IPv4/IPv6 Routing and Deployment Workshop

July 10-14, 2012, Karachi, Pakistan

In conjunction with

APNIC

Presenter

- Champika Wijayatunga
  - Training Unit Manager, APNIC
  - champika@apnic.net

- GZ Kabir
  - Systems Integrator
  - gzkabir@bdcom.com

- Vivek Nigam
  - Internet Resource Analyst, APNIC
  - vivek@apnic.net

Agenda

- Internet Fundamentals
- Internet Resource Management
- IP addressing and IP routing basics
- Introduction to IPv6 and Protocol Architecture
- IPv6 Addressing and Sub-netting
- IPv6 Host Configuration
- IPv4/IPv6 Deployment Plan – Case Study
- IPv4/IPv6 Deployment in IGP – Case Study
- IPv4 to IPv6 Transition Technologies
- IPv4/IPv6 Deployment in EGP – Case Study
- IPv6 DNS

OSPF and BGP

APNIC

What is OSPF Protocol?

OSPF:

- Most commonly used **IP** routing protocol used to do routing within an interior gateway.
- It's an open standard link state routing protocol used by both enterprise and service provider network
- Developed by OSPF working group of IETF
- Specification is defined in RFC1247(v2) RFC 2740 (V3)

Key Features Of OSPF Protocol?

- It's a link state routing protocol
- Designed for TCP/IP protocol environment
- Supports variable length subnet mask
- Fast convergence
- Supports hierarchical architecture and can scale up to very large routing infrastructure (OSPF Area)
- Incremental updates & route authentication

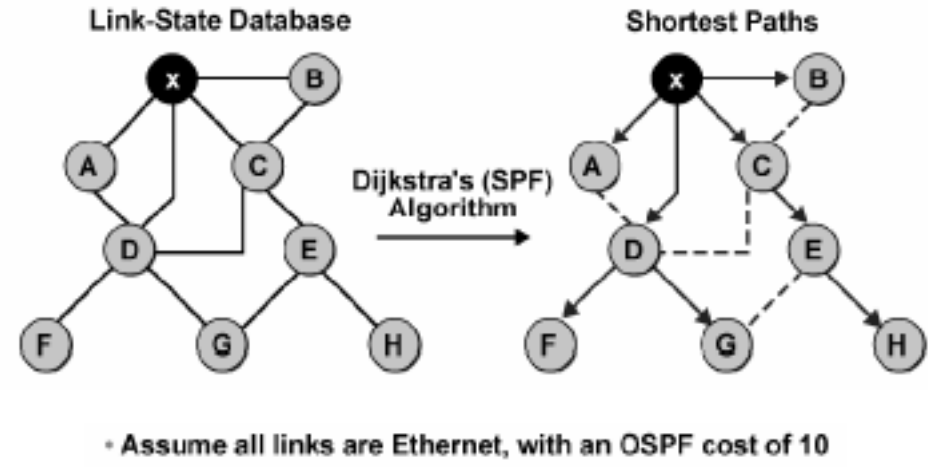
What is Link State Routing

- Do not send full routing table on periodic interval
- Maintain three tables to collect routing information
  - Neighbor table
  - Topology Table
  - Routing table
- Use Shortest Path First (SPF) algorithm to select best path from topology table
- Send very small periodic (Hello) message to maintain link condition
- Send triggered update instantly when network change occur

Link State Data Structure

- Neighbor Table
  - List of all recognized neighboring router to whom routing information will be interchanged
- Topology Table
  - Also called LSDB which maintain list of routers and their link information i.e network destination, prefix length, link cost etc
- Routing table
  - Also called forwarding table contain only the best path to forward data traffic

Shortest Path First (SPF) Tree



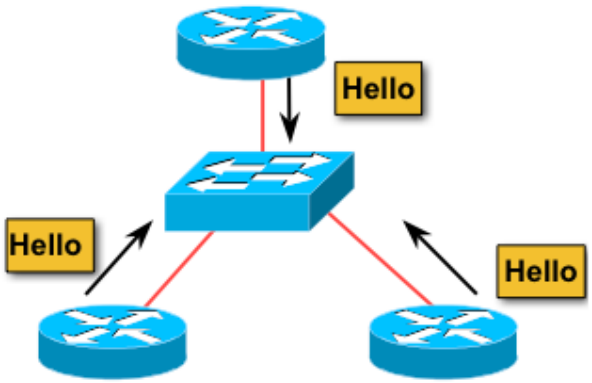
- Every router in an OSPF network maintain an identical topology database
- Router place itself at the root of SPF tree when calculate the best path

Diagram source: www.cisco.com

Basic OSPF Operation

- Neighbor discovery
  - Send L3 multicast message (hello) to discover neighbors
- Exchanging topology table (LSDB)
  - Send L3 multicast message (DBD packets)
- Use SPF algorithm to select best path
  - Each router independently calculates best path from an identical topology database of an OSPF network or area
- Building up routing table
  - All the SPF selected best paths are installed in routing table for the traffic to be forwarded

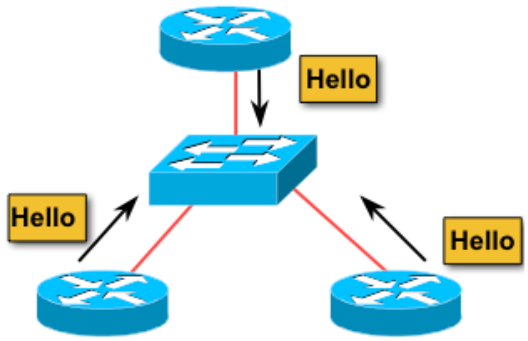
OSPF Neighbor Discovery Process



- Use IP packet to send hello message. At start routers are at OSPF **Down State**
- Use multicast address 224.0.0.5/FF02::5 to make sure single IP packet will be forwarded to every router within OSPF network. Router now at OSPF **Init State**

Diagram source: www.cisco.com

OSPF Neighbor Discovery Process



- All neighboring router with OSPF enabled receive the hello packet
- Checks contents of the hello message and if certain information match it reply (Unicast) to that hello with sending its router ID in the neighbor list.
- This is OSPF **Two-way State**

Contents Of A Hello Packet

- Required information to build up adjacency:
  - Router ID of sending router
  - Hello and dead interval time *
  - List of neighbors
  - Router priority
  - Area ID *
  - DR & BDR IP
  - Authentication information (If any) *

* Need to match to create neighbor relationship

Discovering Network Information

- After creating 2-way neighbor relationship neighboring routers will start exchanging network related information
- At this stage they will decide who will send network information first. Router with the highest router ID will start sending first. This stage is called OSPF **Exstart Stage**
- Then they will start exchanging link state database. This stage is **Exchange Stage**

Adding Network Information

- When router receive the LSDB it perform following action:
  - Acknowledge the receipt of DBD by sending Ack packet (LSAck)
  - Compare the information it received with the existing DB (if any)
  - If the new DB is more up to date the router send link state request (LSR) for detail information of that link. This is **Loading Stage**
- When all LSR have been satisfied and all routers has an identical LSDB this stage is OSPF **Full Stage**

Maintaining Routing Information

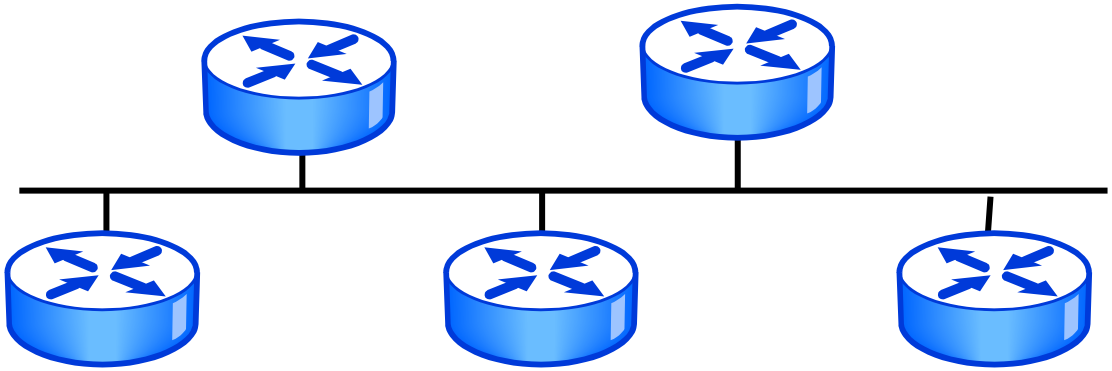
- Send periodic updates (Hello) to all neighbors to make sure link with the neighbor is active. I.e 10 sec for LAN
- Send triggered (Instant) update if any network information changed
- Maintain link state sequence number to make sure all information are up-to-date
- Sequence number is 4-byte number that begins with 0x80000001 to 0x7fffffff

OSPF Packet Types

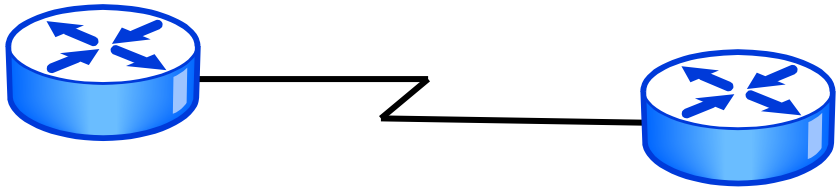
- OSPF use following five packet types to flow routing information between routers:
  - 1: hello [every 10 sec]
    - Hello Builds adjacencies between neighbors
  - 2: DBD [Database Descriptor Packet]
    - DBD for database synchronization between routers
  - 3: LSR [Link State Request Packet]
    - Requests specific link-state records from router to router
  - 4: LSU [Link State Update Packet]
    - Sends specifically requested link-state records
  - 5: LSAck [Link State Ack Packet]
    - Acknowledges the above packet types

OSPF Network Topology

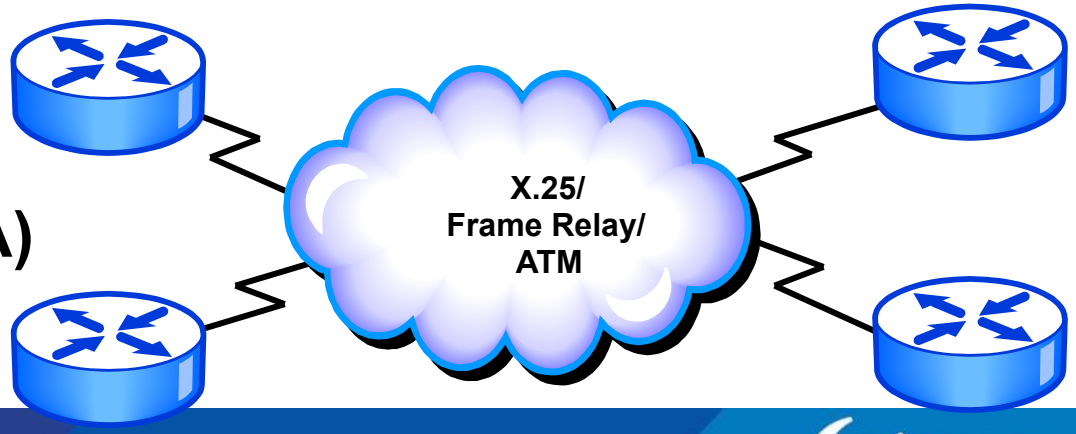
Broadcast Multi-access



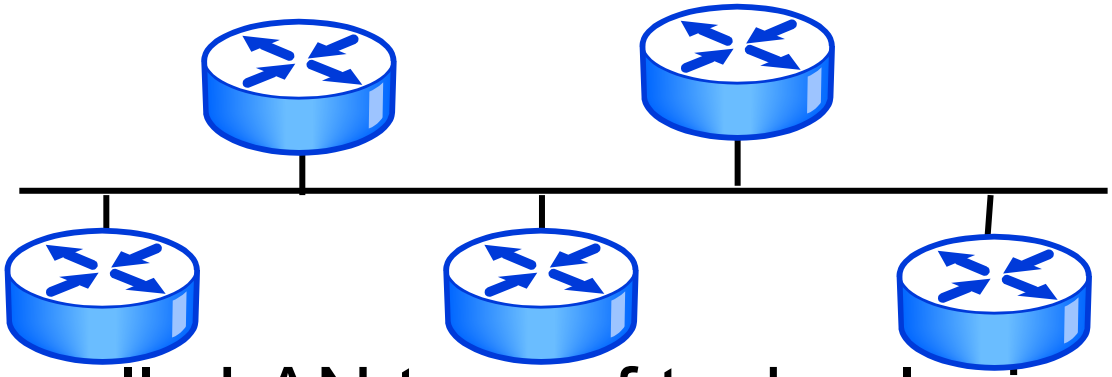
Point-to-Point



Non Broadcast Multi-access (NBMA)



Broadcast Multi-access Network



- Generally LAN type of technologies like Ethernet or Token Ring
- Neighbor relationship are created automatically
- DR/BDR election is required
- Default OSPF hello is 10 sec dead interval is 40 sec

Broadcast Multi-access Network

- Broadcast network use flooding process to send routing update
- Broadcast network use DR/BDR concept to reduce routing update traffic in the LAN
- Packet sent to DR/BDR use 224.0.0.6/FF02::6 multicast address
- Packets from DR to all other routers use 224.0.0.5/FF02::5 multicast address
- All neighbor routers form full adjacencies relation with the DR and BDR only

DR/BDR Election Process

- Router with the highest priority value is the DR, Second highest is BDR
- In the event of tie router with the highest IP address on an interface become DR and second highest is BDR
- DR/BDR election can be manipulated by using router-ID command.
- In practice loopback IP address is used as router ID and the highest IP address will become DR, Second highest is BDR

Questions?

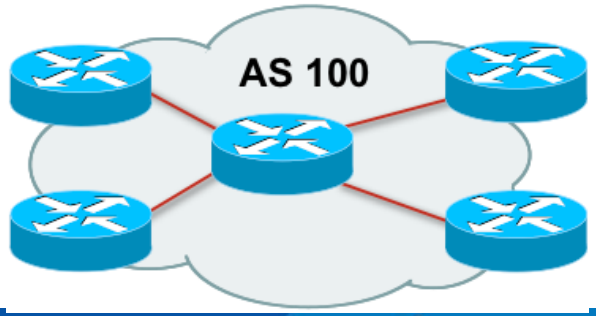
What is Border Gateway Protocol?

BGP:

- A path vector routing protocol to exchange routing information between different Autonomous System (AS)
- ASes are the building block of BGP operational unites
- AS is a collection of routers with a common routing policy
- Specification is defined in RFC4271

What is an Autonomous System (AS)

- An AS is a collection of networks with same routing policy
- Usually under a single administrative control unit
- A public AS is identified by a unique number called AS number
- Around 41442 ASes are visible on the Internet now



BGP features

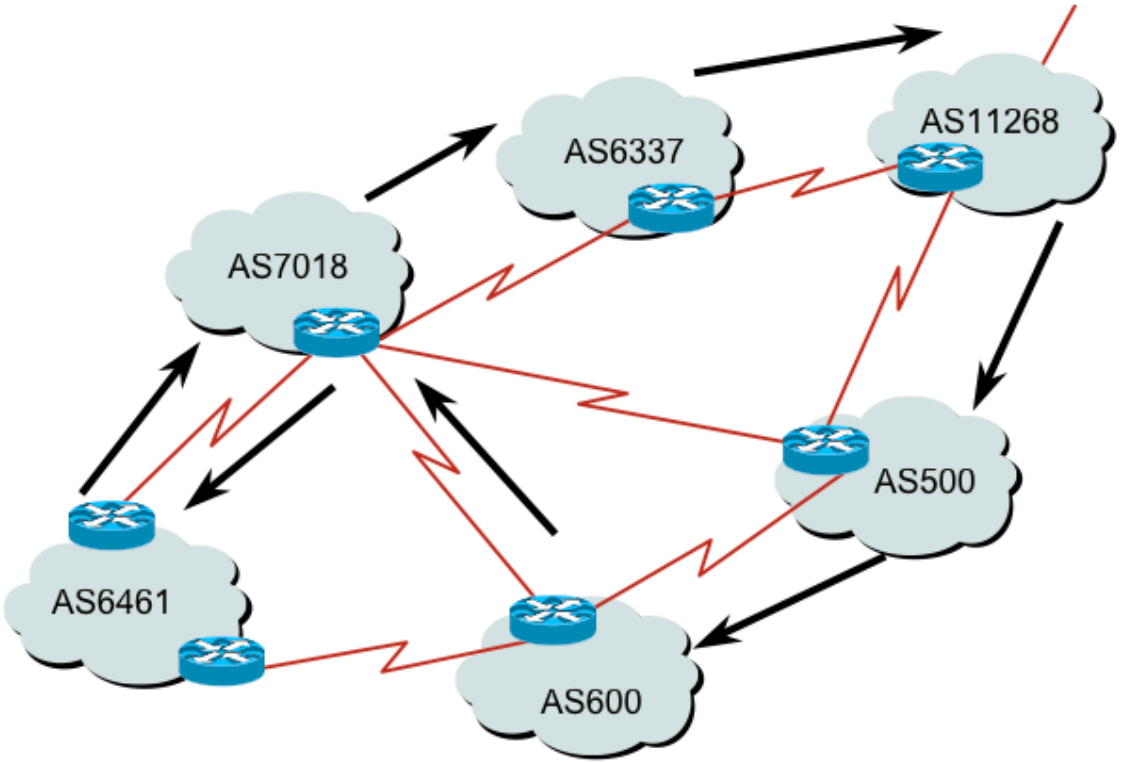
- Path Vector Routing Protocol
- Send incremental updates to peers
- Runs over TCP –Port 179
- Select path based on routing policy/ organization's business requirement
- Support Classless Inter Domain Routing (CIDR) concept
- Widely used in today's Internet Backbone
- Current BGP version is MP-BGP

What is Path Vector Routing Protocol

- A path vector routing protocol is used to span different autonomous systems
- It defines a route as a collection of a number of AS that it passes through from source AS to destination AS
- This list of ASes are called AS path and used to avoid routing loop
- AS path is also used to select path to destination

What is AS path?

An AS path example:




```
12.6.126.0/24 207.126.96.43  1021  0 6461 7018 6337 11268 i
```

AS Path

Source: www.cisco.com

BGP Traffic Arrangement Definition

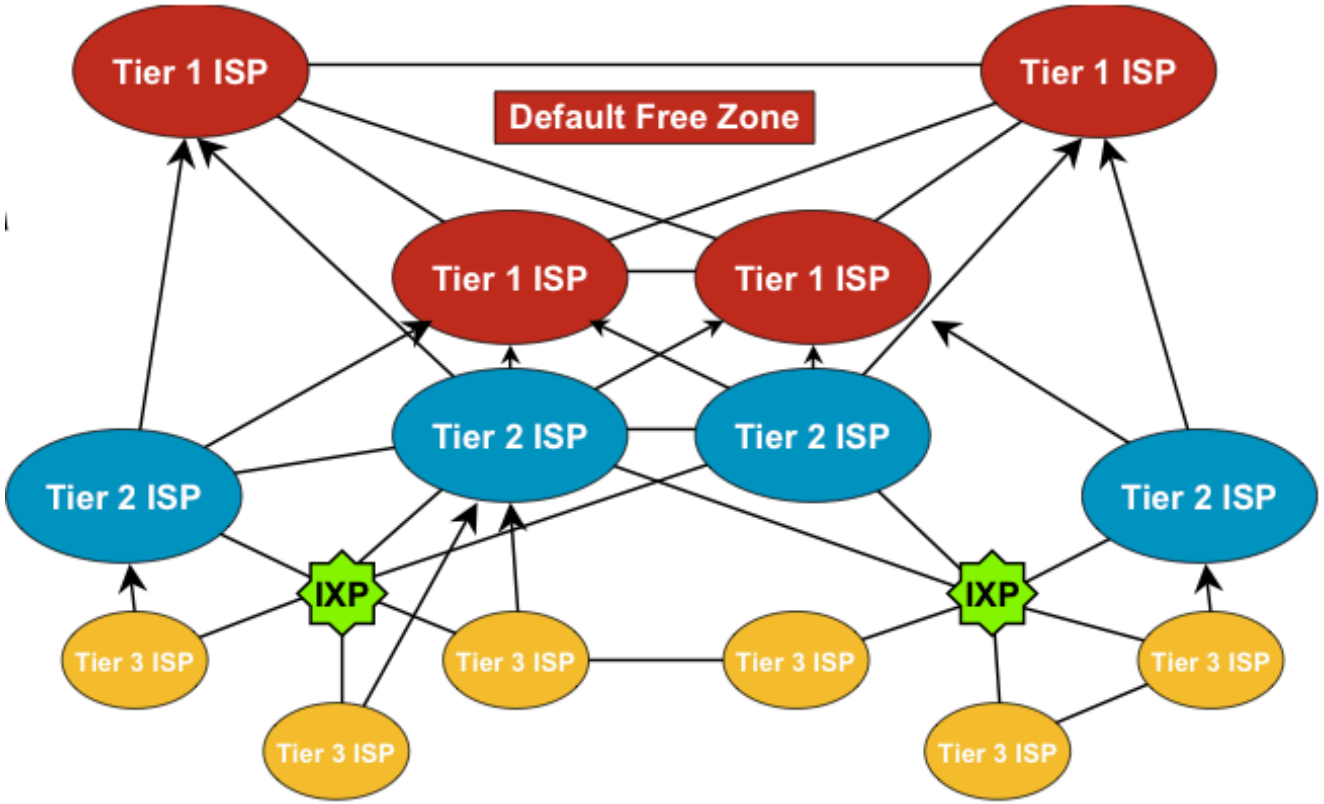
- Transit
  - Forwarding traffic through the network usually for a fee
  - I.e Internet service from upstream ISP
- Peering
  - Exchanging traffic without any fee
  - I.e Connection in an IXP
- Default
  - Where to send traffic if there is no explicit route match in the routing table

What is Default Free Zone?

- Default free zone is made up of Tire One ISP routers which have explicit routing information about every part of the Global Internet
- So there is no need of default route
- If there is no destination network match, then that prefix is still not announced/ used by any ISP yet

ISP Hireracial Connection

Connectivity Diagram:



Source: www.cisco.com

BGP General Operation

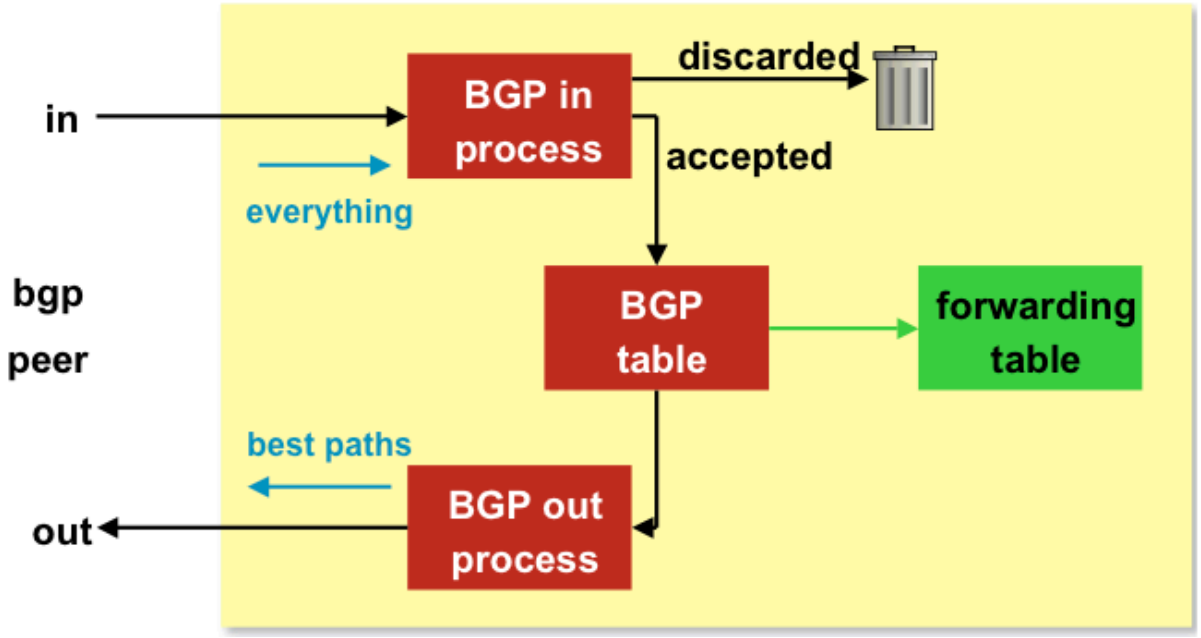
- BGP maintain 3 database i.e Neighbor Table, BGP Table and Forwarding Table
- Learns multiple paths via internal and external BGP speakers
- Picks the best path and installs them on the forwarding tables
- Best path is sent to external BGP neighbors
- Policies are applied by influencing the best path 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
  - Announce “best path” information to peers
- Best path installed in forwarding table if:
  - Prefix and prefix length are equal
  - Lowest protocol distance

Constructing the Forwarding Table

Flowchart:



Source: www.cisco.com

BGP Terminology

- Neighbor
  - Any two routers that have formed a TCP connection to exchange BGP routing information are called peers or neighbors
- iBGP
  - iBGP refers to the BGP neighbor relationship within the same AS.
  - The neighbors do not have to be directly connected.
- eBGP
  - When BGP neighbor relationship are formed between two peers belongs to different AS are called eBGP.
  - EBGP neighbors by default need to be directly connected.

Building Neighbor Relationship

- After adding BGP neighbor:
  - Both router establish a TCP connection and send open message
  - If open message is accepted then both send keepalive message to each other to confirm open message
  - After both confirm open message by sending keepalive message they establish BGP neighbor relationship and exchange routing information

BGP message type

- Open Message
  - To establish BGP neighbor relationship
- Keepalive message
  - Only contain message header to maintain neighbor relationship. Sent every periodic interval
- Update message
  - Contain path information. One update message contain one path information. Multiple path need multiple update message to be sent
- Notification message
  - Sent when an error condition occur and BGP connection closed immediately

BGP Open message

- Open message contain:
  - BGP Version number
  - AS number of the local router
  - BGP holdtime in second to elapse between the successive keepalive message
  - BGP router ID which is a 32 bit number. Usually an IPv4 address is used as router ID
  - Optional parameters i.e types, length and value encoded. An example optional parameter is session authentication info

BGP Keepalive Message

- Send between BGP peers after every periodic interval (60 Sec)
- It refresh hold timer from expiration (180sec)
- A keepalive message contain only the message header

BGP Update Message

- An update message contain:
  - Withdrawn routes: a list contain address prefix that are withdrawn from service
  - Path attributes: includes AS path, origin code, local pref etc
  - Network-layer reachablity information: includes a list of address prefix reachable by this path

BGP Notification message

- Only sent when an error condition occur and detected in a network and BGP connection is closed immediately
- Notification message contain an error code, an error subcode, and data that are related to that error

BGP Neighbor Relationship States

- BGP neighbor goes through following steps:
  - **Idle:** Router is searching its routing table to reach the neighbor
  - **Connect:** Router found route and completed TCP three-way handshake
  - **Open Sent:** Open message sent with the parameter for BGP session
  - **Open Confirm:** Router receive agreement on the parameter to establish BGP session
  - **Established:** Peering is established and routing information exchange began

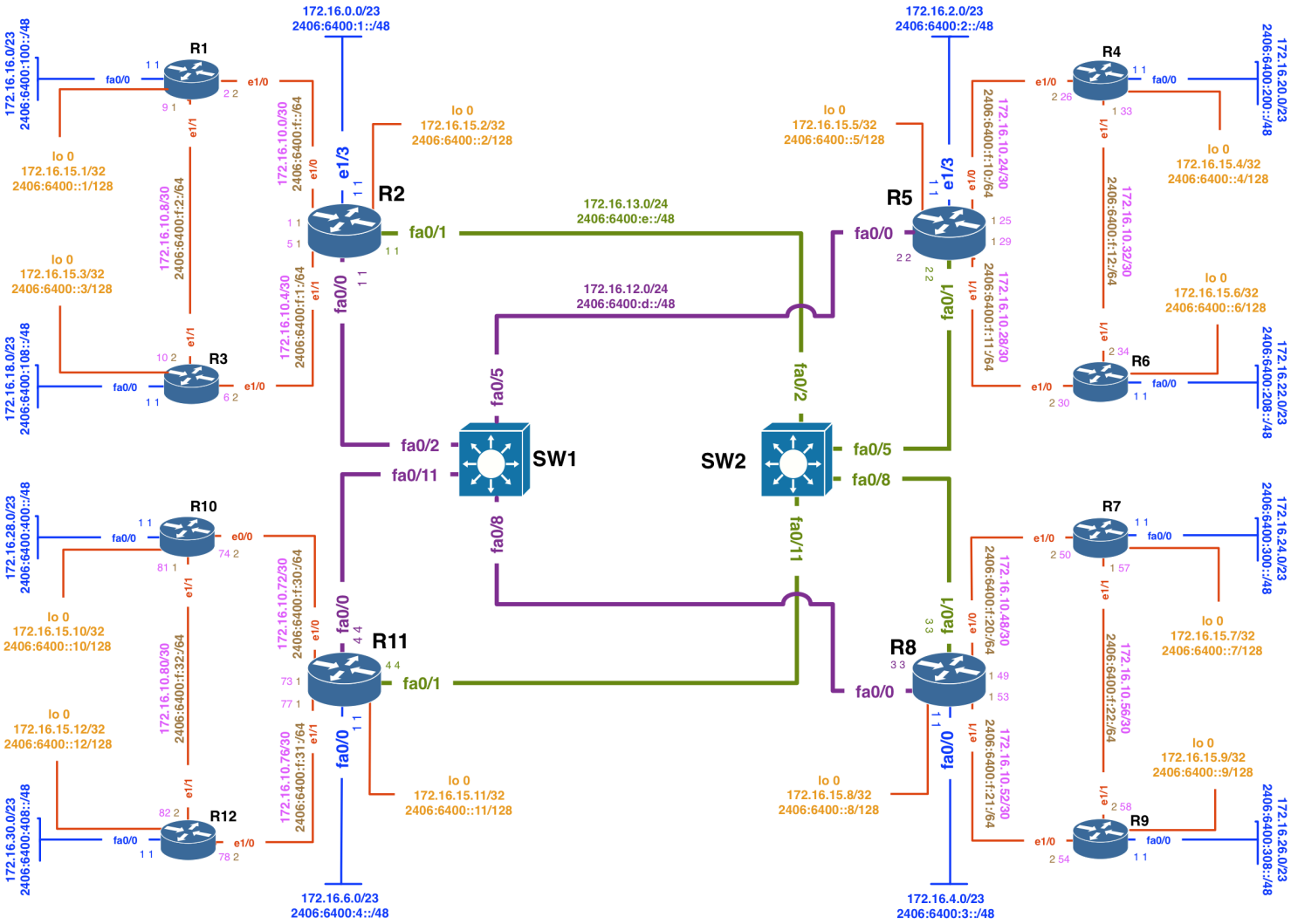
Troubleshoot BGP Neighbor Relation

- Idle:
  - The router can not find address of the neighbor in its routing table
- Active:
  - Router found address of the neighbor in its routing table sent open message and waiting for the response from the neighbor
- Cycle between Active/Idle
  - Neighbor might peer with wrong address
  - Does not have neighbor statement on the other side
  - BGP open message source IP address does not match with remote side neighbor statement or no route to source IP address

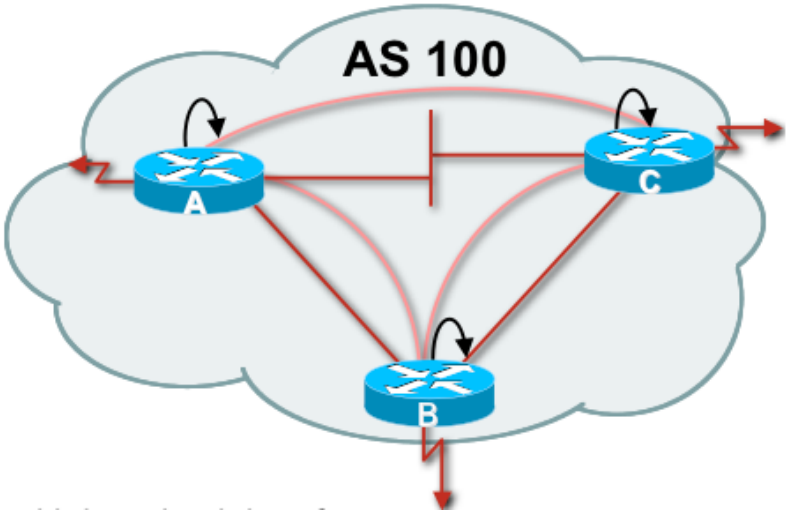
iBGP Peering

- BGP peer within the same AS
- Not required to be directly connected
- iBGP peering require full mesh peering
  - Within an AS all iBGP speaker must peer with other iBGP speaker
  - They originate connected network
  - Pass on prefixes learned from outside AS
  - They do not forward prefixes learned from other iBGP peer

Training ISP IPV6 Addressing Pan



iBGP Peering with Loopback Interface



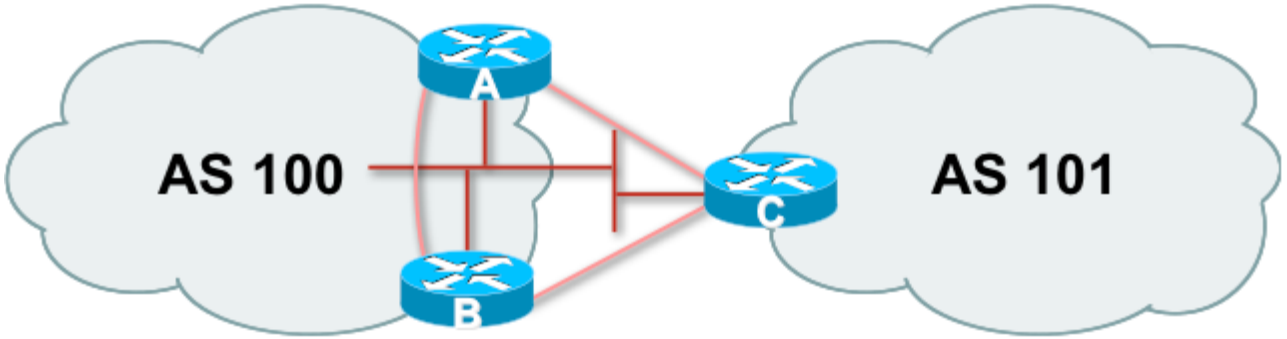
- If iBGP speakers has multiple connection then it is advisable to peer with loopback
- Connected network can go down which might loose iBGP peering
- Loopback interface will never go down

Diagram source: www.cisco.com

iBGP Neighbor Update Source

- This command allows the BGP process to use the IP address of a specified interface as the source IP address of all BGP updates to that neighbor
- A loopback interface is usually used as it will never goes down as long as the router is operational
- All BGP message will use the referenced interface as source of the messages

eBGP Peering



- Peering with BGP speaker in different AS
- Peers should be directly connected and share same WAN link
- eBGP neighbors are usually routed through connected network

Diagram source: www.cisco.com

BGP Next Hop Behavior

- BGP is an AS-by-AS routing protocol not a router-by router routing protocol.
- In BGP, the next hop does not mean the next router it means the IP address to reach the next AS
  - I.e router A advertise 150.10.0.0/16 and 160.10.0.0/16 to router B in eBGP with next hop 150.10.1.1
  - Router B will update Router C in iBGP keepi the next hop unchanged

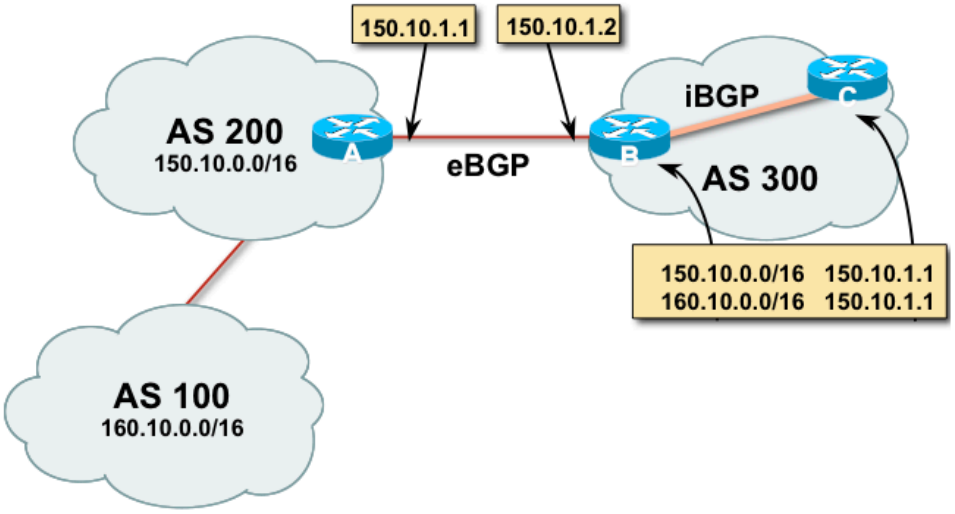
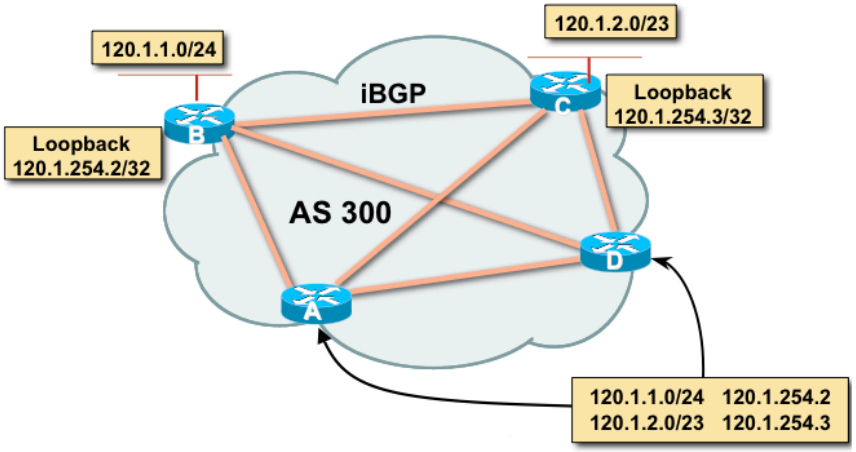


Diagram source: www.cisco.com

iBGP Next Hop



- Next hop is iBGP router loopback address
- Recursive route look-up
- Loopback address need to announce through IGP (OSPF)

Diagram source: www.cisco.com

BGP Synchronous Rule

- BGP do not use or advertise any route to an external neighbor learned by iBGP until a matching route has been learned from an IGP i.e OSPF or static
- It ensure consistency of information throughout the AS
- Avoid black hole route within an AS
- It is safe to turn off if all routers with in the AS run full-mesh iBGP
- Advisable to disable this feature (BCP)

BGP Attributes

BGP metrics are called path attributes. Here is the classifications BGP attributes:

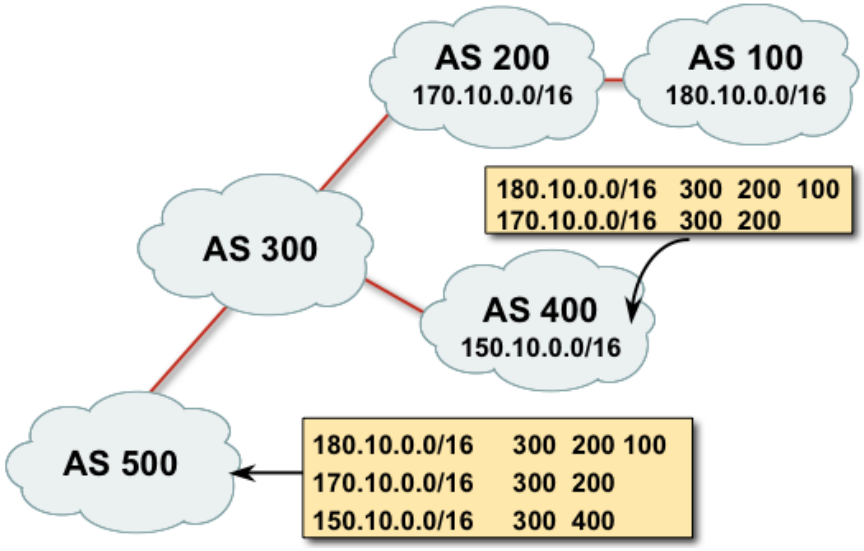
Well-known mandatory
- AS-Path
- Next-hop
- Origin

Well-known discretionary
- Local preference
- Atomic aggregate

Optional transitive
- Community
- Aggregator

Optional non-transitive
- Multi-exit-discriminator (MED)

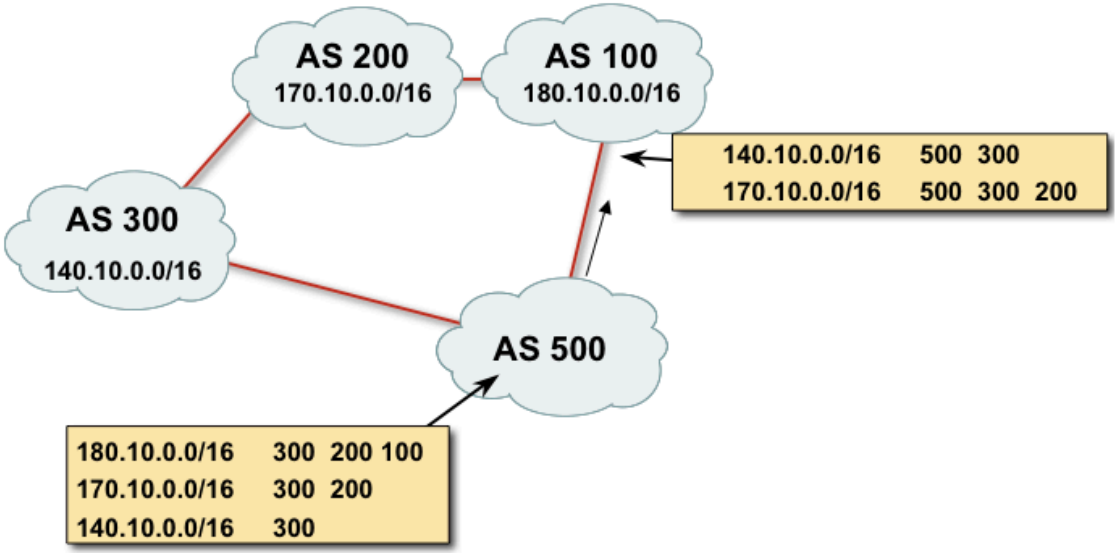
AS Path Attribute



- Sequence of ASes a route has traversed
- Used for
  - Loop detection
  - Path metrics where the length of the AS Path is used as in path selection

Source: www.cisco.com

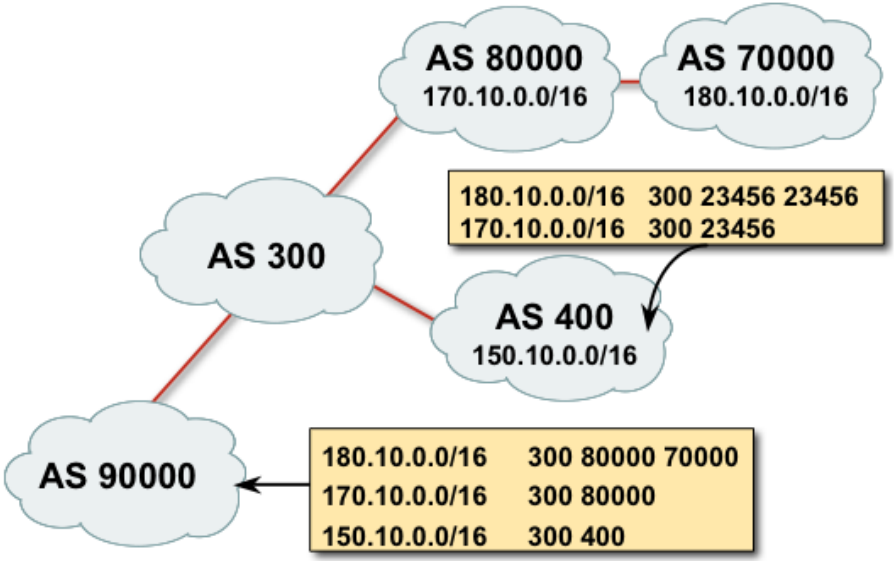
AS Path Loop Detection



- 180.10.0.0/16 is not accepted by AS100 as the prefix has AS100 in its AS-PATH
- This is loop detection in action

Source: www.cisco.com

AS Path Attribute (2 byte and 4 byte)



- Internet with 16-bit and 32-bit ASNs
  - 32-bit ASNs are 65536 and above
  - AS-PATH length maintained

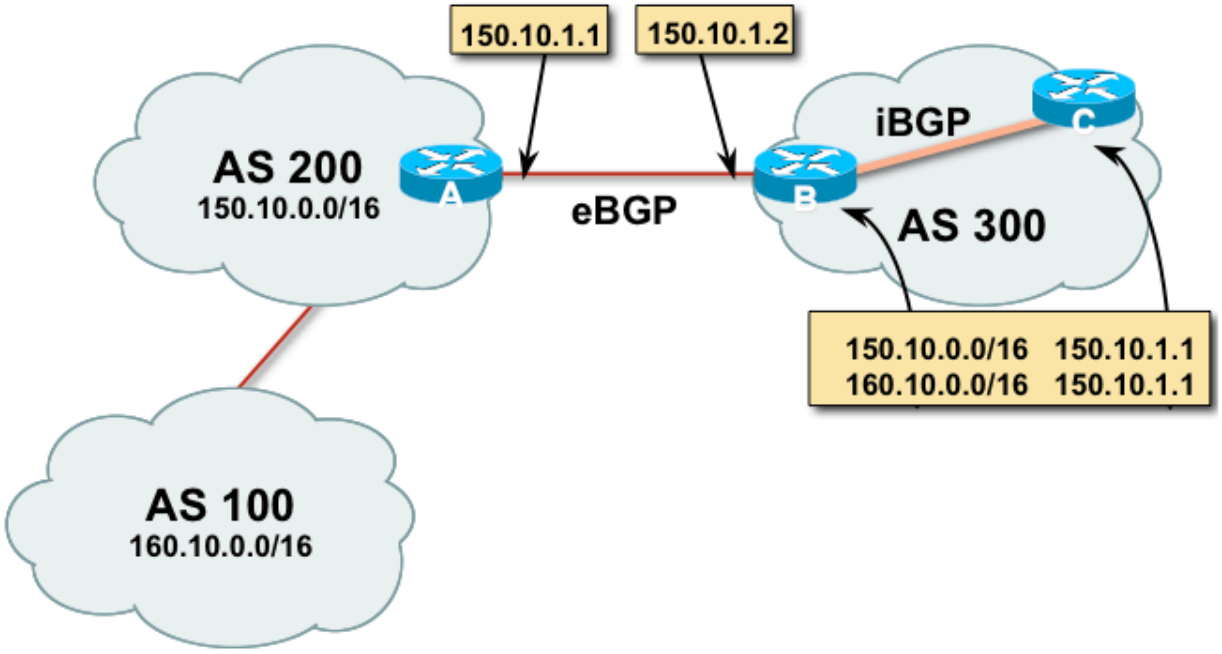
Source: www.cisco.com

AS Path and AS4 Path Example

Router5:

```
Network  Next Hop Metric LocPrf Weight Path
*> 2001::/32 2406:6400:F:41::1 0 23456 38610 6939 i
i  2406:6400:D::5 0 100   0 45192 4608 4826 6939 I
*> 2001:200::/32    2406:6400:F:41::1 0 23456 38610 6939
  2500 i
* i   2406:6400:D::5 0 100 0 45192 4608 4826 6939 2500 i
```

eBGP Next Hop



- T
  - Router A advertise 150.10.0.0/16 and 160.10.0.0/16 to router B in eBGP with next hop 150.10.1.1 (Change it to own IP)
  - Router B will update Router C in iBGP keeping the next hop unchanged
- Well known mandatory attribute

Next Hop Best Practice

- IOS default is for external next-hop to be propagated unchanged to iBGP peers
  - This means that IGP has to carry external next-hops
  - Forgetting means external network is invisible
  - With many eBGP peers, it is unnecessary extra load on IGP
- ISP Best Practice is to change external next-hop to be that of the local router
  - neighbor x.x.x.x next-hop-self

Next Hop Self Configuration

- Next hop default behavior can be changed by using next-hop-self command
- Forces all updates for this neighbor to be advertised with this router as the next hop
- The IP address used for next-hop-self will be the same as the source IP address of the BGP packet

BGP Origin Attribute

- The origin attribute informs all autonomous systems how the prefix introduced into BGP
- Well known mandatory attribute
- Three values: IGP, EGP, incomplete
  - IGP generated by BGP network statement
  - EGP generated by EGP
  - Incomplete redistributed from another routing protocol

BGP Origin Attribute Example

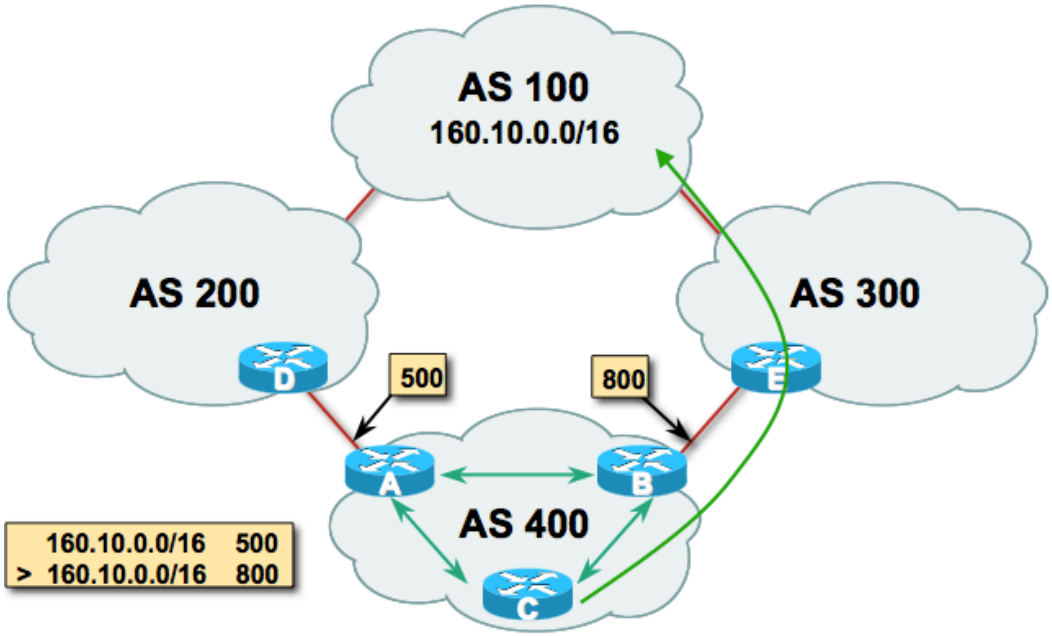
```
Status codes: s suppressed, d damped, h history, *
  valid, > best, i - internal,
                r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

Network Next Hop Metric LocPrf Weight Path
*> 2001::/32  2406:6400:F:41::1 0 23456 38610 6939 i
* i         2406:6400:D::5  0 100 0 45192 4608 4826
  6939 i
```

BGP Local Preference Attribute

- Local preference is used to advertise to IBGP neighbors only about how to leave their AS (Outbound Traffic).
- Paths with highest preference value are most desirable
- Local preference attribute is well-known and discretionary and is passed only within the AS
- Cisco Default Local Pref is 100
-

BGP Local Preference Attribute



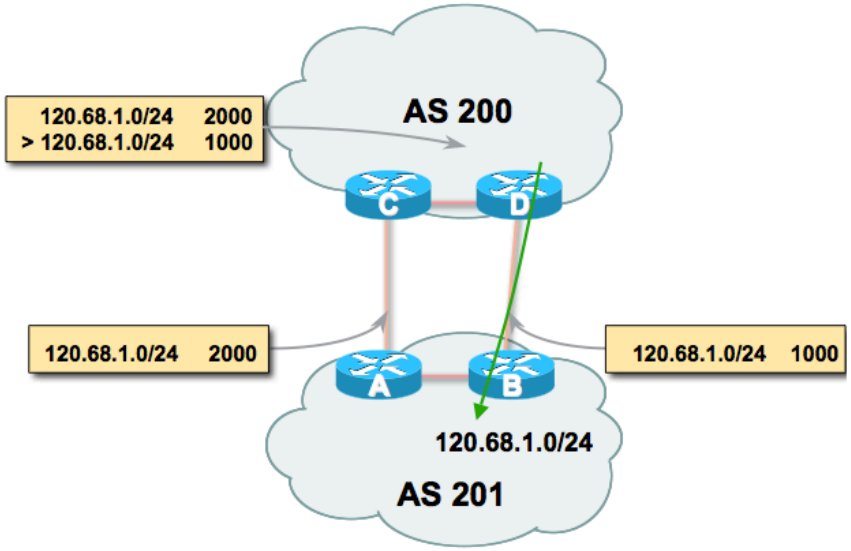
- For destination 160.10.0.0/16 Router A advertise local pref 500 and Router B advertise local pref 800 in iBGP
- 800 will win best path (Router B)

Diagram source: www.cisco.com

BGP MED Attribute

- MED is used to advertise to EBGP neighbors about how to exit their AS to reach networks owned by this AS (Incoming traffic).
- MED is sent to EBGP neighbors only.
- The paths with the lowest MED value are the most desirable
- The MED attribute is optional and non transitive

BGP MED Attribute



- For prefix 120.68.1.0/24 Router B send MED 1000 and router A send MED 2000 to eBGP neighbor
- Incoming traffic from AS200 will choose Router B since lowest MED will win

Diagram source: www.cisco.com

BGP Community Attribute

- Community is a tagging technique to mark a set of routes
- Upstream service provider routers can then use these flags to apply specific routing polices (i.e local preference etc) within their network
- Represented as two 16 bit integers (RFC1998)
- Common format is <local-ASN>:xx
- I.e 0:0 to 0:65535 and 65535:0 to 65535:65535 are reserved
- Very useful in applying policies within and between ASes
- Optional & transitive attribute

BGP Route Selection Process

- Step 1: Prefer highest weight (local to router)
- Step 2: Prefer highest local preference (global within AS)
- Step 3: Prefer route originated by the local router
- Step 4: Prefer shortest AS path
- Step 5: Prefer lowest origin code (IGP < EGP < incomplete)
- Step 6: Prefer lowest MED (from other AS)
- Step 7: Prefer EBGP path over IBGP path
- Step 8: Prefer the path through the closest IGP neighbor
- Step 9: Prefer oldest route for EBGP paths
- Step 10: Prefer the path with the lowest neighbor BGP router ID

Questions?