



Introduction to BGP

BGP for ISPs



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

Acknowledgements

- ❑ This material originated from the Cisco ISP/IXP Workshop Programme developed by Philip Smith & Barry Greene
- ❑ Use of these materials is encouraged as long as the source is fully acknowledged and this notice remains in place
- ❑ Bug fixes and improvements are welcomed
 - Please email *workshop (at) bgp4all.com*

Philip Smith

BGP Videos

- ❑ NSRC has made a video recording of this presentation, as part of a library of BGP videos for the whole community to use:
 - https://learn.nsrc.org/bgp#intro_to_bgp

The screenshot shows the NSRC (Network Startup Resource Center) website. The main content area is titled 'BGP for All' and features a video player for 'BGP for All: Internet Routing'. Below the video player, there are sections for 'BGP Case Studies' and 'Communities'. The 'Introduction to Routing' section lists various topics such as 'Internet Routing', 'Routing Protocols', 'Introduction to IS-IS', 'OS-OS Adjacencies', 'Best Configuration Practices for IS-IS on Cisco IOS', 'OS-OS Authentication, Default Routes and IPv6', 'Introduction to OSPF', 'OSPF Areas', 'OSPF Adjacencies', 'Best Configuration Practices for OSPF on Cisco IOS', 'OSPF Authentication, Default Routes and IPv6', 'Comparing OSPF and IS-IS', 'Choosing between OSPF and IS-IS', 'Migrating from OSPF to IS-IS', 'Migration Plan', and 'Finalizing Migration'. The 'Introduction to BGP' section lists 'Introduction to Border Gateway Protocol', 'Peering and Peering Autonomous Systems', 'How BGP works', 'Supporting Multiple Protocols', 'BGP versus EBGP', 'Setting up EBGP', and 'Setting up iBGP'.

-
- 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

- BGP is classified as a *path vector* routing protocol (see RFC 1322)

Path Vector Protocol

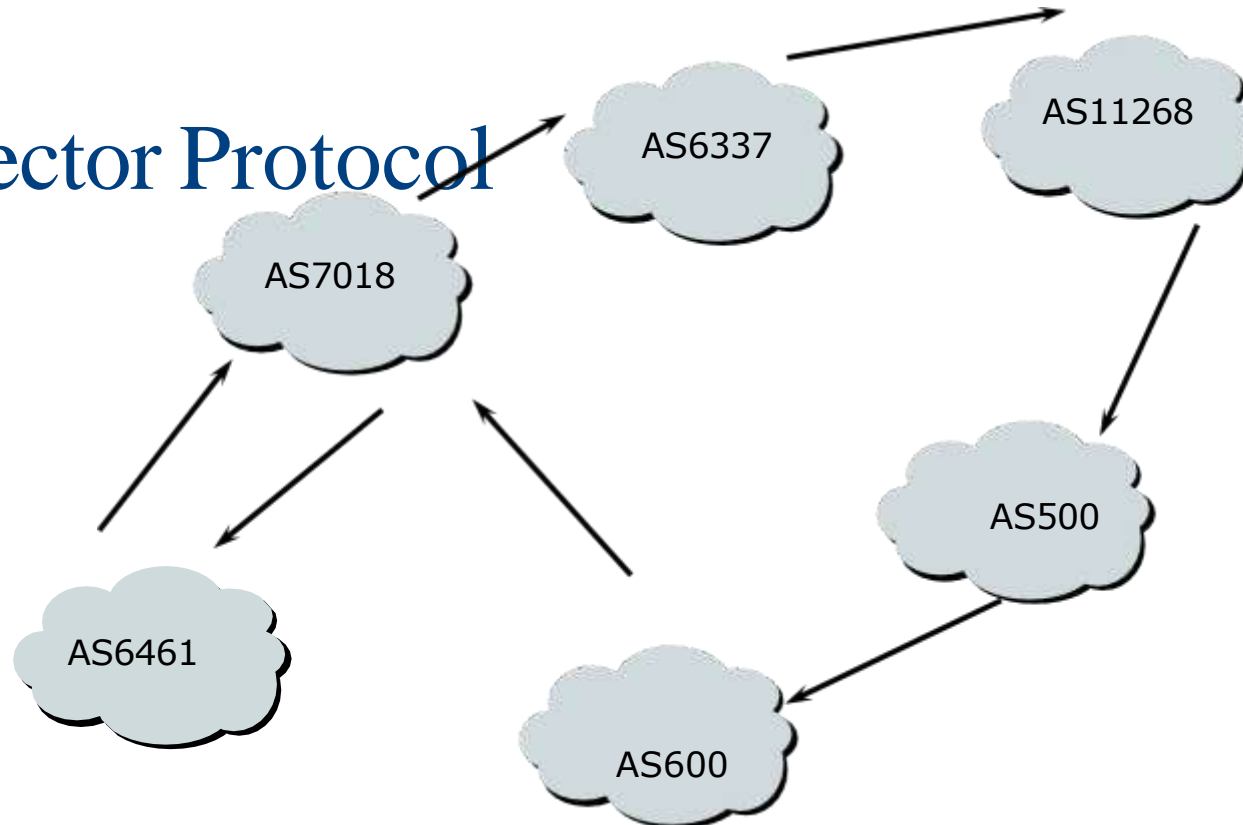
- A path vector protocol defines a route as a pairing between a destination and the attributes of the path to that destination.

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



AS Path

Path Vector Protocol



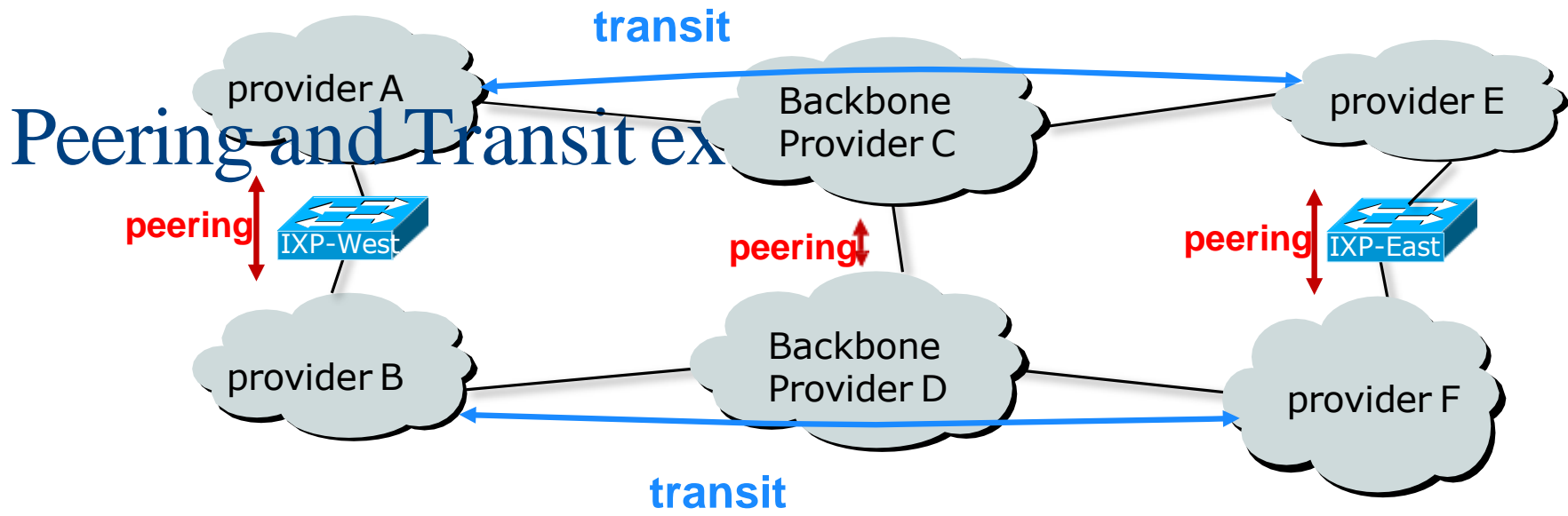
Definitions

- **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

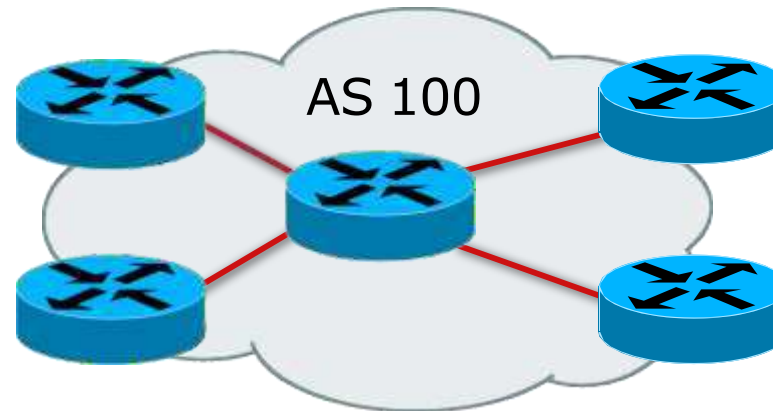
NB: is not related to where an ISP is in the hierarchy



Peering and Transit ex

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

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 (ASN)

- 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 41200 16-bit ASNs are visible on the Internet
 - (this number is dropping slightly as 32-bit ASN numbers increase)
- Each RIR has also received a block of 32-bit ASNs
 - Out of 36400 assignments, around 30300 are visible on the Internet (June 2021)
- 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 displays ASNs in plain notation by default
 - Dot notation is optional (and **NOT** recommended):

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

- This command sets the local autonomous system number

Conf `set routing-options autonomous-system 131076`

- All BGP configuration is then carried out under:

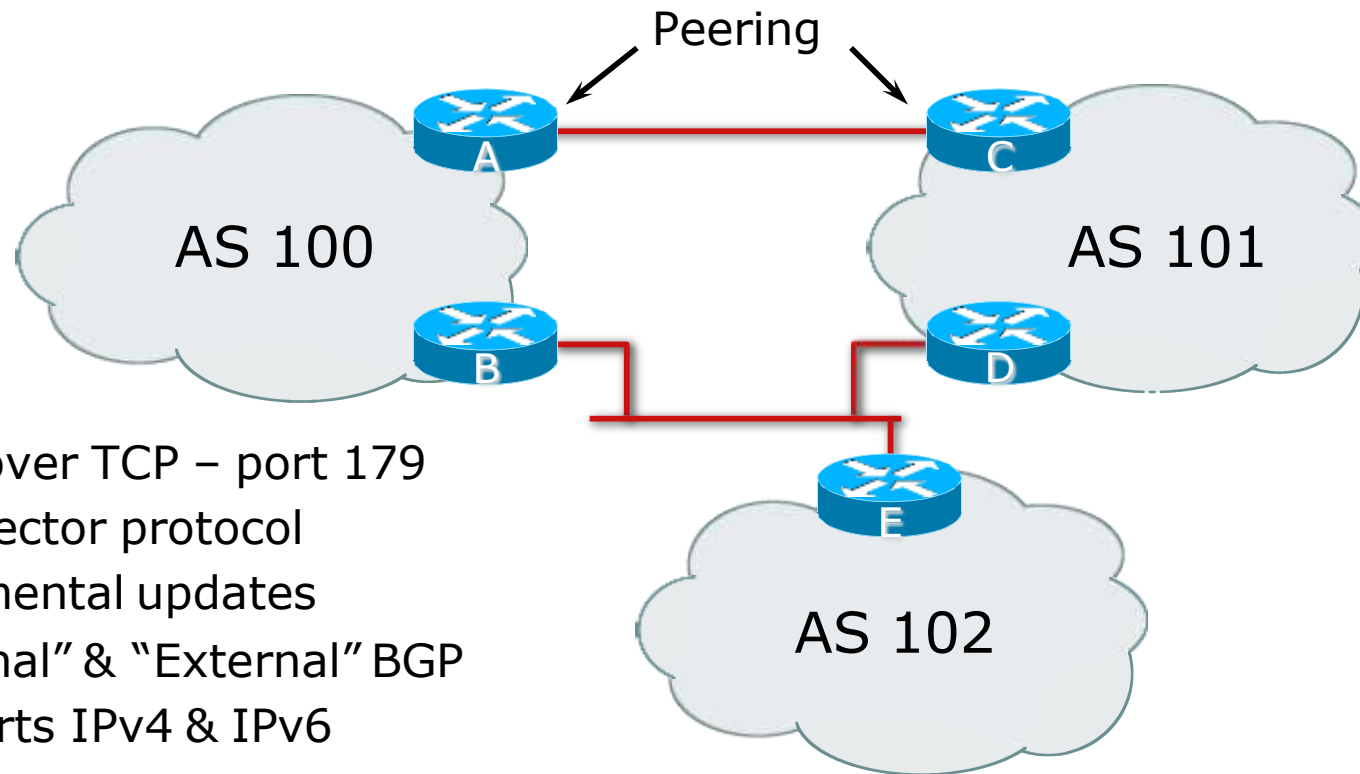
```
edit protocols bgp
```

- JunOS displays ASNs in plain notation by default

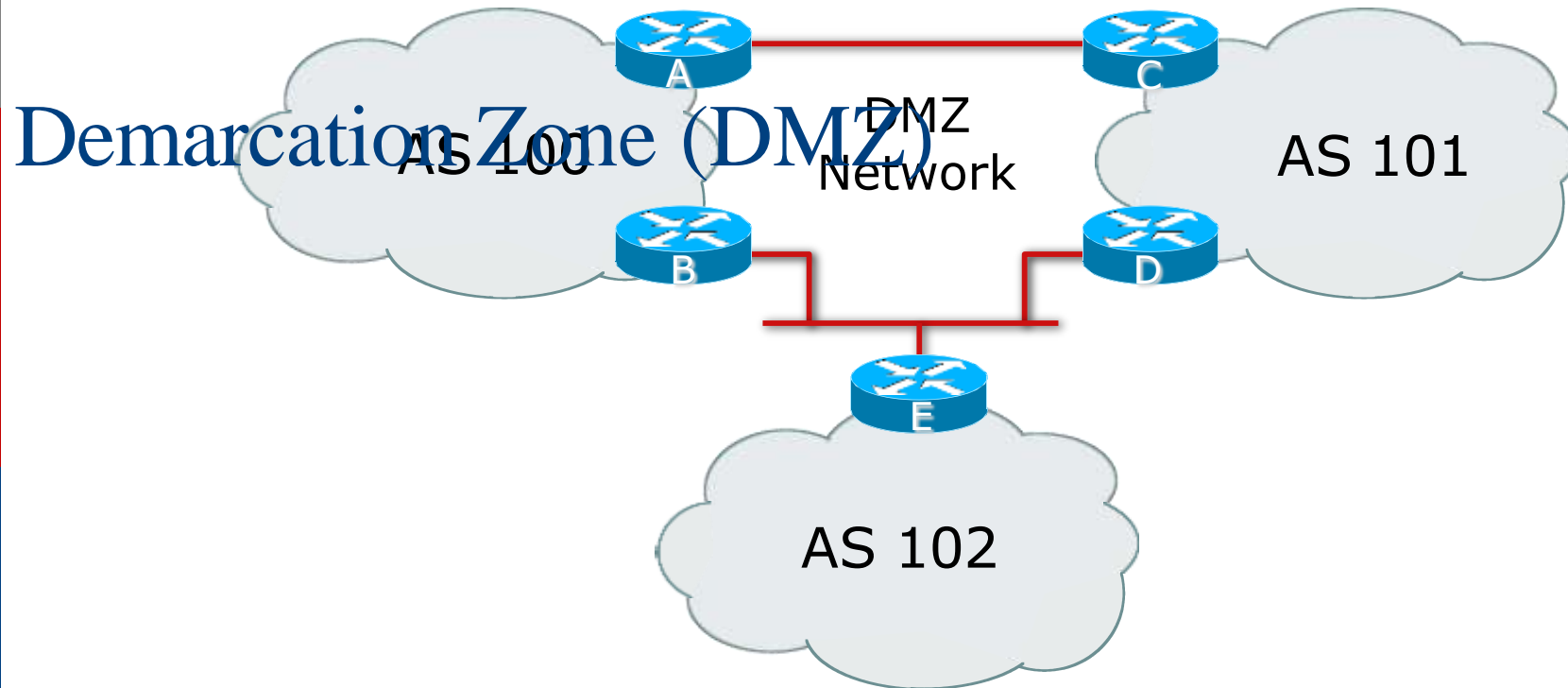
- Dot notation is optional (and **NOT** recommended):

```
set routing-options autonomous-system asdot-notation 2.4
```

BGP Basics



- ❑ Runs over TCP – port 179
- ❑ Path vector protocol
- ❑ Incremental updates
- ❑ “Internal” & “External” BGP
- ❑ Supports IPv4 & IPv6



Demarcation Zone (DMZ)

- DMZ is the link or network shared between ASes

-
- Learns multiple paths via internal and external BGP

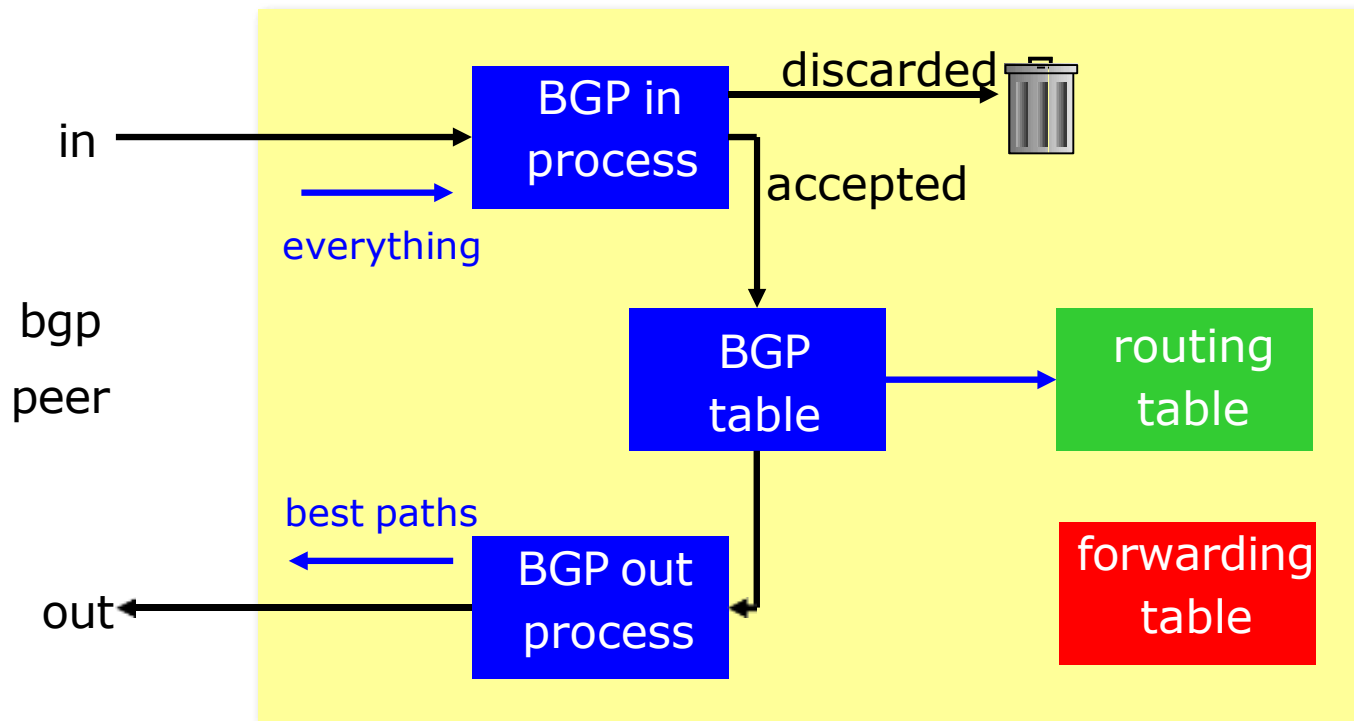
BGP General Operation

- 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 (after best path selection)
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
 - The IP address of the next router must belong to 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
 - Most other implementations do not
 - We need to remove this assumption in Cisco IOS
- ```
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

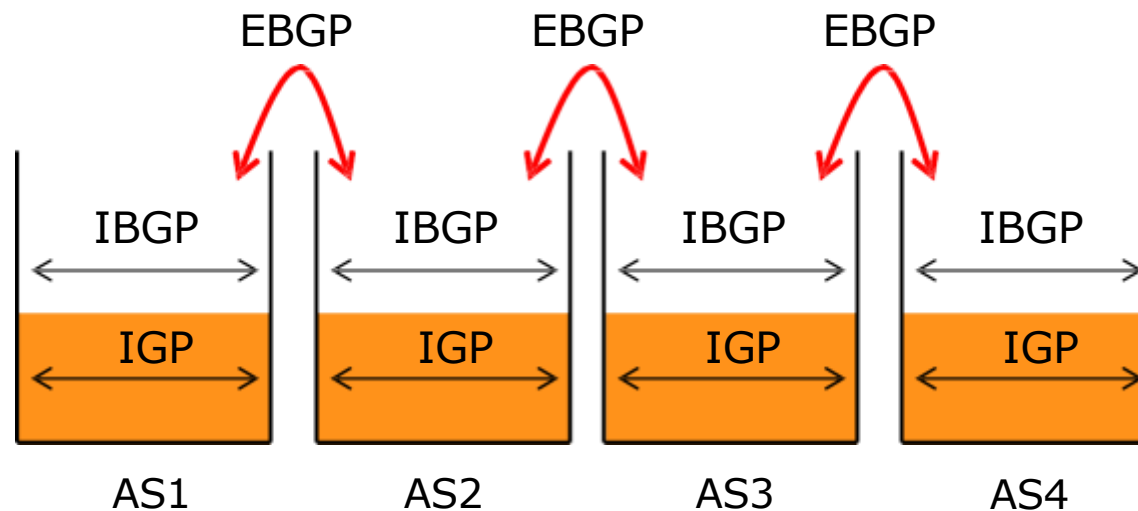
---

- 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

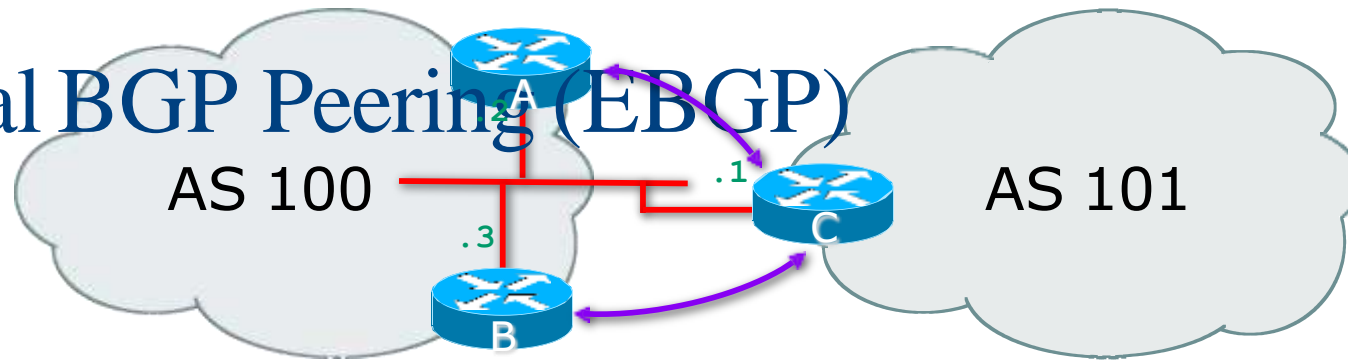


- Model representation

## BGP/IGP model used in ISP networks



## External BGP Peering (EBGP)



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

□ Router A in AS100

# Configuring External BGP

```
interface FastEthernet 5/0
 ip address 102.102.10.2 255.255.255.240
!
router bgp 100
 address-family ipv4
 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
 neighbor 102.102.10.1 activate
!
```

ip address on ethernet interface

Local ASN

Select IPv4 or IP v6

Remote A SN

Inbound and outbound filters

ip address of Router C ethernet interface

- Router C in AS101

## Configuring External BGP

```

interface FastEthernet 1/1/0
ip address 102.102.10.1 255.255.255.240
!
router bgp 101
address-family ipv4
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
neighbor 102.102.10.2 activate
!

```

ip address on ethernet interface

Local ASN

Select IPv4 or IP v6

Remote A SN

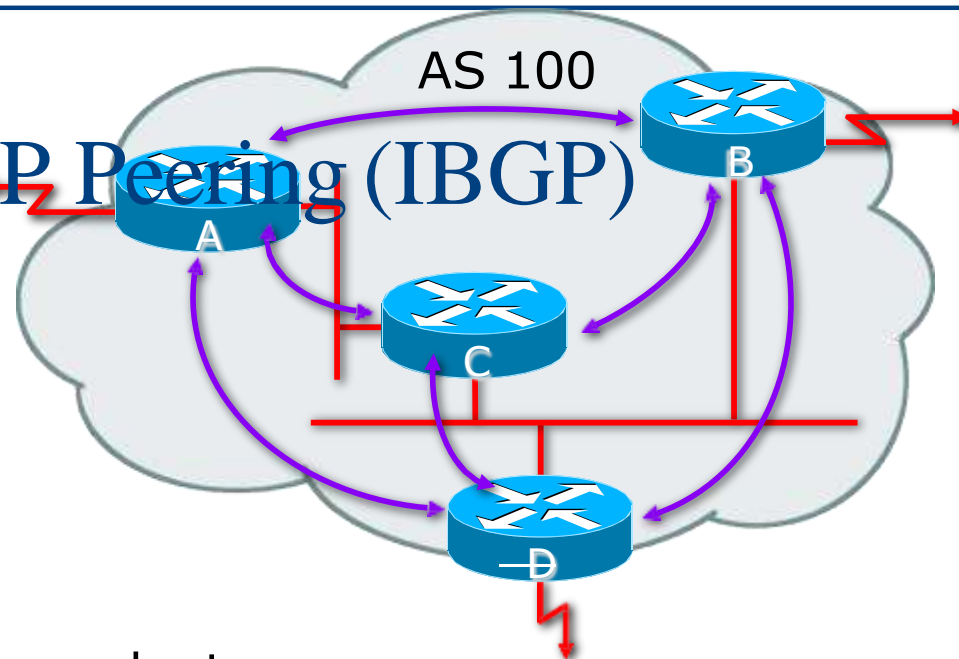
Inbound and outbound filters

ip address of Router A ethernet interface

## Internal BGP (IBGP)

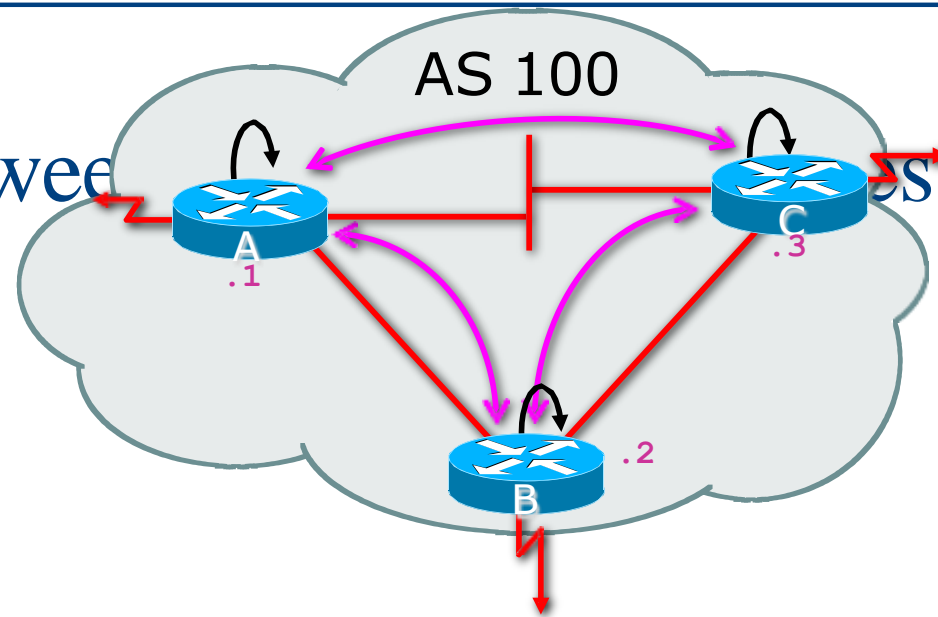
- ❑ BGP peer within the same AS
- ❑ Not required to be directly connected
  - IGP takes care of inter-BGP speaker connectivity
- ❑ IBGP speakers must be fully meshed:
  - They originate connected networks
  - They pass on prefixes learned from outside the AS
  - **They do not pass on prefixes learned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

## Internal BGP Peering (IBGP)



- ❑ Topology independent
- ❑ Each IBGP speaker must peer with every other IBGP speaker in the AS as per ↔

## Peering between



- ❑ 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
 address-family ipv4
 network 100.100.1.0 mask 255.255.255.0
 neighbor 105.3.7.2 remote-as 100
 neighbor 105.3.7.2 update-source loopback0
 neighbor 105.3.7.2 activate
 neighbor 105.3.7.3 remote-as 100
 neighbor 105.3.7.3 update-source loopback0
 neighbor 105.3.7.3 activate
!
```

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
 address-family ipv4
 network 100.100.1.0 mask 255.255.255.0
 neighbor 105.3.7.1 remote-as 100
 neighbor 105.3.7.1 update-source loopback0
 neighbor 105.3.7.1 activate
 neighbor 105.3.7.3 remote-as 100
 neighbor 105.3.7.3 update-source loopback0
 neighbor 105.3.7.3 activate
!
```

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
 address-family ipv4
 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 named *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
 address-family ipv4
 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
 address-family ipv4
 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
 address-family ipv4
 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
 address-family ipv4
 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
```

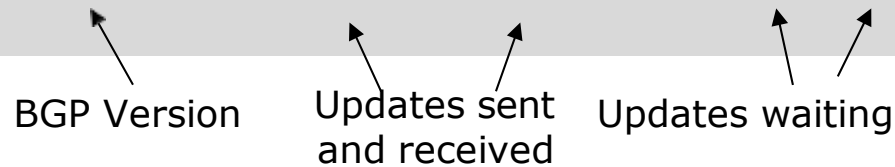
- ❑ Requires more specific prefix in BGP table before aggregate is announced
- ❑ `summary-only` keyword
  - Optional keyword which ensures that only the summary is announced (the more specific routes are suppressed)

# Summary

## BGP neighbour status (Cisco IOS IPv4)

```
Router6>show ip bgp summary
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
...
```



# Summary

## BGP neighbour status (Cisco IOS 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
```



Neighbour Information



BGP Messages Activity

# Summary

## BGP neighbour status (JunOS)

```

philip@R6> show bgp summary
Groups: 1 Peers: 14 Down peers: 0
Table Tot Paths Act Paths Suppressed History Damp State Pending
inet.0 20 20 0 0 0 0
inet6.0 20 20 0 0 0 0

Peer AS InPkt OutPkt OutQ Flaps Last Up/Dwn State|#Active/Received/Accepted/Damped..
10.0.15.241 10 1067980 202487 0 0 9w1d 4:32:05 Establ inet.0: 10/10/10/0
10.0.15.242 10 204577 1001705 0 0 9w1d 4:32:09 Establ inet.0: 3/3/3/0
10.0.15.243 10 277630 1886656 0 0 9w1d 4:32:06 Establ inet.0: 4/4/4/0
...
2001:DB8::1 10 416832 202568 0 0 9w1d 4:30:46 Establ inet6.0: 10/10/10/0
2001:DB8::2 10 204605 411166 0 0 9w1d 4:34:47 Establ inet6.0: 3/3/3/0
2001:DB8::3 10 277568 729073 0 0 9w1d 1:03:31 Establ inet6.0: 2/2/2/0
...

```

AS Number

Updates sent  
and received

Updates waiting

Address Family

# Summary

## BGP Table (Cisco IOS IPv4)

```
Router6>sh ip bgp
BGP table version is 18, 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    |
| *>i 10.0.2.128/26 | 10.0.15.251 | 0      | 100    | 0      | i    |
| *>i 10.0.2.192/26 | 10.0.15.252 | 0      | 100    | 0      | i    |
| *>i 10.0.3.0/26   | 10.0.15.253 | 0      | 100    | 0      | i    |
| *>i 10.0.3.64/26  | 10.0.15.254 | 0      | 100    | 0      | i    |

# Summary

## BGP Table (Cisco IOS IPv6)

```
Router6>sh bgp ipv6 unicast
BGP table version is 18, 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 2001:DB8:1::/48 | 2001:DB8::1 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:2::/48 | 2001:DB8::2 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:3::/48 | 2001:DB8::3 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:4::/48 | 2001:DB8::4 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:5::/48 | 2001:DB8::5 | 0      | 100    | 0      | i    |
| *> 2001:DB8:6::/48  | ::          | 0      |        | 32768  | i    |
| *>i 2001:DB8:7::/48 | 2001:DB8::7 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:8::/48 | 2001:DB8::8 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:9::/48 | 2001:DB8::9 | 0      | 100    | 0      | i    |
| *>i 2001:DB8:A::/48 | 2001:DB8::A | 0      | 100    | 0      | i    |
| *>i 2001:DB8:B::/48 | 2001:DB8::B | 0      | 100    | 0      | i    |
| *>i 2001:DB8:C::/48 | 2001:DB8::C | 0      | 100    | 0      | i    |
| *>i 2001:DB8:D::/48 | 2001:DB8::D | 0      | 100    | 0      | i    |
| *>i 2001:DB8:E::/48 | 2001:DB8::E | 0      | 100    | 0      | i    |

# Summary

## BGP Table (JunOS)

```
philip@R6> show route protocol bgp terse

inet.0: 14 destinations, 14 routes (14 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

A V Destination P Prf Metric 1 Metric 2 Next hop AS path
? 10.0.0.0/26 B 100 >10.0.15.241 I
unverified
? 10.0.0.64/26 B 100 >10.0.15.241 I
unverified
...
? 10.1.0.0/24 B 100 >10.0.15.242 20 I
unverified
? 10.4.0.0/24 B 100 >10.0.15.241 20 I
unverified
...

inet6.0: 14 destinations, 14 routes (14 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

A V Destination P Prf Metric 1 Metric 2 Next hop AS path
? 2001:DB8:1::/48 B 100 >fe80::82ac:acff:fed2:ea88 I
unverified
? 2001:DB8:2::/48 B 100 >fe80::82ac:acff:fed2:ea88 I
unverified
...
? 2001:DB9::/32 B 100 >fe80::224e:71ff:fe90:2500 20 I
unverified
? 2001:DB9::/32 B 100 >fe80::82ac:acff:fed2:ea88 20 I
unverified
...
```

# Summary

---

- ❑ BGP – path vector protocol
- ❑ Multi-protocol (IPv4 & IPv6)
- ❑ IBGP versus EBGP
- ❑ Stable IBGP – peer with loopbacks
- ❑ Announcing prefixes & aggregates