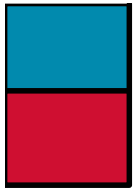
BGP4

Ravi Chandra

cisco Systems

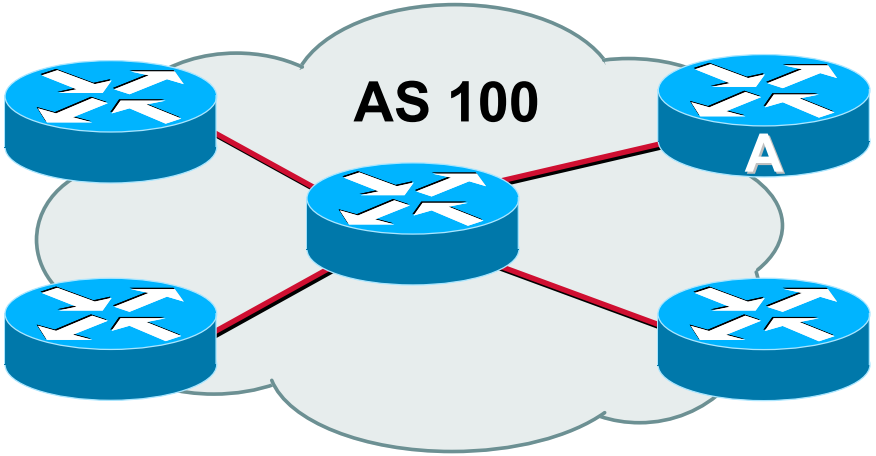


Border Gateway Protocol (BGP)

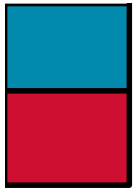
- Introduction to BGP
- BGP Peer Relationship
- BGP Attributes
- Applying Policy with BGP
- Putting it all together

Cisco Systems

Autonomous System (AS)

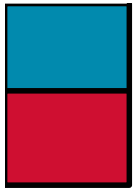


- Collection of networks with same policy
- Single routing protocol
- Usually under single administrative control



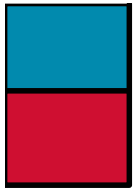
Autonomous System...

- Identified by ‘AS number’
- Examples:
  - service provider
  - multihomed customers
  - anyone needing policy descrimination



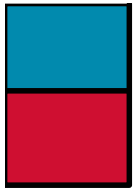
Terminology: Exterior routes

- **Routes learnt from other autonomous systems**



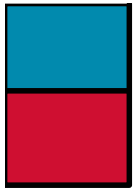
What Is an IGP?

- Interior Gateway Protocol
- Within an Autonomous System
- Carries information about internal prefixes
- Examples—OSPF, ISIS, EIGRP…



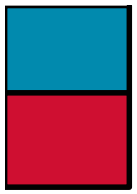
What Is an EGP?

- Exterior Gateway Protocol
- Used to convey routing information between Autonomous Systems
- Decoupled from the IGP
- Current EGP is BGP



Why Do We Need an EGP?

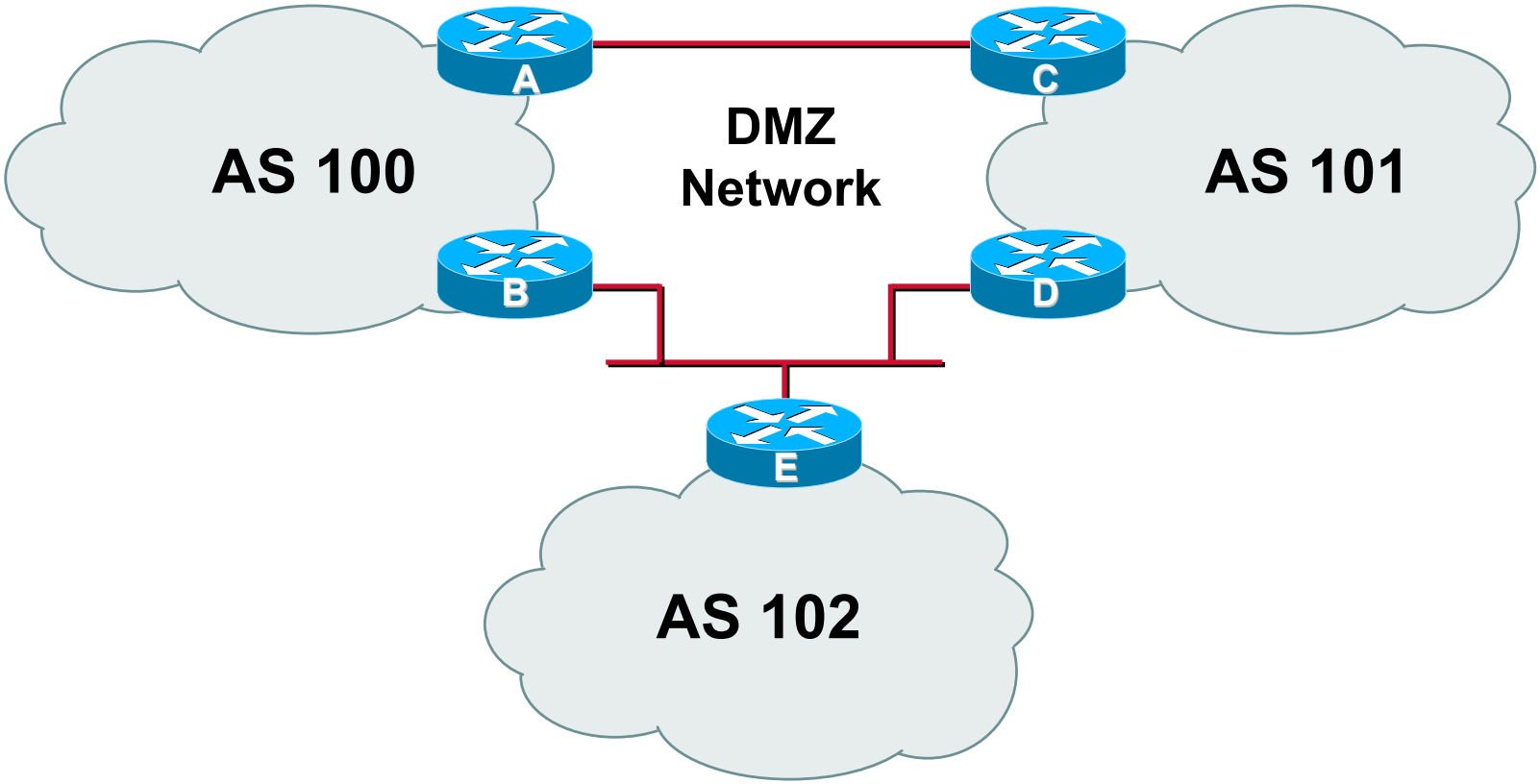
- **Scaling to large network**
  - **Hierarchy**
  - **Limit scope of failure**
- **Policy**
  - **Control reachability to prefixes**



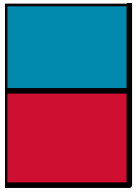
Interior vs. Exterior Routing Protocols

- **Interior**
  - Automatic discovery
  - Generally trust your IGP routers
  - Routes go to all IGP routers
- **Exterior**
  - Specifically configured peers
  - Connecting with outside networks
  - Set administrative boundaries

Demilitarized Zone (DMZ)



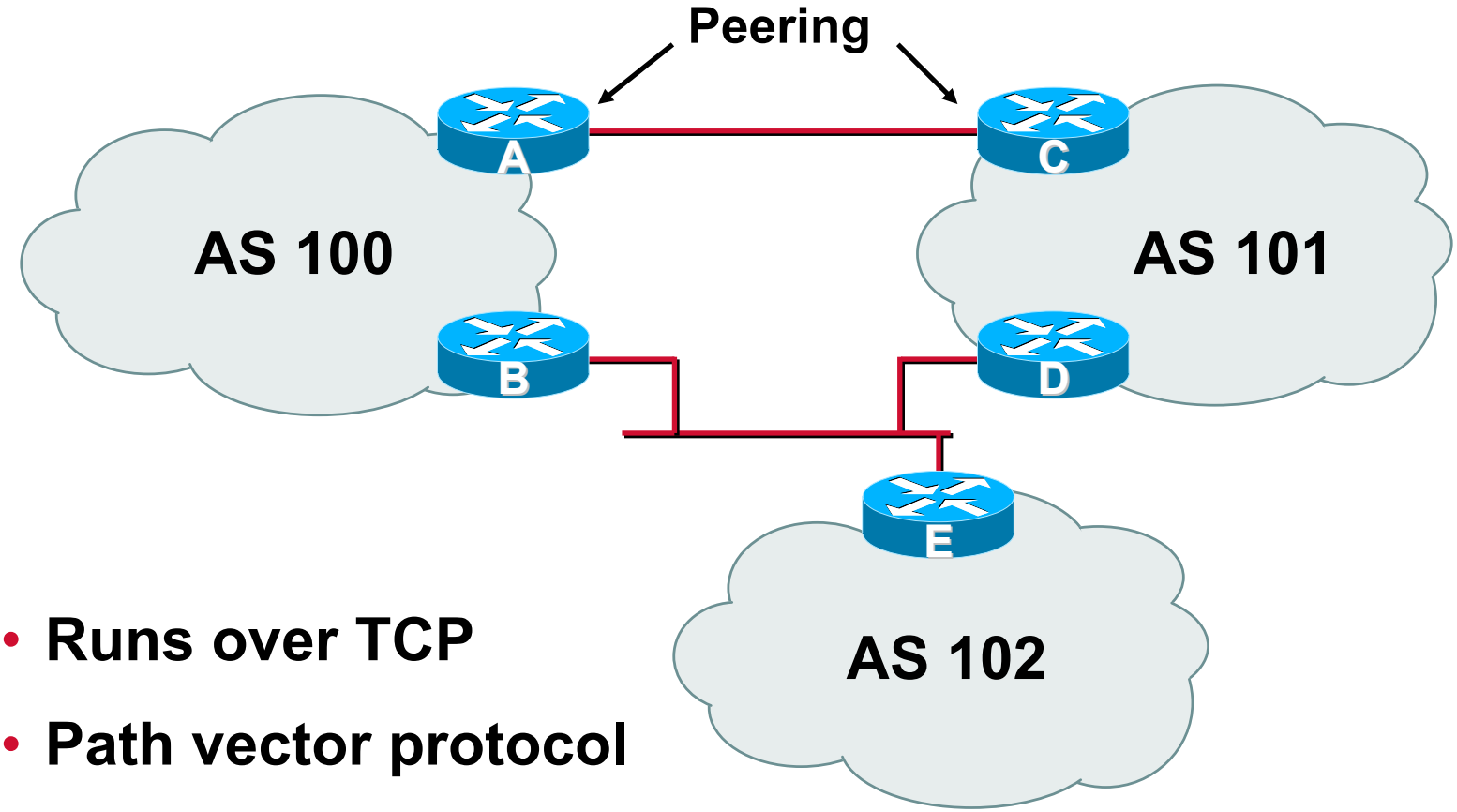
- Shared network between ASs



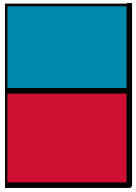
Static Routes

- no path information
- very versatile
- low protocol overhead
- high maintainance
- very bad convergence time
- requires manual configuration

BGP Basics



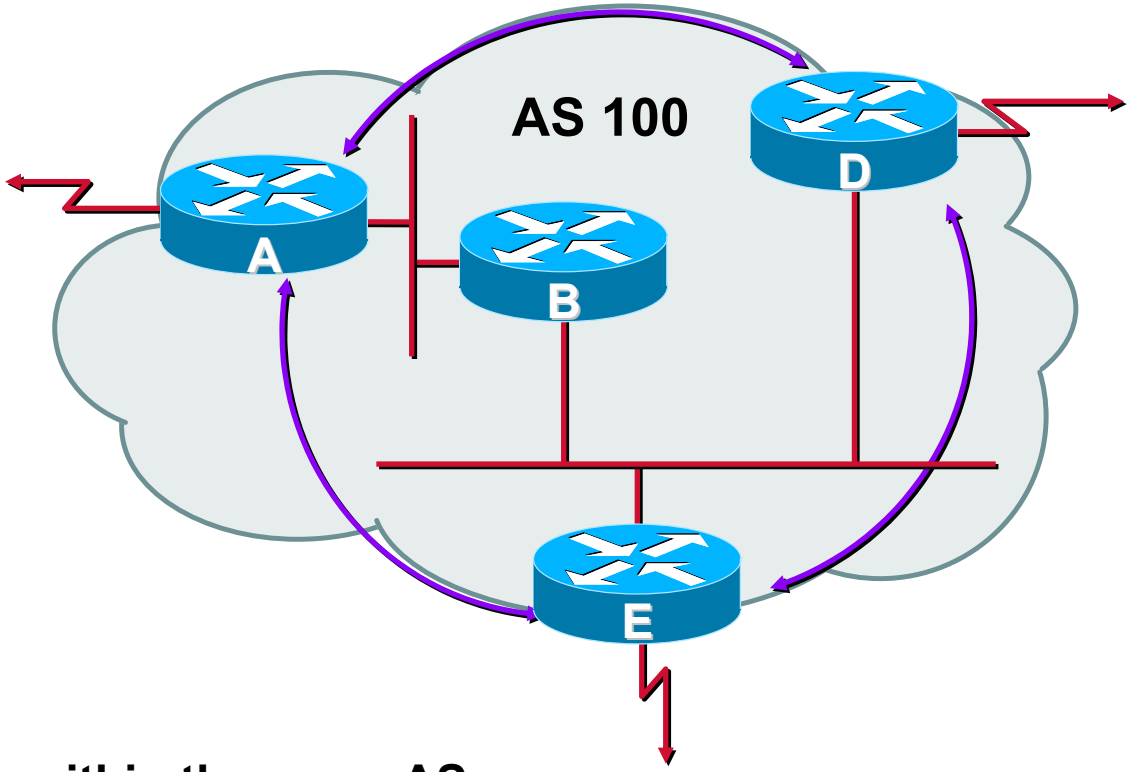
- Runs over TCP
- Path vector protocol
- Incremental update



General Operation

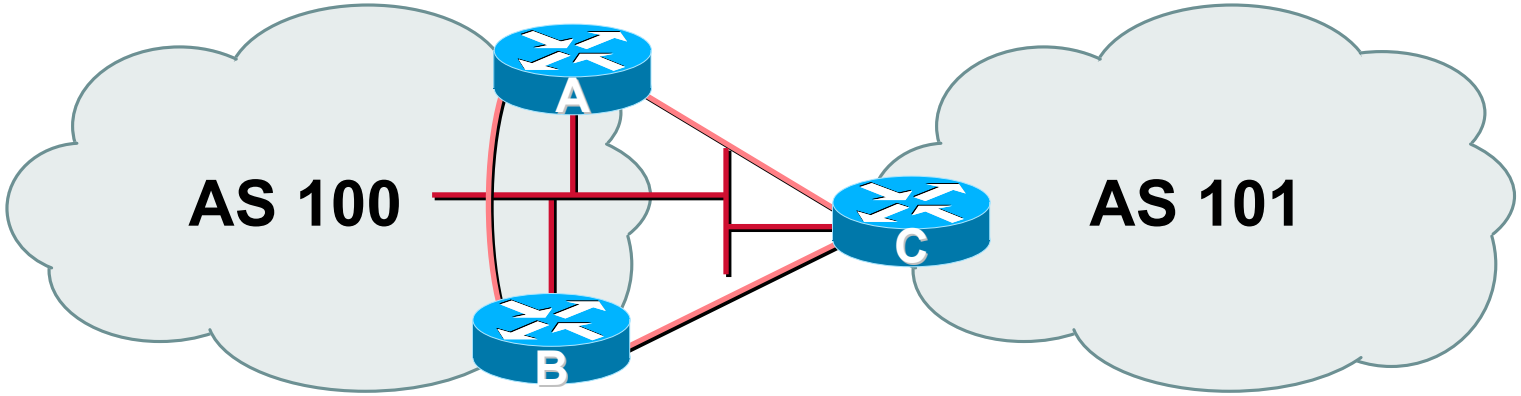
- Learns multiple paths via internal and external BGP speakers
- Picks the best path and installs in the IP forwarding table
- Policies applied by influencing the best path selection

Internal BGP Peering

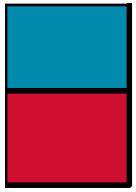


- BGP peer within the same AS
- Not required to be directly connected
- IBGP neighbors should be fully meshed

External BGP Peering



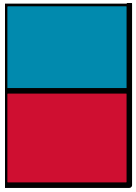
- Between BGP speakers in different AS
- Should be directly connected



Basic BGP commands:

- router bgp <as-number>
- neighbor <ip address> remote-as <as-number>
- show commands
  - show ip bgp summary
  - show ip bgp neighbors

Cisco Systems

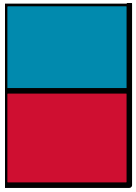


Exercise - 1

- Define a static route
- Bring the route into BGP table
- Verify if route is in the BGP table

Exercise - 2

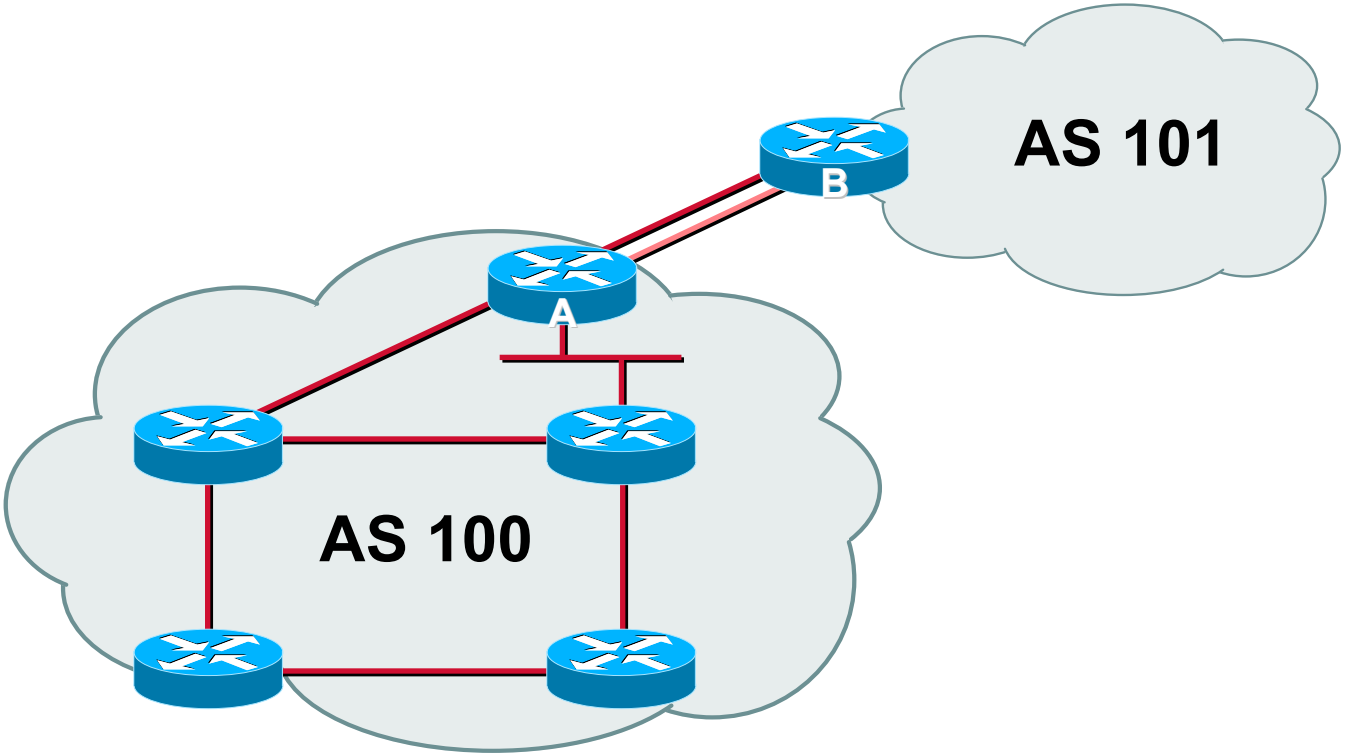
- IBGP peering
- Verify IBGP peering

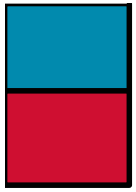


Exercise - 3

- EBGP peering
- Verify EBGP peering

Stub Network

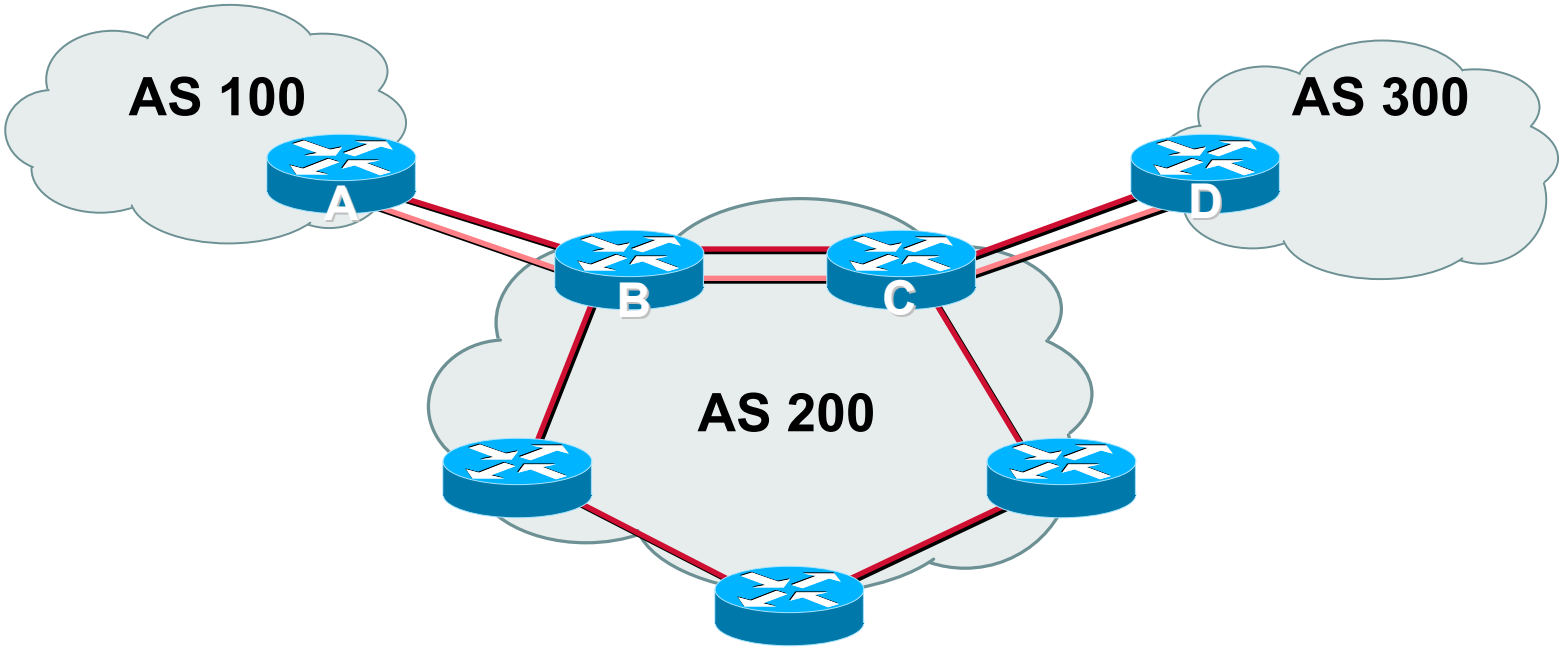


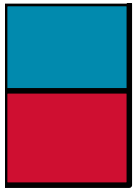


Stub Network

- No need for BGP
- Point default towards the ISP
- ISP advertises the stub network
- Policy confined within ISP policy

Multi-Homed AS

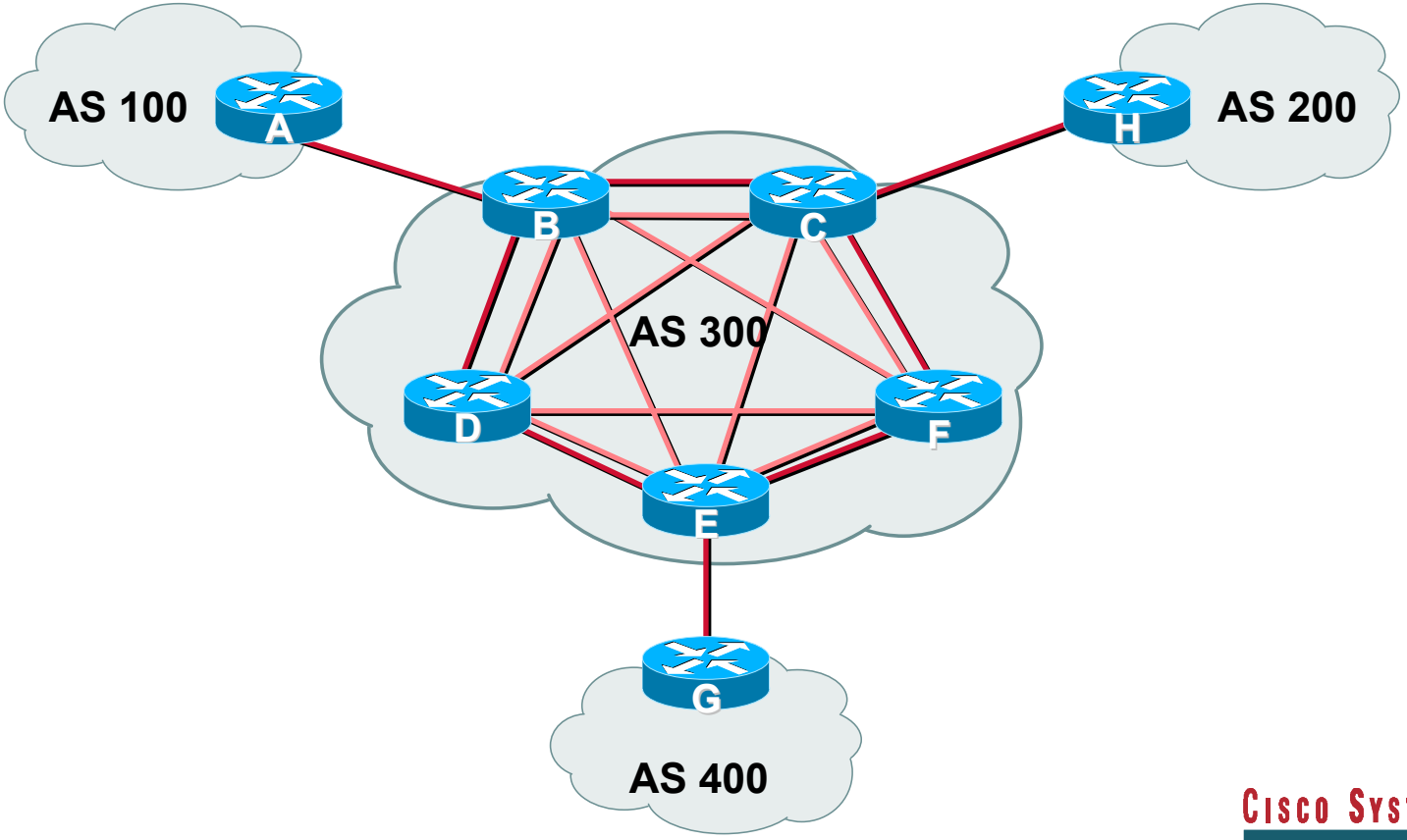


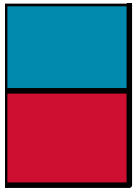


Multihomed AS

- Internal BGP used with IGP
- IBGP only between border gateways
- Only border gateways speak BGP
- Exterior routes must be redistributed into IGP or use defaults

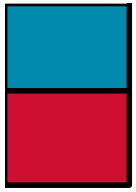
Common Service Provider Network





Service Provider Network

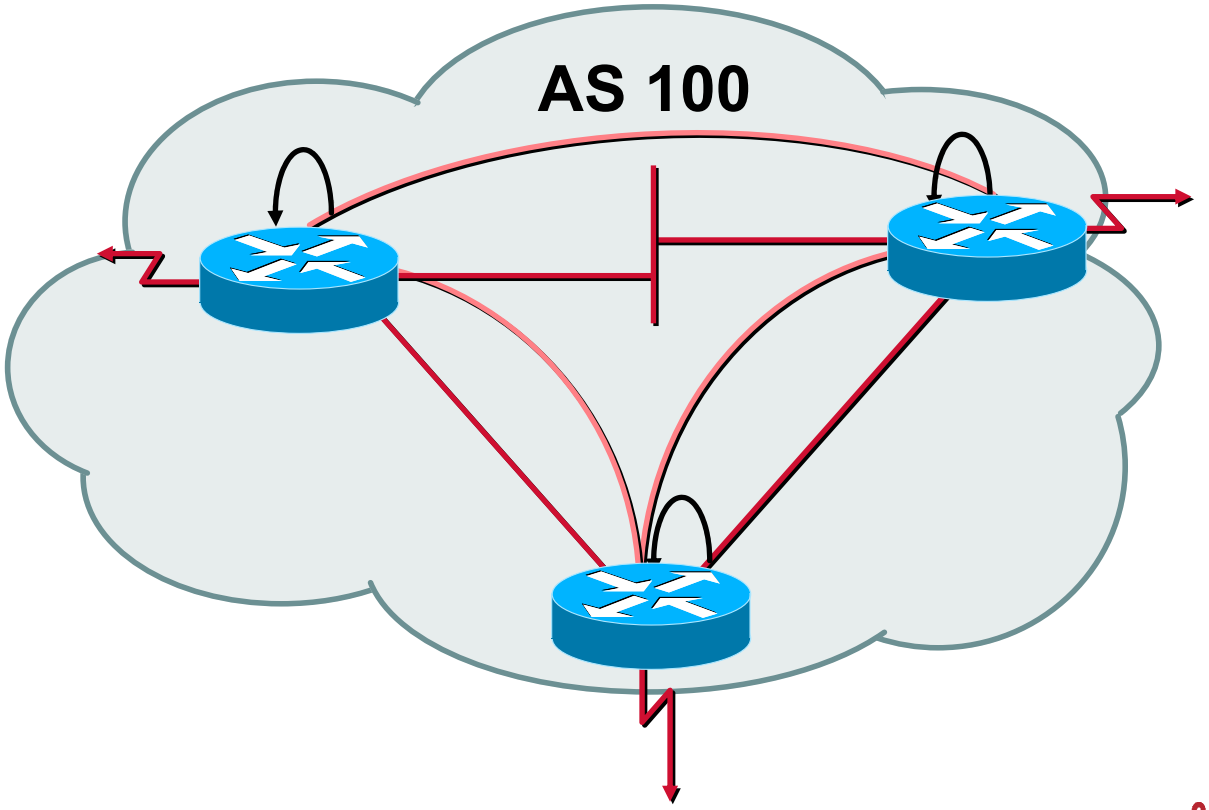
- IBGP used to carry exterior routes
- IGP carries local information only
- Full IBGP mesh is required

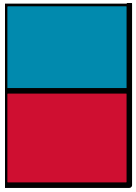


Stable IBGP Peering

- Peer with loop-back address
- IBGP session is not dependent on a single interface
- Loop-back interface does not go down

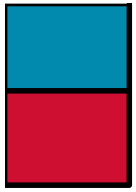
Peering to Loop-Back Address





Exercise - 4

- Stable IBGP peering
- Verify IBGP peering

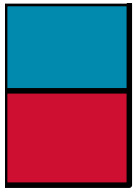


BGP - Update messages

- withdrawn routes
- attributes
- advertised routes

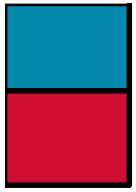
BGP: Update Messages..

- Network reachability information
- network prefix/length
- Example :
  - 131.108/16
    - 131.108.0.0 255.255.0.0
  - 198/8
    - 198.0.0.0 255.0.0.0



BGP Attributes

- What is an attribute?
- AS path
- Next hop
- Local preference
- Multi-Exit Discriminator (MED)



BGP Attributes

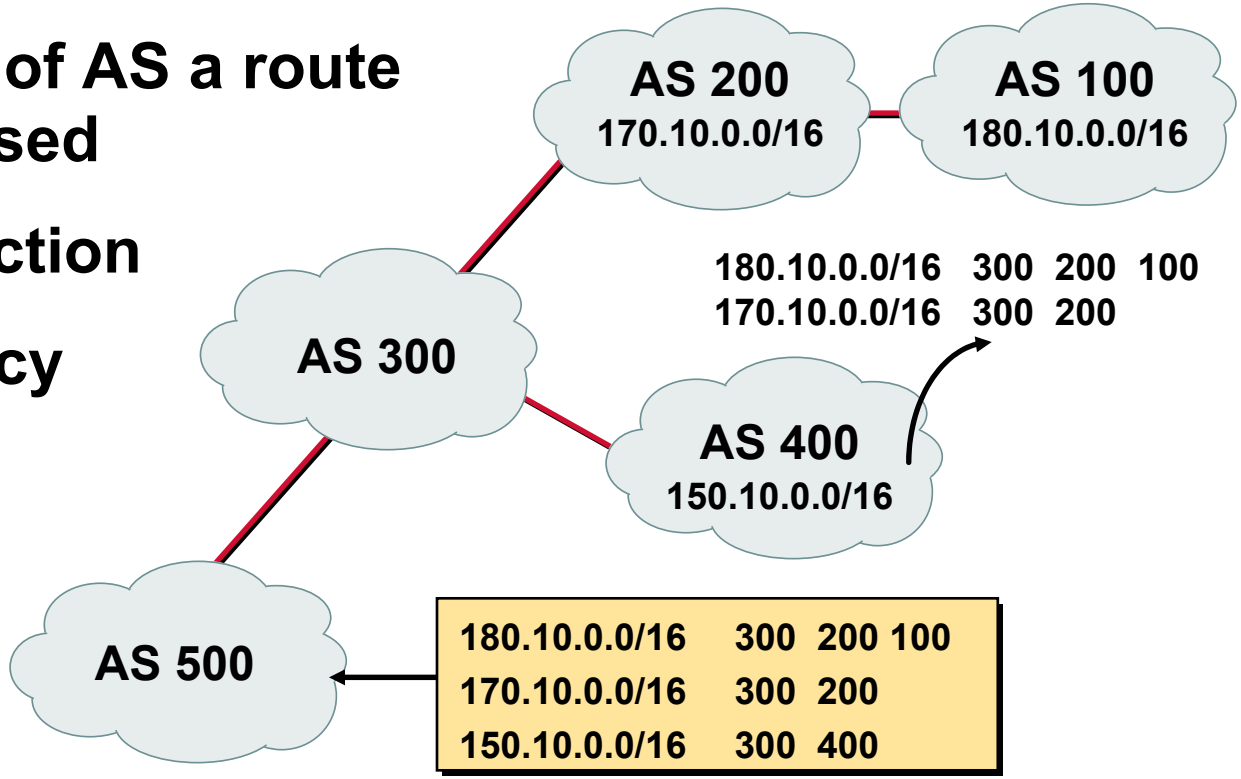
- BGP community
- Others

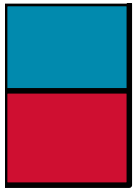
What Is an Attribute?

- Describes the characteristics of prefix
- Transitive or non-transitive
- Some are mandatory

AS-Path

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

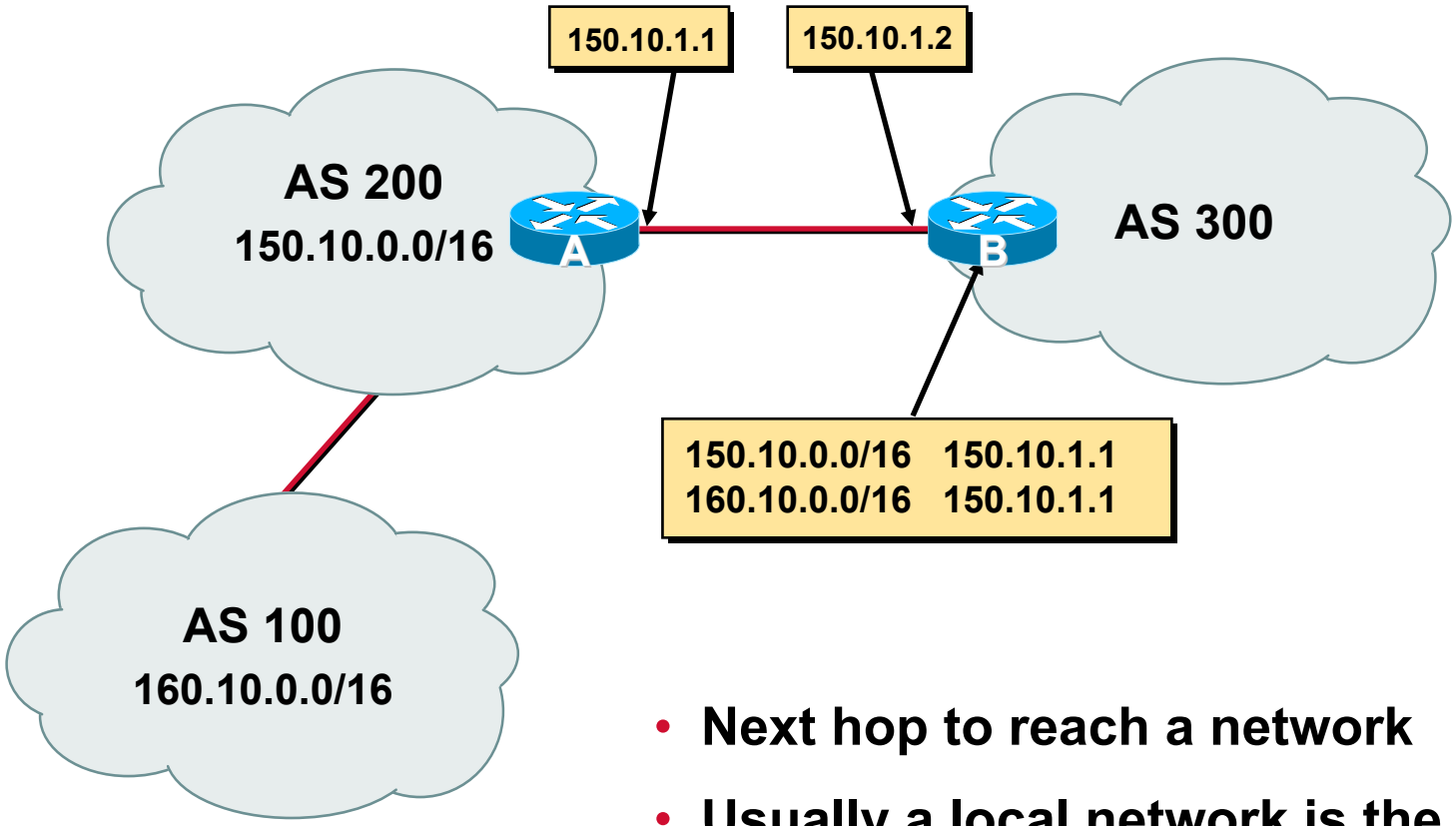




Exercise - 5

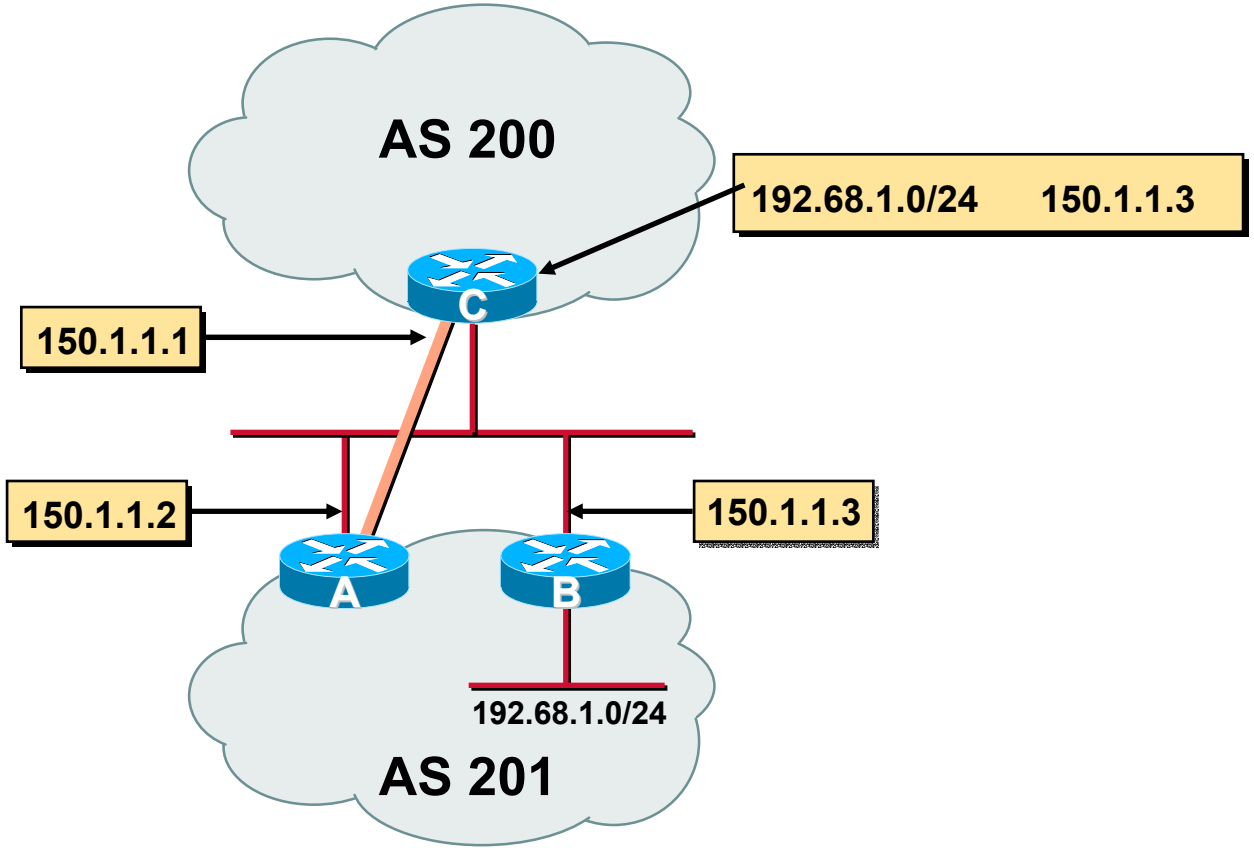
- Look at live routing table

Next Hop



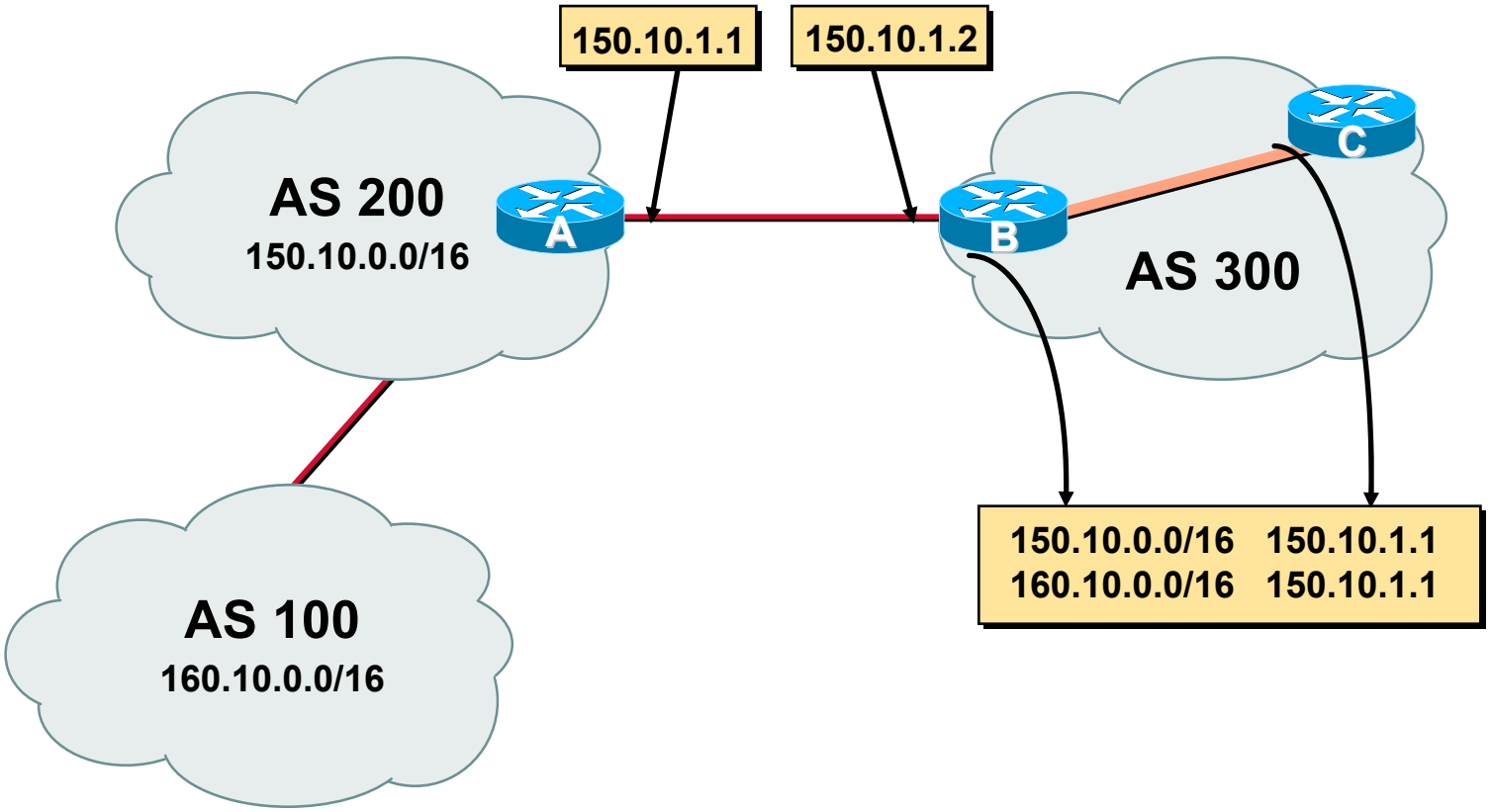
- Next hop to reach a network
- Usually a local network is the next hop in EBGP session

Third Party Next Hop



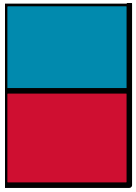
- More efficient

IBGP Next Hop



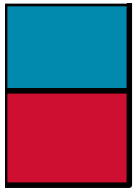
- Next hop not changed

CISCO SYSTEMS



Next Hop-More

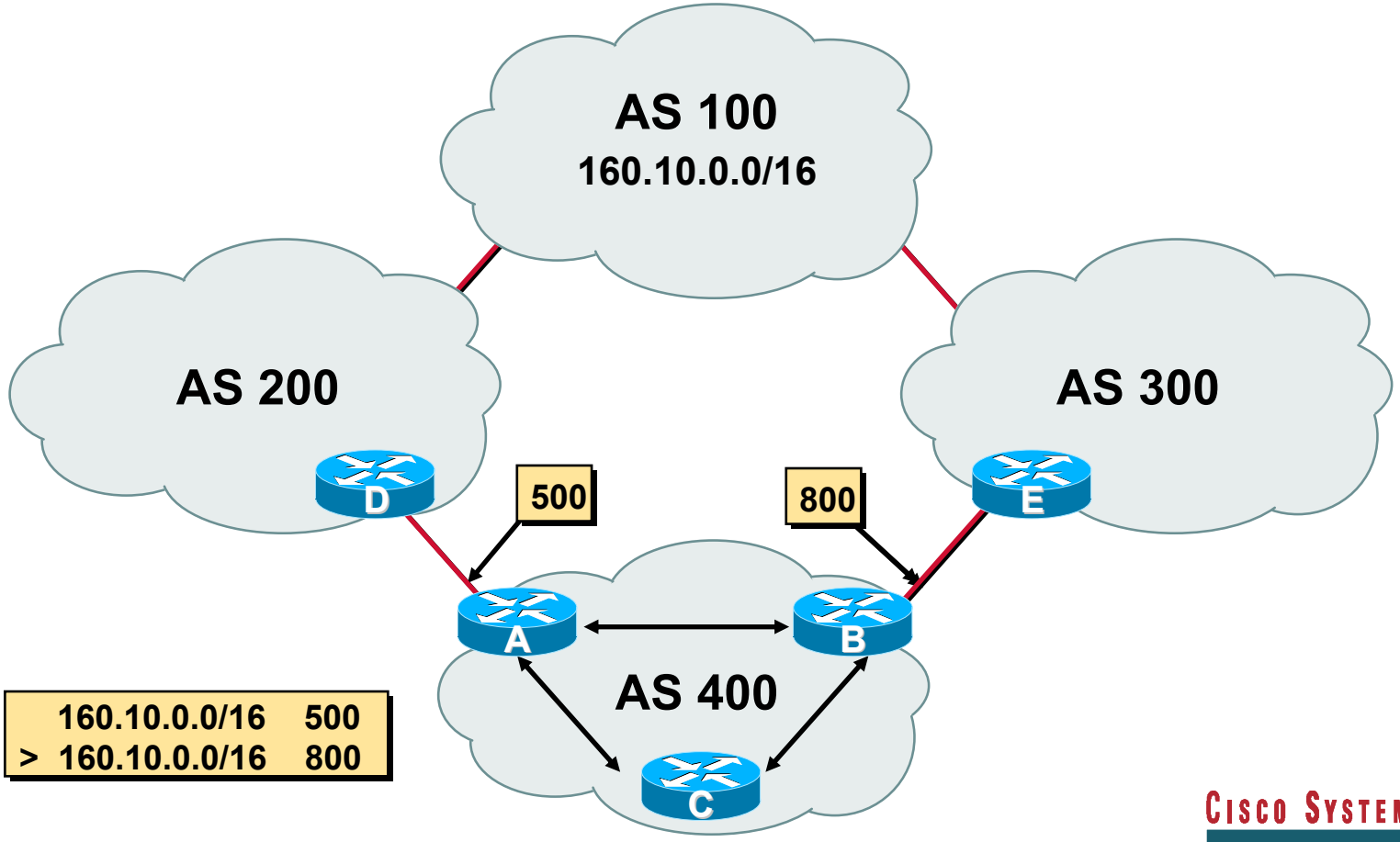
- IGP should carry route to next hops
- Recursive route look-up
- Unlinks BGP from actual physical topology
- Allows IGP to make intelligent fowarding decision

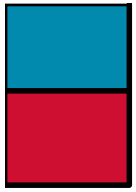


Exercise - 6

- IBGP nexthop exercise

Local Preference





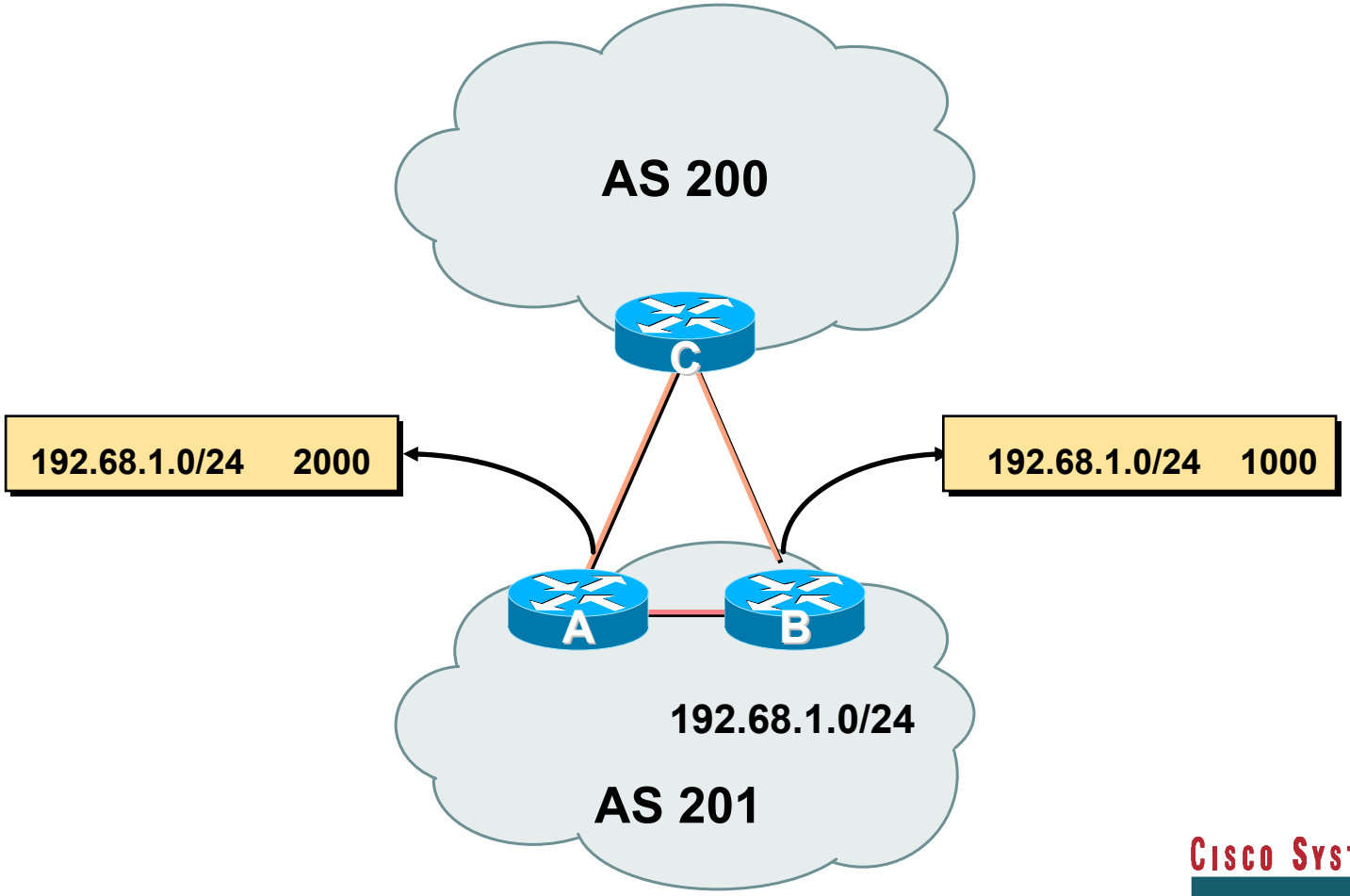
Local Preference

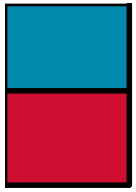
- Local to an AS
- Used to influence BGP path selection
- Path with highest local preference wins

Exercise - 7

- Local Preference exercise

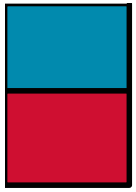
Multi-Exit Discriminator (MED)





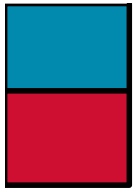
Multi-Exit Discriminator

- 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



Origin

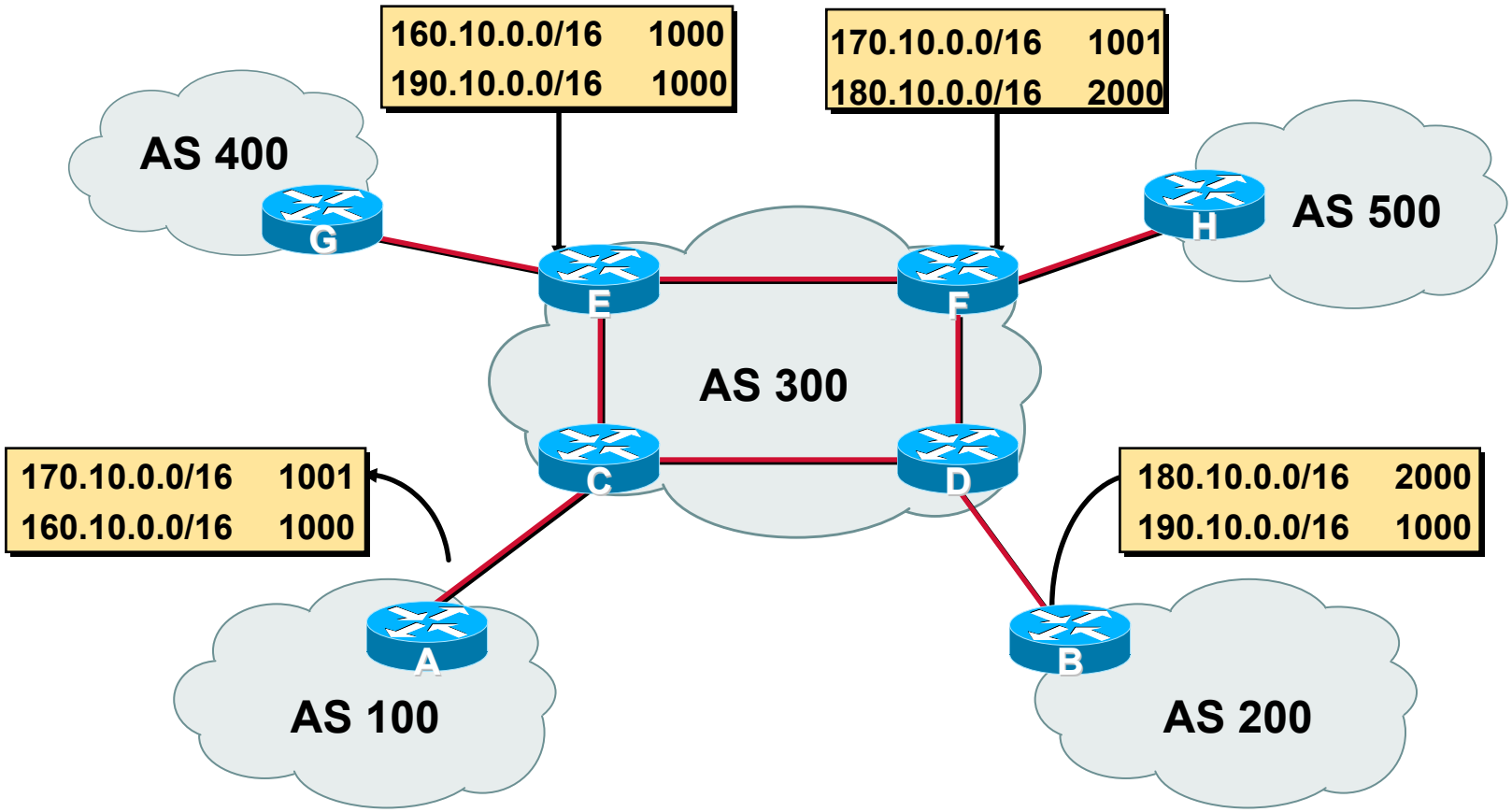
- Conveys the origin of the prefix
- Three values—igp, egp, incomplete
- Influences best path selection

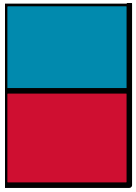


Communities

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

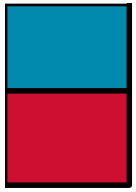
Community


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



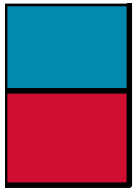
Applying Policy with BGP

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



BGP Path Selection Algorithm

- Do not consider IBGP path if not synchronized
- Do not consider path if no route to next hop
- Highest weight (local to router)
- Highest local preference (global within AS)
- Shortest AS path



BGP Path Selection

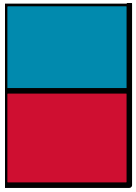
- **Lowest origin code**

  IGP < EGP < incomplete

- **Multi-Exit Discriminator**

  Considered only if paths are from the same AS

- **Prefer EBGP path over IBGP path**



BGP Path Selection

- Path with shortest nexthop metric wins
- Lowest router-id

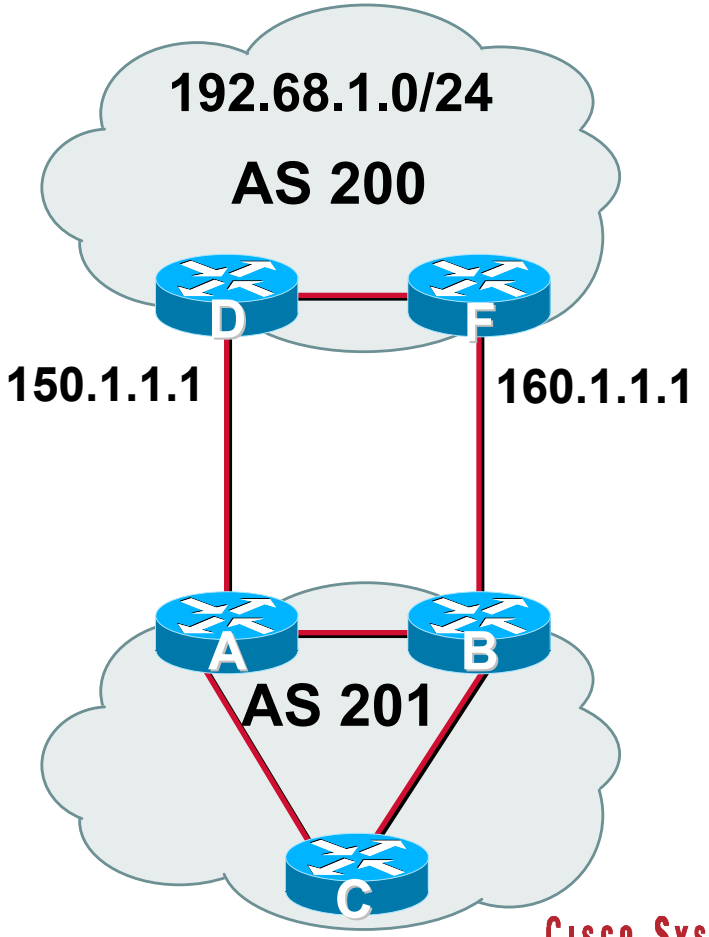
BGP Path Selection

BGP TABLE IN AS-201:
192.68.1.0/24 150.1.1.1
160.1.1.1

A's IP TABLE:
192.68.1.0/24 150.1.1.1

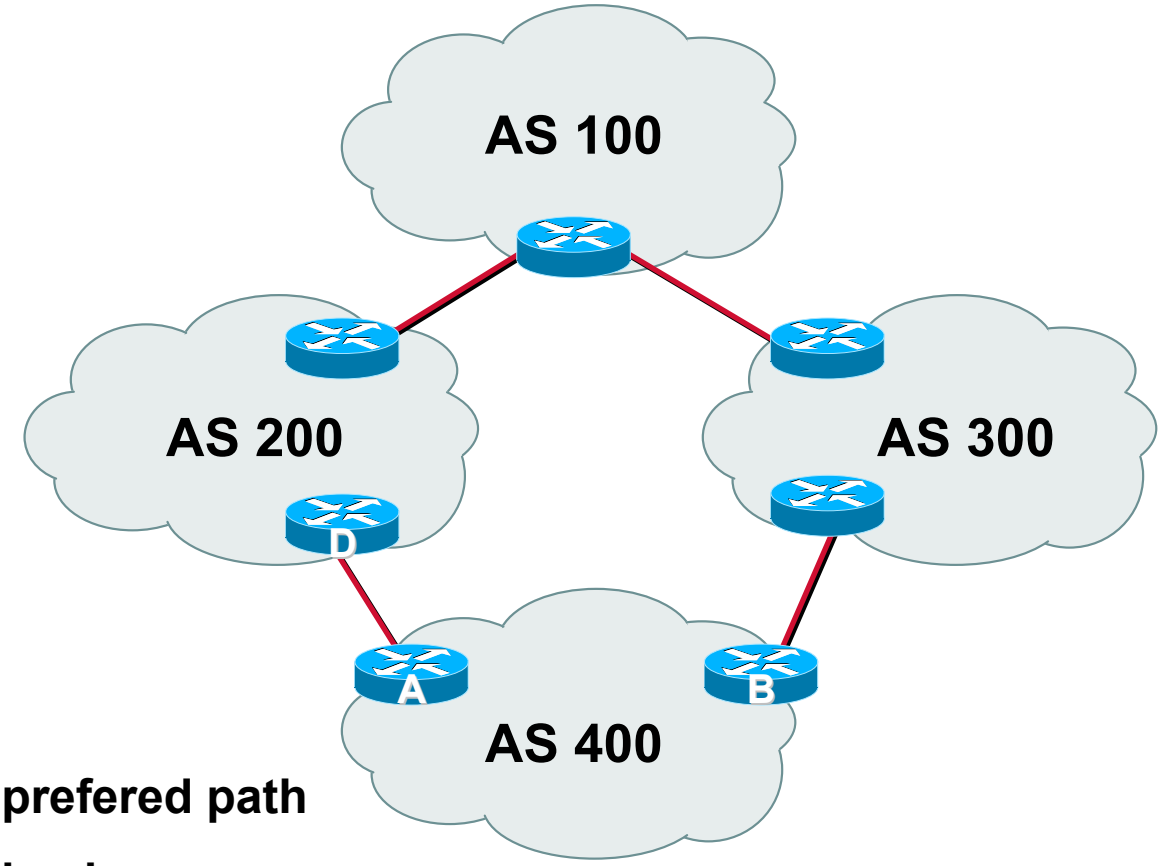
B's IP TABLE:
192.68.1.0/24 160.1.1.1

C's IP TABLE:
Either one depending on IGP metric to nexthop

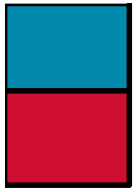


CISCO SYSTEMS

BGP Path Selection—More



- AS 200 prefered path
- AS 300 backup
- Increase AS path length to 300

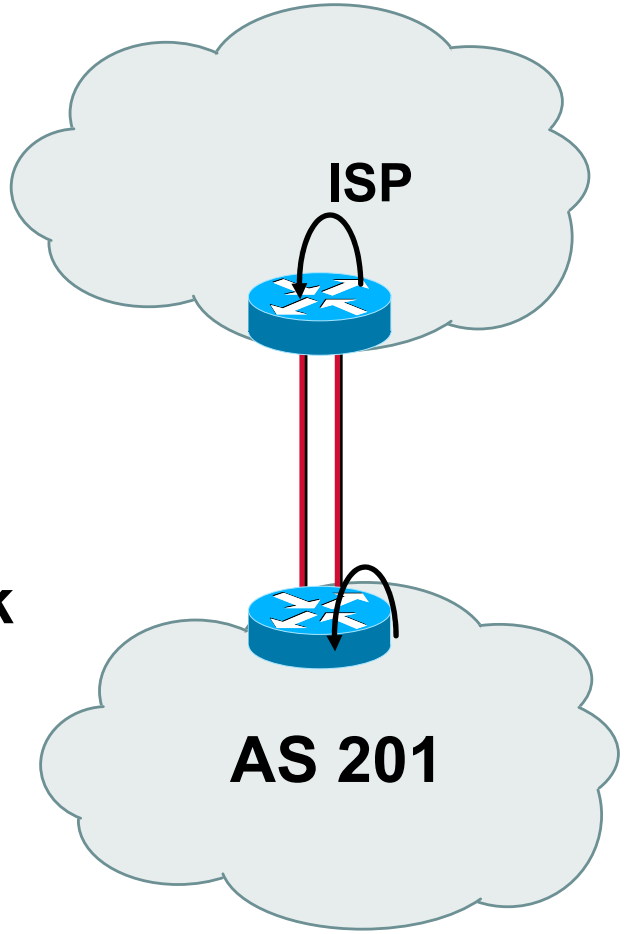


Multi-Homed AS

- **Many situations possible**
  - **Multiple sessions to same ISP**
  - **Secondary for only backup**
  - **Load share between primary and secondary**
  - **Selectively use different ISPs**

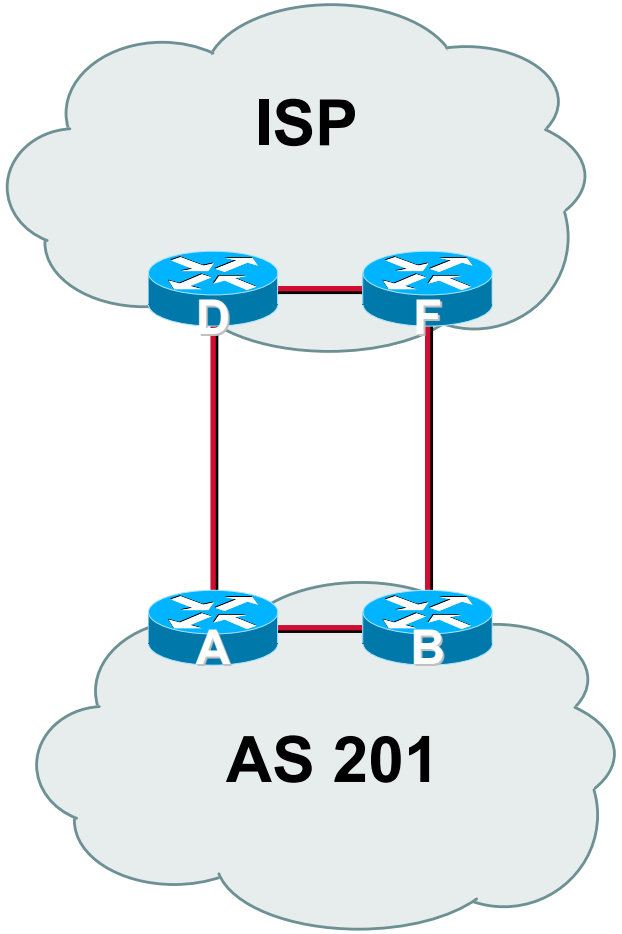
Multiple Sessions to an ISP

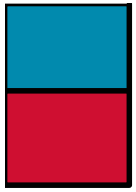
- EBGP to loopback address
- EBGP prefixes learnt with loopback address as nexthop
- Parallel paths to loopback address allows load sharing



Multiple Sessions to an ISP

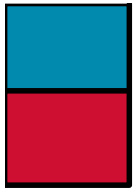
- Simplest scheme is to use defaults
- Learn/advertise prefix for better control





Multiple Session to ISPs

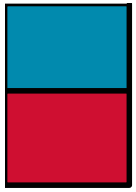
- Difficult to achieve load sharing
- Point default towards one ISP
- Learn selected prefixes from second ISP
- Modify the number of prefixes learnt to acheive acceptable loadsharing



Putting it all together

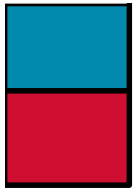
- Your network is going to grow at an exponential rate
- Design to scale... but be prepared to reorganize from scratch
- Don’t be afraid of change!

  - Most network redisigns are only configuration changes



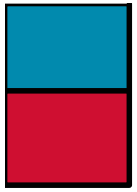
Putting it all together

- Requirements for IGPs for backbones
- IGP connects your backbone together, not your client's routes
- Must
  - converge quickly
- Should
  - carry netmask information



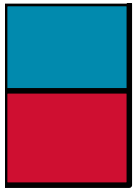
Putting it all together.. Connecting to a customer

- Static routes
  - you control directly
  - no route flaps
- Shared routing protocol or leaking
  - You must filter your customers info
  - route flaps
- BGP for multihomed customers



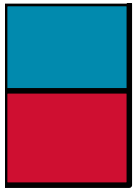
Putting it all together Building your backbone

- Keep it simple
- redundancy is good, but expensive
- use an IGP that carrys mask information
- use an IGP that converges quickly
- use OSPF, ISIS or EIGRP



Putting it all together
Connecting to other ISPs

- Use BGP4
- advertise only what you serve
- take back as little as you can



Putting it all together
The internet exchange

- Long distance connectivity is expensive
- Connect to several providers at a single point

Q&A