Demystifying **BGP**

The Core Routing Protocol of the Internet

Ken Propes

This presentation can also be found at
http://self2015.grimoi.re/demystifying_bgp/bgp.pdf

Preamble

- BGP, Border Gateway Protocol, RFC 4271

For IP Networking, the most scalable protocol

  - Ties the Global Internet together
- Path Vector Protocol
- Event Driven, Incremental Updates
- BGP scope is more than pure Layer 3 Routing
  - IP Virtual Private Networks (IP VPN)
  - Layer 2 VPNs (VPLS, VLL, BGP/MPLS)
  - Support Large Scale MPLS Networks
- Varieties of BGP (BGP4, eBGP, iBGP, MP-BGP…)

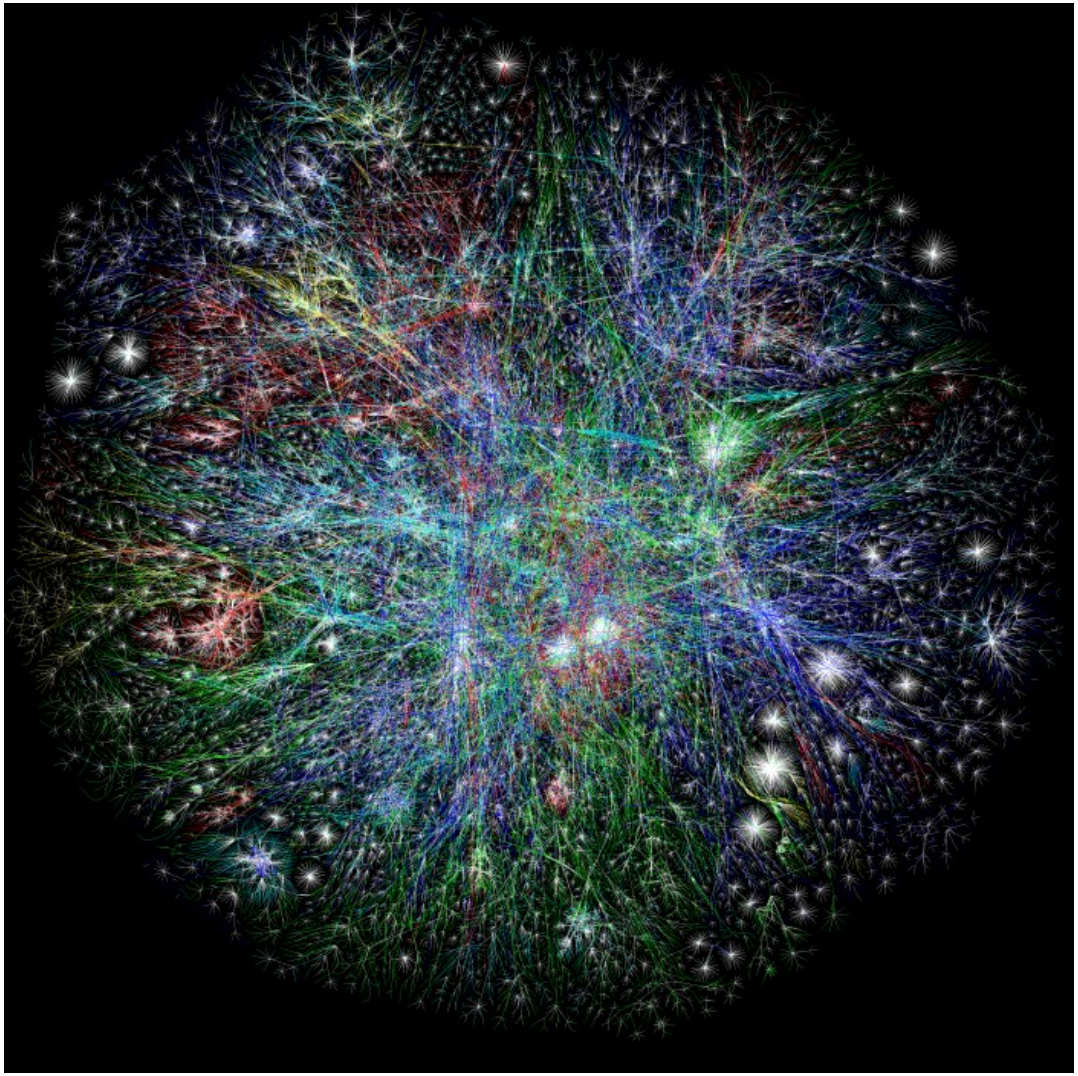
- Our focus will be on eBGP, for IPv4, & The Internet

Global Internet Connections

Massive number of connections

Map of the internet in 2003

Red: Asia Pacific
Green: Europe/MidEast/Africa
Blue: North America
Yellow: Latin America/Caribbean
Cyan: RFC1918 IP Addresses
White: Unknown

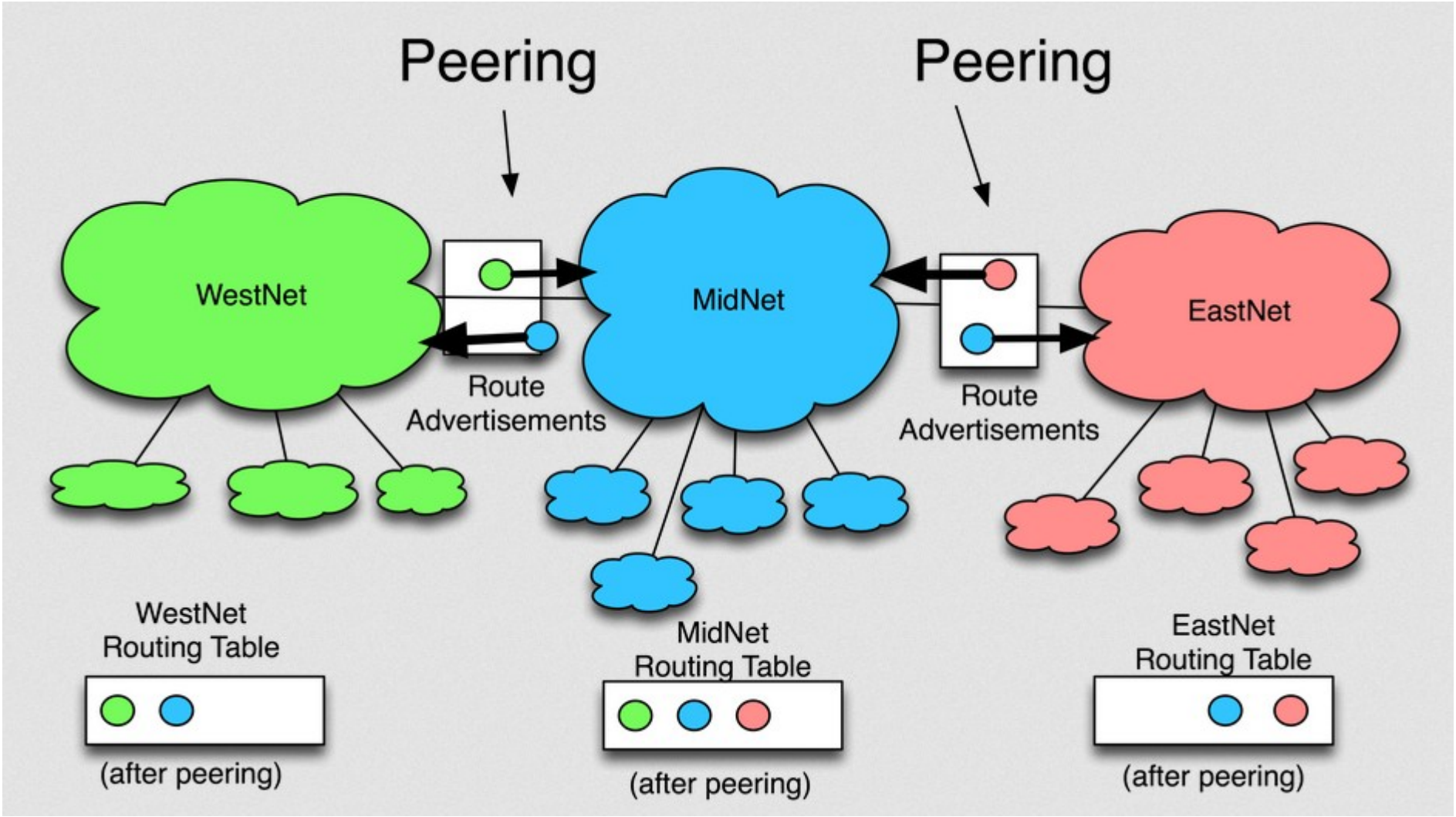


How Big is the Global Internet?

- Advertising all Addresses & Prefixes is not possible, not desirable
- Requires a mechanism to simplify & summarize

- But first, how do the major Internet Service Providers interconnect with each other?
  - Tier 1 ISP's like L3, ATT, Verizon, CenturyLink, NTT, Deutsche Telekom, Orange, etc.
  - Tier 2 ISP's
  - Large Enterprises

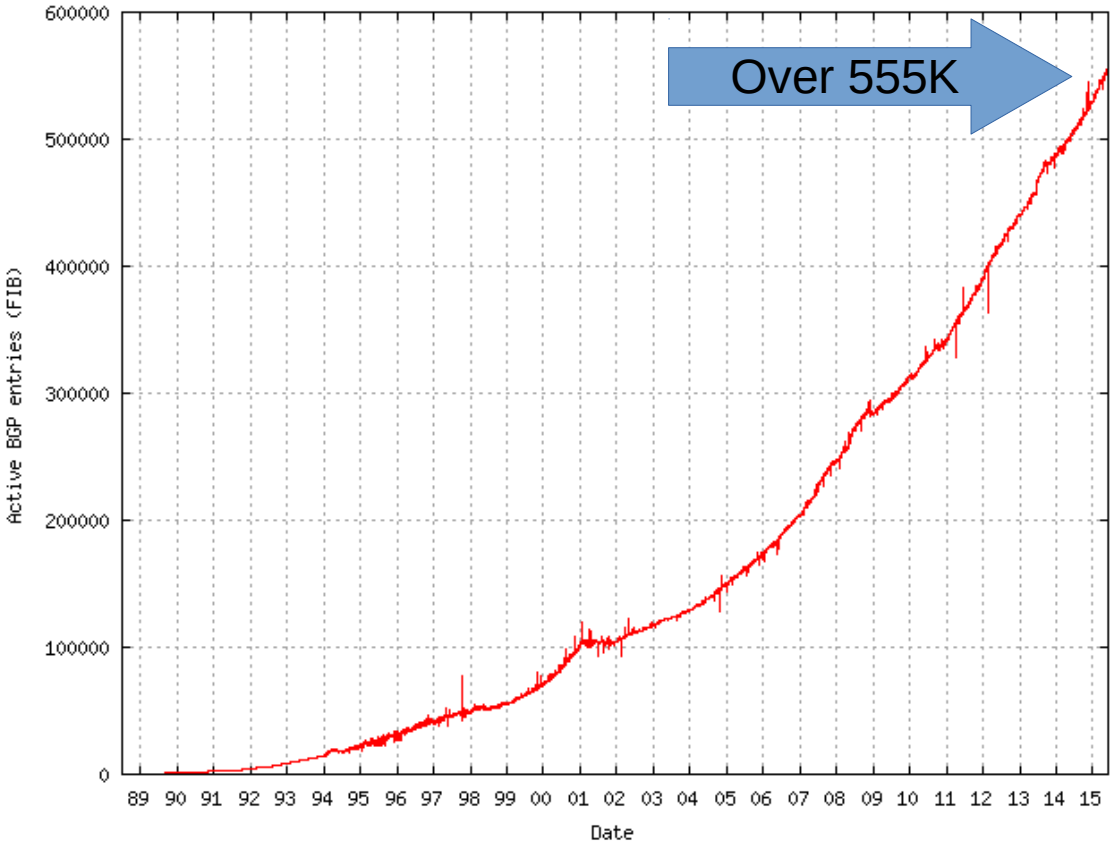
Peering of ISPs
Top Tier connect at Internet Exchange points



Source: DrPeering.net ©2014. http://drpeering.net/img/PeeringDiagram.png.

BGP Entries keep Growing

Active BGP Forwarding Information Base continues to grow:



Source: The CIDR Report. http://www.cidr-report.org/as2.0/#General_Status

Autonomous Systems
The Cornerstone of BGP

- An Autonomous System is Defined in RFC 1930
  - Collection of IP Routing Prefixes
  - Administered by one or more Network Operators
  - On behalf of a Single Entity (typically ISP or Large Organization)
  - Presents a Clearly Defined Routing Policy to other AS's
- Think Tier 1 ISP's, Tier 2 ISP's or Large Enterprises like Exxon, etc.
  - Simplifies and Consolidates a Large ISP's Territory
  - Each Autonomous System has a unique AS number ASN
  - Helps reduce number of advertised prefixes between AS's

But where do AS Numbers come from?

AS Numbers (ASN's)

- Public: Unique ASN's
  - For Exchanging Routes with other AS's
  - Assigned by IANA RIR's Regional Authorities
    - ARIN in North America
    - RIPEN CC in Europe
- Private: 64,512 to 65,534
  - Not routed to Global Internet
    - Internal Network Use only
    - Like Private IP Addresses
    - Not visible in the internet
- 16 bit & 32-bit ASNs

  (0 to 4,294,967,295)

Allocation of AS Numbers by RIR's

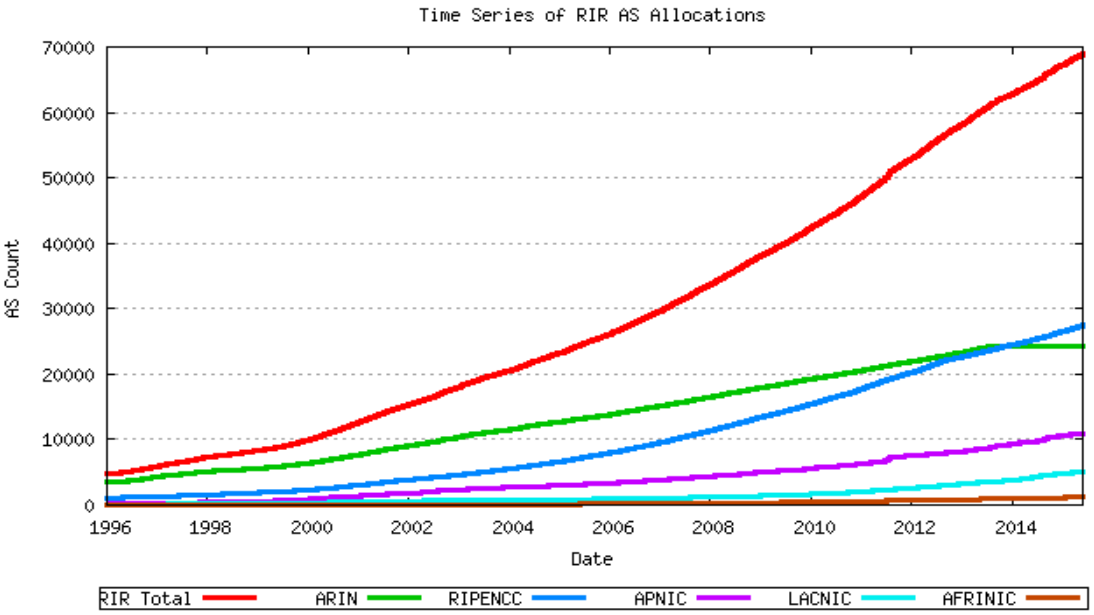


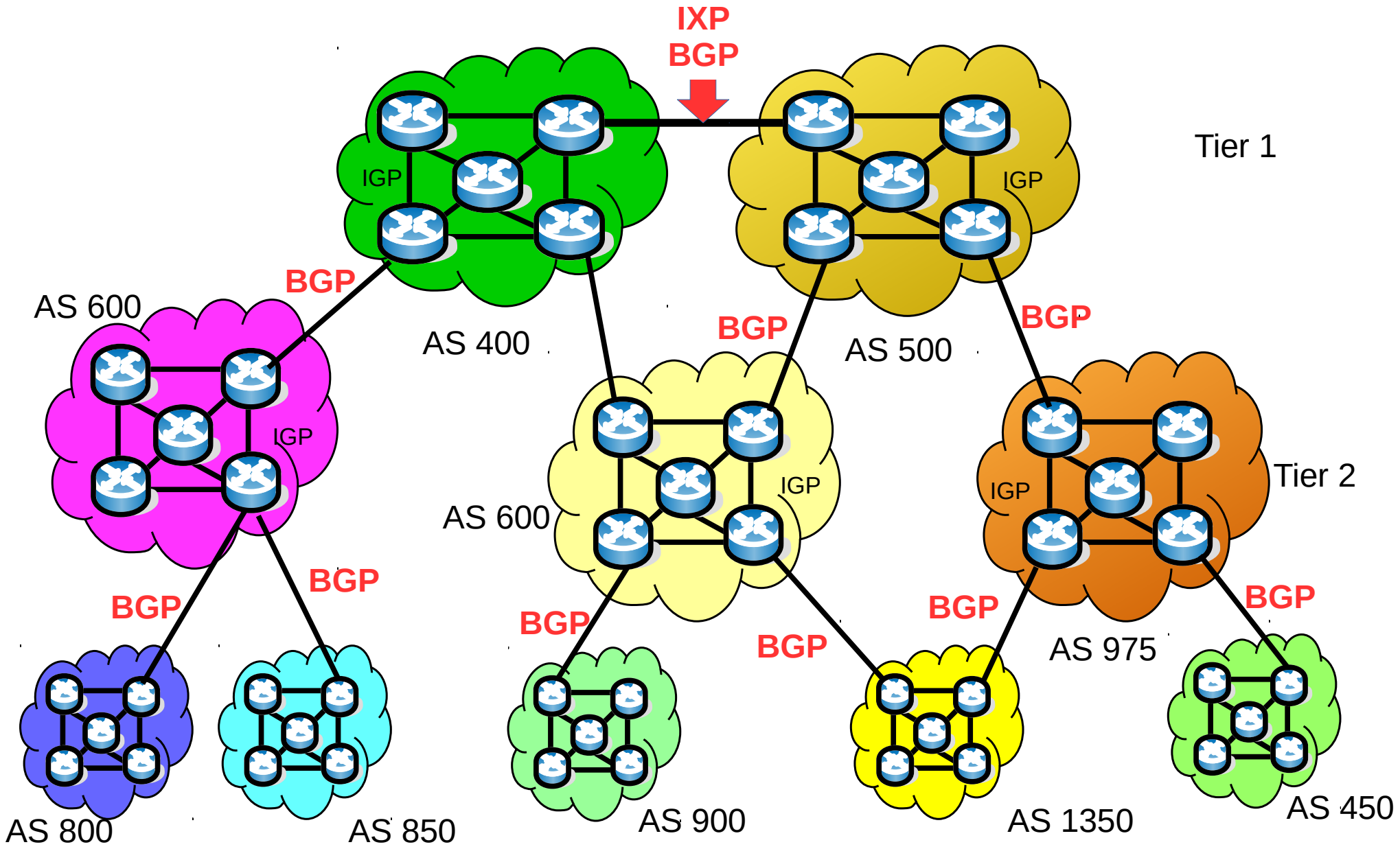
Figure 5 - Cumulative RIR AS assignments per RIR

Source: http://www.potaroo.net/tools/asn32/index.html

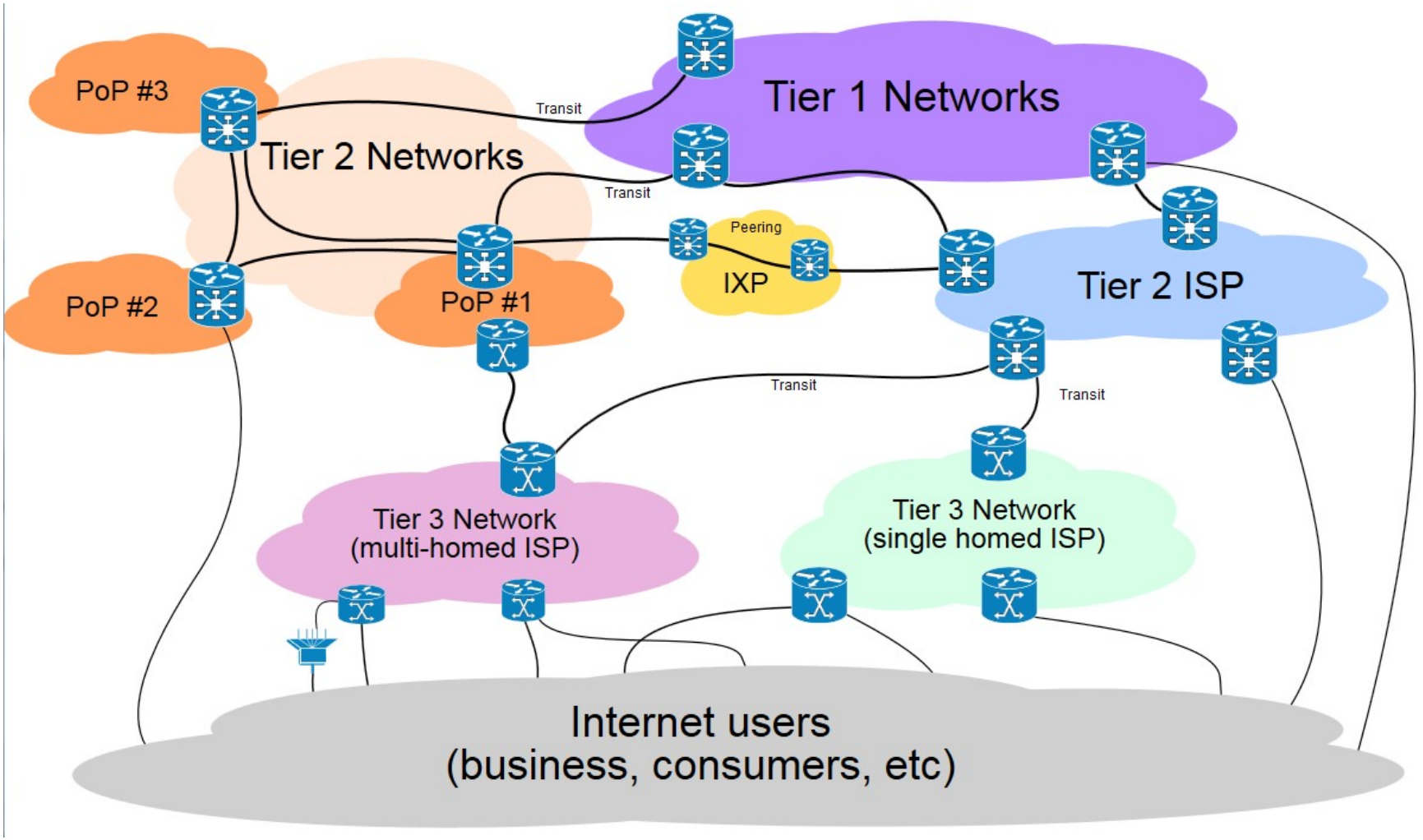
By 2014, > 47K unique ASNs in the Internet

Peering of ISPs

Top Tier connect to other AS's (ISPs and Enterprises)



Routing across the Internet involves several tiers of Internet service providers.

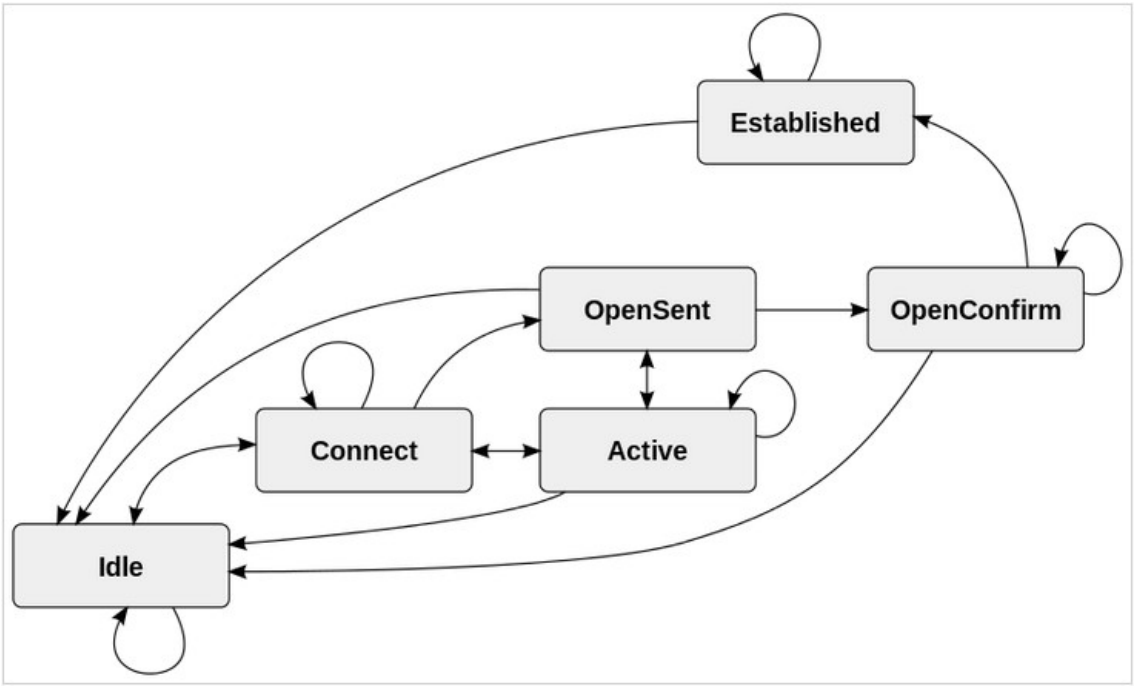


BGP Concepts

- BGP Peering uses a **Finite State Machine** to Establish & Maintain Communications between Autonomous System Neighbors
- How does BGP Communicate with Neighbors?
  - BGP Uses TCP Connection on Port 179
  - Agreed Collection of Specific Attributes are exchanged
  - Exchanges Messages: Updates and KeepAlives
  - Timers to detect faults and outages
- BGP runs Finite State Machine

BGP Finite State Machine

Each BGP interface sets up and maintains communications between AS's for building a loop-free topology of the Internet



Established is the desired state

Open Sent – almost there
Open Confirm- Both Peers talking

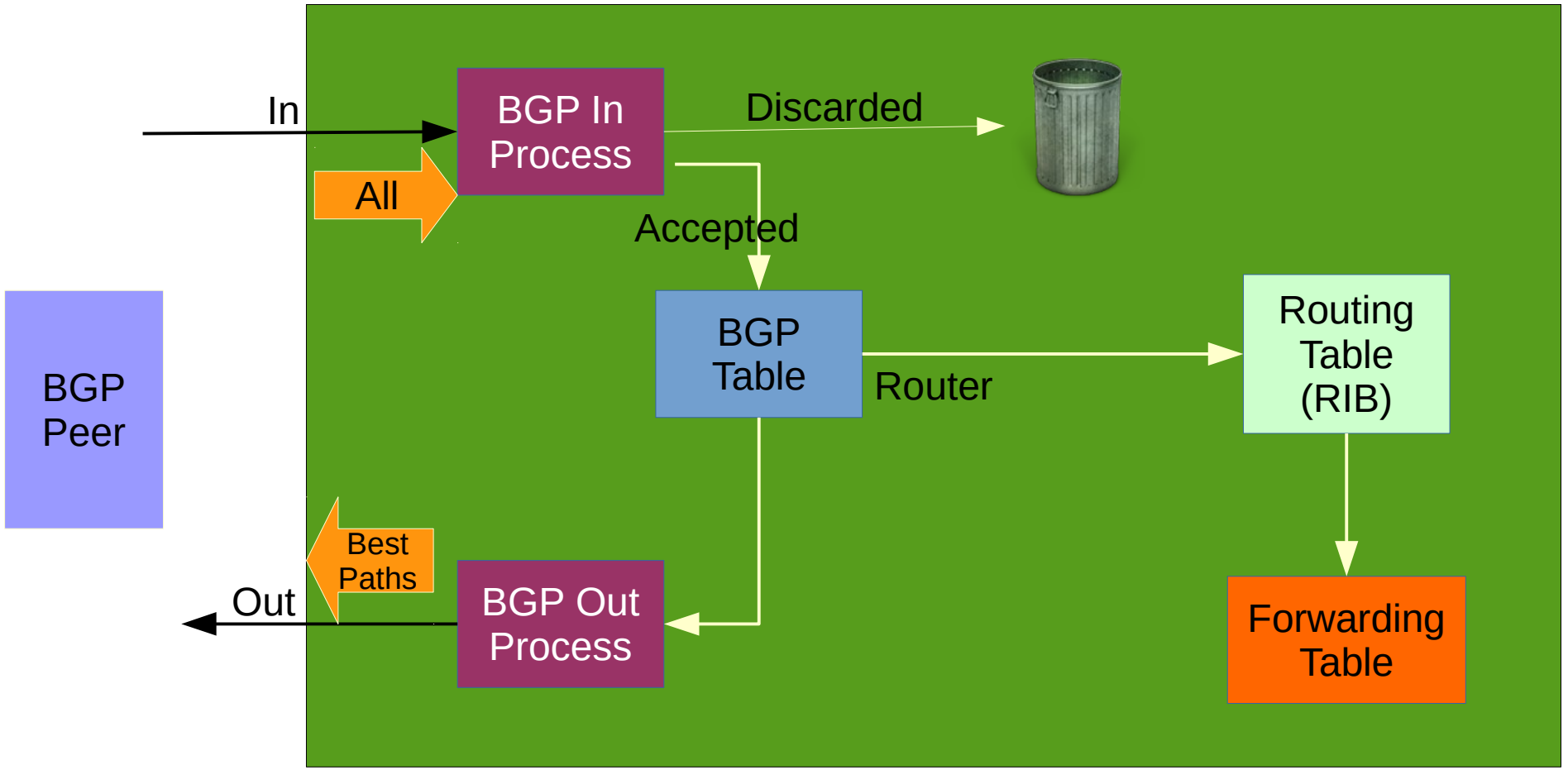
Connect – TCP negotiations
Active – Cant get TCP session

Idle - The beginning point or Timers have expired: Start over

BGP Router Operation

- Learns multiple paths through internal and external BGP “Speakers”
- Selects the best path
- Installs Best Path in its routing table (Routing Information Base – RIB)
- Best Path is used to forward traffic
- Only the Best Path is sent to BGP Neighbors
- Policies are applied which help selecting the Best Path

BGP Tables


In
BGP In Process
Discarded
All
Accepted
BGP Peer
BGP Table
Router
Routing Table (RIB)
Best Paths
Out
BGP Out Process
Forwarding Table

Each Router Must be Configured for Successful Neighbor Peering

- Information exchanged & negotiated at start up to complete BGP Handshake
- Capabilities
  - BGP Version number (usually BGP 4)
  - AS Number
  - Hold Time (time between update messages)
  - BGP Router ID (Router Loopback IP address)
  - Optional Parameters (Time, Length, Value)
- Handshake completed – Messages Exchanged

BGP Messages

BGP Messages can be from 19 to 4096 bytes

Type	Name	Purpose
1	Open	Exchange Application Parameters
2	Update	Transfer Routing Information or Withdraw NRLI (Network Layer Reachability Information)
3	Notification	Sent when BGP encounters an error
4	Keep Alive	Maintains the BGP Session Sent at approx 1/3 of Hold Time parameter
5	Route Refresh (RFC 2918)	Request Route Refreshment during Open exchange. Extend Protocol to request peer to resend Prefix info to minimize databases and avoid adversely affecting a Neighbor session.

BGP Session Established? Now Let's Learn Routes

- Network Layer Reachability Information (NRLI)
- Updates
- Local Routing Information Base (Lo-RIB)
- Adjacent RIB for each Neighbor (Adj-RIB)
  - Adj-RIB-In for incoming Updates
  - Adj-RIB-Out for NLRI sent to the Neighbor

- Decision Factors for Best Path populates Lo-RIB

What's in BGP Update Messages

Consists of 3 parts

- Withdrawn Prefixes
  - Variable list of Paths that are no longer valid
  - Update Messages can contain only Withdrawn routes
- Path
  - Attributes that all the subsequent prefixes in the update messages share
- Network Layer Reachability Information
  - Actual Prefixes announced with attributes listed in Path
  - Multiple NLRI's can be sent in an Update Message
  - Neighbors can re-send the same NLRI with new or updated Path Attributes as necessary.

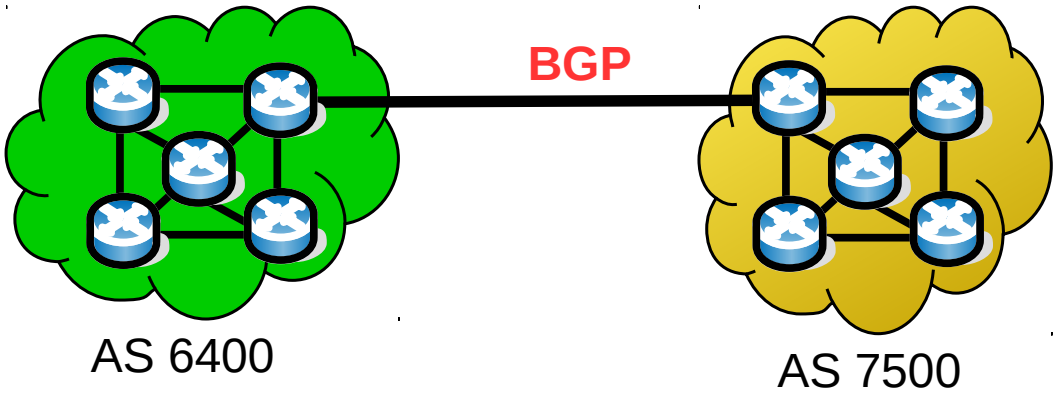
Attributes in BGP Update Messages

- 10 Attributes, Some are Mandatory, some Optional
- 3 well known Mandatory Attributes in every Update
  - **Origin** (where did you learn this NLRI/Prefix?)
    - Your local IGP is better than from eBGP
  - **AS Path** (where did you come from?)
    - ID's all the AS's through which the message has passed
    - Considered as the “Hop Count” of the Path
    - Used for Loop Detection
  - **Next Hop** (where do I send packets with this prefix?)
    - The IP address of the Border Router toward destination

Best Path Decision Factors

Choose Routes in this Order of Decisions

1) Local Preference (within an AS)
2) Routes Originated from the Local Router
3) Shortest Autonomous System Path
4) Lowest Origin Code (from IGP < from BGP < Incomplete)
5) Path with lowest MED (Multi-Exit Discriminator)
6) Prefer External Paths (eBGP) to Internal Paths (iBGP)
7) Path through the closest IGP Neighbor (shortest internal AS path)
8) Select Oldest Route to minimize route flapping
9) Route with Lowest Neighbor BGP Router ID value
10) Prefer Router with Lowest Neighbor IP Address



A bit about iBGP – Within an AS
Route Reflectors & Confederations

- iBGP requires fully meshed network among speakers
  - To avoid routing loops, IBGP does not advertise routes learned from an internal BGP peer to other internal BGP peers. For this reason, BGP cannot propagate routes throughout an AS by passing them from one router to another
- Larger fully meshed networks face the N-squared problem: n*(n-1)/2
  - 5 Routers = 10 links, 10 Routers = 45 links, 20 Routers = 190 links
  - As a network grows, the full mesh requirement becomes increasingly expensive to manage.
- Route Reflector is one solution (RFC 2796)
  - Divide backbone into multiple clusters, one route reflector per cluster
  - Single IGP to carry next hop and local routes
- Confederations is another solution (RFC 3065)
  - Divide the AS into sub-AS's, Usually a single IGP

Now You Know

- When BGP is used
- Who requires BGP
- How BGP Communicates
- What BGP Communicates
- Where BGP is running in the Internet
- Why you should not try this at home

BGP Demystified

Thanks for Coming !

Questions?