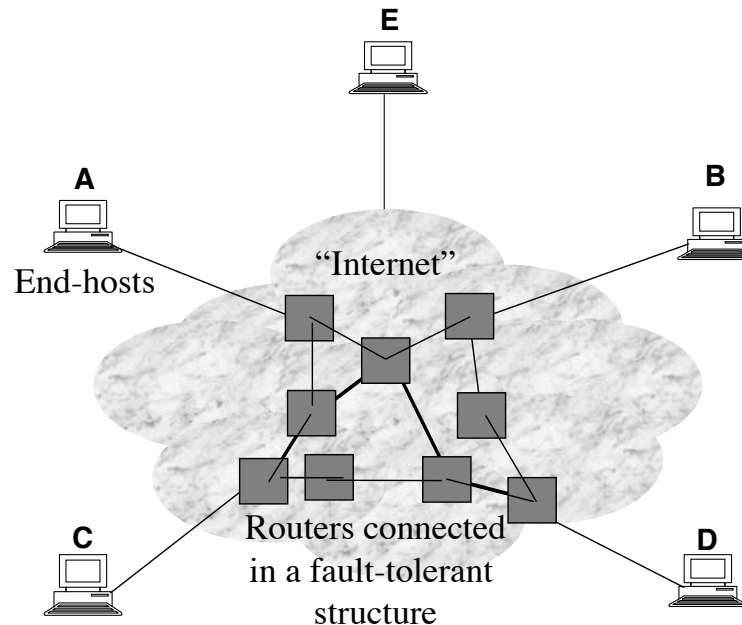# COSC 301
Network Management and Security

Lecture 23: Exterior Routing and BGP

# Today's Focus

How routing between different administrative domains works in Internet?

Idealized View of the Internet

- E
- A
  - End-hosts
- "Internet"
- B
- C
  - Routers connected in a fault-tolerant structure
- D

- This abstraction is quite misleading
- The TRUE story
  - The Internet service is provided by a large number of commercial enterprises, generally in competition with each other.
  - Global connectivity is achieved through cooperation between profitable commercial enterprises.

The Real Internet Structure

AS 2

AS 5

AS 4

AS 3

AS 1
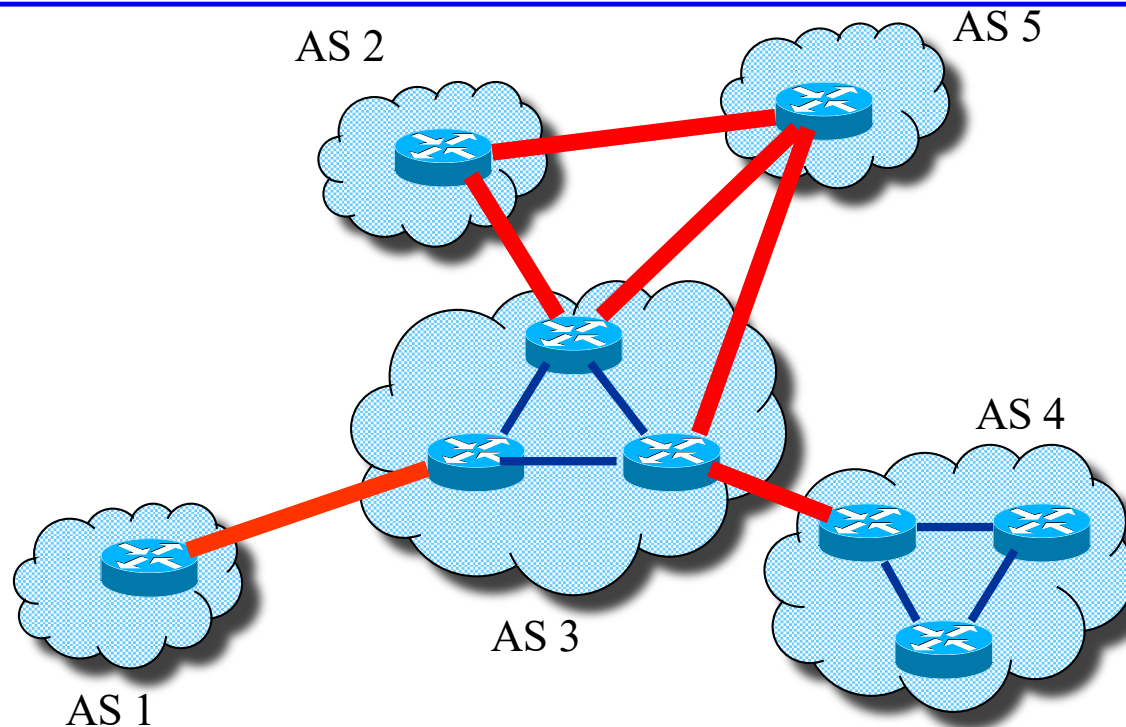
- Internet is composed of over 50000 Autonomous Systems (ASs)
  - Superlinear growth
- Each public AS is identified a globally unique number
  - IANA allocates AS Numbers to Regional Internet Registries (RIRs)
  - The RIRs further allocate or assign AS Numbers to network operators
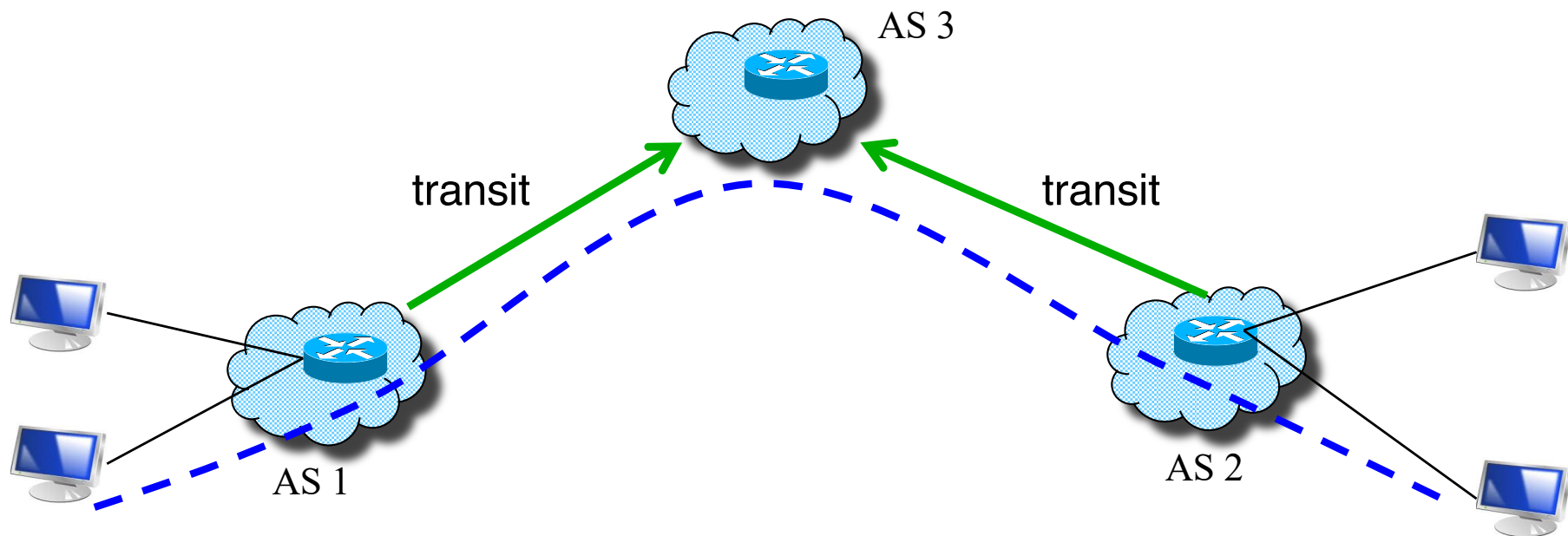 - 16 bit values

Inter-AS Relationships (1)

- **Peering**: two or more ASs interconnect directly with each other to exchange traffic
  - often done without charging for the interconnection or the traffic.

peering

AS 1

AS 2
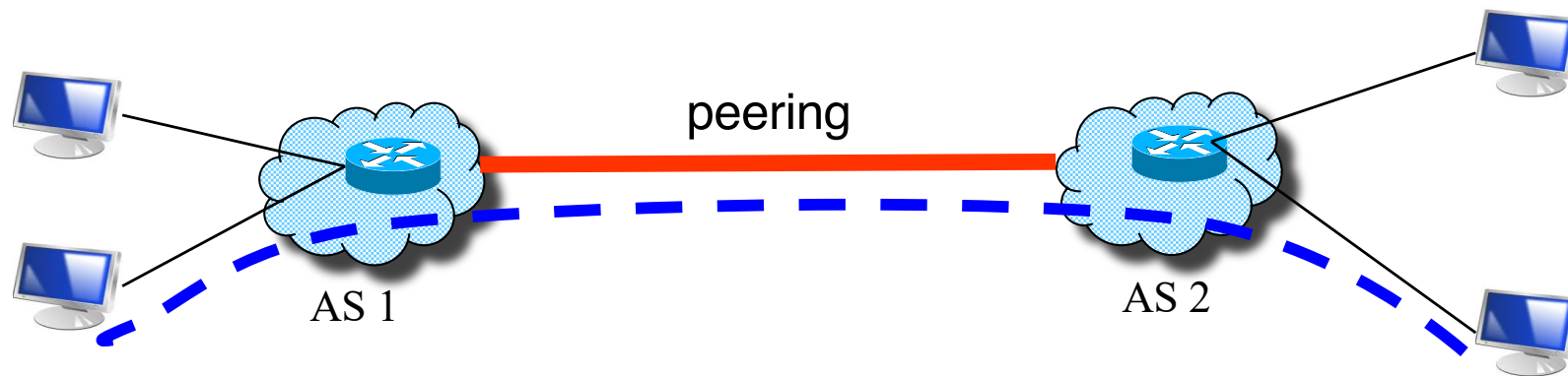
Inter-AS Relationships (2)

- **Transit**: one AS agrees to carry the traffic that flows between ASs

  – The transit provider receives a "transit fee" for the service

Routing between ASs

- Distance vector or link state?

  - No universal routing metric

- Problems with distance vector

  - Bellman-Ford algorithm may not converge

- Problems with link state

  - Metric used by routers not the same – loops

  - LS database too large – entire Internet
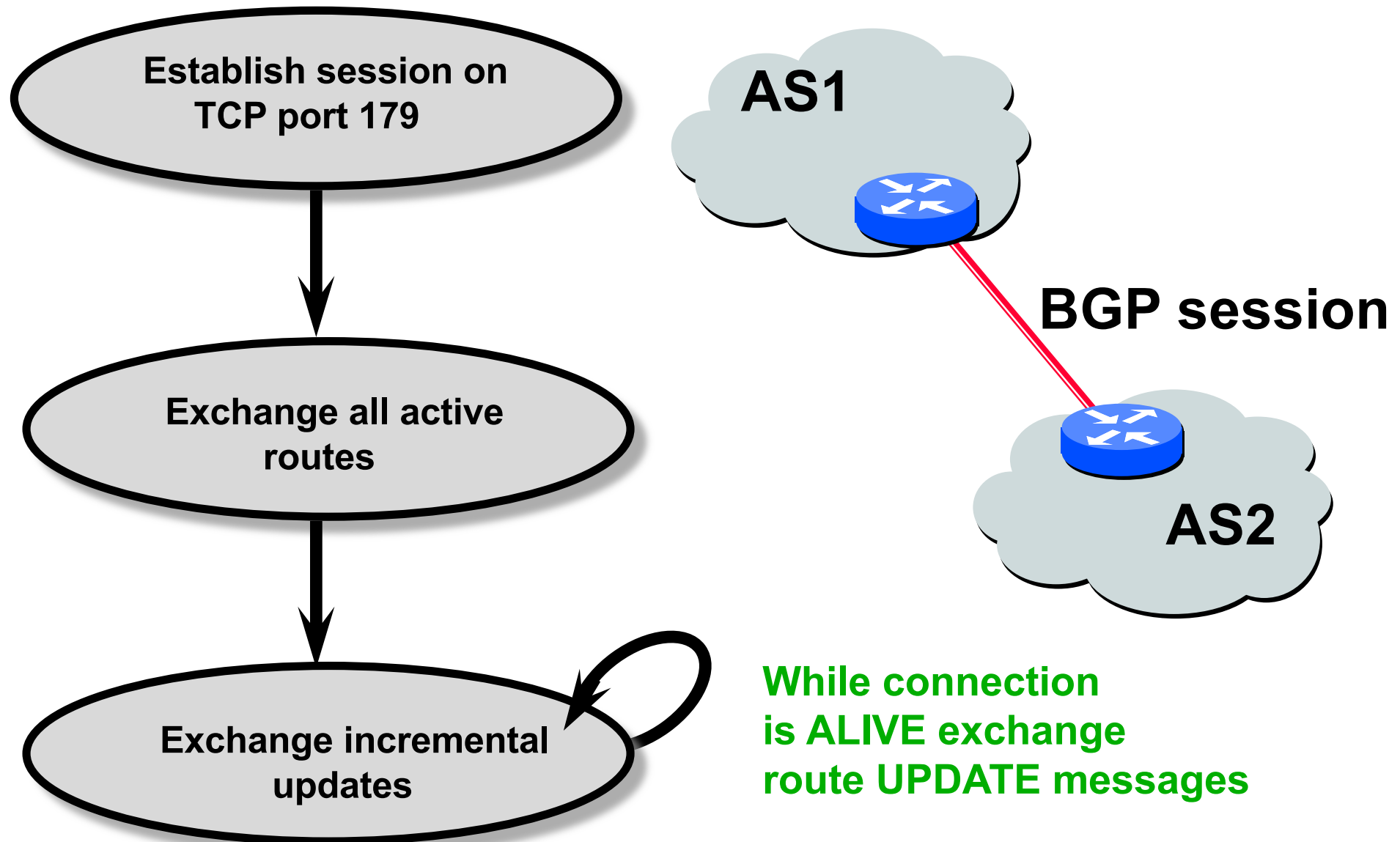
  - May expose policies to other ASs

BGP

- <u>B</u>order <u>G</u>ateway <u>P</u>rotocol
  - Is a **<u>Policy-Based</u>** routing protocol
  - Is the **<u>de facto inter-domain routing protocol</u>** of today's global Internet (current version BGP4)

- BGP is classified as a path vector routing protocol
  - makes routing decisions based on paths, network policies, or rule-sets configured by a network administrator
  - 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

BGP Operations

Establish session on TCP port 179

↓

Exchange all active routes

↓

Exchange incremental updates

AS1

BGP session

AS2

While connection
is ALIVE exchange
route UPDATE messages

Four Types of BGP Messages

- ## Open
  - Establishes a peering session on port 179.

- ## Keep Alive
  - sends 19-byte keep-alive messages every 60 seconds to maintain the connection

- ## Notification
  - Used for error notification
  - Shuts down a peering session

- ## Update
  - Announces new routes or withdraws previously announced routes.

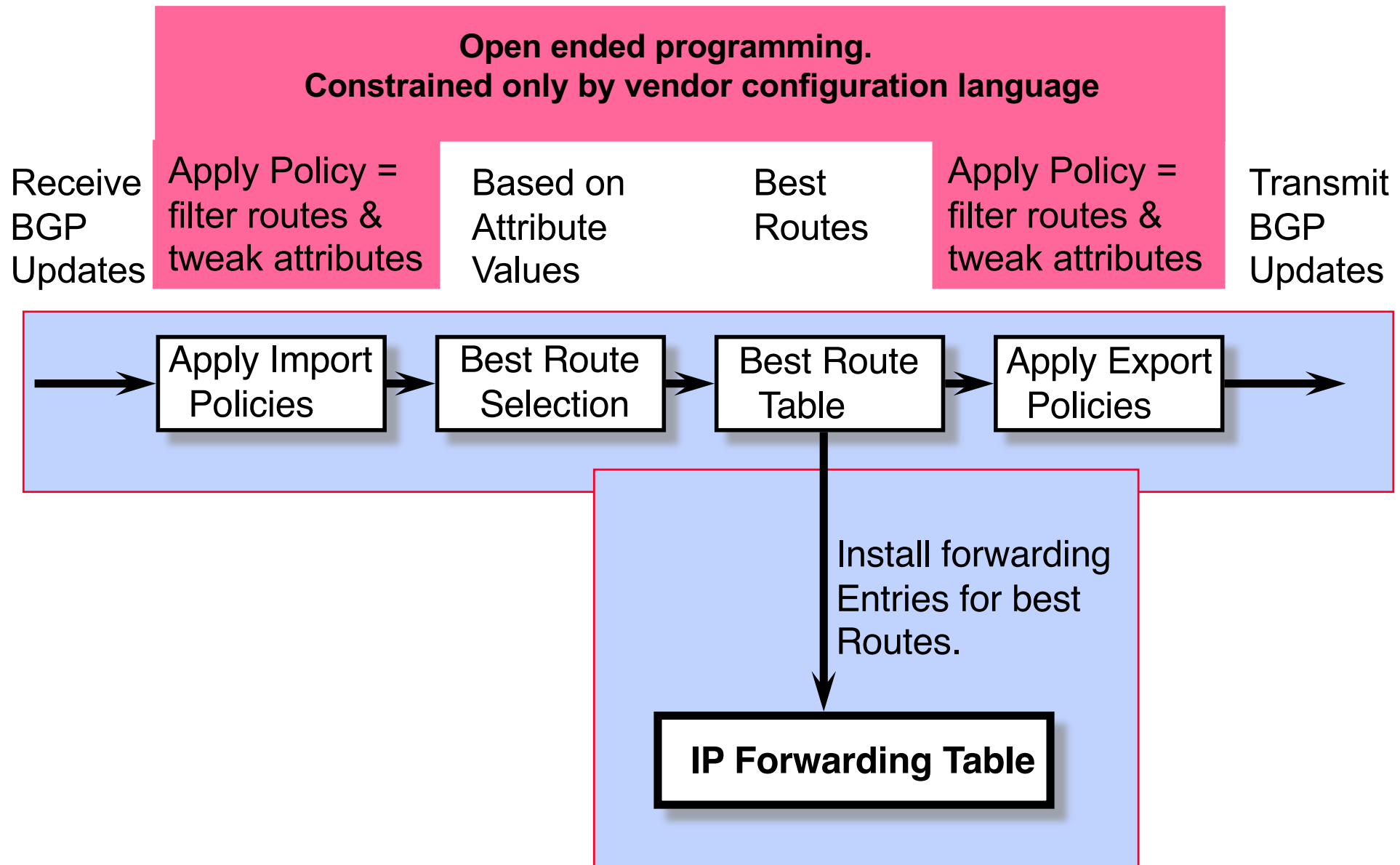# Two Types of BGP Neighbour Relationships

- eBGP: External Neighbor in a different AS

- iBGP: Internal Neighbor in the same AS

AS1

AS2
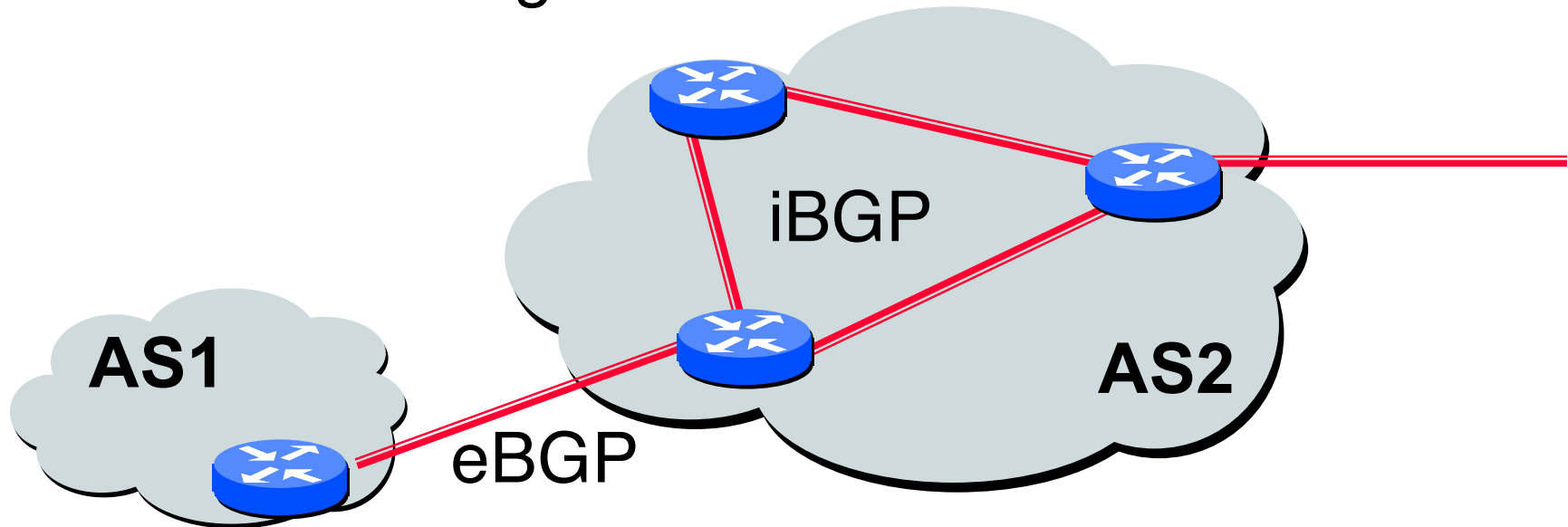
iBGP

eBGP

- The main difference is the way to propagate routes
  - New routes from an eBGP is typical redistributed all other iBGP and eBGP peers.
  - New routes learned by an iBGP are advertised to only iBGP peers.

BGP Route Processing

**Open ended programming.
Constrained only by vendor configuration language**

Receive BGP Updates

Apply Policy = filter routes & tweak attributes

Based on Attribute Values

Best Routes

Apply Policy = filter routes & tweak attributes

Transmit BGP Updates

Apply Import Policies → Best Route Selection → Best Route Table → Apply Export Policies

Install forwarding Entries for best Routes.

IP Forwarding Table

Import/Export Routes

- ## Import Routes
  - When a router hears many possible routes to a destination network, it needs to decide which route to install in its forwarding tables.
  - Order of route preference: customer > peer > provider

- ## Export Routes
  - Each AS needs to make decisions on which routes to export to its neighboring ISPs using BGP.
    - No ISP wants to act as transit for packets that it isn't somehow making money on.
  - Transit customer routes: Highest priority
  - Transit provider routes: Most likely not (no money earned)
  - Peer routes: only selected routes to other peering ISPs.

Best Route Selection

- ## Use BGP attributes

192.0.2.0/24 pick me!

192.0.2.0/24 pick me!

192.0.2.0/24 pick me!

192.0.2.0/24 pick me!
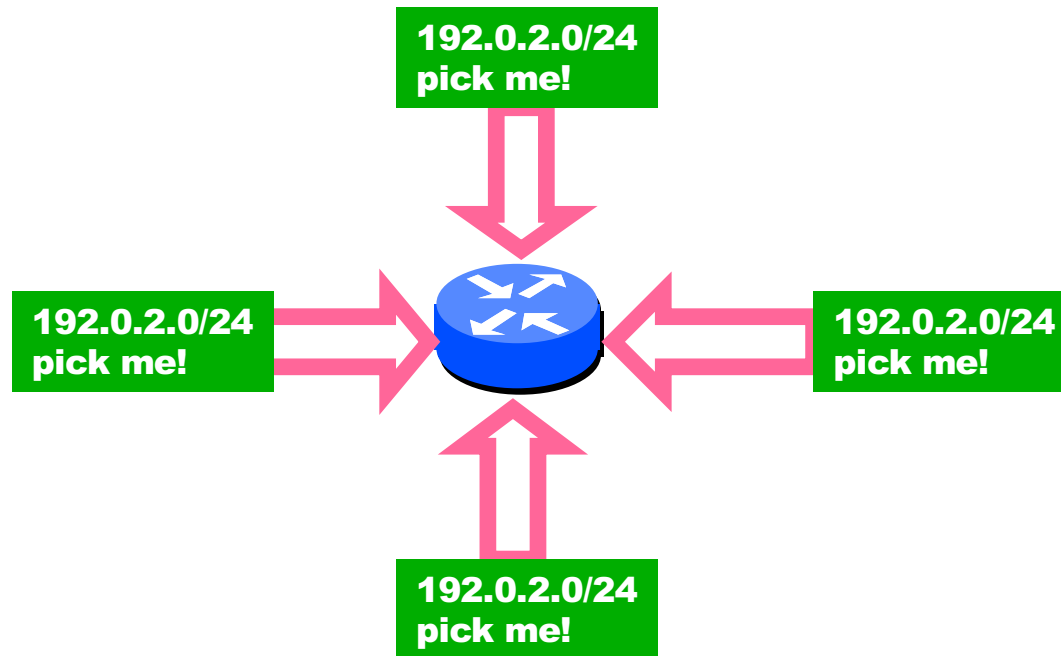
Given multiple routes to the same prefix, a BGP speaker must pick at most <u>one</u> best route

BGP Attributes

```
Value       Code                                Reference
-----       ----------------------------------  ----------
   1        ORIGIN                              [RFC1771]
   2        AS_PATH                             [RFC1771]
   3        NEXT_HOP                            [RFC1771]
   4        MULTI_EXIT_DISC                     [RFC1771]
   5        LOCAL_PREF                          [RFC1771]
   6        ATOMIC_AGGREGATE                    [RFC1771]
   7        AGGREGATOR                          [RFC1771]
   8        COMMUNITY                           [RFC1997]
   9        ORIGINATOR_ID                       [RFC2796]
  10        CLUSTER_LIST                        [RFC2796]
  11        DPA                                    [Chen]
  12        ADVERTISER                          [RFC1863]
  13        RCID_PATH / CLUSTER_ID              [RFC1863]
  14        MP_REACH_NLRI                       [RFC2283]
  15        MP_UNREACH_NLRI                     [RFC2283]
  16        EXTENDED COMMUNITIES                  [Rosen]
...
 255        reserved for development
```
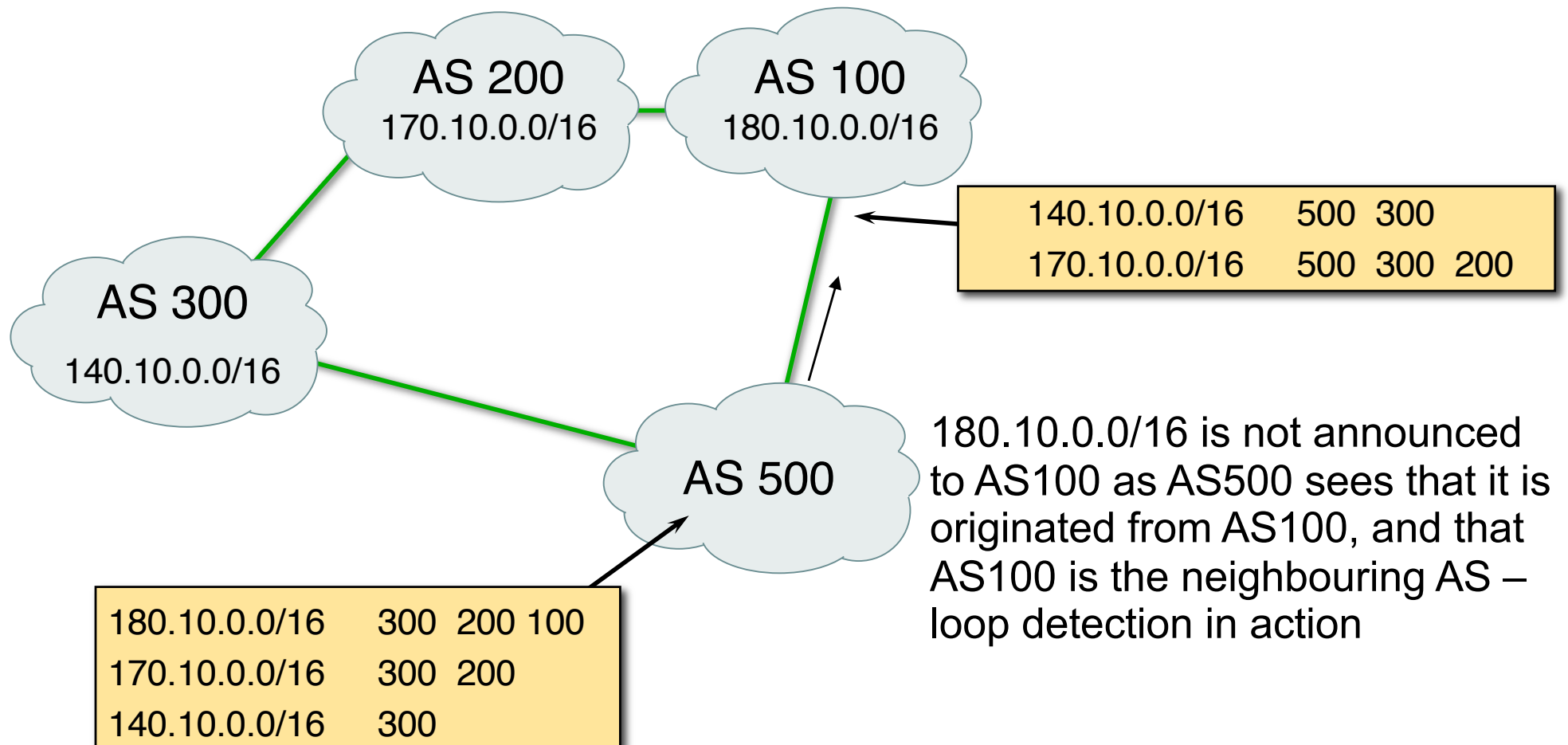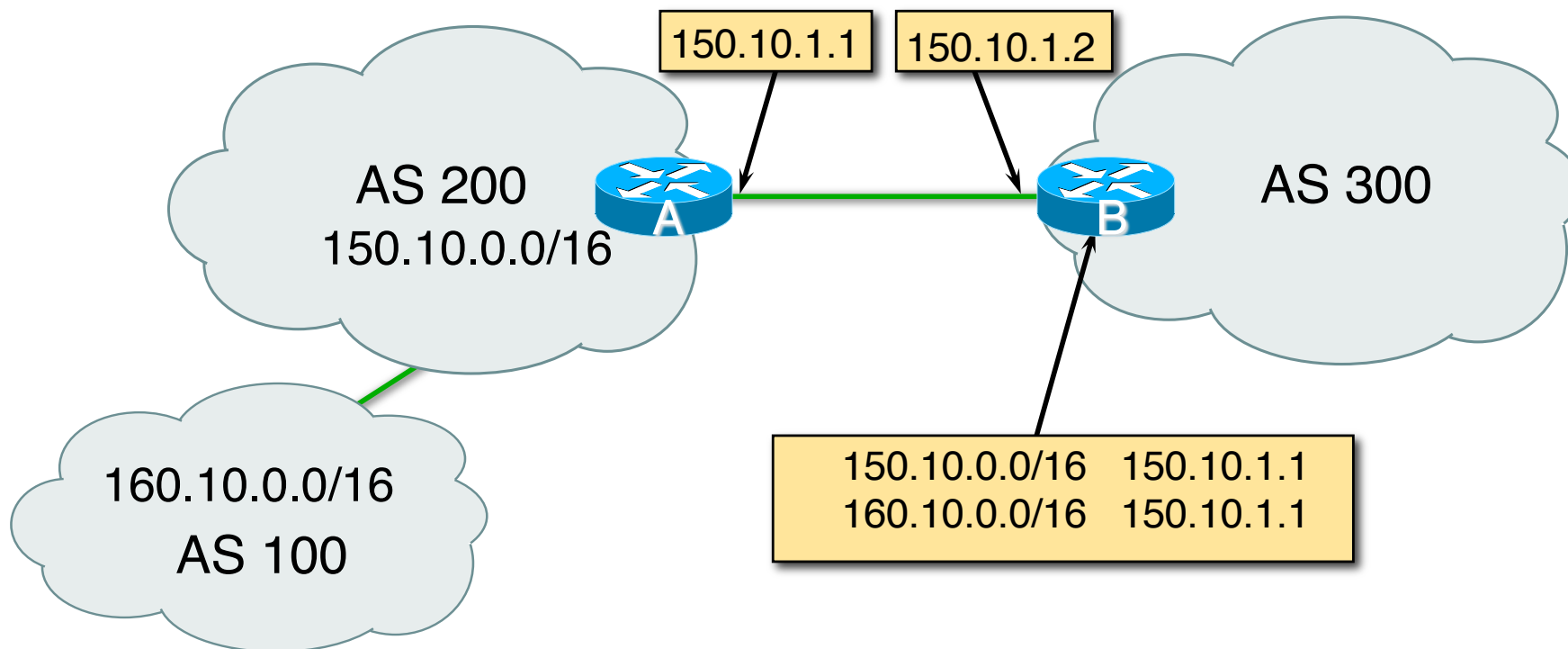
Most important attributes

From IANA: http://www.iana.org/assignments/bgp-parameters

AS-Path and Loop Detection

- AS_PATH: sequence of AS identifiers that the route advertisement has traversed.

AS 200
170.10.0.0/16

AS 100
180.10.0.0/16

AS 300
140.10.0.0/16

AS 500

| 140.10.0.0/16 | 500 300 |
| 170.10.0.0/16 | 500 300 200 |

180.10.0.0/16	300 200 100
170.10.0.0/16	300 200
140.10.0.0/16	300

180.10.0.0/16 is not announced to AS100 as AS500 sees that it is originated from AS100, and that AS100 is the neighbouring AS – loop detection in action

Next_Hop

- IP address of the next-hop router along the path to the destination.
  - On eBGP sessions, the next hop is set to the IP address of the border router of the neighbouring AS.
  - On iBGP sessions, the next hop is not modified.

150.10.1.1   150.10.1.2

AS 200
150.10.0.0/16

A   B

AS 300

160.10.0.0/16
AS 100

150.10.0.0/16   150.10.1.1
160.10.0.0/16   150.10.1.1

Local Preference

- The first criteria used to select routes
- Not attached on routes learned via eBGP sessions, but assigned by the import policy of these sessions.

AS 4

AS 3

AS 2

AS 1

local pref = 80

local pref = 90

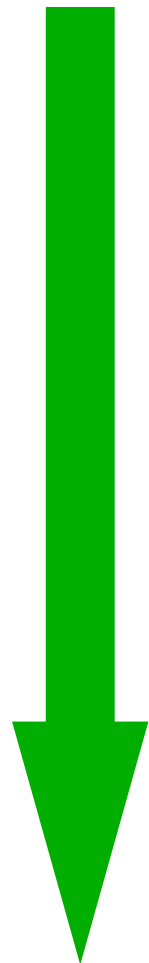local pref = 100

13.13.0.0/16
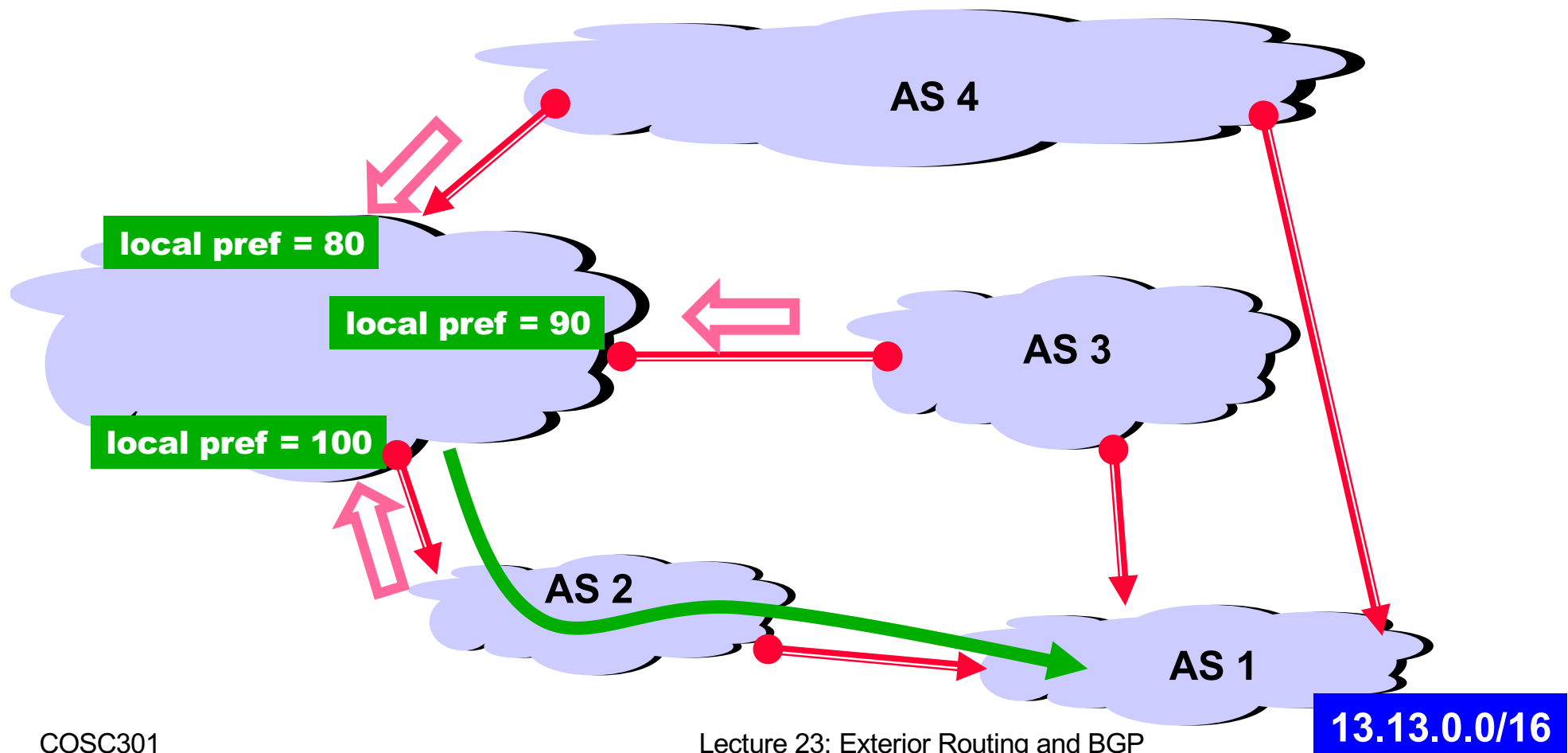
Multi-Exit Discriminator (MED)

- Neighboring AS sets the MED values to indicate which router it prefers to receive traffic for destination.
- Used for comparing two or more routes from the same neighboring AS.

AS 200

192.68.1.0/24   2000

192.68.1.0/24   1000

C

A   B

AS 201   192.68.1.0/24
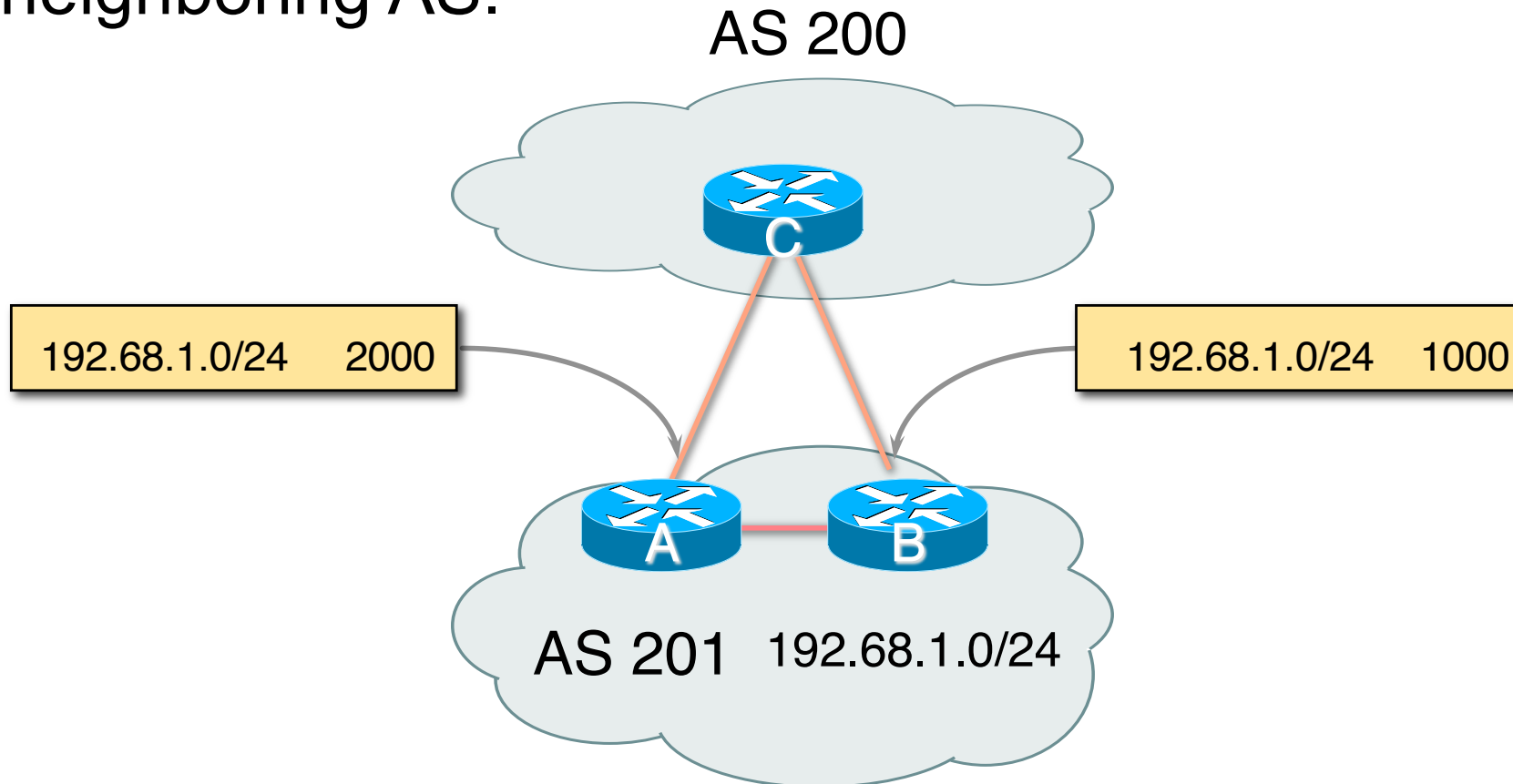
Best Route Selection Criteria

Highest Local Preference	**Enforce relationships**
Shortest ASPATH **Lowest MED** **i-BGP < e-BGP** **Lowest IGP cost to BGP egress**	**traffic engineering**
Lowest router ID	**Throw up hands and break ties**

BGP Problem and Migration

- Internal BGP scalability
  - Full mesh connectivity
  - Route reflectors and confederations.

- Instability
  - The routing tables are adjusted continually to reflect actual changes
  - Route flap damping

- Routing table growth
  - Route summarization

- Load-balancing
  - Locater/Identifier Separation Protocol

Summary

- Inter-AS relationships
  - Peering
  - Transit

- BGP
  - eBGP and iBGP
  - Route import and export
  - Best route selection

Resources

- Hari Balakrishnan, and Nick Feamster, Interdomain Internet Routing, http://nms.csail.mit.edu/6.829-f05/lectures/L4-routing.pdf

- Mike Pennington, BGP Deployment & Scalability, www.pennington.net/tutorial/bgp_001/BGP_Overview.ppt