The Border Gateway Protocol (BGP)

Internet Structure

- Internet is divided into Autonomous Systems (ASs)
- AS - a collection of one or more networks under a single technical administration
- *technical administration*- refers to aspects of the n/w like, routing policies etc.

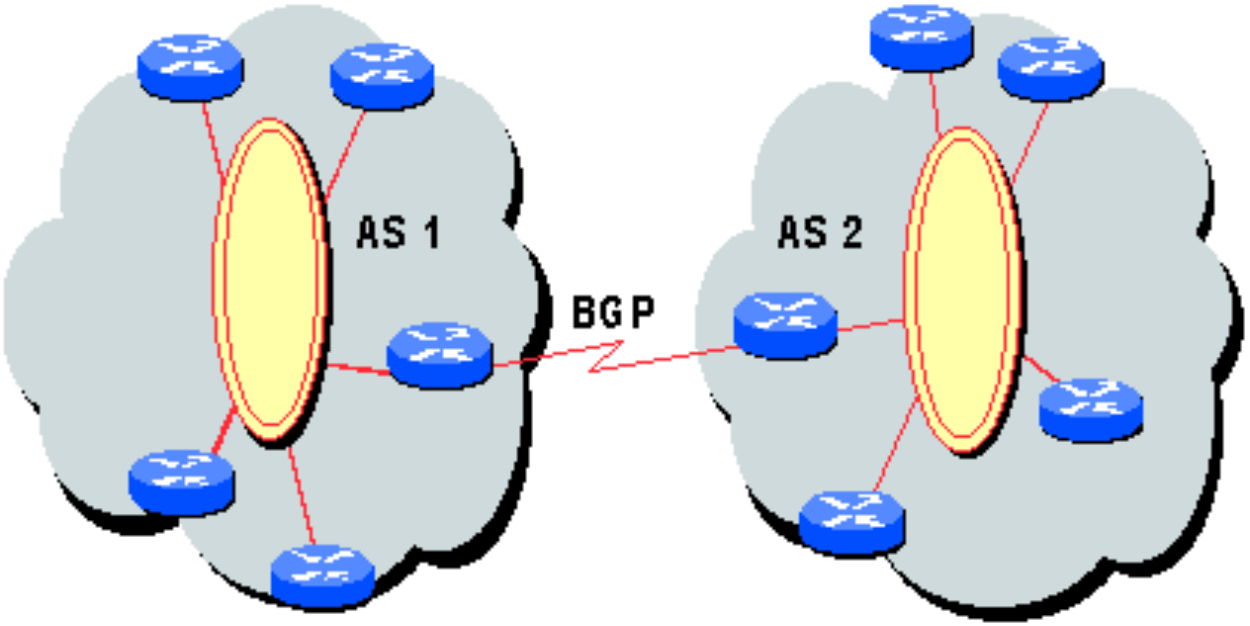
BGP Operation

- External BGP is the *de facto* Inter-AS routing protocol
- Used for exchanging route information between ASs
- Conveys information about AS path topology
- iBGP is widely used for internal routing by large cloud providers

Routing between AS

- Intra-AS routing driven mostly by performance considerations
- Inter-AS routing depends on policy issues, economics etc
- IXPs provide efficient local peering opportunities and usually involve the use of BGP route servers…

BGP Between AS's



BGP Basics

- Current version, BGP 4, defined in RFC 1771
- Runs over TCP (port 179)
- Path Vector protocol
  - Exchange entire path information
  - Prevents loops
- Internet routing (between AS's) depends on the BGP protocol

How does BGP work?

- How is it configured?
- How is route information exchanged using BGP?
- What are the attributes of the exchanged routes?
- How is this information processed? And further distributed?

BGP Configuration

- Identify BGP “speakers”, assign AS number and define neighbours (peers)
- External BGP peers are directly connected
- Direct connection is not necessary for Internal BGP peers

BGP Message Exchange

- BGP peers form a TCP connection, use the OPEN message to establish BGP connection;
- Connections kept open by KEEPALIVE messages sent periodically;
- Initially exchange routing tables, further modifications are incremental;

BGP Message Exchange (II)

- Modifications (Route additions and withdrawls) are made by UPDATE messages;
- Errors are reported by NOTIFICATION messages;
- Most interesting stuff happens in the processing of UPDATE messages.

The UPDATE message

- Used to transfer routing information between peers (NLRI)
  - Network Layer Reachability Information (NLRI) exchanged between BGP routers using UPDATE messages.
  - NLRI message is composed of a LENGTH and an IP PREFIX e.g., 91.100.56.0/21
  - Also contains Path Attributes information

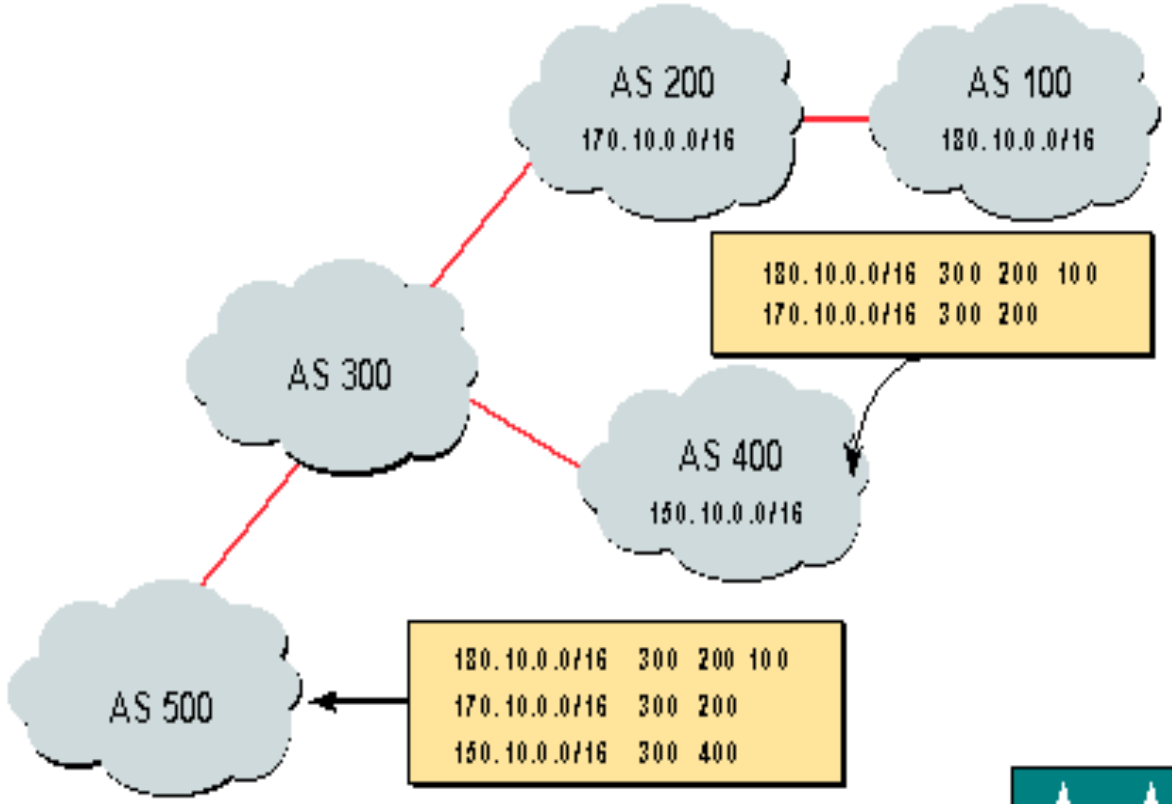
Routing Information Storage

- Adj-RIBs-In
  - Store messages that have been learned from inbound UPDATE messages
- Loc-RIB
  - Contains local routing information of a BGP speaker
- Adj-RIBs-Out
  - Store routing information local router has chosen for distribution

Path Attributes

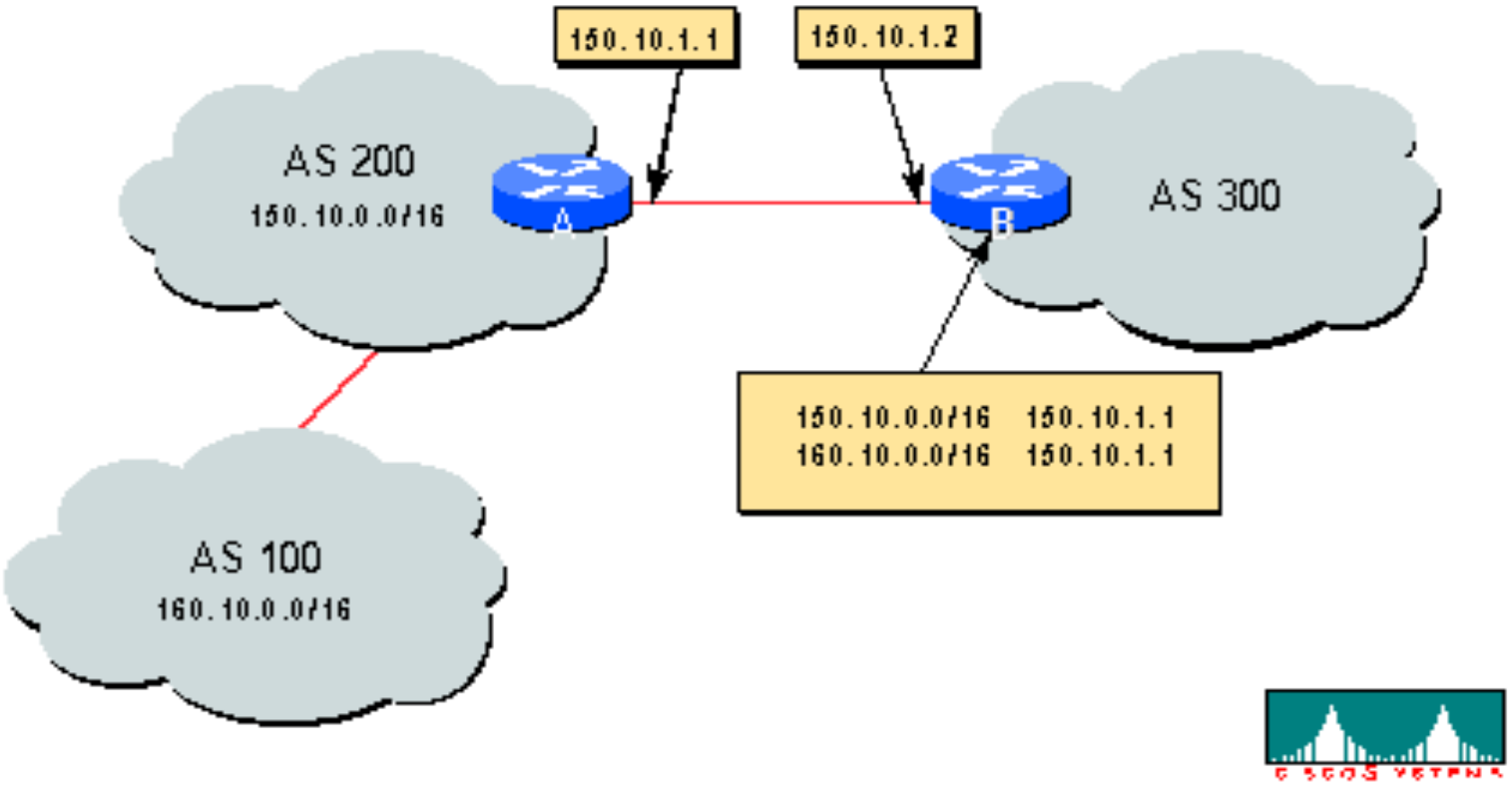
- Origin
  - defines the origin of a path
- AS_PATH
- NEXT_HOP
- MULTI_EXIT_DISC
- LOCAL_PREF
- Community String

AS-Path



Composed of a sequence of AS path segments or the set of ASs traversed

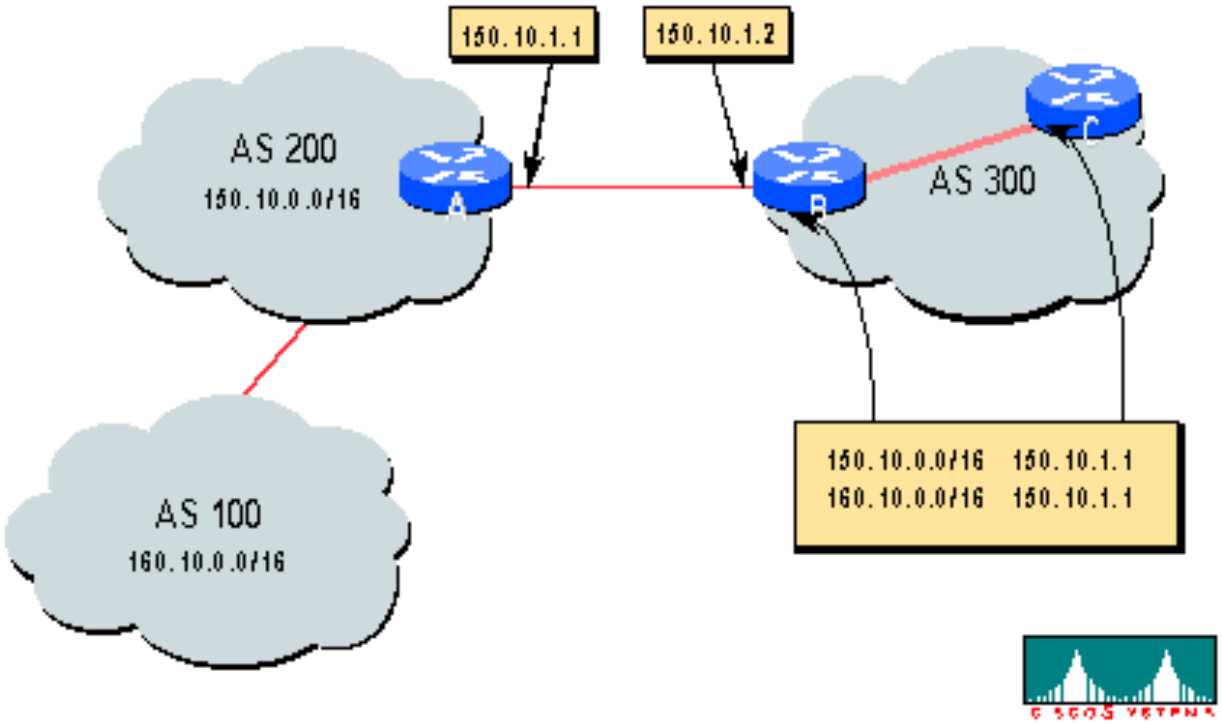
eBGP and Next Hop



IP address of the Border router that should be used for the next hopxxx

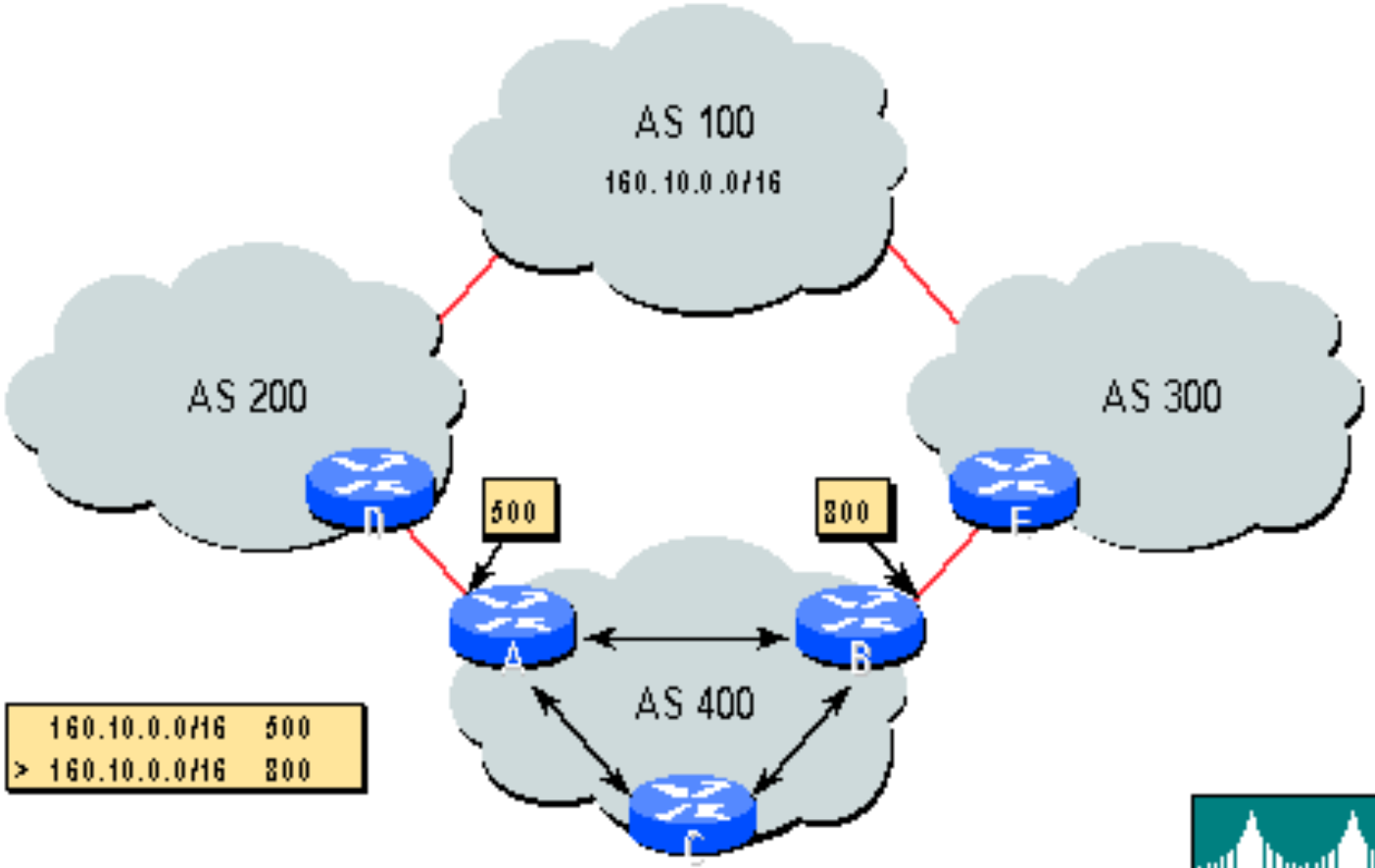
c

iBGP and Next Hop



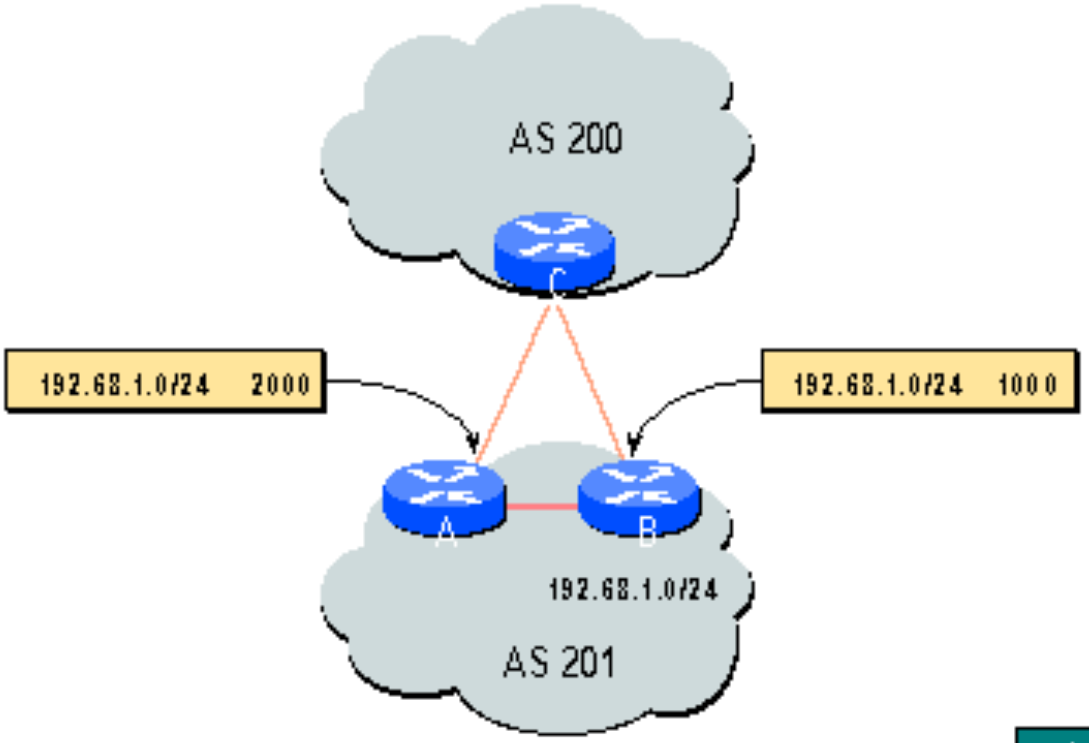
Routers generally use the next-hop-self attribute to set next hop attribute for received updates to its own IP address override the next hop received via eBGP

Local Preference



Used to inform other BGP speakers in the same AS about preferences for a particular route

Multi-Exit Discriminator



Used to discriminate among multiple exit points in neighboring ASs

Community Attribute

- Optional Attribute e.g., 312:2666
- Method to group destination in a certain community
- Apply routing policy destinations (accept, prefer, redistribute, reject) using these communities

Other Issues!

- Achieving Stability
  - BGP/IGP interaction
  - Route Flap dampening
  - Max Prefixes (for each peer)
  - Simulate configurations first (especially the inbound and outbound route filters)
  - Configuration mistakes (and poor testing) can lead to unexpected consequences…