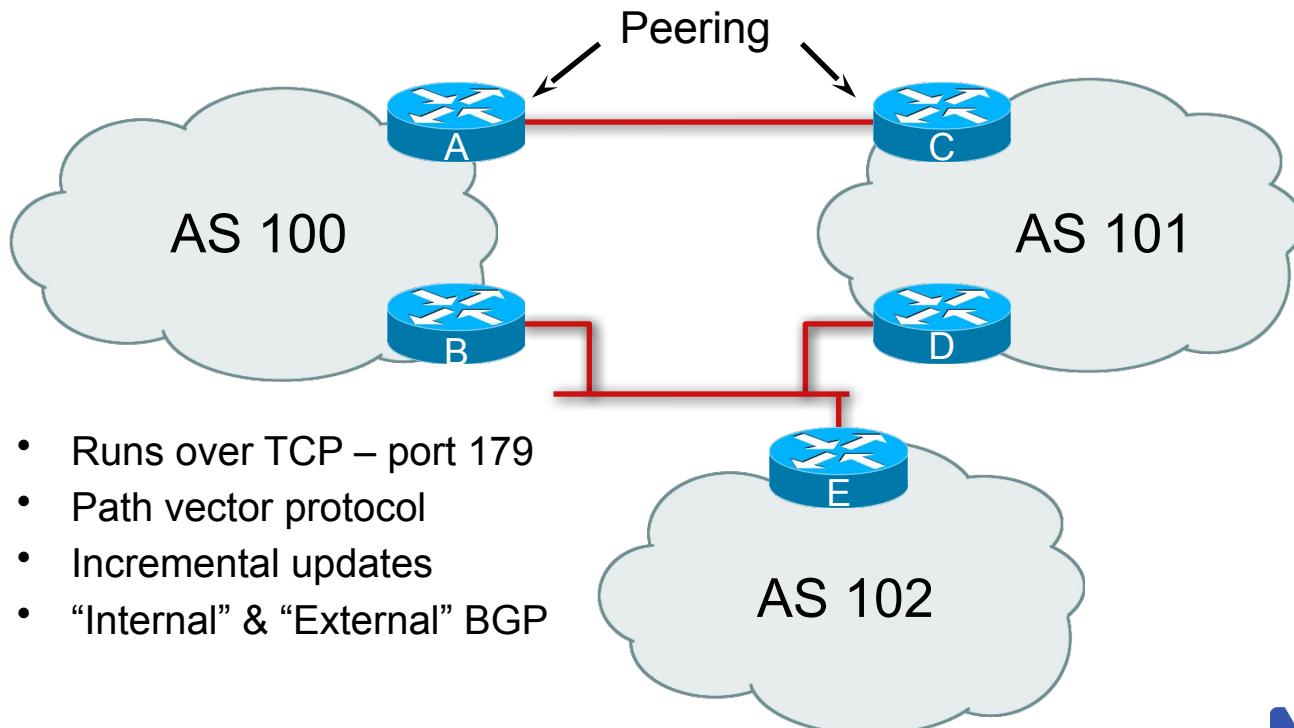
```
router bgp 2.4
```

- IOS will display ASNs in plain notation by default

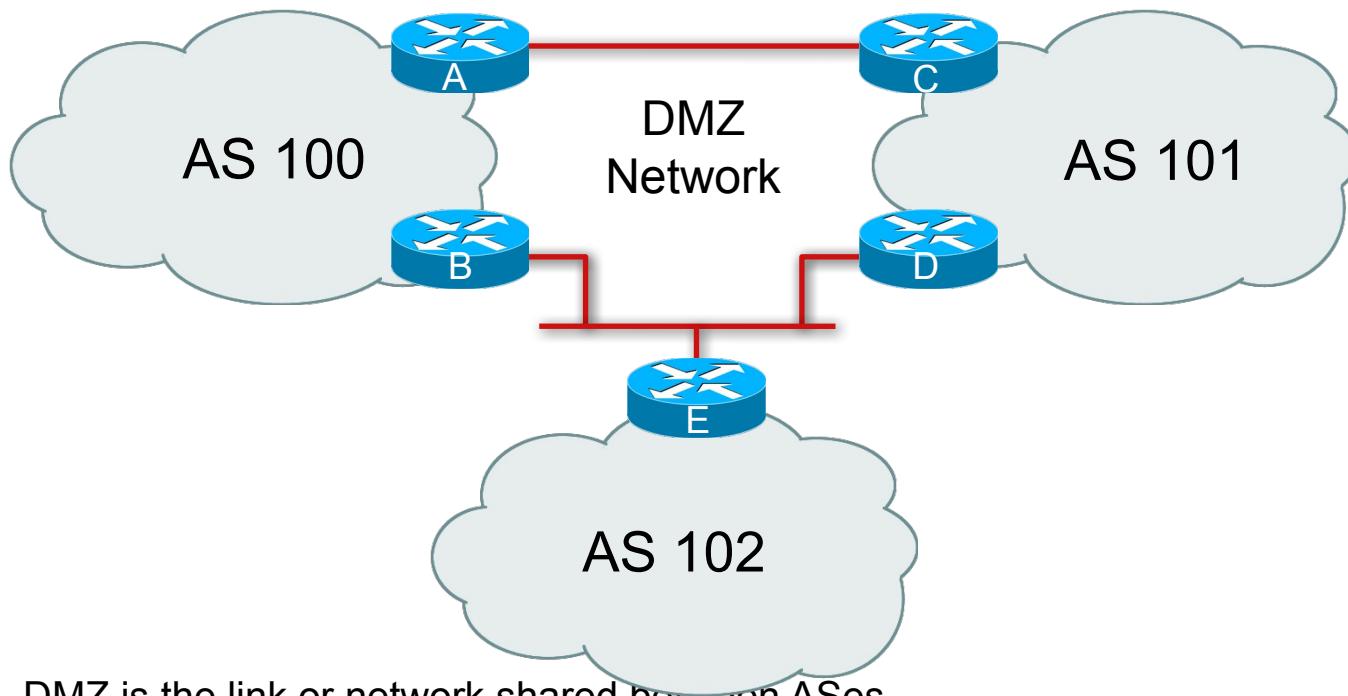
- Dot notation is optional:

```
router bgp 2.4
bgp asnotation dot
```

BGP Basics



Demarcation Zone (DMZ)



- DMZ is the link or network shared between ASes

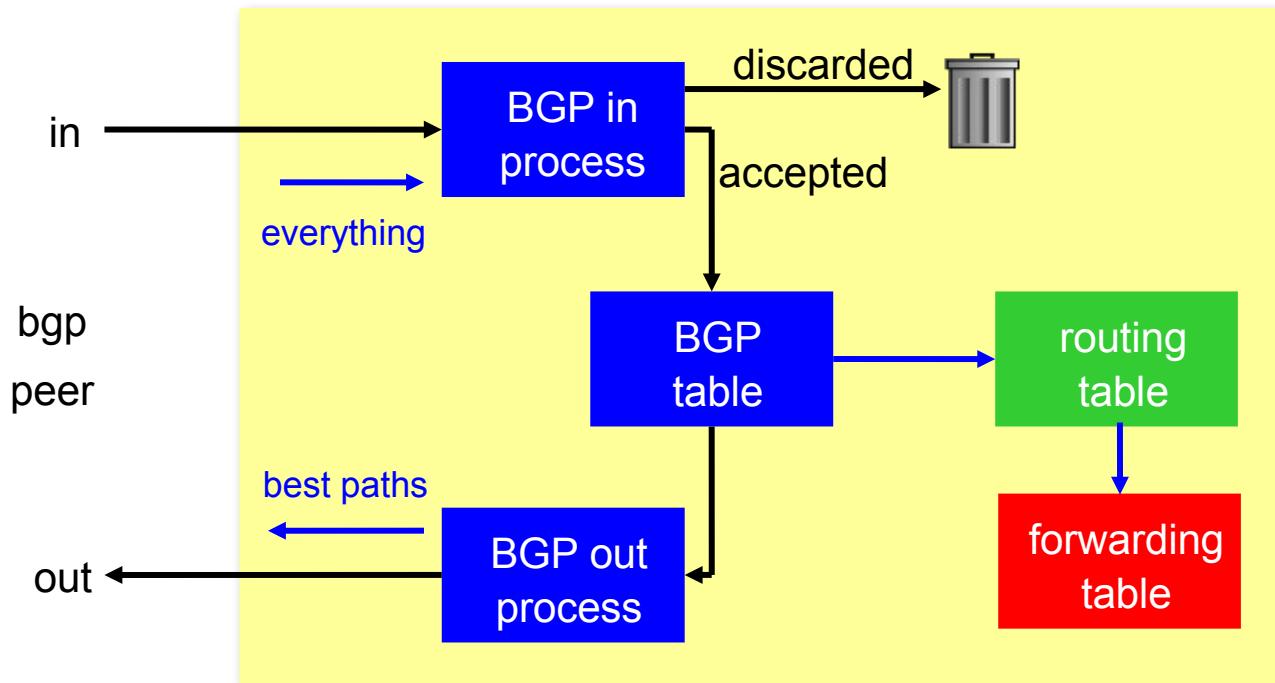
BGP General Operation

- Learns multiple paths via internal and external BGP speakers
- Picks the best path and installs it in the routing table (RIB)
- Best path is sent to external BGP neighbours
- Policies are applied by influencing the best path selection

Constructing the Forwarding Table

- BGP “in” process
 - Receives path information from peers
 - Results of BGP path selection placed in the BGP table
 - “best path” flagged
- BGP “out” process
 - Announces “best path” information to peers
- Best path stored in Routing Table (RIB) if:
 - Prefix and prefix length are unique, and
 - Lowest “protocol distance”
- Best paths in the RIB are installed in forwarding table (FIB)

Constructing the Forwarding Table



Supporting Multiple Protocols

- RFC4760
 - Defines Multi-protocol Extensions for BGP4
 - Enables BGP to carry routing information of protocols other than IPv4
 - e.g. MPLS, IPv6, Multicast etc
 - Exchange of multiprotocol NLRI must be negotiated at session startup
- RFC2545
 - Use of BGP Multiprotocol Extensions for IPv6 Inter-Domain Routing
 - Address family for IPv6

Supporting Multiple Protocols

- Independent operation
 - One RIB per protocol
 - IPv6 routes in BGP's IPv6 RIB
 - IPv4 routes in BGP's IPv4 RIB
 - Each protocol can have its own policies
- NEXTHOP
 - Address of the next router must be of the same address family as that of the local router

Supporting Multiple Protocols

- Cisco IOS assumes that all BGP neighbours will exchange IPv4 unicast prefixes
 - We need to remove this assumption

```
router bgp 100
  no bgp default ipv4-unicast
```

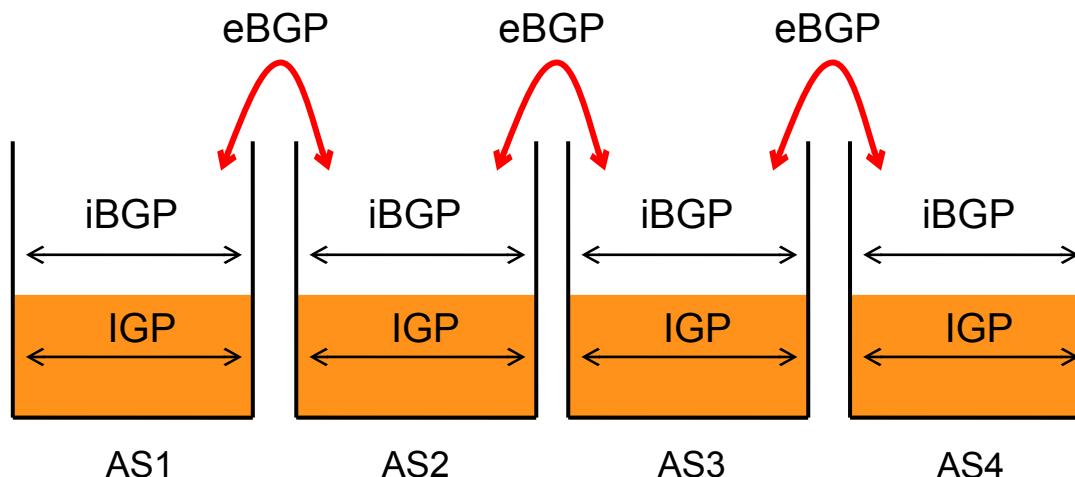
- For operational simplicity, the desire is for:
 - IPv4 neighbours to exchange IPv4 unicast prefixes
 - IPv6 neighbours to exchange IPv6 unicast prefixes
- Failure to do this results in:
 - IPv6 neighbours appearing to be set up to exchange IPv4 unicast prefixes
 - Cluttered configuration
 - Confusing troubleshooting and diagnosis

eBGP & iBGP

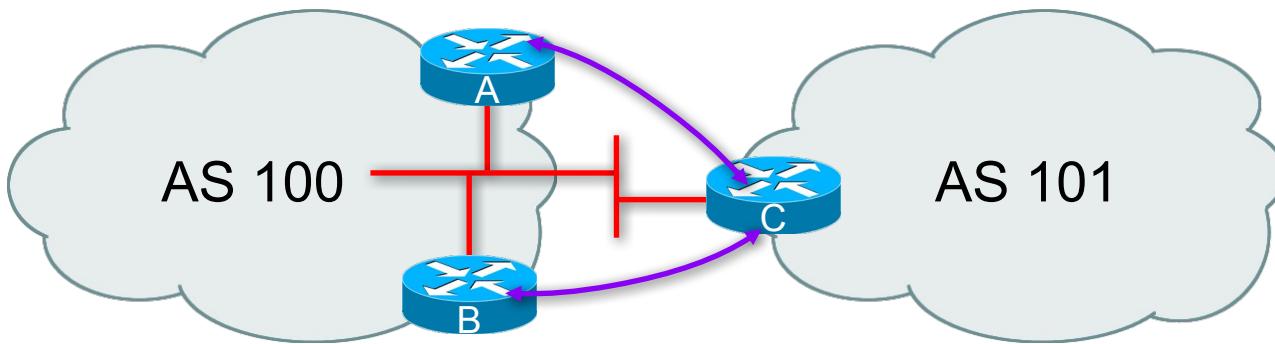
- BGP is used
 - Internally (iBGP)
 - Externally (eBGP)
- iBGP used to carry
 - Some/all Internet prefixes across ISP backbone
 - ISP's customer prefixes
- eBGP used to
 - Exchange prefixes with other ASes
 - Implement routing policy

BGP/IGP model used in ISP networks

- Model representation



External BGP Peering (eBGP)

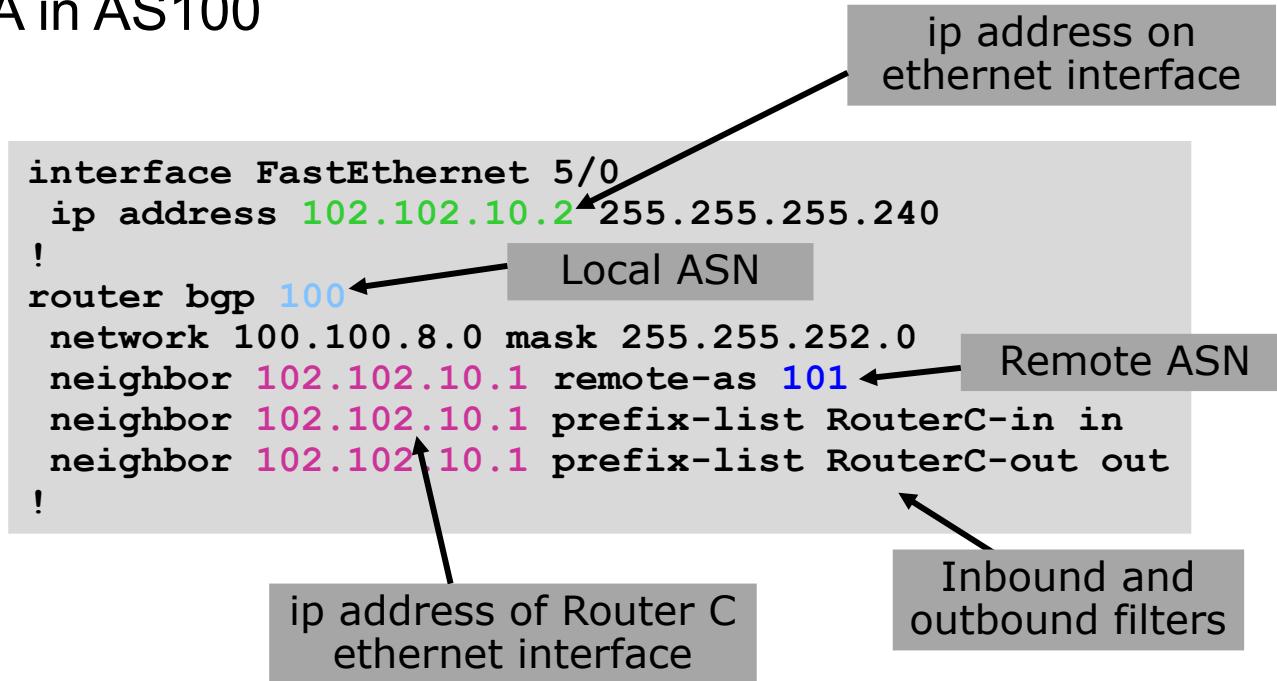


- Between BGP speakers in different AS
- Should be directly connected
- Never run an IGP between eBGP peers

Configuring External BGP

- Router A in AS100

```
interface FastEthernet 5/0
  ip address 102.102.10.2 255.255.255.240
!
router bgp 100
  network 100.100.8.0 mask 255.255.252.0
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list RouterC-in in
  neighbor 102.102.10.1 prefix-list RouterC-out out
!
```



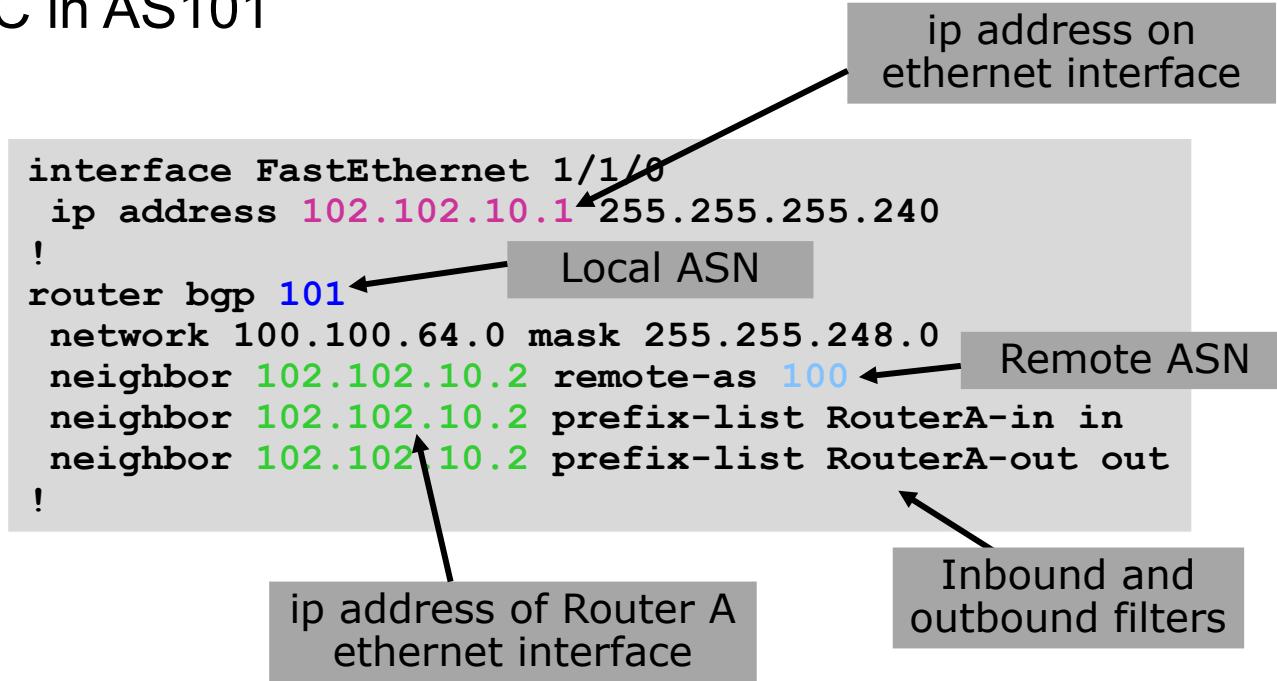
The diagram shows a configuration block for Router A in AS100. Annotations with arrows point to specific parts of the configuration:

- An arrow points from the text "ip address on ethernet interface" to the line "ip address 102.102.10.2 255.255.255.240".
- An arrow points from the text "Local ASN" to the number "100" in the "router bgp 100" command.
- An arrow points from the text "Remote ASN" to the number "101" in the "neighbor 102.102.10.1 remote-as 101" command.
- An arrow points from the text "ip address of Router C ethernet interface" to the IP address "102.102.10.1" in the "neighbor" command.
- An arrow points from the text "Inbound and outbound filters" to the two "prefix-list" commands: "RouterC-in in" and "RouterC-out out".

Configuring External BGP

- Router C in AS101

```
interface FastEthernet 1/1/0
  ip address 102.102.10.1 255.255.255.240
!
router bgp 101
  network 100.100.64.0 mask 255.255.248.0
  neighbor 102.102.10.2 remote-as 100
  neighbor 102.102.10.2 prefix-list RouterA-in in
  neighbor 102.102.10.2 prefix-list RouterA-out out
!
```



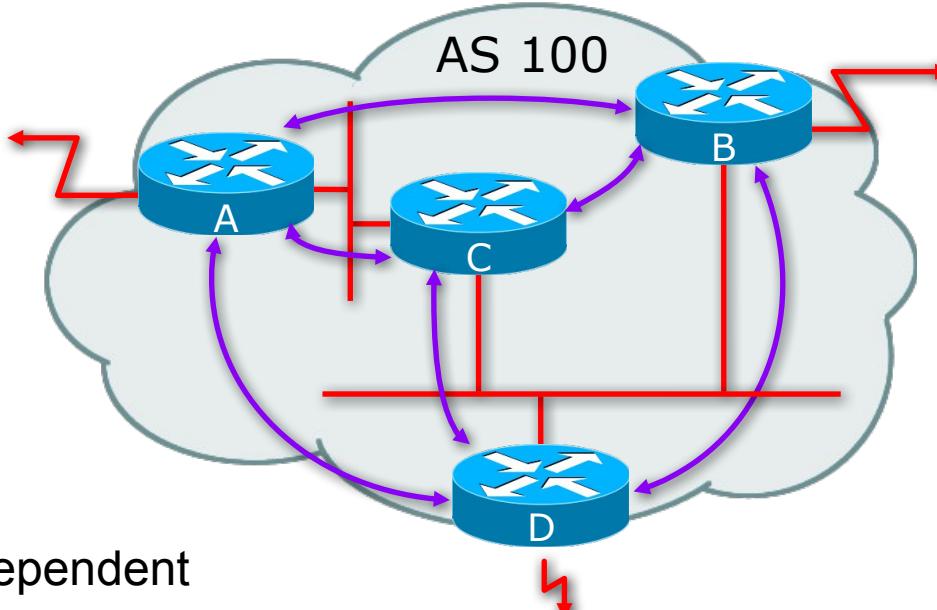
The diagram shows annotations for the BGP configuration on Router C:

- ip address on ethernet interface**: Points to the line `ip address 102.102.10.1 255.255.255.240`.
- Local ASN**: Points to the number `101` in the `router bgp 101` command.
- Remote ASN**: Points to the number `100` in the `neighbor 102.102.10.2 remote-as 100` command.
- Inbound and outbound filters**: Points to the two `prefix-list` commands: `neighbor 102.102.10.2 prefix-list RouterA-in in` and `neighbor 102.102.10.2 prefix-list RouterA-out out`.
- ip address of Router A ethernet interface**: Points to the IP address `102.102.10.2` in the `neighbor` command.

Internal BGP (iBGP)

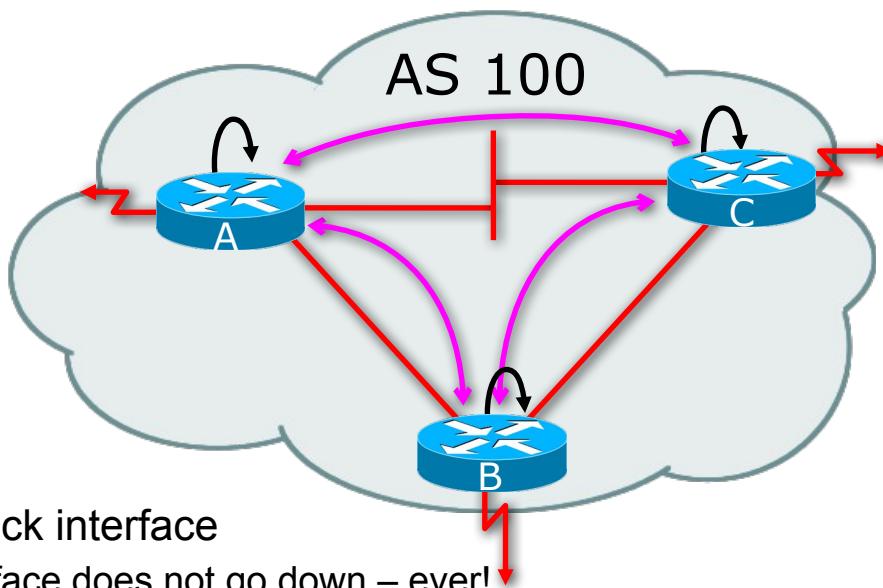
- BGP peer within the same AS
- Not required to be directly connected
 - IGP takes care of inter-BGP speaker connectivity
- iBGP speakers must be fully meshed:
 - They originate connected networks
 - They pass on prefixes learned from outside the ASN
 - **They do not pass on prefixes learned from other iBGP speakers**

Internal BGP Peering (iBGP)



- Topology independent
- Each iBGP speaker must peer with every other iBGP speaker in the AS

Peering between Loopback Interfaces



- Peer with loop-back interface
 - Loop-back interface does not go down – ever!
- Do not want iBGP session to depend on state of a single interface or the physical topology

Configuring Internal BGP

- Router A in AS100

```
interface loopback 0
  ip address 105.3.7.1 255.255.255.255
!
router bgp 100
  network 100.100.1.0
  neighbor 105.3.7.2 remote-as 100
  neighbor 105.3.7.2 update-source loopback0
  neighbor 105.3.7.3 remote-as 100
  neighbor 105.3.7.3 update-source loopback0
!
```

ip address on loopback interface

Local ASN

Local ASN

ip address of Router B loopback interface

Configuring Internal BGP

- Router B in AS100

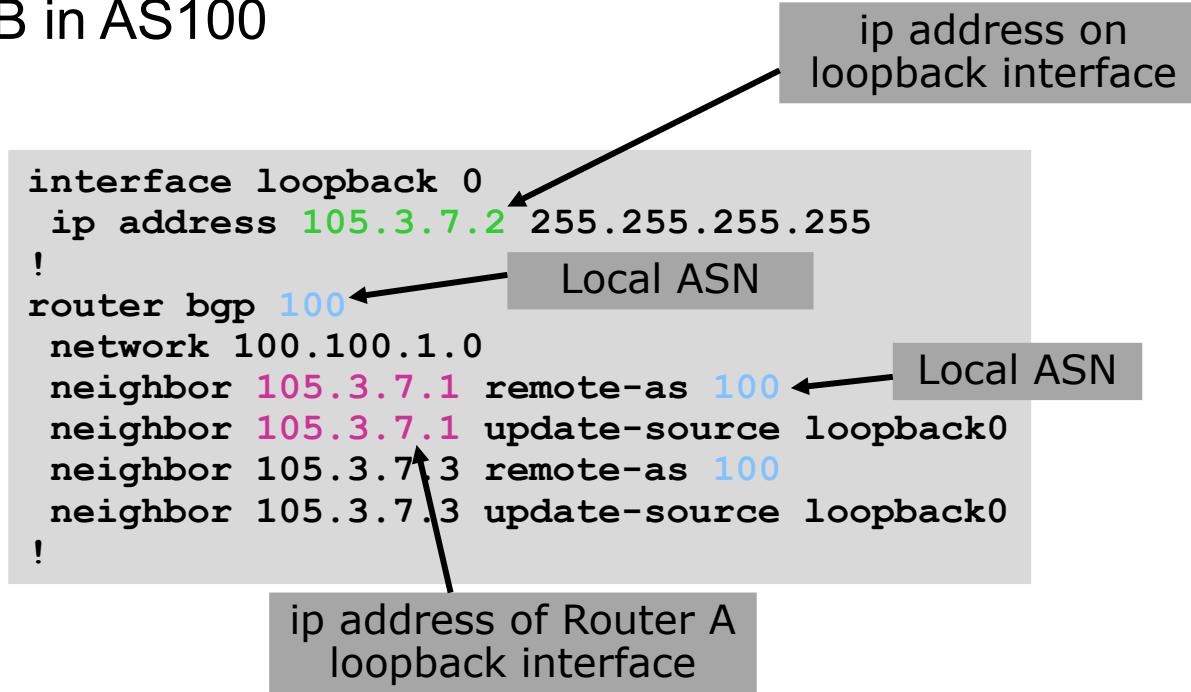
```
interface loopback 0
  ip address 105.3.7.2 255.255.255.255
!
router bgp 100
  network 100.100.1.0
  neighbor 105.3.7.1 remote-as 100
  neighbor 105.3.7.1 update-source loopback0
  neighbor 105.3.7.3 remote-as 100
  neighbor 105.3.7.3 update-source loopback0
!
```

ip address on loopback interface

Local ASN

Local ASN

ip address of Router A loopback interface



Inserting prefixes into BGP

- Two ways to insert prefixes into BGP
 - `redistribute static`
 - `network` command

Inserting prefixes into BGP – redistribute static

- Configuration Example:

```
router bgp 100
  redistribute static
    ip route 102.10.32.0 255.255.254.0 serial0
```

- Static route must exist before redistribute command will work
- Forces origin to be “incomplete”
- Care required!

Inserting prefixes into BGP – redistribute static

- Care required with redistribute!
 - redistribute <routing-protocol> means everything in the <routing-protocol> will be transferred into the current routing protocol
 - Will not scale if uncontrolled
 - Best avoided if at all possible
 - redistribute normally used with route-maps **and** under tight administrative control

Inserting prefixes into BGP – network command

- Configuration Example

```
router bgp 100
  network 102.10.32.0 mask 255.255.254.0
  ip route 102.10.32.0 255.255.254.0 serial0
```

- A matching route must exist in the routing table before the network is announced
- Forces origin to be “IGP”

Configuring Aggregation

- Three ways to configure route aggregation
 - **redistribute static**
 - **aggregate-address**
 - **network** command

Configuring Aggregation – Redistributing Static

- Configuration Example:

```
router bgp 100
  redistribute static
    ip route 102.10.0.0 255.255.0.0 null0
```

- Static route to “null0” is called a pull up route
 - Packets only sent here if there is no more specific match in the routing table
 - Care required – see previously!

Configuring Aggregation – Network Command

- Configuration Example

```
router bgp 100
  network 102.10.0.0 mask 255.255.0.0
  ip route 102.10.0.0 255.255.0.0 null0
```

- A matching route must exist in the routing table before the network is announced
- Easiest and best way of generating an aggregate

Configuring Aggregation – aggregate-address command

- Configuration Example:

```
router bgp 100
  network 102.10.32.0 mask 255.255.252.0
  aggregate-address 102.10.0.0 255.255.0.0 [summary-only]
  !
  ip route 102.10.32.0 255.255.252.0 null 0
```

- Requires more specific prefix in BGP table before aggregate is announced
- **summary-only** keyword
 - Optional keyword which ensures that only the summary is announced (the more specific routes are suppressed)

Summary

BGP neighbour status (IPv4)

```
Router6>sh ip bgp sum
BGP router identifier 10.0.15.246, local AS number 10
BGP table version is 16, main routing table version 16
7 network entries using 819 bytes of memory
14 path entries using 728 bytes of memory
2/1 BGP path/bestpath attribute entries using 248 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 1795 total bytes of memory
BGP activity 7/0 prefixes, 14/0 paths, scan interval 60 secs
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
10.0.15.241	4	10	9	8	16	0	0	00:04:47	2
10.0.15.242	4	10	6	5	16	0	0	00:01:43	2
10.0.15.243	4	10	9	8	16	0	0	00:04:49	2
...									

BGP Version

Updates sent
and received

Updates waiting

Summary

BGP neighbour status (IPv6)

```
Router1>sh bgp ipv6 unicast summary
BGP router identifier 10.10.15.224, local AS number 10
BGP table version is 28, main routing table version 28
18 network entries using 2880 bytes of memory
38 path entries using 3040 bytes of memory
9/6 BGP path/bestpath attribute entries using 1152 bytes of memory
4 BGP AS-PATH entries using 96 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 7168 total bytes of memory
BGP activity 37/1 prefixes, 95/19 paths, scan interval 60 secs
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
2001:DB8::2	4	10	185	182	28	0	0	02:36:11	16
2001:DB8::3	4	10	180	181	28	0	0	02:36:08	11
2001:DB8:0:4::1	4	40	153	152	28	0	0	02:05:39	9



Summary BGP Table

```
Router6>sh ip bgp
BGP table version is 16, local router ID is 10.0.15.246
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*>i 10.0.0.0/26	10.0.15.241	0	100	0	i
*>i 10.0.0.64/26	10.0.15.242	0	100	0	i
*>i 10.0.0.128/26	10.0.15.243	0	100	0	i
*>i 10.0.0.192/26	10.0.15.244	0	100	0	i
*>i 10.0.1.0/26	10.0.15.245	0	100	0	i
*> 10.0.1.64/26	0.0.0.0	0		32768	i
*>i 10.0.1.128/26	10.0.15.247	0	100	0	i
*>i 10.0.1.192/26	10.0.15.248	0	100	0	i
*>i 10.0.2.0/26	10.0.15.249	0	100	0	i
*>i 10.0.2.64/26	10.0.15.250	0	100	0	i
...					

Summary

- BGP4 – path vector protocol
- Multi-protocol (IPv4 and IPv6)
- iBGP versus eBGP
- Stable iBGP – peer with loopbacks
- Announcing prefixes & aggregates

Questions?