# Precept 6

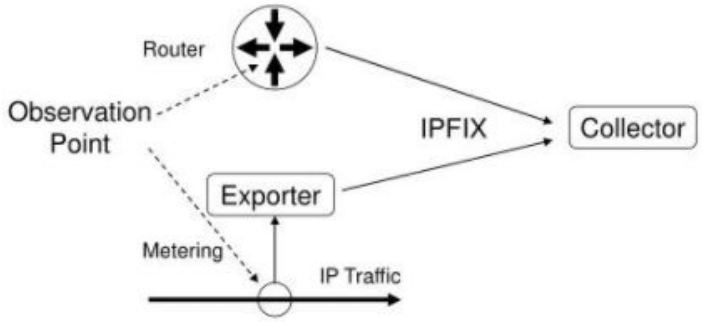
**Main topics:**

- Assignment 3: Passive Network Measurement
- Due Friday, March 24th 11:59pm ET
- Review / Deep Dive on BGP

# How was your assignment 3 data collected?

**1) Traffic Measurement with IPFIX (IP Flow Information eXport) protocol**

- Goal: develop common IP traffic flow reporting protocol to be available on most routers
- NetFlow - a proprietary form of IPFIX.




**2) Interdomain Routing Measurement with BGP Routing Tables**

- To understand the state of Internet routing, many routers "dump" BGP routing tables periodically into a static file.
- Contain information about each IP prefix, all BGP routes that the router learns for each prefix, and the "best" BGP route that the router selects.
- Analyzing the BGP routing tables can provide information about where traffic to different IP prefixes is destined.

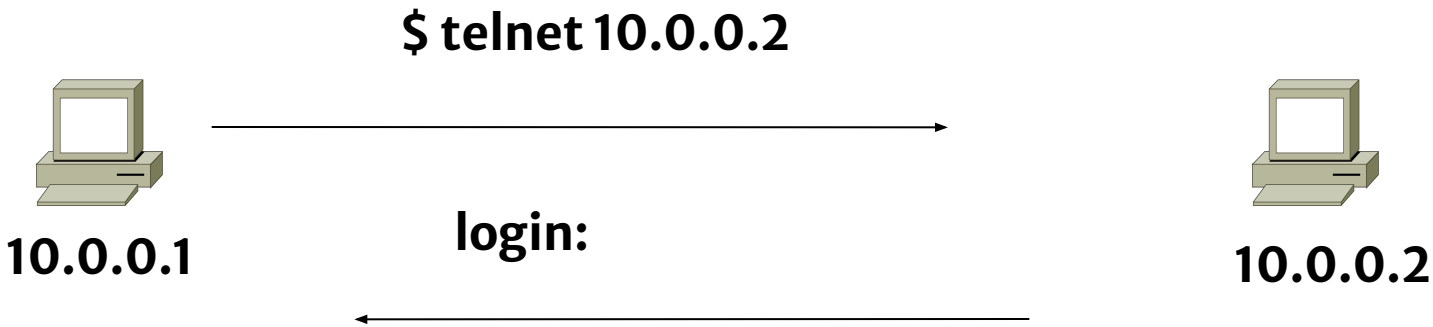
# Network Flows

- Packets or frames that have a common properties.
  - Examples: IP source/destination, Port source/destination, L4 protocol, VLANid
- Creation and expiration policy
  - what conditions start and stop a flow.
- Network Flows contain:
  - Counters – packets, bytes, time.
  - Routing information – AS, network mask, interfaces.
  - Peers – flow source and destination

# Network Flows

- Unidirectional or bidirectional flows
  - Bidirectional flows can contain other information such as round trip time, TCP behavior.
- Application flows look past the headers to classify packets by their contents.
- Aggregated flows – flows of flows.

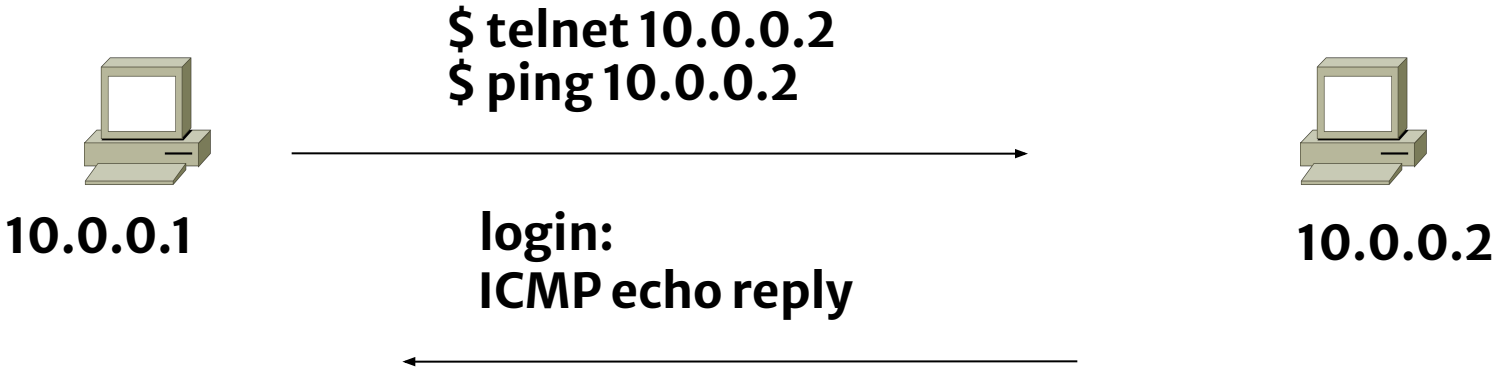
# Unidirectional Flow with Source/Destination IP Key




**Active Flows**

| Flow | Source IP | Destination IP |
|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 |
| 2 | 10.0.0.2 | 10.0.0.1 |

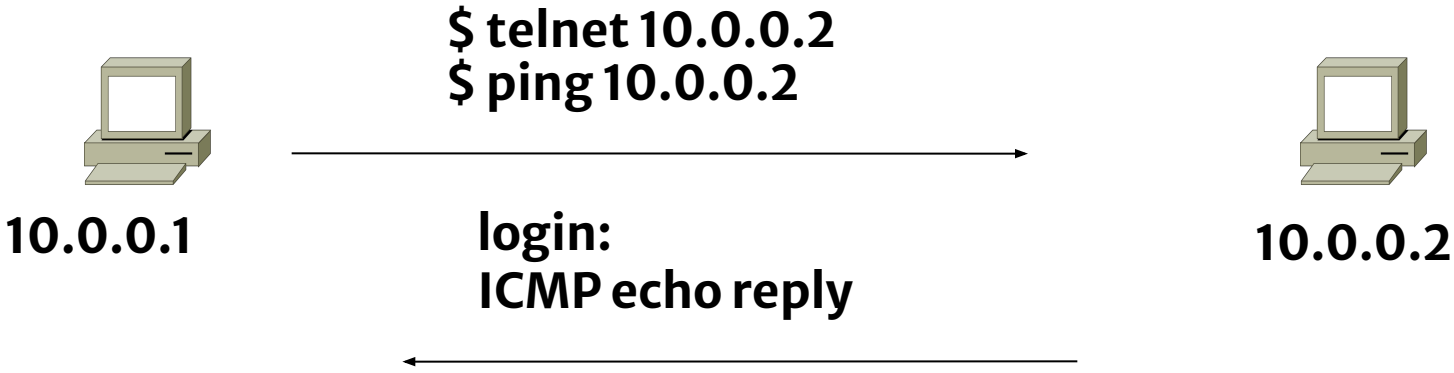
# Unidirectional Flow with Source/Destination IP Key




**Active Flows**

| Flow | Source IP | Destination IP |
|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 |
| 2 | 10.0.0.2 | 10.0.0.1 |

# Unidirectional Flow (IP, Port, Protocol Key)




**Active Flows**

| Flow | Source IP | Destination IP | protocol | srcPort | dstPort |
|---|---|---|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 | TCP | 32000 | 23 |
| 2 | 10.0.0.2 | 10.0.0.1 | TCP | 23 | 32000 |
| 3 | 10.0.0.1 | 10.0.0.2 | ICMP | 0 | 0 |
| 4 | 10.0.0.2 | 10.0.0.1 | ICMP | 0 | 0 |

# Aggregated Flow

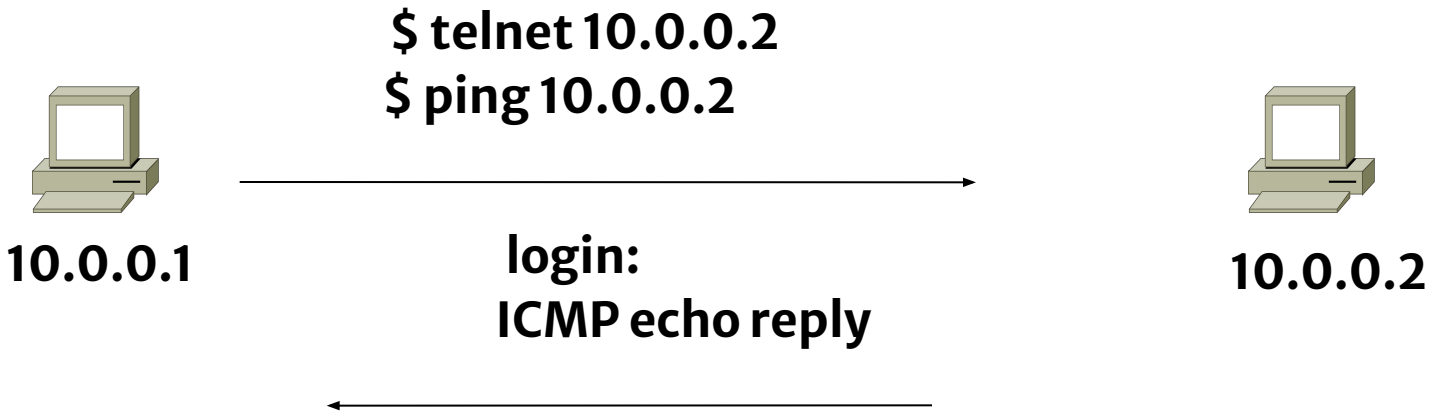
**Main Active flow table**

| Flow | Source IP | Destination IP | protocol | srcPort | dstPort |
|---|---|---|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 | TCP | 32000 | 23 |
| 2 | 10.0.0.2 | 10.0.0.1 | TCP | 23 | 32000 |
| 3 | 10.0.0.1 | 10.0.0.2 | ICMP | 0 | 0 |
| 4 | 10.0.0.2 | 10.0.0.1 | ICMP | 0 | 0 |

**Source/Destination IP Aggregate**

| Flow | Source IP | Destination IP |
|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 |
| 2 | 10.0.0.2 | 10.0.0.1 |

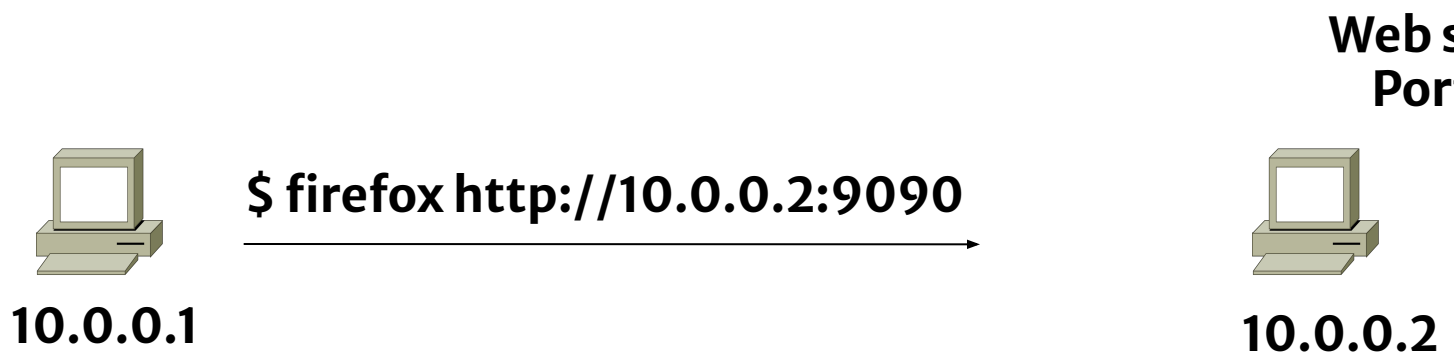
# Bidirectional Flow (IP, Port, Protocol Key)




**Active Flows**

| Flow | Source IP | Destination IP | protocol | srcPort | dstPort |
|---|---|---|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 | TCP | 32000 | 23 |
| 2 | 10.0.0.1 | 10.0.0.2 | ICMP | 0 | 0 |

# Application Flow




**Active Flows**

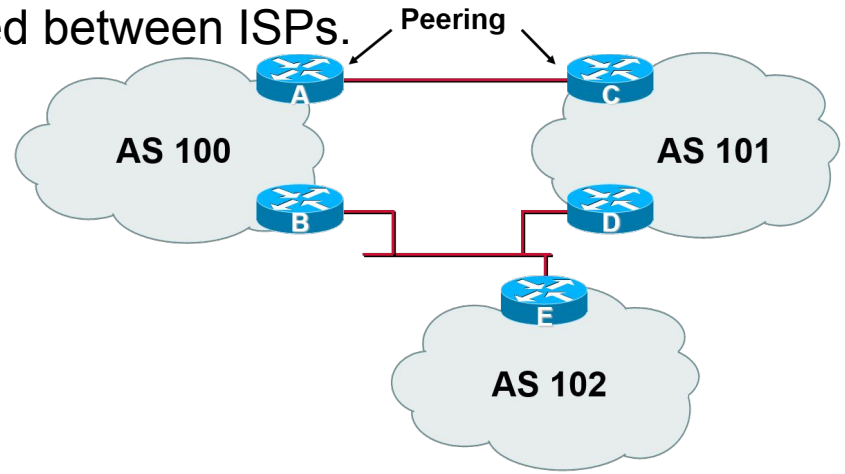
| Flow | Source IP | Destination IP | Application |
|---|---|---|---|
| 1 | 10.0.0.1 | 10.0.0.2 | HTTP |

# Analyze BGP Routing Tables

- Which networks (i.e., autonomous systems) are responsible for sending or receiving traffic to the network.
  - Why do we need to know this?
- The RouteViews project allows real-time information about the global routing system from the perspectives of several different ASs.
- RouteViews servers act as software BGP routers, obtaining their BGP routing information via BGP sessions.
- The main difference between the RouteViews servers and other BGP-speaking routers is that the RouteViews servers do not forward any real Internet traffic.

# BGP Basics

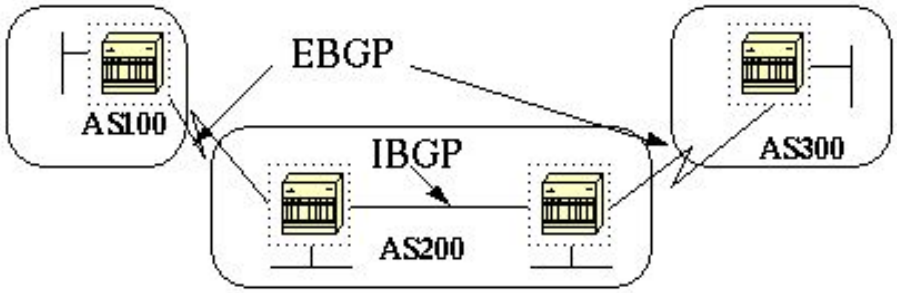
- BGP - internet exterior gateway protocol used between ISPs.
- The characteristics:
  - Run over TCP
  - Path vector protocol
  - Incremental updates




- BGP speaker - A router that advertises BGP messages and establishes peer relationships with other BGP speakers to exchange routing information.
- BGP can be configured to run on a router in the following two modes:
  - iBGP (internal BGP) - When a BGP speaker peers with another BGP speaker that resides in the same AS
  - eBGP (external BGP) - When a BGP speaker peers with a BGP speaker that resides in a different AS
- Peers (neighbors) - Any two routers that have formed a TCP connection in order to exchange BGP routing information

# EBGP and IBGP

- If an AS has multiple BGP speakers, it could be used as a transit service for other ASs.
- It is necessary to ensure reachability for networks within an AS before sending the information to other external ASs.
  - Internal BGP peering between routers inside an AS
  - Redistributing BGP information to Internal Gateway Protocols running in the AS.
- EBGP (Exterior BGP)
  - When BGP is running between routers belonging to two different ASs
  - Should be directly connected
- IBGP (Interior BGP)
  - BGP running between routers in the same AS.
  - Not required to be directly connected
  - IBGP neighbors should be fully meshed




**AS200 is a transit autonomous system for AS100 and AS300**

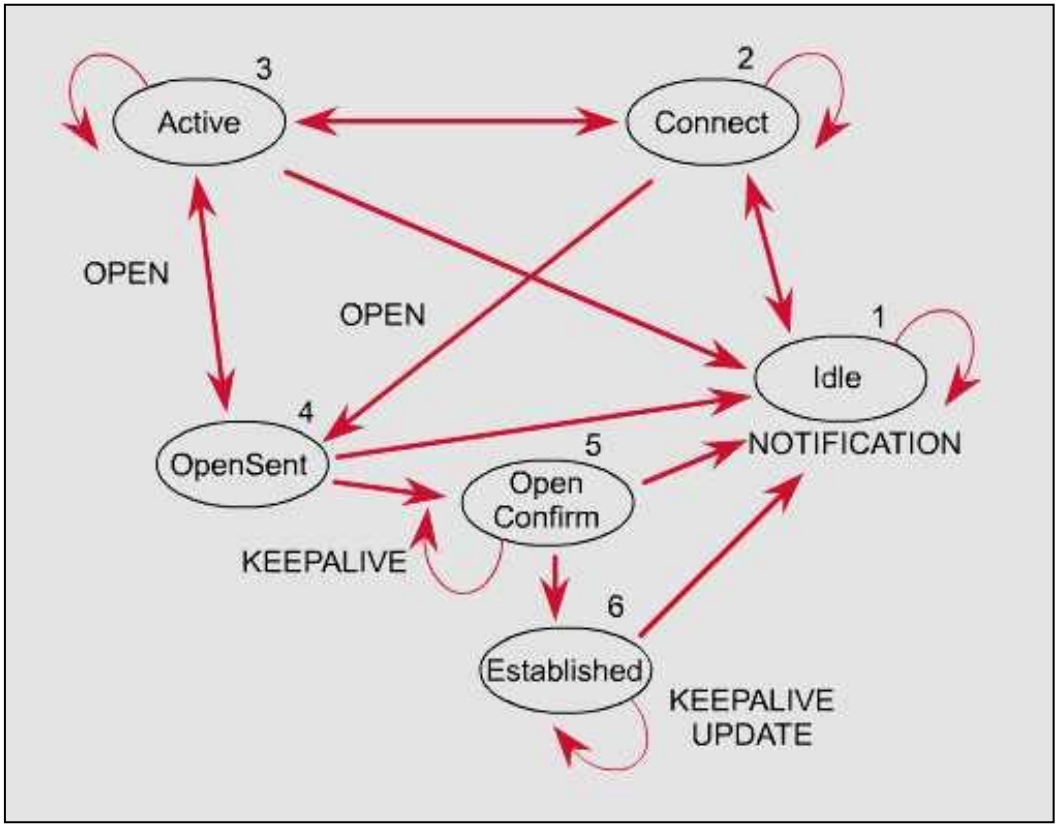
# General Operation

- Learns multiple paths via internal and external BGP speakers
- Picks the best path and installs in the IP forwarding table
- Policies applied by influencing the best path selection

# General Operation

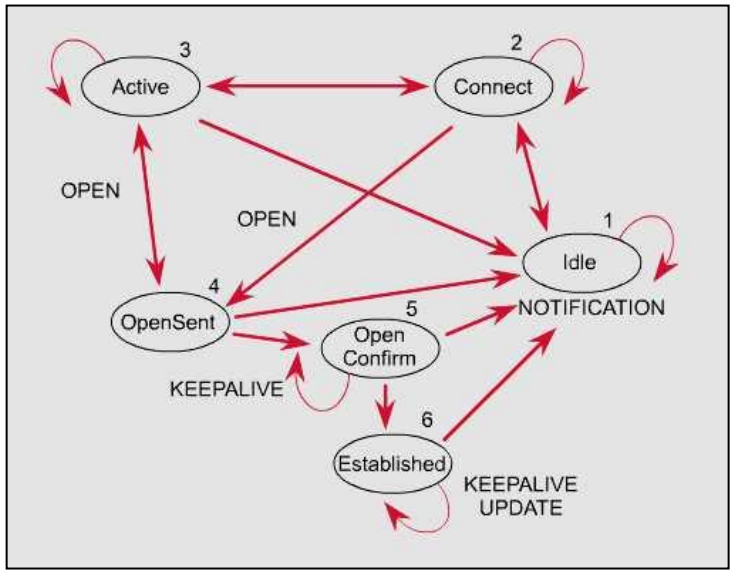
- Information exchange between peers
  - BGP peers will initially exchange their full BGP routing tables.
  - From then on incremental updates are sent as the routing table changes.
    - Update message – path attribute information
  - BGP keeps a version number of the BGP table and it should be the same for all of its BGP peers.
  - The version number will change whenever BGP updates the table due to some routing information changes.
  - Keepalive packets ensure that the connection is alive between the BGP peers

# BGP FSM




- The BGP neighbor negotiation process proceeds through various states, or stages, which can be described in terms of a finite-state machine (FSM).

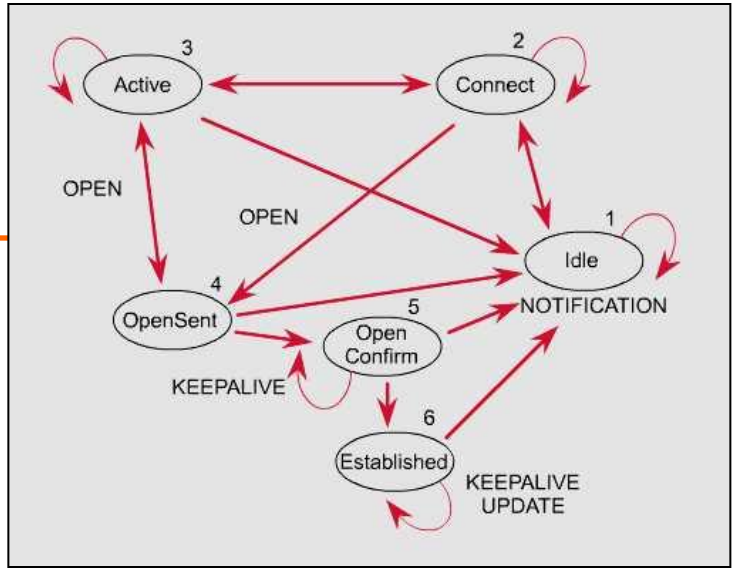
# BGP FSM




BGP FSM includes six states:

1. **Idle**
2. **Connect**
3. **Active**
4. **OpenSent**
5. **Open Confirm**
6. **Established**

# BGP FSM

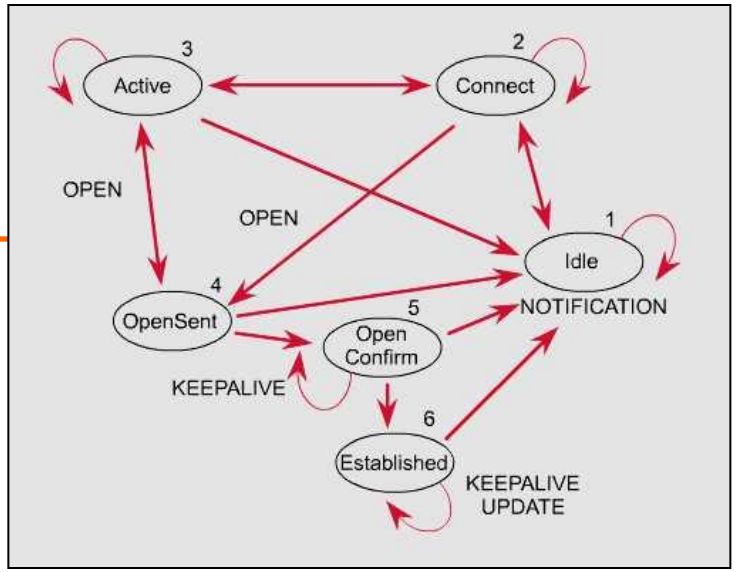



## Idle State

- BGP always begins in the Idle state, in which it refuses all incoming connections.
- When Start event occurs, the BGP process:
  - Initializes all BGP resources
  - Starts the ConnectRetry timer
  - Initializes a TCP connection to the neighbor
  - Listens for a TCP initialization from the neighbor
  - Changes its state to **Connect**

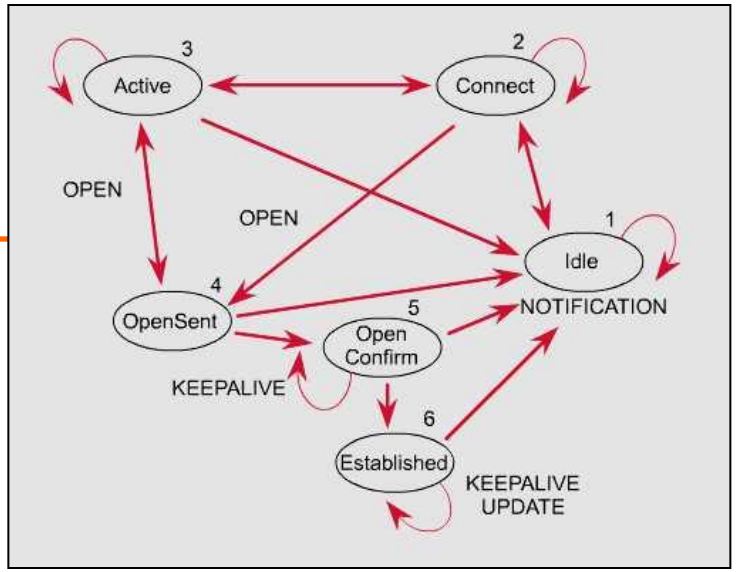
# BGP FSM

## Connect State




- In this state, the BGP process is waiting for the TCP connection to be completed.
- If the connection is **successful**, the BGP process:
  - Clears the ConnectRetry timer
  - Completes initialization
  - Sends an **Open message** to the neighbor to identify itself and to specify its BGP operational parameters
  - Transitions to the **OpenSent** state
- If the connection is **unsuccessful**, the BGP process:
  - Continues to listen for a connection to be initiated by the neighbor
  - Resets the ConnectRetry timer
  - Transitions to the **Active** state

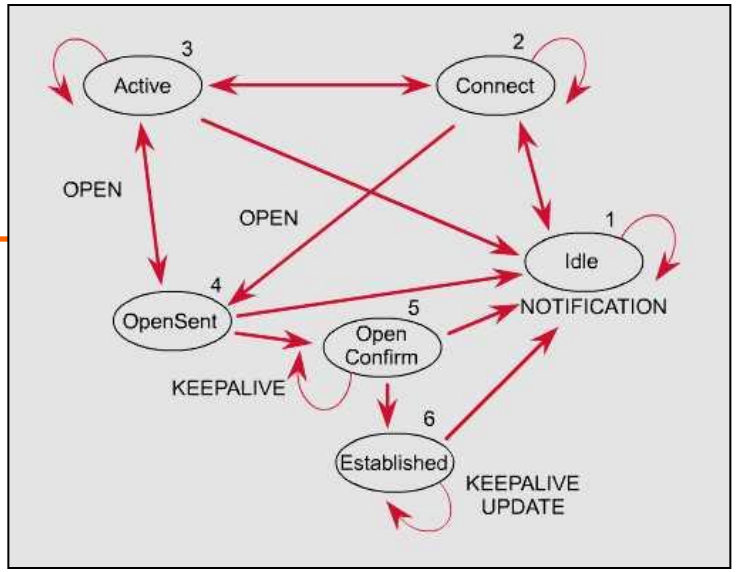
# BGP FSM




## Active State

- In this state, the BGP process is trying to initiate a TCP connection with the neighbor.
- If the TCP connection is **successful**:
  - Clears the ConnectRetry timer
  - Completes initialization
  - Sends an **Open message** to the neighbor
  - Transitions to the **OpenSent** state
- If the ConnectRetry timer expires while BGP is in the Active State, the BGP process:
  - Transitions back to the **Connect** state
  - Resets the ConnectRetry timer

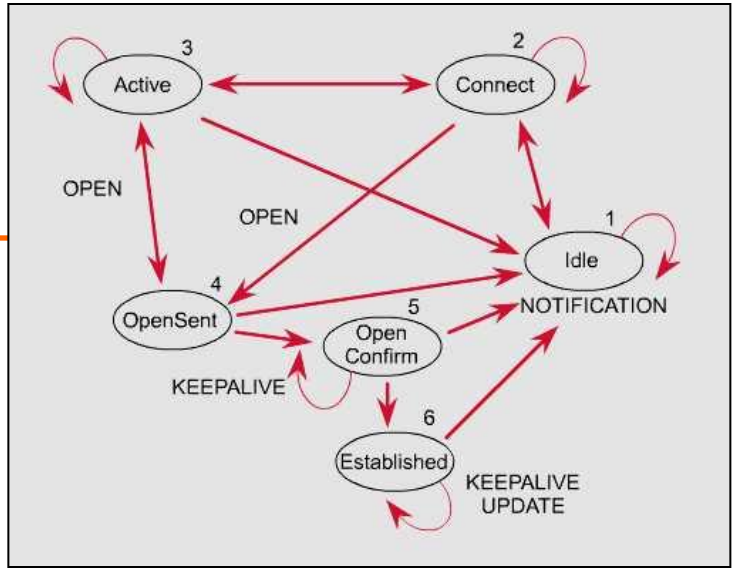
# BGP FSM




## OpenSent State

- In this state an **Open message** has been sent and BGP is waiting to hear an Open message from its neighbor.
- When an **Open message** is received, all its fields are checked.
- **If errors** exist, a **Notification message** is sent and the state transitions to **Idle**.
- **If no errors** exist, a **Keepalive message** is sent and the Keepalive timer is set, the peer is determined to be internal or external, and state is changed to **OpenConfirm**.

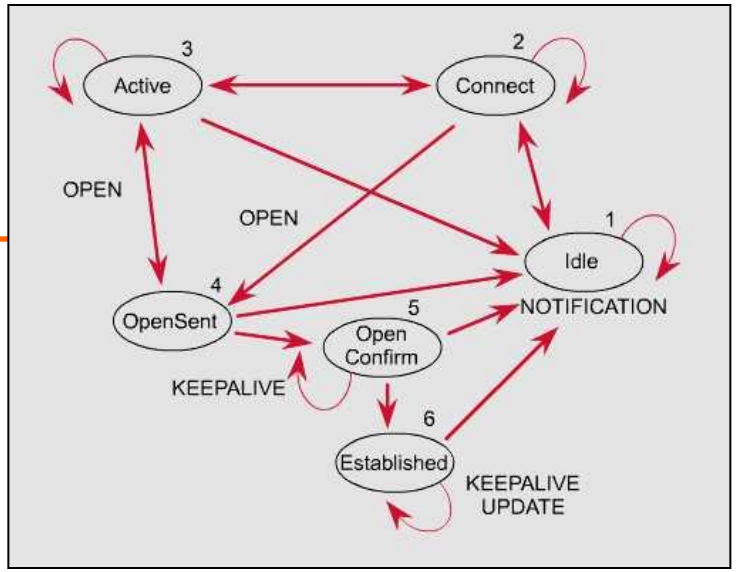
# BGP FSM




## OpenConfirm State

- In this state, the BGP process waits for a **Keepalive** or **Notification message**.
- If a **Keepalive message** is received, the state transitions to **Established**.
- If a **Notification message** is received, or a TCP disconnect is received, the state transitions to **Idle**.

# BGP FSM




## Established State

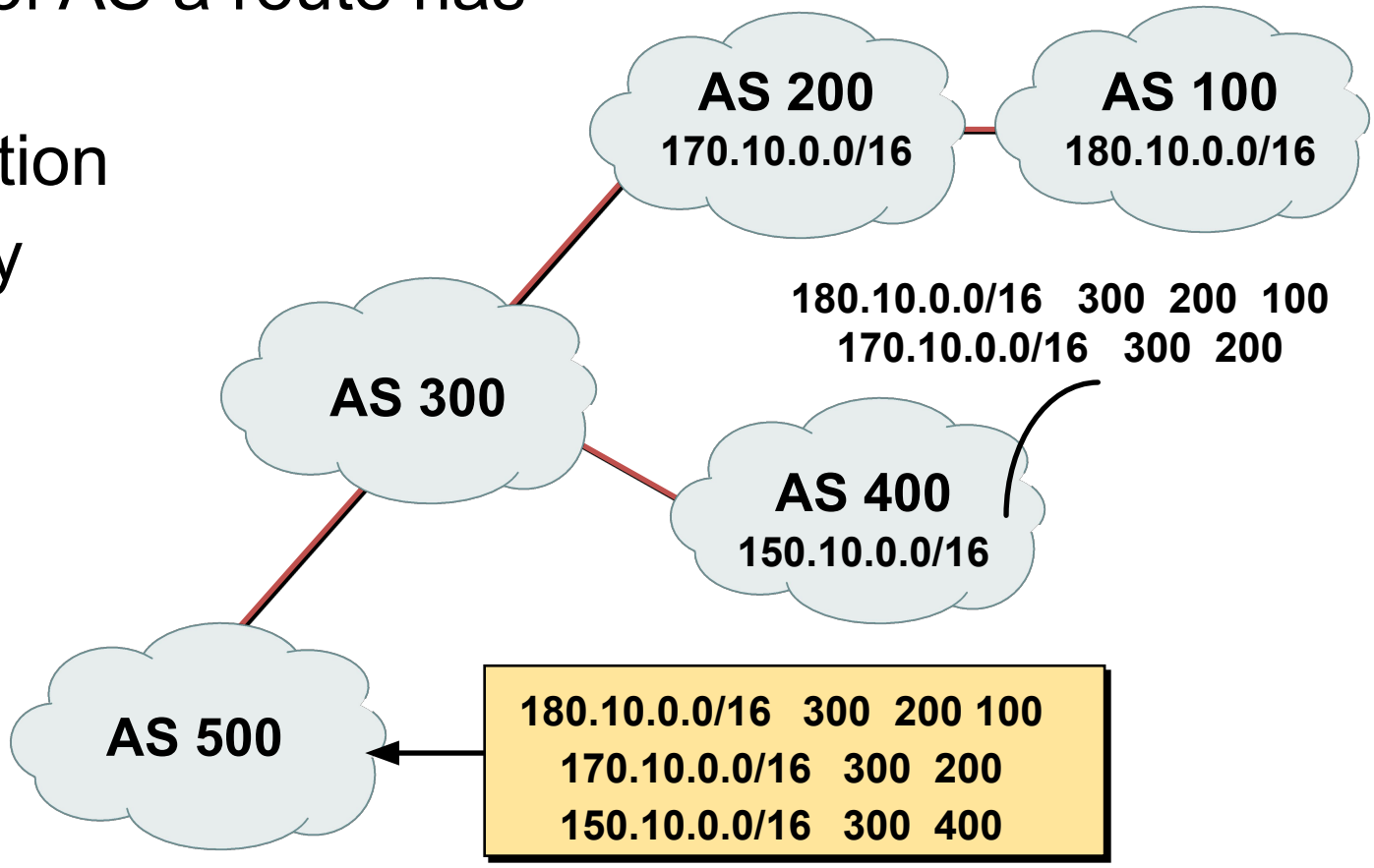
- In this state, the BGP connection is fully established and the peers can exchange **Update, Keepalive** and **Notification messages**.
- If an **Update** or **Keepalive message** is received, the Hold timer is restarted.
- If a **Notification message** is received, the state transitions to **Idle**.

# BGP - Update messages

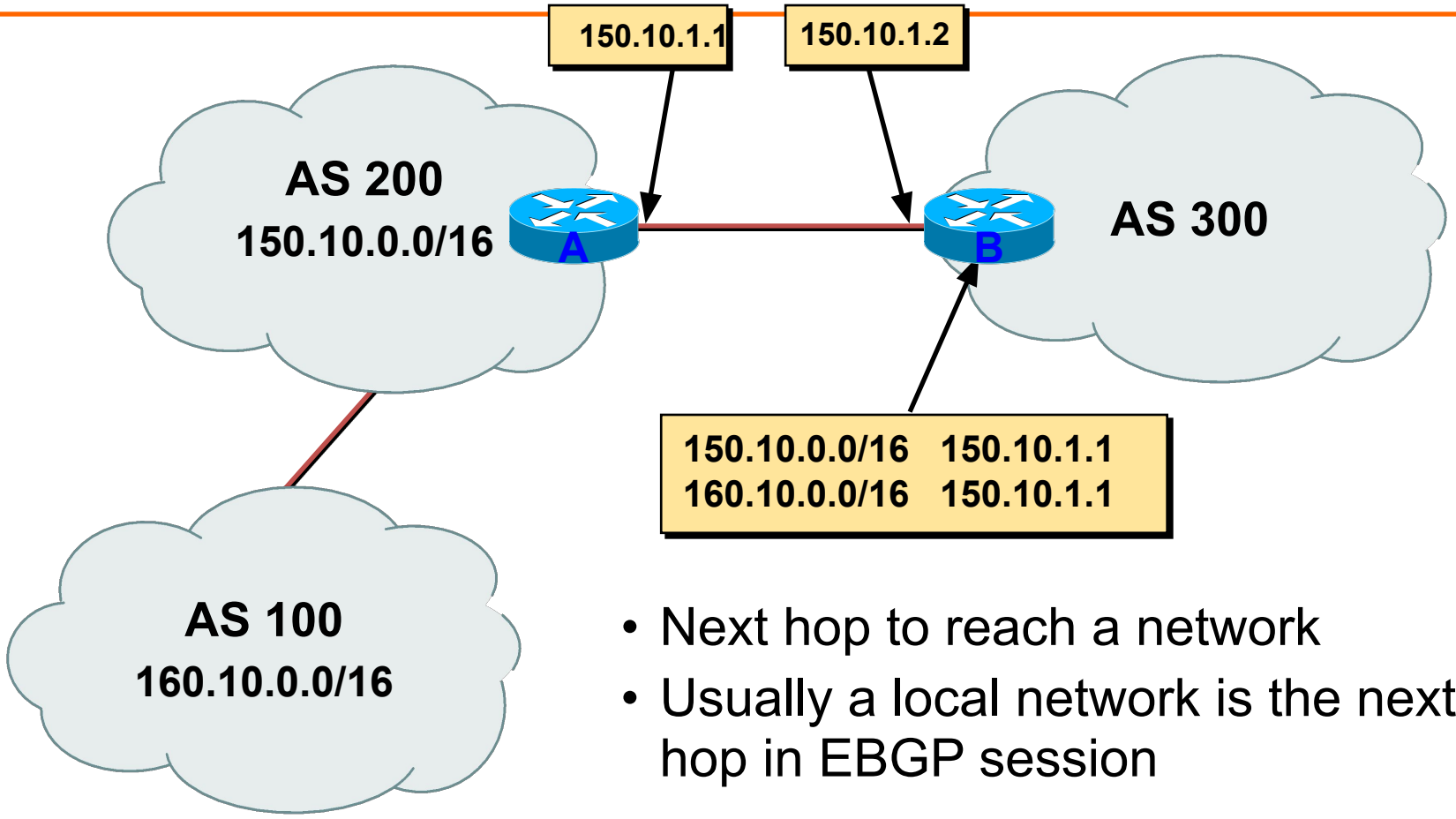
- BGP path attributes
  - AS path
  - Next hop
  - Local preference
  - Multi-Exit Discriminator (MED)
  - BGP community

# AS-Path

- Sequence of AS a route has traversed
- Loop detection
- Apply policy

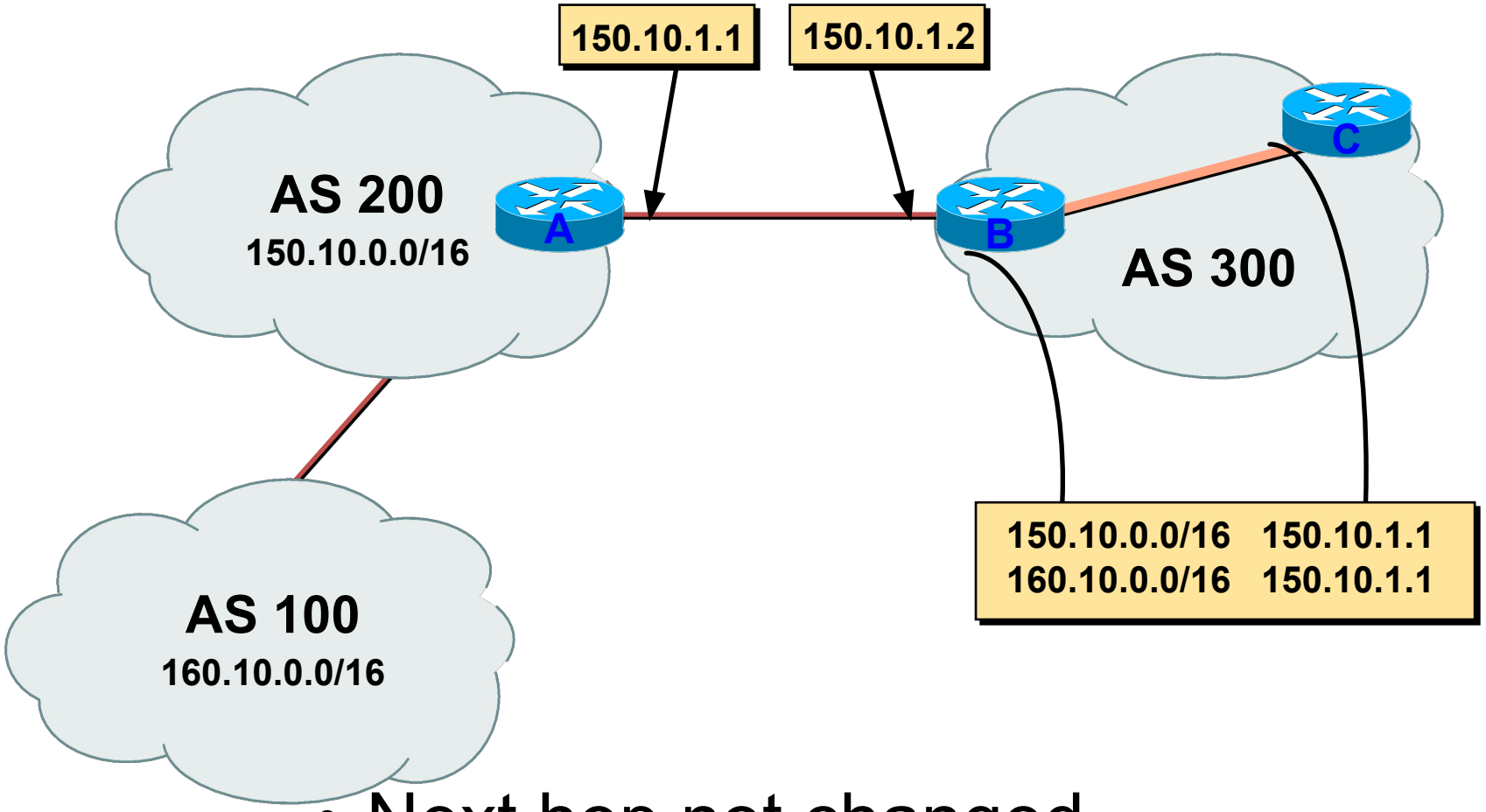


# Next Hop



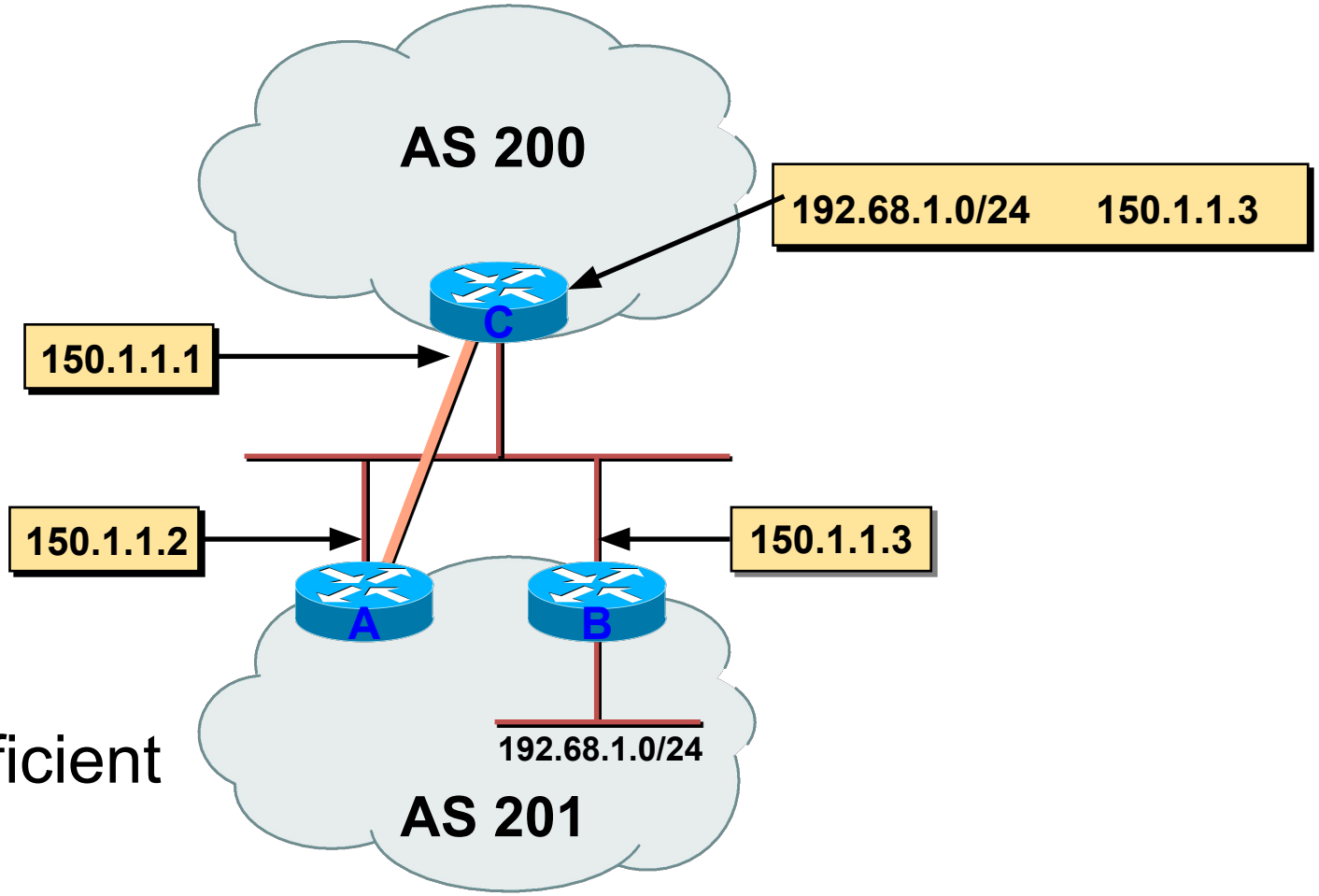

- Next hop to reach a network
- Usually a local network is the next hop in EBGP session

# IBGP Next Hop



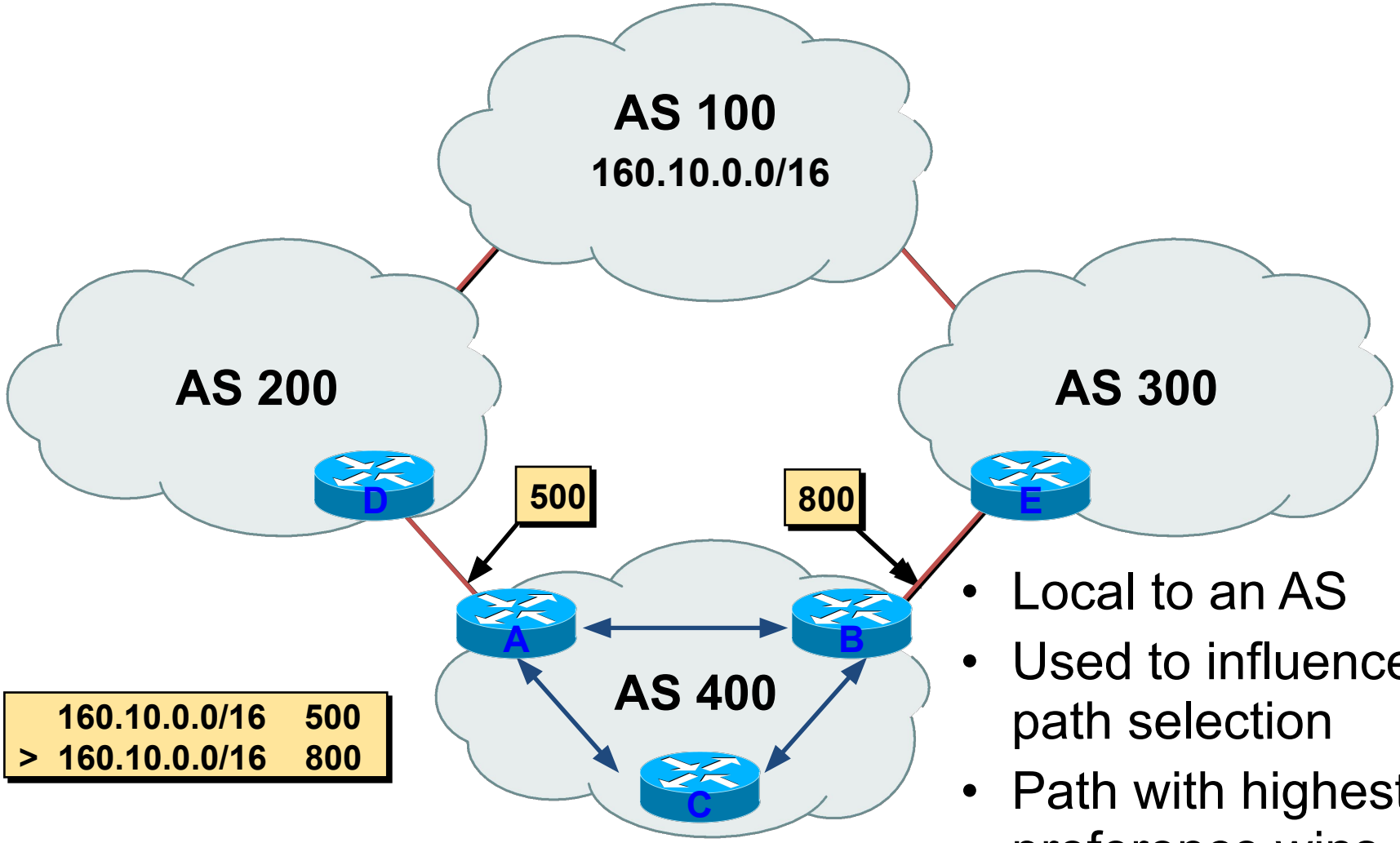

- Next hop not changed

# Third Party Next Hop



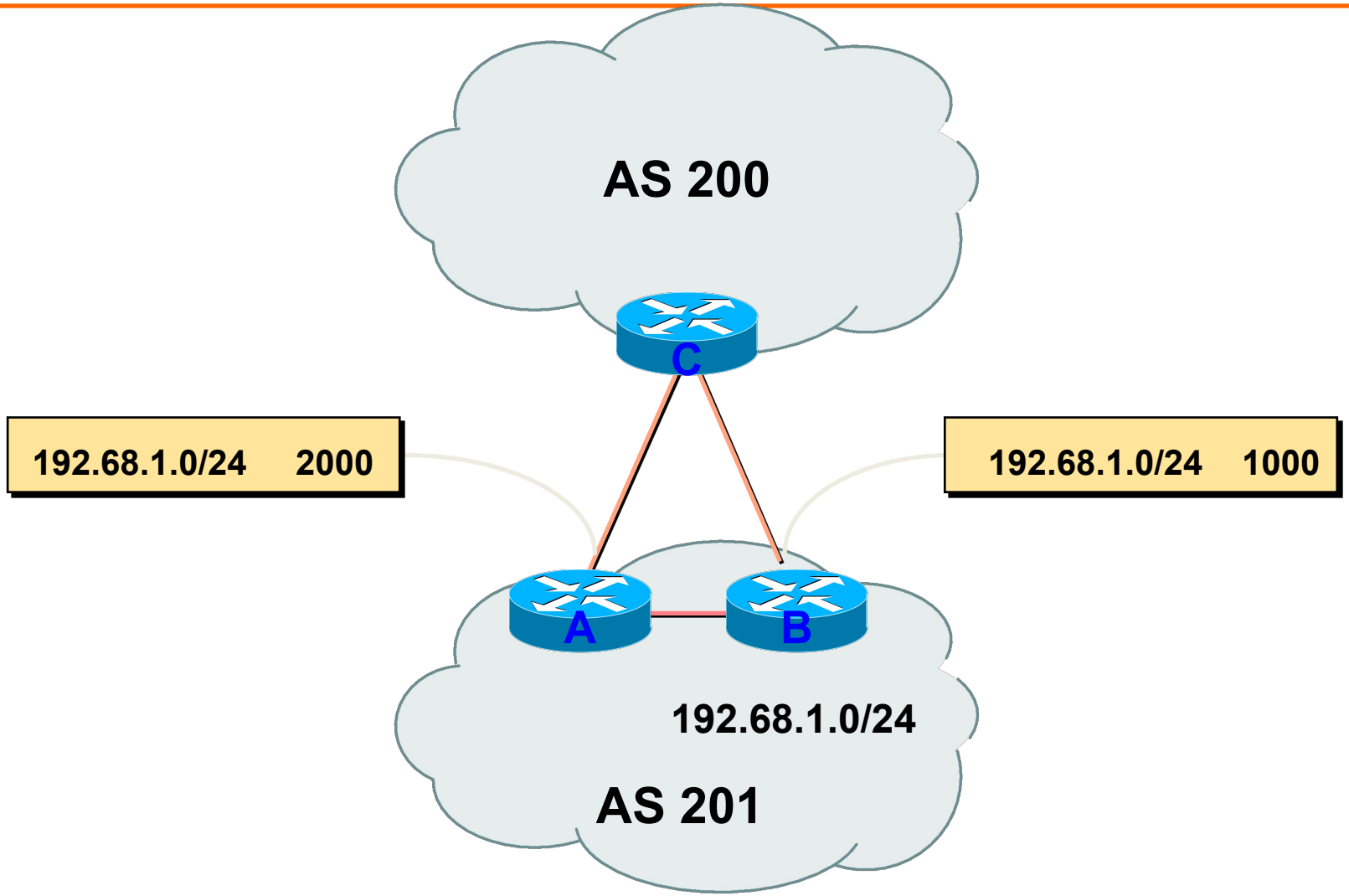

- More efficient

# Local Preference




- Local to an AS
- Used to influence BGP path selection
- Path with highest local preference wins

# Multi-Exit Discriminator (MED)



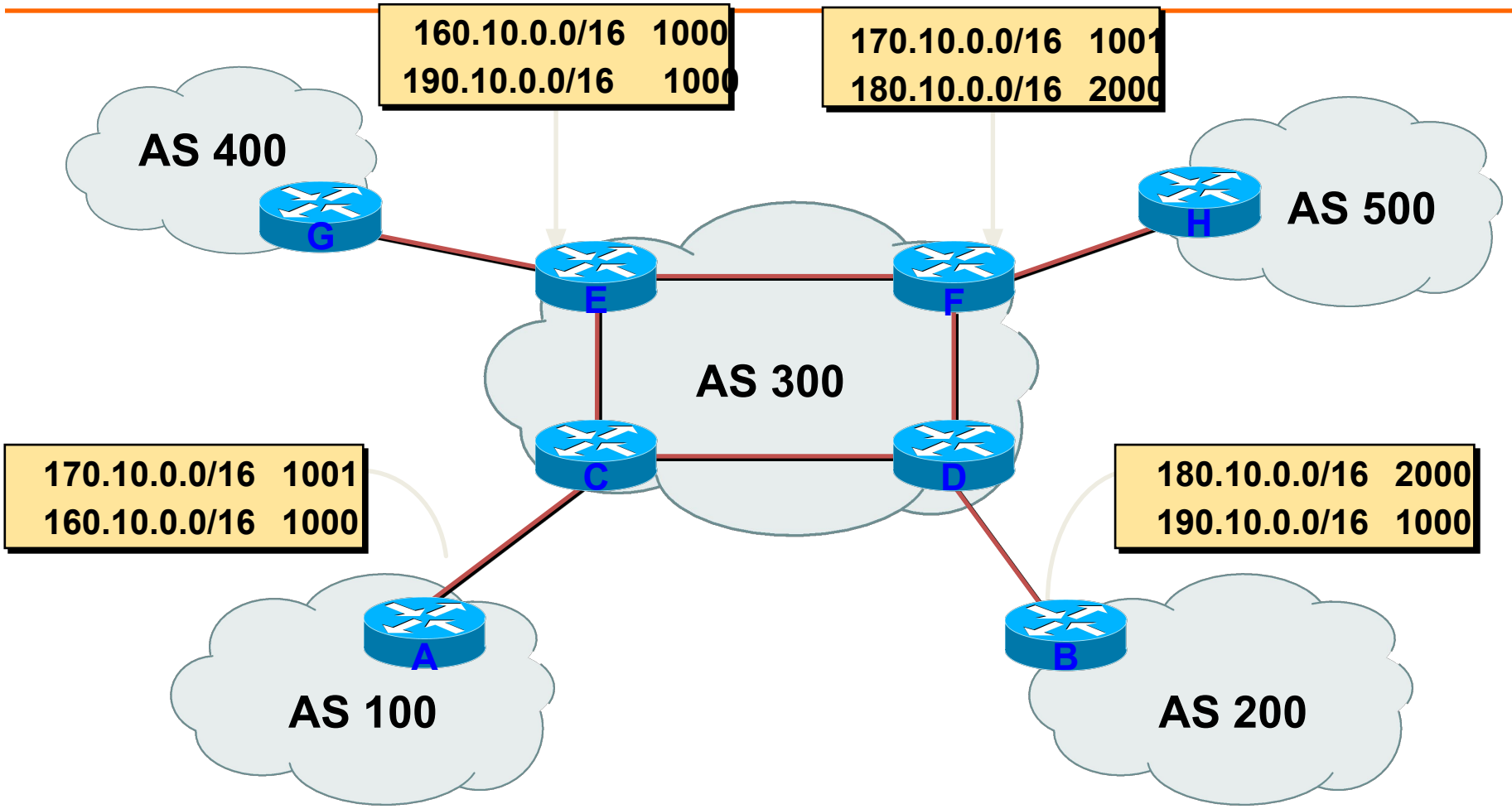
# Multi-Exit Discriminator

- Non-transitive
- Used to convey the relative preference of entry points
- Influences best path selection
- Comparable if paths are from same AS
- IGP metric can be conveyed as MED

# Communities

- BGP attribute
- Used to group destinations
- Each destination could be member of multiple communities
- Community attribute carried across ASs
- Useful in applying policies

# Community


160.10.0.0/16 1000
190.10.0.0/16 1000
170.10.0.0/16 1001
180.10.0.0/16 2000
AS 400
G
E
F
H
AS 500
AS 300
C
D
170.10.0.0/16 1001
160.10.0.0/16 1000
180.10.0.0/16 2000
190.10.0.0/16 1000
A
B
AS 100
AS 200

# Applying Policy with BGP

- Decision Process
  - Phase I - Calculating the degree of preference for each route based on the local preference attribute
  - Phase II - Choosing the best route with highest degree of preference
  - Phase III – dissemination to peers in neighboring ASs, route aggregation and information reduction

- Policy-based on AS path, community or the prefix
- Rejecting/accepting selected routes
- Set attributes to influence path selection

# Overlapping Routes

- BGP speaker may transmit routes with overlapping NLRI Information
- Overlap occurs when a set of destinations are identified in non-matching routes
- Destinations are always identified by IP prefixes
- More specific prefix route gets precedence.

# Breakout rooms

- What would the uses for flow be?
- Why measure BGP?
- Why is end-to-end paths are significantly longer than necessary