



BGP For The Enterprise



Keith Bogart

Cisco CCIE #4923



Course Objectives

- + Understand the differences between BGP and IGPs
- + Configure, monitor and troubleshoot BGP peering and prefix exchange
- + Control BGP path selection
- + Configure BGP Filtering & Summarization
- + Explain BGP Security Mechanisms

+ Course Prerequisites

- **An understanding of IPv4 addressing & subnetting**
- **Familiarity with Cisco IOS CLI**
- **Familiarity with Wireshark**



Let's Get Started!



Introduction To BGP



Intro to the Border Gateway Protocol

- + Before one can begin to understand BGP you know understand some of its terminology
 - + Autonomous Systems
 - + Peers
 - + Prefix & NLRI
- + BGP is an Exterior Gateway Protocol (EGP)
- + BGP's main strengths are:
 - + Scalability
 - + Flexibility
- + BGP's main weakness is convergence



Comparison Between IGPs and BGP

+ Similarities with IGPs

- + BGP needs to form neighborhood (eg, "peers") like IGPs.
- + BGP needs to advertise prefixes, just like IGPs.
- + BGP also advertises Next Hops for those prefixes.

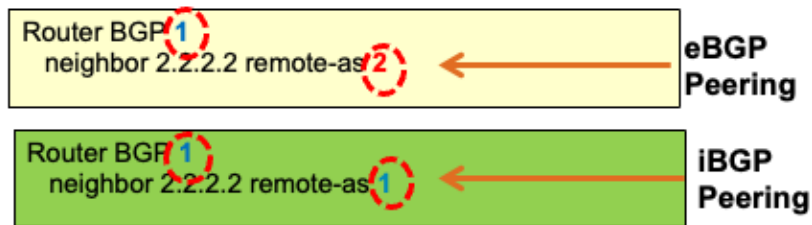
+ Differences than IGPs

- + Neighbor IP address may not be on a common subnet for BGP.
- + BGP uses **TCP (179)** and unicast...IGPs do not.
- + IGPs describe a route using a single numeric value (metric) whereas BGP associates several descriptive characteristics with each prefix



BGP Peer Types

- + There are two types of neighbors in BGP: internal BGP (iBGP) and external BGP (eBGP).
- + A BGP router behaves differently in several ways depending on whether the peer (neighbor) is an iBGP or eBGP peer.



Overview of iBGP & eBGP Differences

- + Peer establishment
 - + eBGP imposes certain rules/restrictions not imposed by iBGP
- + Prefix exchange
 - + BGP updates received from external peers can be forwarded to any other type of peer.
 - + BGP updates received from internal peers can ONLY be forwarded to external peers.
- + Update modification
 - + Certain BGP Path Attributes have forwarding restrictions







Autonomous Systems



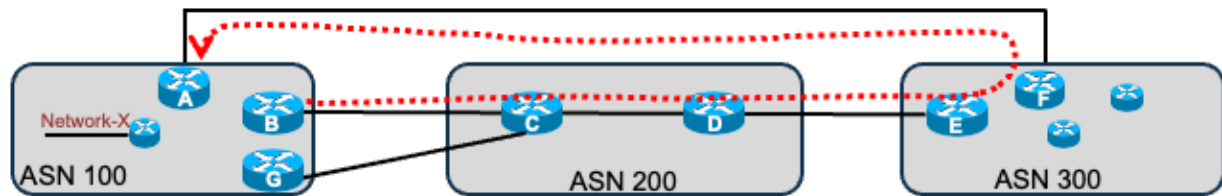
Overview of BGP Autonomous Systems

- + BGP messages include the ASN of the advertising router
- + ASNs are provided to an organization in one of three ways:
 - + Directly from their local Regional Internet Registry (RIR)
 - + From an Internet Service Provider (ISP)
 - + Local configuration (when BGP will only be used internally)
- + Originally, ASNs were 16-bit values
- + In 1995 the IANA allocated a block of 32-bit ASN values to RIRs
- + All ASN values assigned since 2006 have been 32-bit values



Usage of BGP ASNs

- + BGP uses autonomous system values in several different ways:
 - + Determination of iBGP or eBGP peering
 - + Loop detection and prevention
 - + Best path algorithm for received routes
 - + Attempts at influencing eBGP peers and their path selection



ASN Categories

- + ASNs are divided into three categories;
 - + Reserved ASNs;
 - + For documentation purposes only
 - + For 2-byte to 4-byte ASN transition
 - + Private ASNs – For private, internal use
 - + Public ASNs – For exchanging BGP updates between ASNs
- + Each of the three categories has a range of values in both the 2-byte (old) ASN allocation scheme and current 4-byte allocation scheme

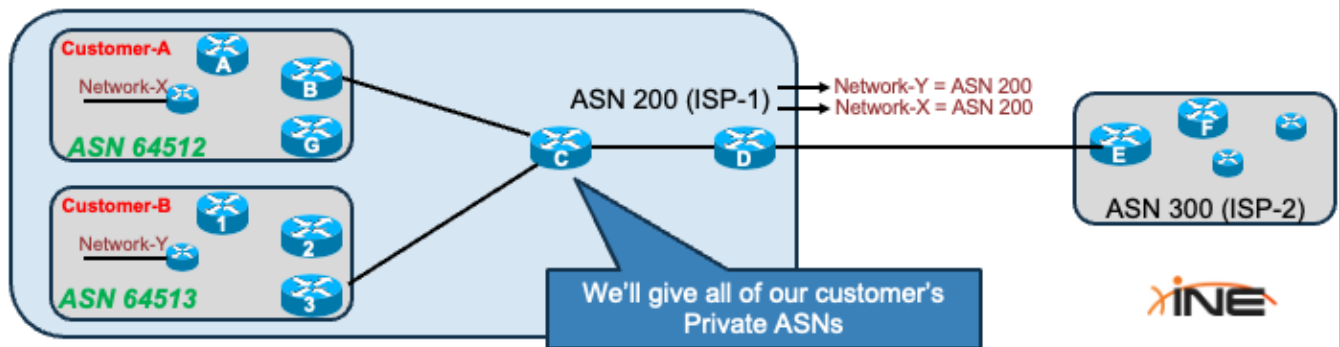
ASN Allocations

- + Reserved ASNs
 - + 24,256 reserved for representing 4-byte ASNs to older (legacy) 2-byte ASN devices
 - + 64496 to 64511 (2-byte ASNs)
 - + 65536 to 65551 (4-byte ASNs)
- + Private ASNs
 - + 64512 to 65534
 - + 65535 is reserved for special use (avoid using if possible).
- + Public ASNs are all other values



Why So Private?

- + There are a few reasons one might choose to utilize Private ASNs
 - + Your ISP tells you to
 - + You want to use BGP internally with no intention of using it with an ISP
 - + You want to incorporate eBGP within your internal ASN and avoid iBGP restrictions



Simple BGP Configuration

- + Initial BGP configuration on a Cisco router involves providing (at minimum) the following information:
 - + Local ASN of the router
 - + IPv4 address of at least one BGP peer
 - + IPv4 reachability to peer's address (if not connected)
 - + ASN value of the BGP peer

```
Router BGP 100  
neighbor 2.2.2.2 remote-as 1.2
```



4-Byte ASN Configuration

- + During initial BGP peering, routers advertise and negotiate the use of 2-byte or 4-byte BGP ASN capability.
- + If both routers support 4-byte ASNs then BGP messages carry ASN within a 32-bit field.
- + Network administrators can configure local 4-byte ASNs in two different ways;
 - + ASPlain (65536 to 4294967295)
 - + ASDot (1.0 to 65535.65535)

$$65538 = \overset{65536}{\boxed{00000000 \ 00000001}} + \overset{2}{\boxed{00000000 \ 00000010}}$$

$65538 = 1.2$



- You might ask, “Why does ASPlain configuration start with ASN 65536?” This is because this was the first 4-byte ASN allocated. All other ASNs prior to this were allocated as older 2-byte ASNs.
- When a 4-byte BGP router is peering with a legacy 2-byte BGP router, the 4-byte ASN is “hidden” and instead the router presents itself as ASN 23456 however all enterprise-level routers and multilayer switches produced over the past decade have been capable of 4-byte ASNs so it’s unlikely you’ll ever run across a situation in which you need to know the details of 2-byte to 4-byte translations.
- When issuing BGP “show” commands on Cisco devices, you may need to use the BGP router command, “bgp asnotation dot” if you want the ASN output in your “show bgp” commands to be represented in ASDot format.





BGP Message Types



BGP Message Types

- + All BGP messages carried within IP/TCP Headers

IP Header		
TCP Header		
Marker (All "Fs") 16-bytes	Length (2-bytes)	Type (1 byte)
BGP Data		

- + BGP uses four types of messages for its operation:
 - + Open
 - + Update
 - + Keepalive
 - + Notification



BGP Open Message

- + BGP Open Message:
 - + Used in Neighbor Establishment
 - + BGP values and capabilities are exchanged.

Marker (All "Fs") 16-bytes		Length (2-bytes)		Type = 1
Version = 4	My AS#	Hold Time	Router-ID	
Optional Parameters Length		BGP Capabilities		



- The default-hold time on Cisco routers is 180-seconds. This is 3x the keepalive interval.

BGP Open Sniffer Trace

```

* Ethernet II, Src: Cisco_c5:f9:00 (00:1b:d4:c5:f9:00), Dst: Cisco_dc:5a:c8 (00:1b:d4:dc:5a:c8)
* Internet Protocol Version 4, Src: 1.1.1.1 (1.1.1.1), Dst: 1.1.1.2 (1.1.1.2)
* Transmission Control Protocol, Src Port: 54748 (54748), Dst Port: bgp (179), Seq: 1, Ack: 1, Len: 57
* Border Gateway Protocol - OPEN Message
  Marker: ffffffffffffffffffffffffffffffff
  Length: 57
  Type: OPEN Message (1)
  Version: 4
  My AS: 1
  Hold Time: 180
  BGP Identifier: 1.1.1.1 (1.1.1.1)
  Optional Parameters Length: 28
* Optional Parameters
  * Optional Parameter: Capability
    Parameter Type: Capability (2)
    Parameter Length: 6
  * Capability: Multiprotocol extensions capability
    Type: Multiprotocol extensions capability (1)
    Length: 4
    AFI: IPv4 (1)
    Reserved: 00
    SAFI: Unicast (1)

```



BGP Update Message

- + BGP Update Message:
 - + Informs neighbors about withdrawn routes, changed routes, and new routes.
 - + Used to exchange PAs and the associated prefix/length (NLRI) that use those attributes.

Marker (All "Fs") 16-bytes		Length (2-bytes)	Type = 2
Unfeasible Routes Length	Withdrawn Routes (if any)		
Total Path Attributes Length	Path Attributes (TLV)		
NLRI Prefix Length		NLRI Prefix	

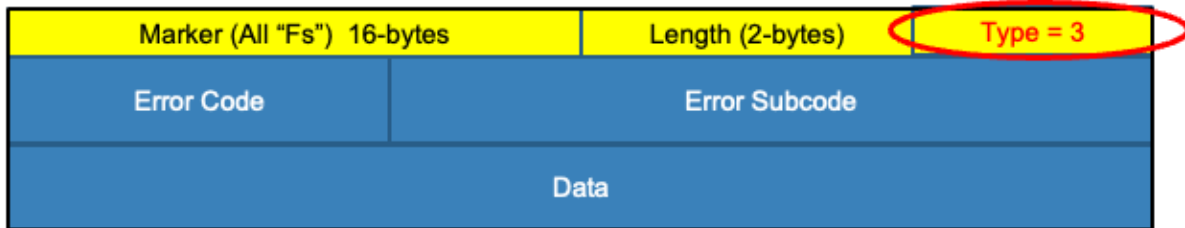
BGP Update Sniffer Trace

```
* Ethernet II, Src: Cisco_c5:f9:00 (00:1b:d4:c5:f9:00), Dst: Cisco_dc:5a:c8 (00:1b:d4:dc:5a:c8)
* Internet Protocol Version 4, Src: 1.1.1.1 (1.1.1.1), Dst: 1.1.1.2 (1.1.1.2)
* Transmission Control Protocol, Src Port: 54748 (54748), Dst Port: bgp (179), Seq: 96, Ack: 77, Len: 77
* Border Gateway Protocol - UPDATE Message
  Marker: ffffffffffffffffffffffffffffffffff
  Length: 54
  Type: UPDATE Message (2)
    Unfeasible routes length: 0 bytes
    Total path attribute length: 27 bytes
  * Path attributes
    * ORIGIN: IGP (4 bytes)
    * AS_PATH: 1 (9 bytes)
    * NEXT_HOP: 1.1.1.1 (7 bytes)
    * MULTI_EXIT_DISC: 0 (7 bytes)
  * Network layer reachability information: 4 bytes
    * 11.11.11.0/24
      NLRI prefix length: 24
      NLRI prefix: 11.11.11.0 (11.11.11.0)
```



BGP Notification Message

- + BGP Notification message:
 - + Used to signal a BGP error; typically results in a reset to the neighbor relationship



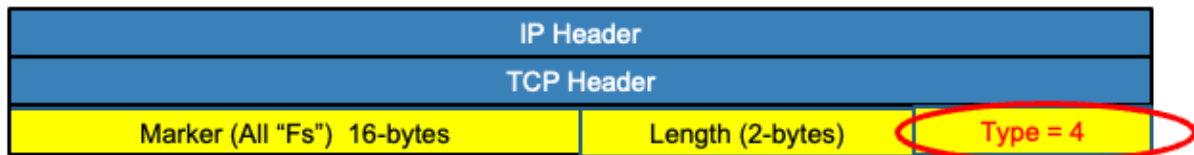
BGP Notification Sniffer Trace

```
» Ethernet II, Src: Cisco_dc:5a:c8 (00:1b:d4:dc:5a:c8), Dst: Cisco_c5:f9:00 (00:1b:d4:c5:f9:00)
» Internet Protocol Version 4, Src: 1.1.1.2 (1.1.1.2), Dst: 1.1.1.1 (1.1.1.1)
» Transmission Control Protocol, Src Port: 15292 (15292), Dst Port: bgp (179), Seq: 66, Ack: 77, Len: 23
» Border Gateway Protocol - NOTIFICATION Message
  Marker: ffffffffffffffffffffffffffffffffff
  Length: 23
  Type: NOTIFICATION Message (3)
  Major error Code: OPEN Message Error (2)
  Minor error Code (Open Message): Bad Peer AS (2)
  Data: 0001
```



BGP Keepalive Message

- + BGP Keepalive message:
 - + Sent on a periodic basis to maintain the neighbor relationship. The lack of receipt of a Keepalive message within the negotiated Hold timer causes BGP to bring down the neighbor connection.
 - + Default interval = Every 60-seconds



- Sent every 60-seconds by default

BGP Keepalive Sniffer Trace

```
■ Ethernet II, Src: Cisco_c5:f9:00 (00:1b:d4:c5:f9:00), Dst: Cisco_dc:5a:c8 (00:1b:d4:dc:5a:c8)
■ Internet Protocol Version 4, Src: 1.1.1.1 (1.1.1.1), Dst: 1.1.1.2 (1.1.1.2)
■ Transmission Control Protocol, Src Port: bgp (179), Dst Port: 15292 (15292), Seq: 58, Ack: 66, Len: 19
■ Border Gateway Protocol - KEEPALIVE Message
  Marker: ffffffffffffffffffffffffffffffffff
  Length: 19
  Type: KEEPALIVE Message (4)
```







BGP Peering



BGP Peering Overview

- + Two types of BGP peers
 - + eBGP peers (peer in different ASN than local ASN)
 - + iBGP peers (peer in same ASN as local ASN)
- + BGP “neighbor” statement is used for both types of peers
- + Very different peering rules apply

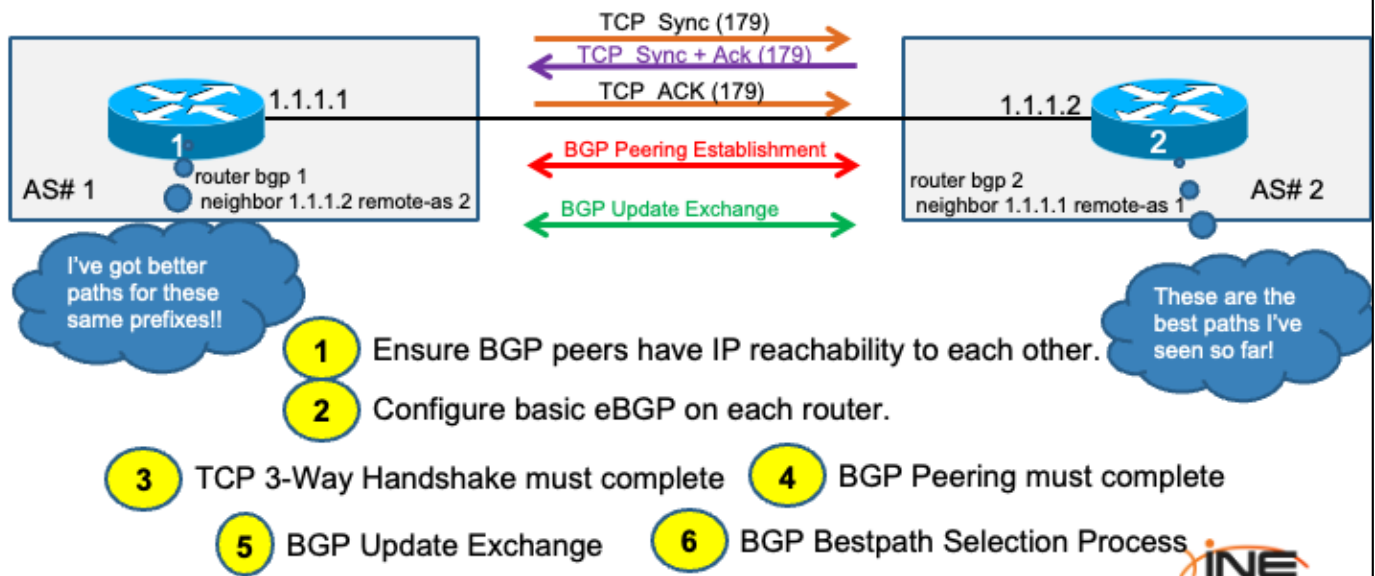


BGP Peering Rules

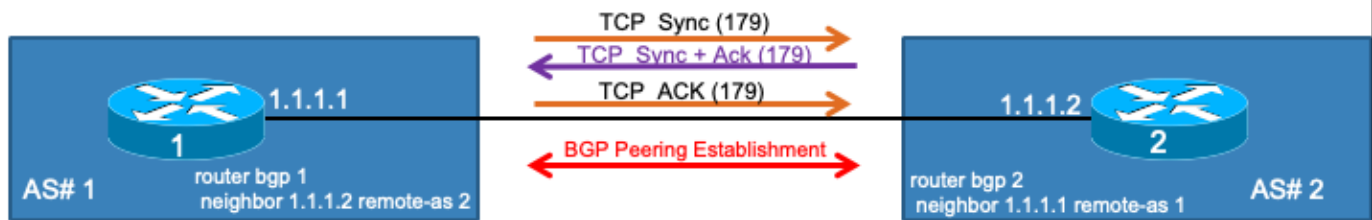
- + eBGP peers are assumed to be directly-connected
 - + Peering will not commence by default if not directly-connected
 - + Default TTL for eBGP packets is "1".
 - + May override this restriction with configuration
- + iBGP peers have no requirement they be directly-connected
 - + If iBGP peers are NOT directly-connected, black-hole routing could occur
- + Both iBGP and eBGP peer addresses must be reachable via an IGP (or directly-connected)
 - + **BGP cannot use itself for peer reachability**



BGP Peering Overview



Peering Sanity Checks



- 1 Source IP address of incoming TCP connection must be from an expected/configured BGP peer.
- 2 Peer's advertisement of his BGP AS# must be what we expect.
- 3 If BGP authentication is used, same password must be configured.
- 4 Peers must have unique BGP Router-IDs
- 5 Peers must use the same BGP version.



- BGP-1: This was the original version of BGP, and it was first released in 1989. BGP-1 was a very simple protocol, and it did not support many of the features that are now considered to be essential for BGP.
- BGP-2: BGP-2 was released in 1991, and it was a significant improvement over BGP-1. BGP-2 added support for a number of new features, including path attributes, which are used to convey information about BGP routes.
- BGP-3: BGP-3 was released in 1994, and it was a minor update to BGP-2. BGP-3 added support for a few new features, such as route aggregation, which allows BGP routers to advertise multiple routes in a single message.
- BGP-4: BGP-4 is the current version of BGP, and it was released in 2006. BGP-4 is a backward-compatible upgrade to BGP-3, and it supports all of the features of BGP-3. BGP-4 also adds support for a number of new features, such as IPv6 routing and multipath routing.

BGP Router-ID

- + Just like any IGP, BGP elects a Router-ID.
- + The BGP router-ID is elected as follows:
 - + Use the setting of the *bgp router-id <x.x.x.x>* router subcommand.
 - + Choose the highest numeric IP address of any up/up loopback interface, at the time the BGP process initializes.
 - + Choose the highest numeric IP address of any up/up non- loopback interface, at the time the BGP process initializes.



BGP Update-Source

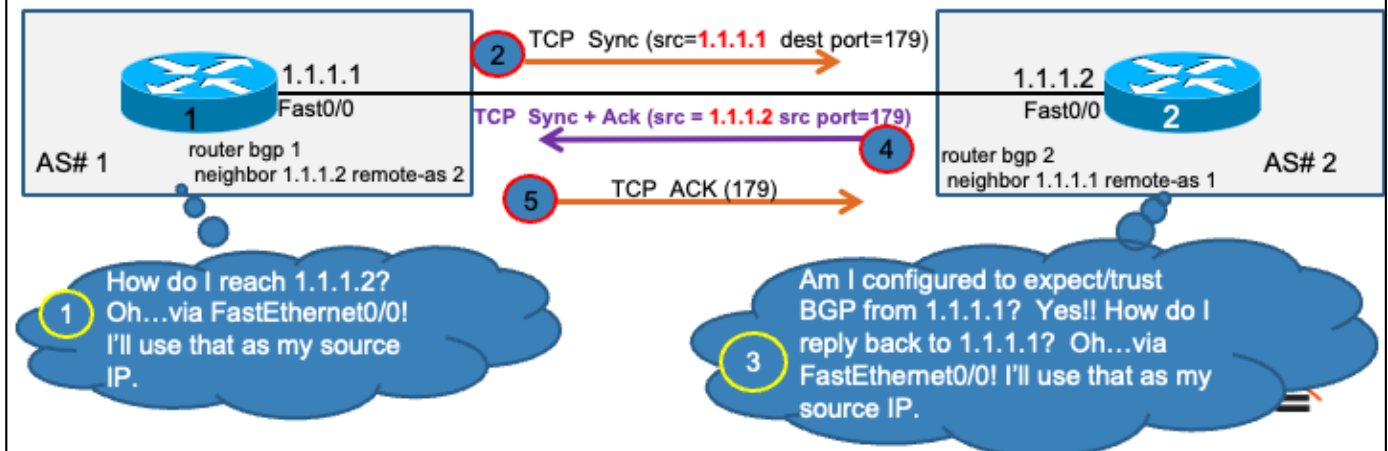
- + TCP Connection must first form between BGP peers.
- + This TCP connection must form before BGP messages flow over this TCP connection.
- + Source IP address used in TCP connection usually must match what your neighbor is expecting from you in his “neighbor” command.
- + The local router tries to form a TCP connection with the IP address defined in the **neighbor remote-as** command.



- With IGPs such as EIGRP and OSPF, a router doesn't care about the source ip address other than verifying that it is from a directly-connected subnet. This is because when another IGP router suddenly multicast/broadcasts, “Hello...here I am!” it is assumed that router is a trusted peer, residing within the same company/Autonomous System.
- -
- BGP does not have this implicit trust. It will only talk to routers it has been TOLD to talk to.

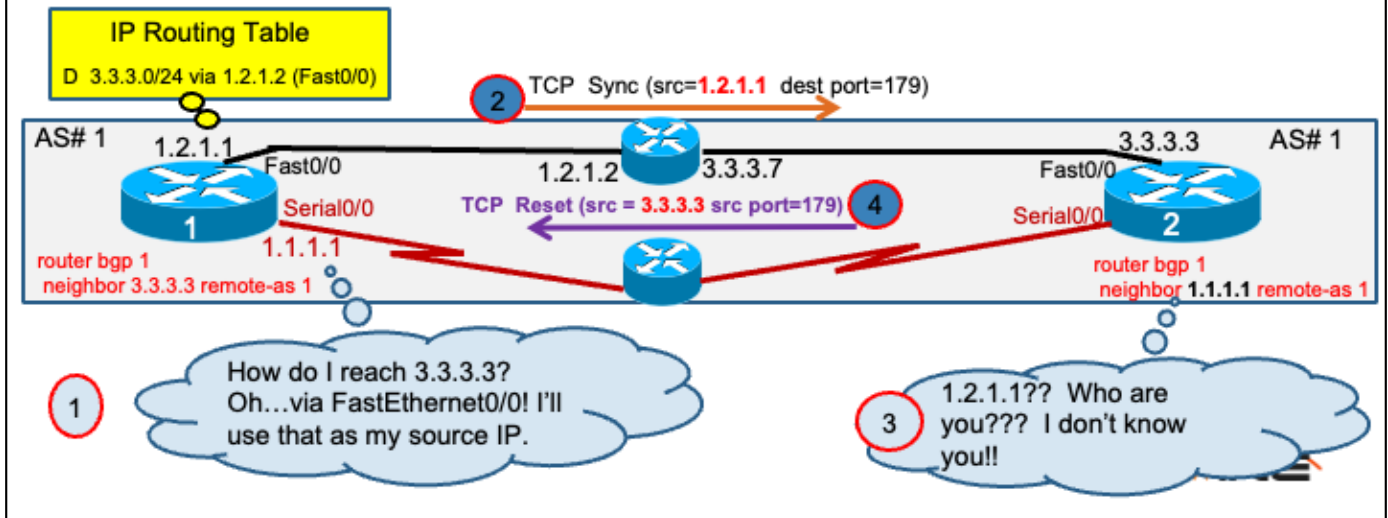
BGP Update-Source

- + When peers are directly-connected, source-IP address of incoming BGP messages is trusted.



Unexpected BGP Source Addresses

+ What if peers are NOT directly connected?

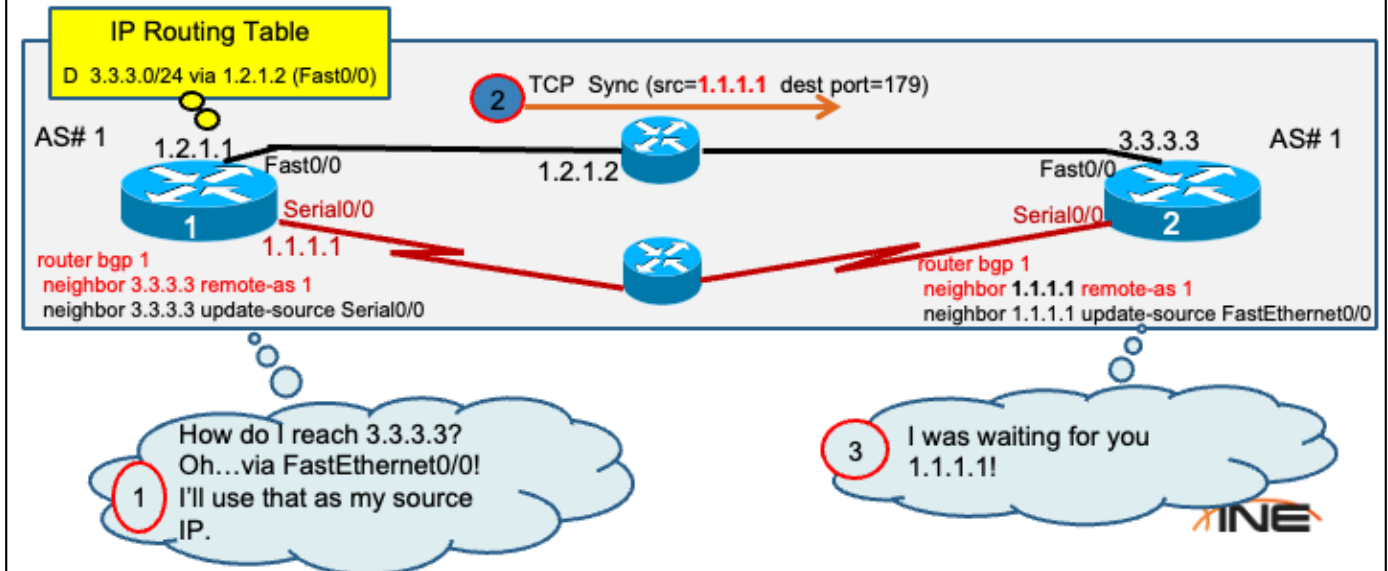


BGP Loopback Peering

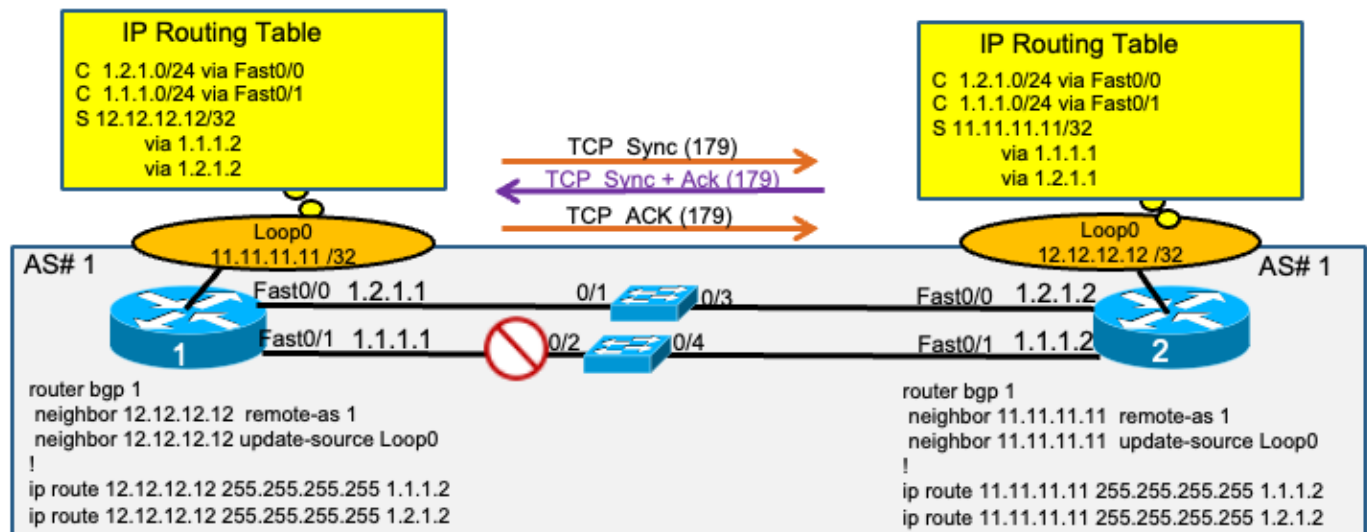
- + Network topology changes can result in BGP packets being sourced from an interface you didn't expect.
- + Non-connected routers are typically peered using Loopback interfaces.
- + Routers must be explicitly configured to use the Loopback address as their source IP address.



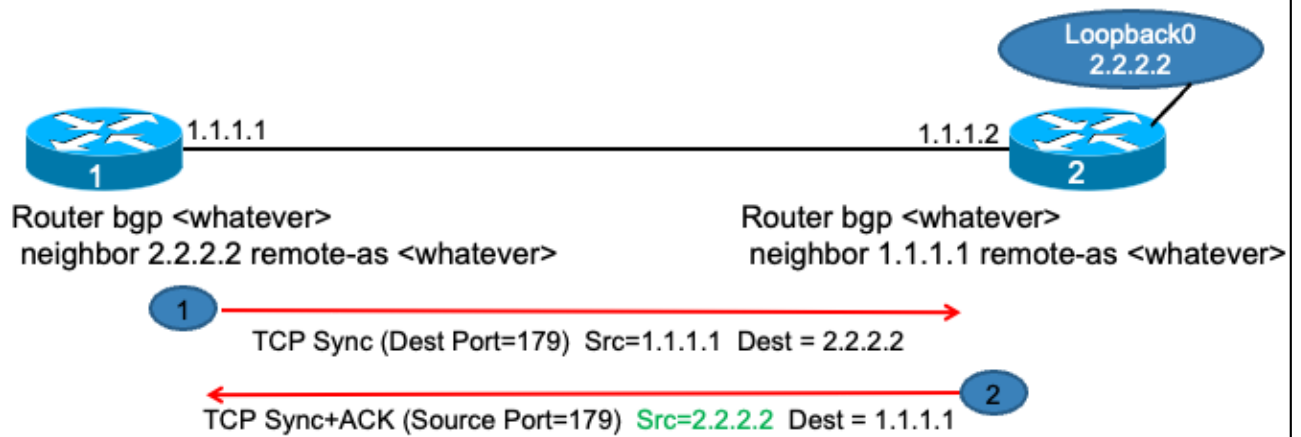
BGP Update-Source



BGP Parallel Links



Cases Where "Update-Source" Is Not Needed

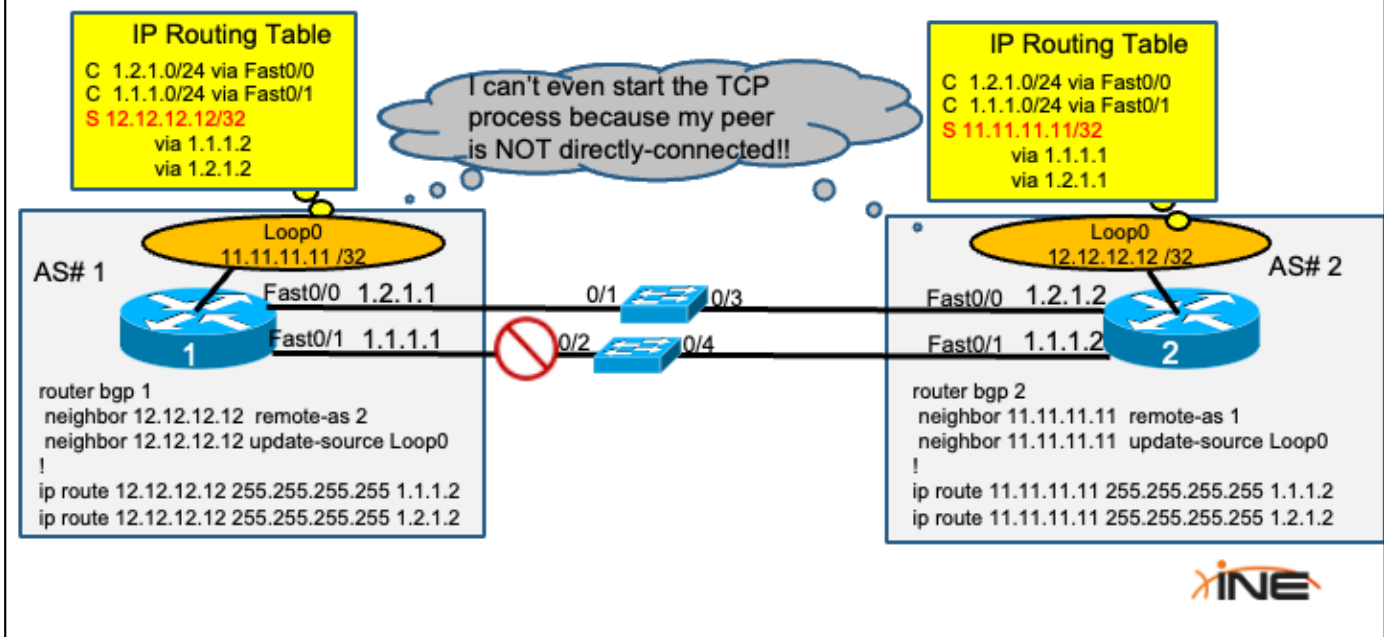


- Notice that in this instance, Router-2 responds using its Loopback Interface IP Address as a source IP...even without "update-source" configured.



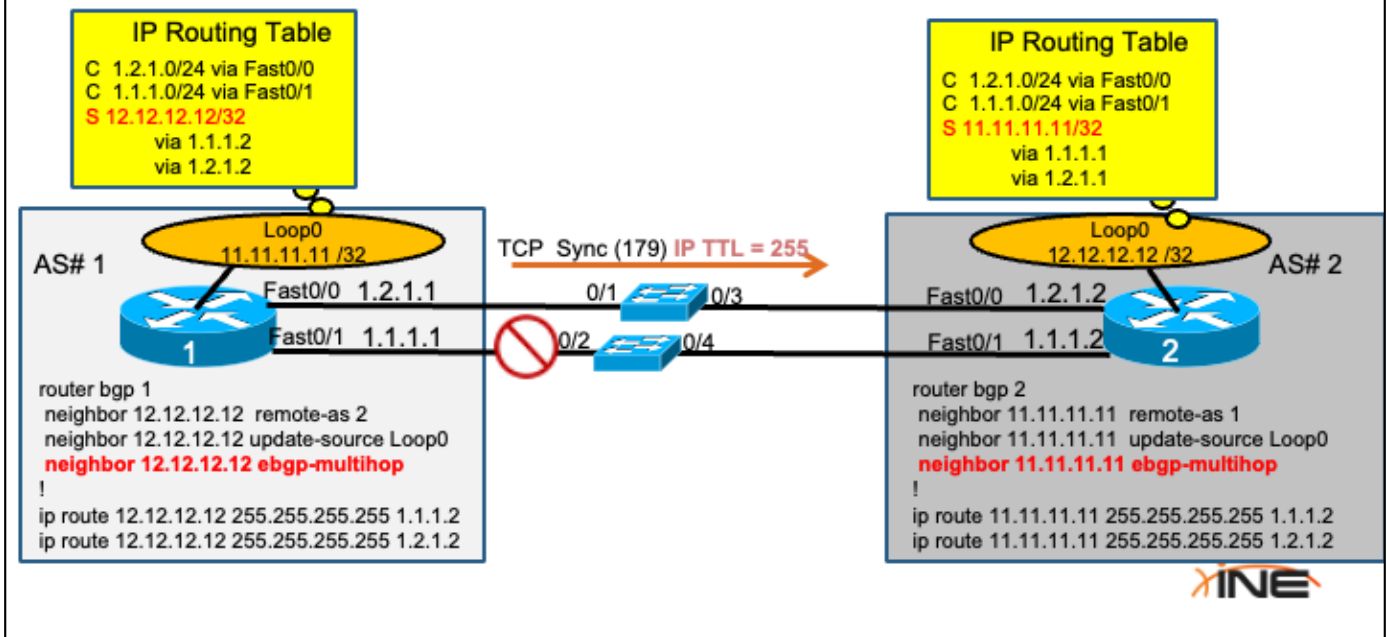
- Normally, you'd think that "update-source loopback0" would be required on Router-2...but in this case it works even without it.
- -
- Not a very elegant design though and people might ask you, "Ummm....WHY did you design it like that??"

eBGP Problem



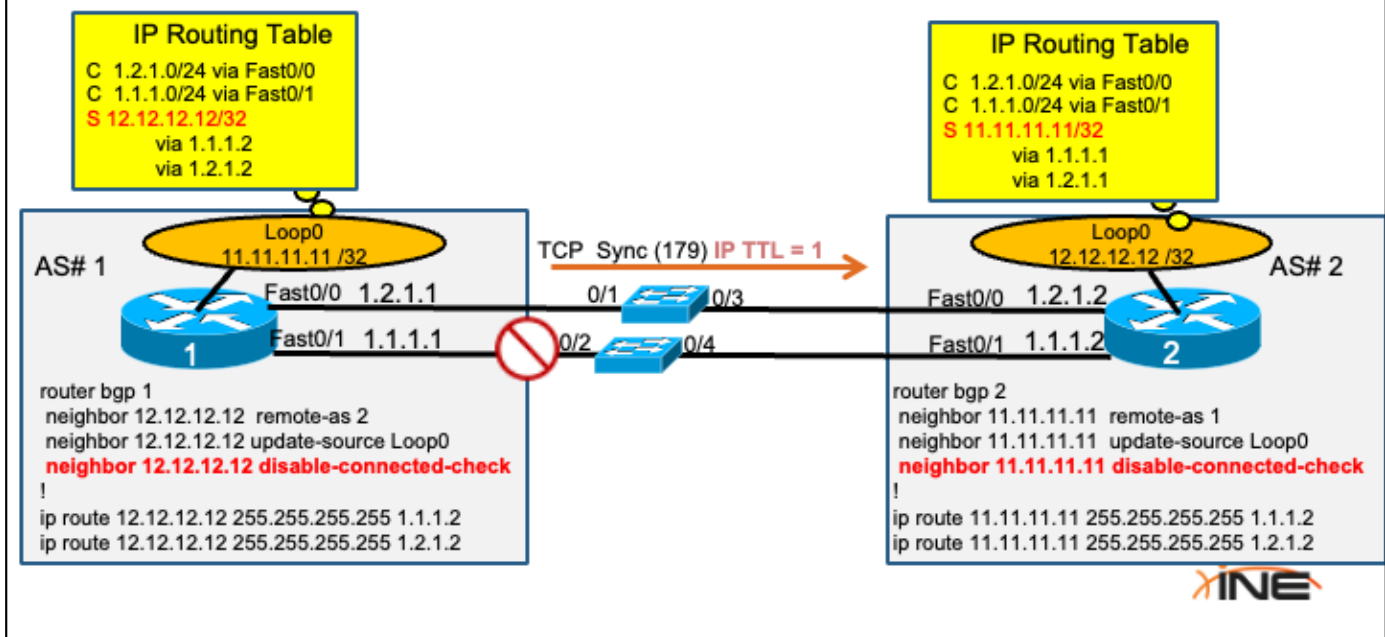
- eBGP has a rule that states external peers must be directly connected. In this case, while the neighbor is physically, directly-connected, the Loopback routes are NOT displayed as directly-connected routes in the IP Routing Table.
- -
- IP packets carrying eBGP data normally have a TTL = 1. If the BGP process determines that a neighbor is not reachable using this TTL it will not even attempt to start the TCP 3-way handshake.
- -
- In this topology, Router-1 has no idea how far away 12.12.12.12 is. The static route provides no indication if that address physically resides on the next-hop...or if the next-hop will simply continue to forward packets to that destination.

eBGP Multihop



- eBGP multihop overrides the default TTL behavior of eBGP setting it to a default value of 255. This value can be reduced as an optional keyword within the ebgp-multihop command.

BGP Disable-Connected-Check



- The command shown above disables normal eBGP rules and allows a router to start a TCP session with a remote peer in another AS that is not directly-connected.
- -
- The main difference between this command and eBGP-multihop is that the command, “disable-connected-check” is only for peering to the Loopback address of a directly-connected peer. It enforces this by setting the TTL = 1 of all outbound TCP/BGP packets.

