Campus Network Design Workshop

Introduction to the Border Gateway Protocol

UNIVERSITY OF OREGON

NSRC
Network Startup Resource Center

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

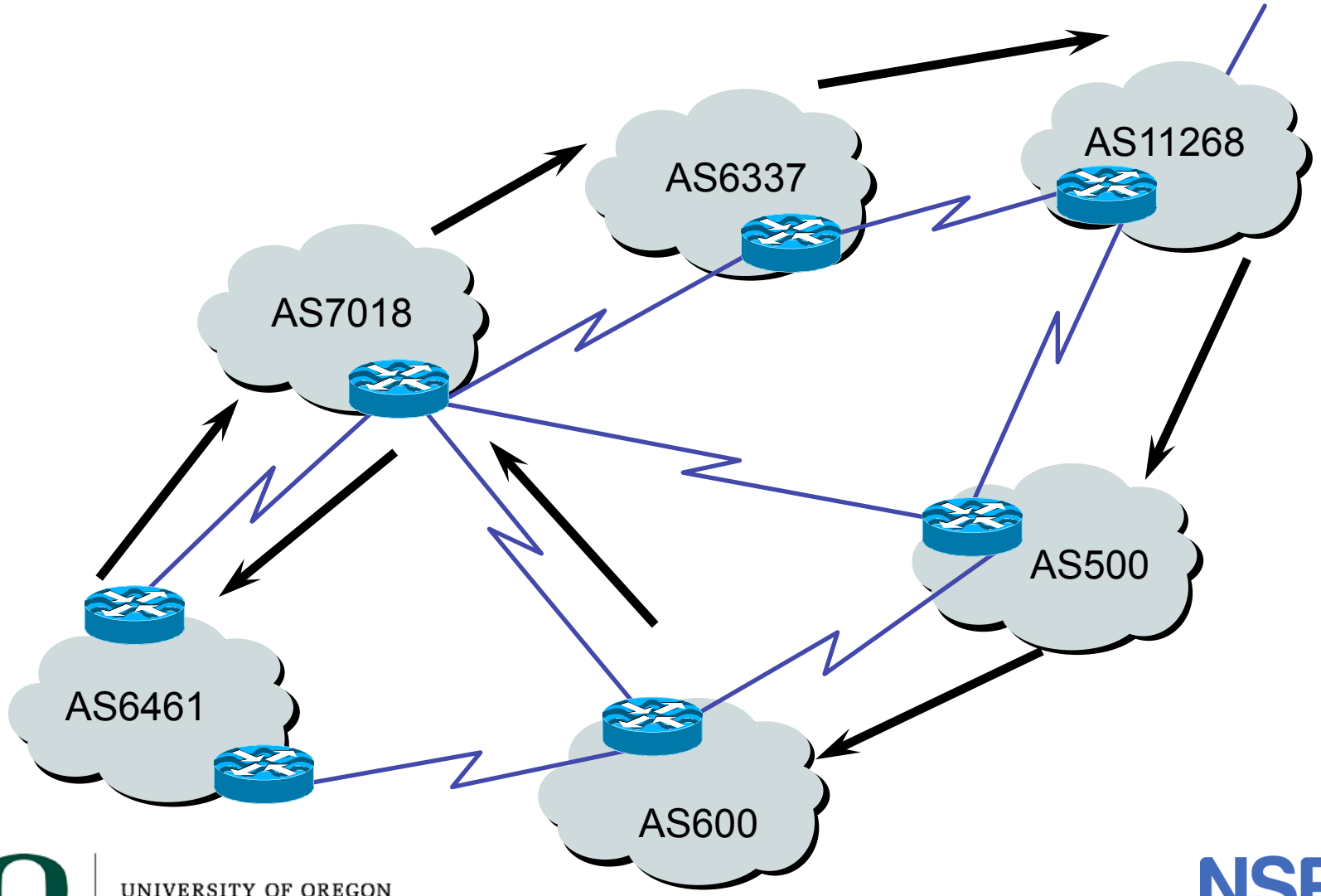
Path Vector Protocol

- 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

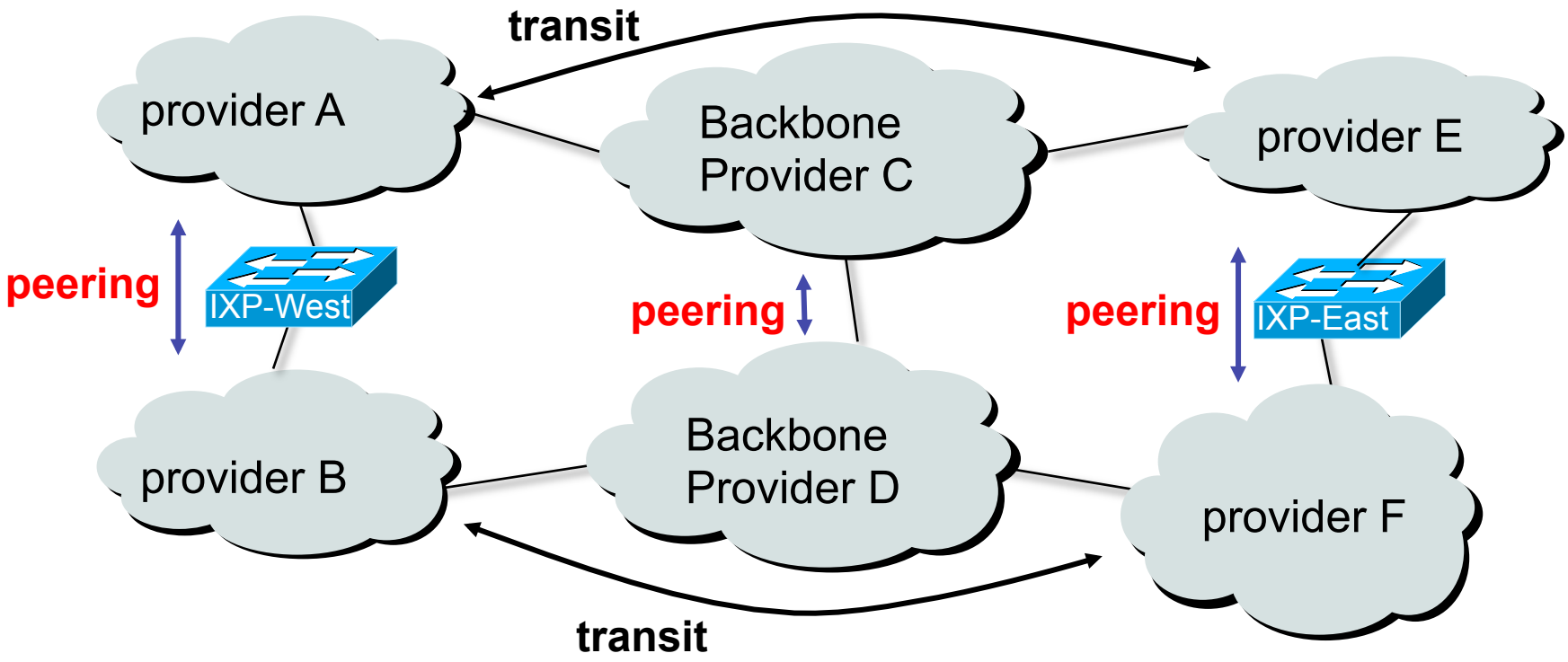
- Transit – carrying traffic across a network
  - (Commercially: for a fee)
- Peering – exchanging routing information and traffic
  - (Commercially: between similar sized networks, and for no fee)
- Default – where to send traffic when there is no explicit match in the routing table

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

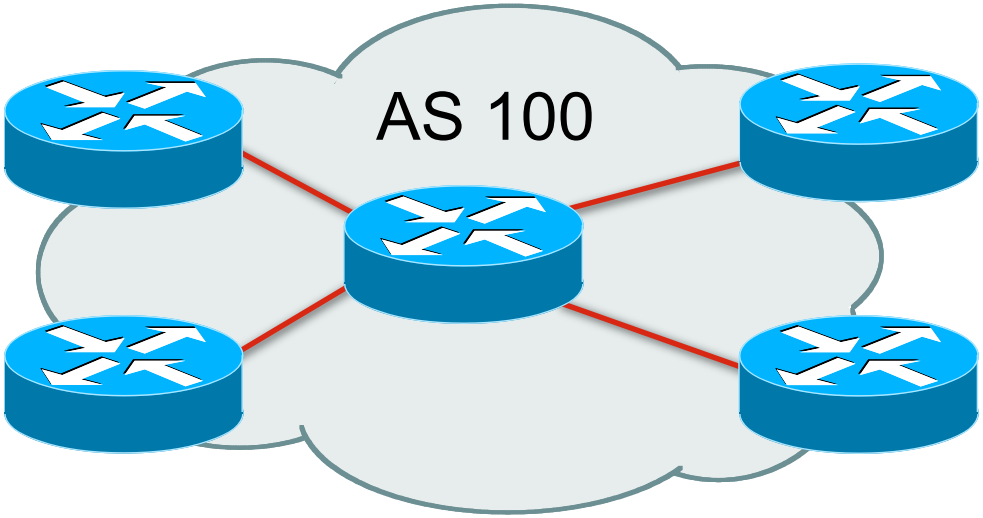
NB: is not related to where an ISP is in the hierarchy

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 (ASN)

- Two ranges
  - 0-65535 (original 16-bit range)
  - 65536-4294967295 (32-bit range – RFC6793)
- 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 – RFC6996)
- 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
- Current 16-bit ASN assignments up to 64395 have been made to the RIRs
  - Around 43000 16-bit ASNs are visible on the Internet
  - 100 left unassigned
- Each RIR has also received a block of 32-bit ASNs
  - Out of 13400 assignments, around 10300 are visible on the Internet
- 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
```

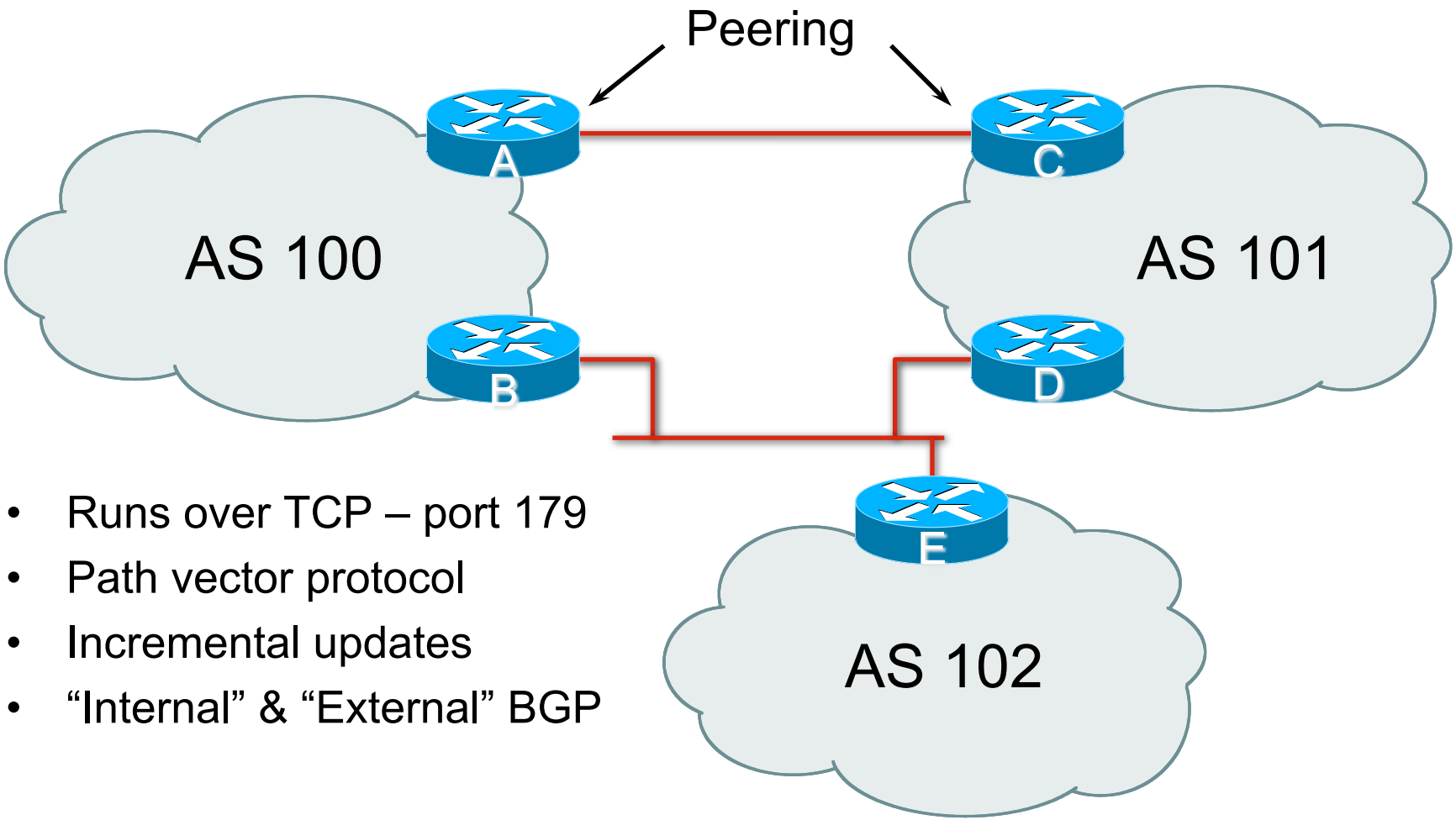
  – Or

```
router bgp 2.4
```

- IOS will display ASNs in plain notation by default
  – Dot notation is optional:

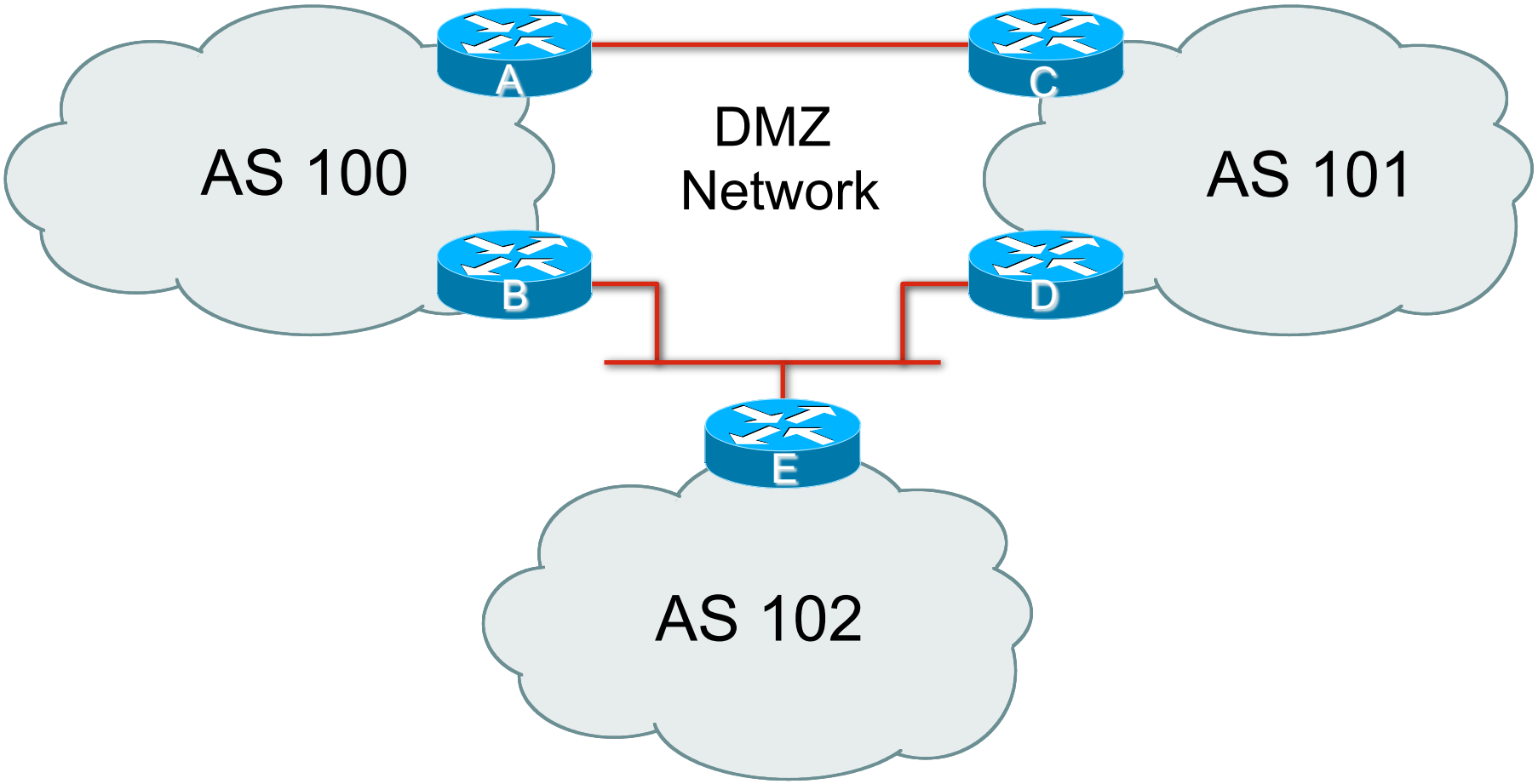
```
router bgp 2.4
 bgp asnotation dot
```

BGP Basics



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

Demarcation Zone (DMZ)



- DMZ is the link or network shared between ASes

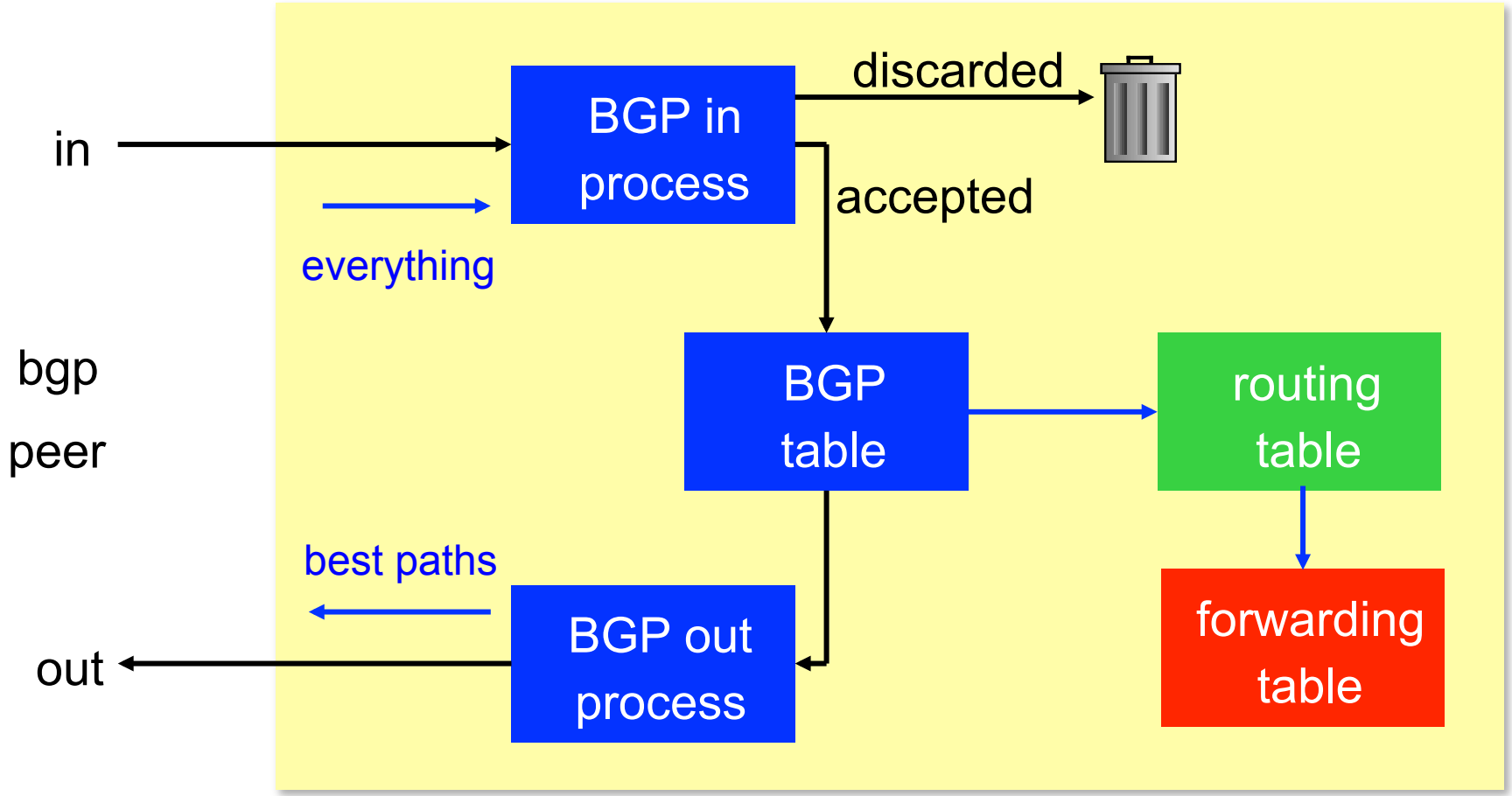
BGP General Operation

- 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

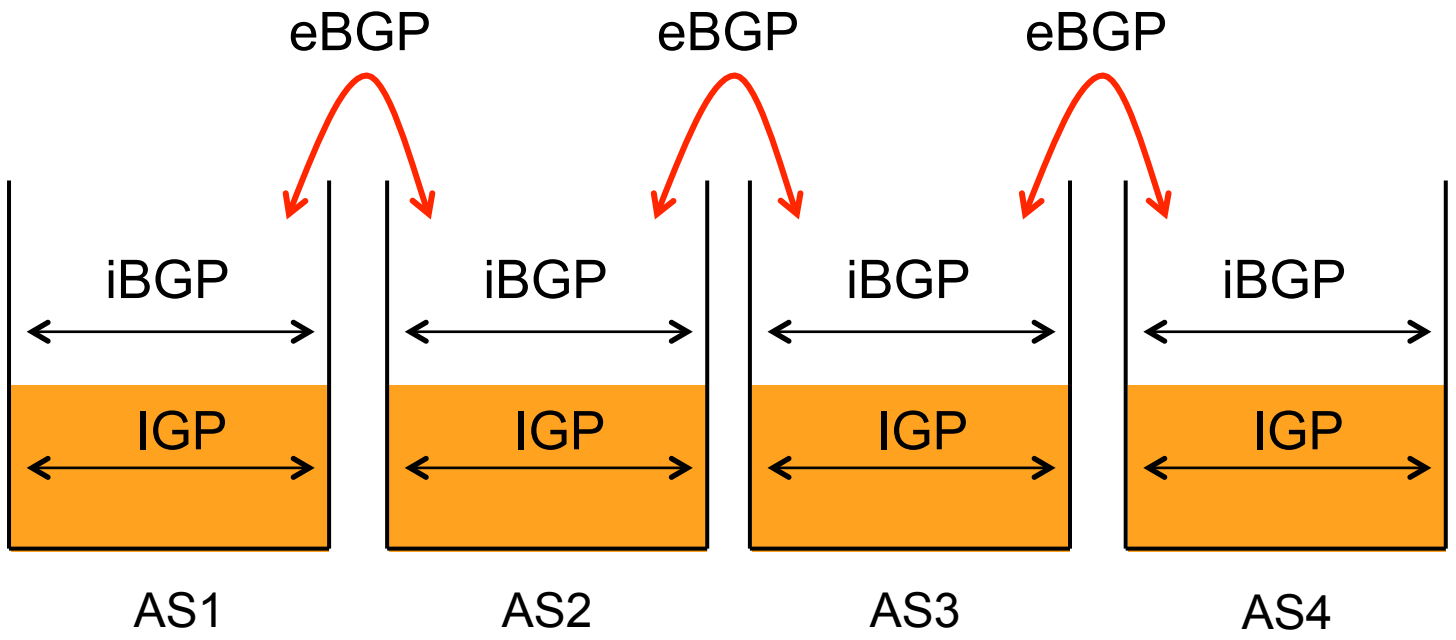


eBGP & iBGP

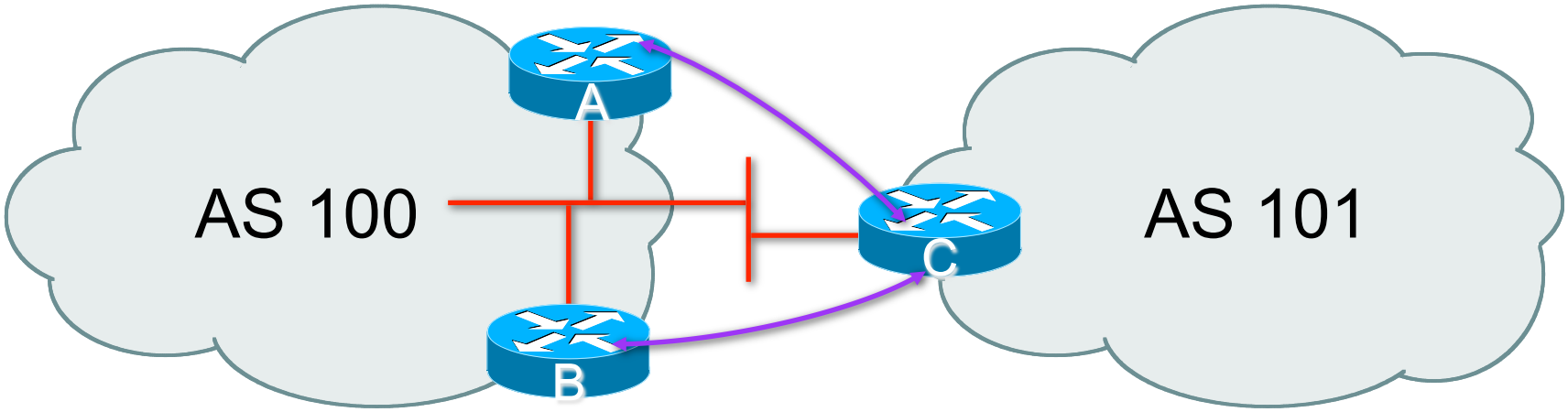
- BGP is used
  - Internally (iBGP)
  - Externally (eBGP)
- iBGP used to carry
  - Some/all Internet prefixes across ISP backbone
  - ISP's customer prefixes
- eBGP used to
  - Exchange prefixes with other ASes
  - Implement routing policy

BGP/IGP model used in ISP networks

- Model representation



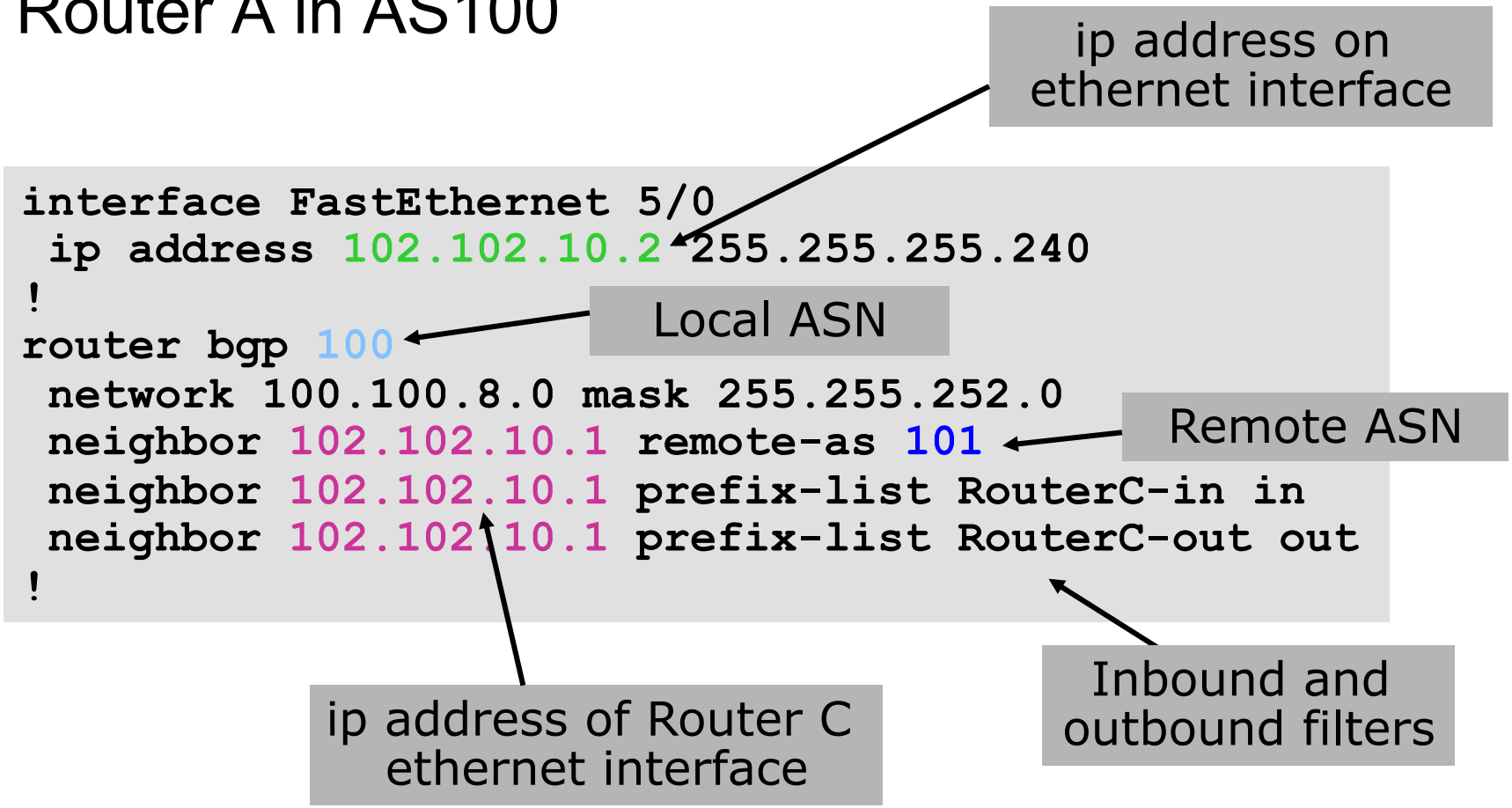
External BGP Peering (eBGP)



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

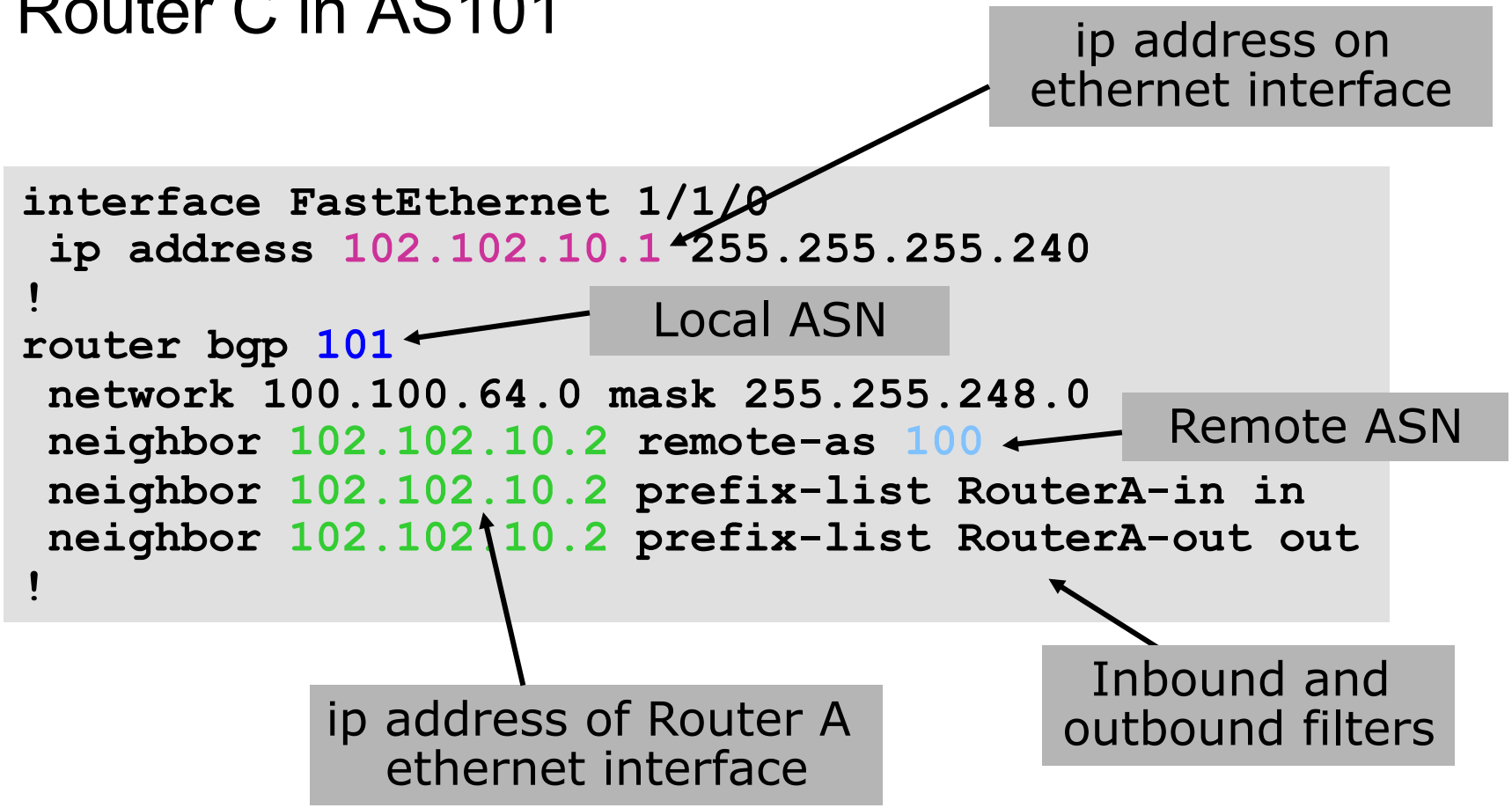
Configuring External BGP

- Router A in AS100



Configuring External BGP

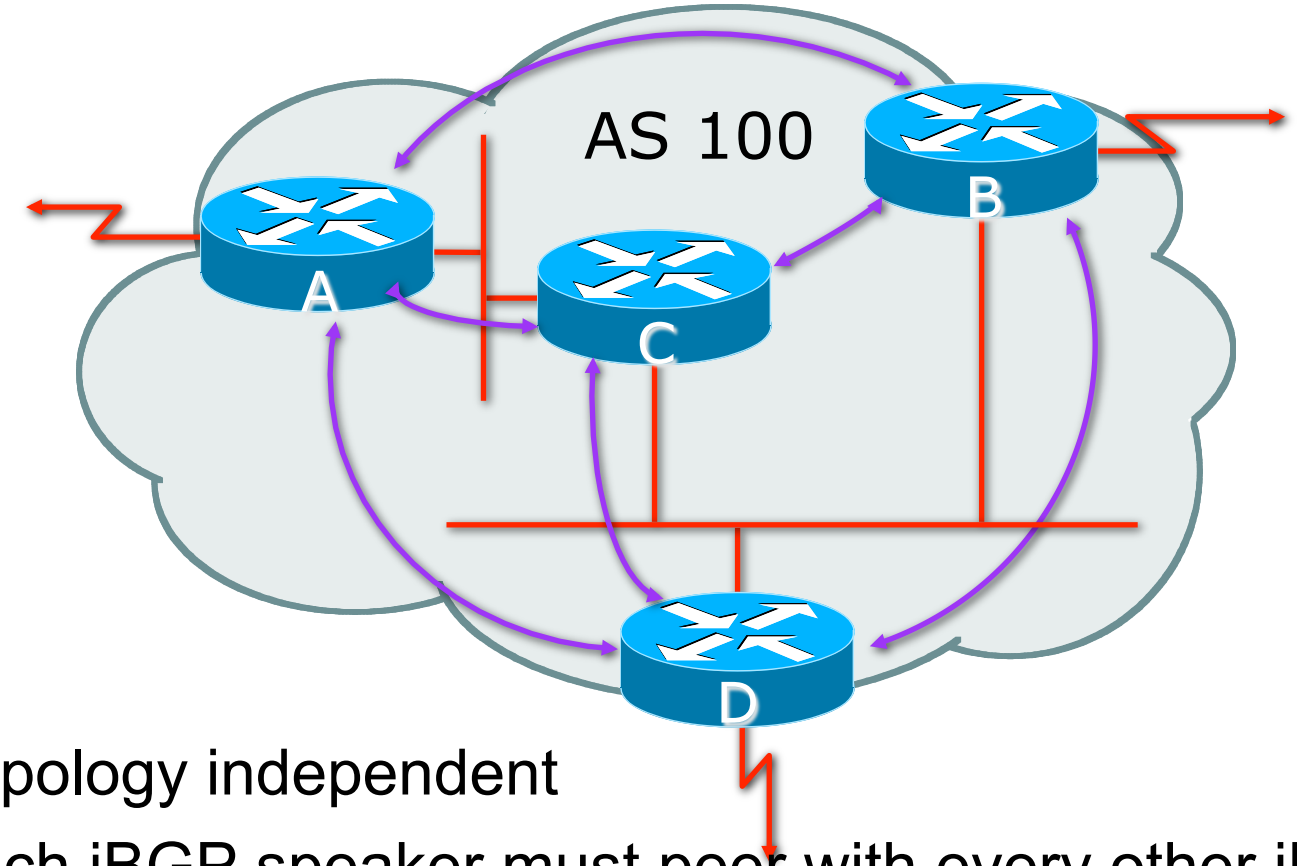
- Router C in AS101



Internal BGP (iBGP)

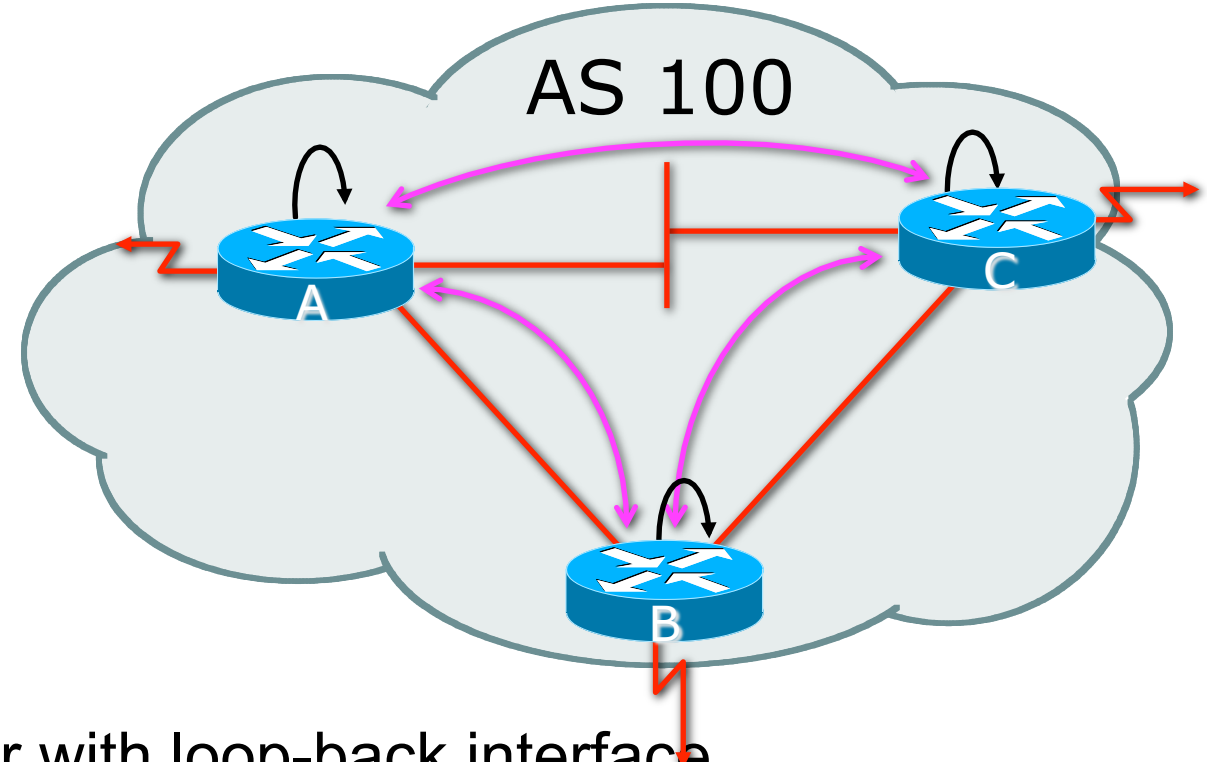
- BGP peer within the same AS
- Not required to be directly connected
  - IGP takes care of inter-BGP speaker connectivity
- iBGP speakers must be fully meshed:
  - They originate connected networks
  - They pass on prefixes learned from outside the ASN
  - They do not pass on prefixes learned from other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

Internal BGP Peering (iBGP)



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

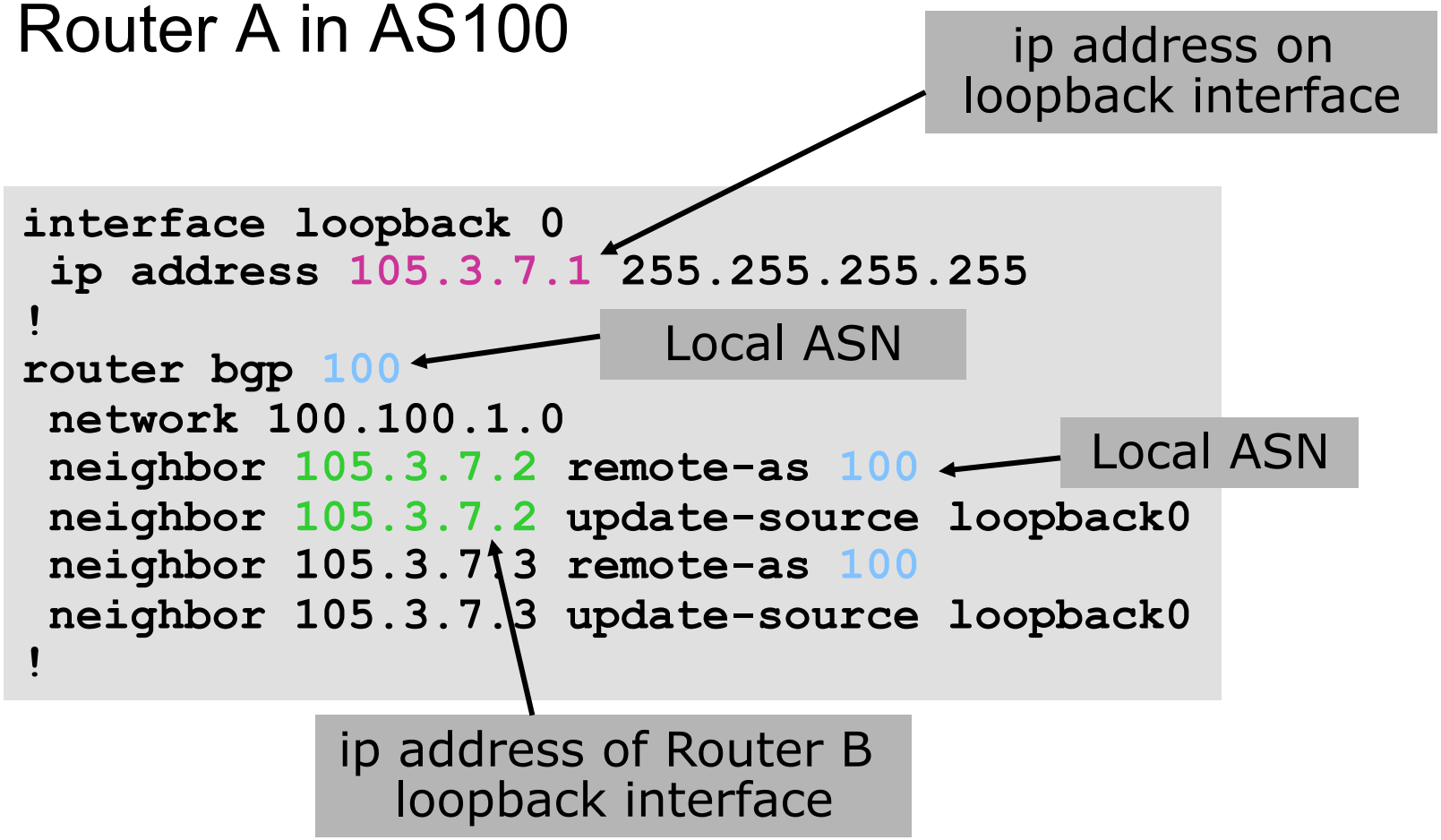
Peering between Loopback Interfaces



- 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

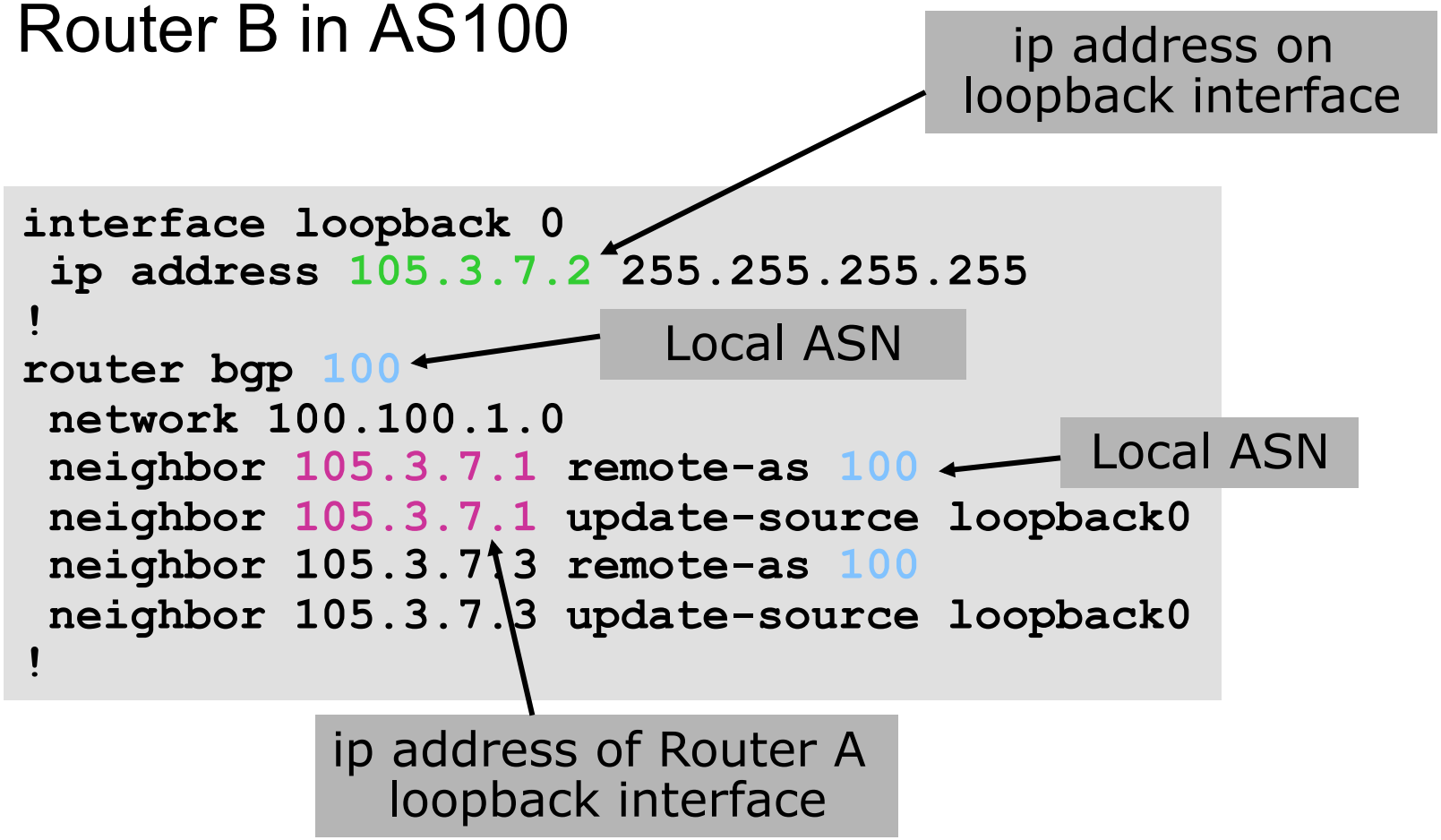
Configuring Internal BGP

- Router A in AS100



Configuring Internal BGP

- Router B in AS100



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

```
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 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
- iBGP versus eBGP
- Stable iBGP – peer with loopbacks
- Announcing prefixes & aggregates

Questions?

UNIVERSITY OF OREGON

NSRC
Network Startup Resource Center