Border Gateway Protocol

Peter Chapin

CIS-3210, Routing and Switching

Vermont Technical College

Autonomous Systems (AS)

- An *autonomous system*…
  - … defines a single *administrative domain*
  - … a region (of the network) over which a single entity controls policy, and the configuration of all network infrastructure devices (routers, etc.)
- Examples
  - The Vermont State College System (VTSU + CCV)
  - The University of Vermont

Interior vs Exterior

- Interior routing
  - Routing done *inside an autonomous system*.
  - Since all routers are configured by the same entity, they can be configured in a consistent way.
  - Various routing protocols available to pick from: RIP, OSPF, EIGRP, etc.
  - It is reasonable for every router to know about the every other “internal” router and subnetwork (as is done by link-state routing protocols such as OSPF). Note: The VSCS uses OSPF internally.
- Exterior routing
  - Routing *between autonomous systems*.
  - Only one routing protocol is used: Border Gateway Protocol (BGP)

What About the VSCS?

- Two ISPs (as of Fall 2022)
  - Consolidated Communications, formerly Fairpoint (primary)
  - Firstlight (secondary)
- The VSCS is considered an autonomous system. So are Fairpoint and Firstlight.
- The VSCS peers at the BGP level with Fairpoint, Firstlight, and UVM.
  - The VSCS is thus a “multi-homed” autonomous system
  - Packets from the VSCS for “the Internet” go through one of these three peers.
  - The actual AS where such packets are sent depends on routing decisions made by BGP routers.

Autonomous System Numbers (Part 1)

- Autonomous Systems are assigned numbers by a higher authority (ARIN; the American Registry for Internet Numbers).
  - VSCS: 54257
  - Fairpoint: 13977
  - Firstlight: 13536
  - UVM: 1351
- ASNs are 16 bit or 32 bit
  - Originally 16 bit but expanded to 32 bits as the network grew.
  - BGP needed updating to handle the larger ASNs
- *ASNs are like addresses for autonomous systems*.

Autonomous System Numbers (Part 2)

- An ASN is not needed if…
  - … an autonomous system has a single connection to the Internet and a single block (one prefix) of IP addresses.
- For example, if the VSCS had only one ISP…
  - … that ISP could include the prefix 155.42.0.0/16 into the collection of IP ranges known to it…
  - … and share that information with other autonomous systems connected to it via BGP.
- However, since there is more than one way to reach 155.42.0.0/16…
  - … the VSCS needs to use BGP to route across the multiple ISPs that it has.

Autonomous System Types

- Single-Homed AS
  - Only one connection to the rest of the Internet (a single ISP).
  - Does not normally need an ASN
- Multi-Homed AS
  - Has multiple connections to the rest of the Internet (multiple ISPs).
  - Does need an ASN
  - The VSCS is like this
- Transit AS
  - An AS that passes traffic between other autonomous systems.

Border Gateway Protocol (Part 1)

- BGP version 4 (the latest) is described by RFC-4721 from Jan 2006.
- BGP can handle multiple network layer (layer 3) protocols
  - Thus, IPv4 and IPv6 are treated as two separate but equal layer 3 protocols.
  - In principle other network protocols could also be supported.
  - See RFC-2858 for more information.
- BGP uses “destination based” routing. Only the destination address is used to make routing decisions.
  - This affects certain policy decisions: can’t send a packet that uses source routing to a neighboring AS to force the packet to go somewhere different than where that neighboring AS would normally send it.

Border Gateway Protocol (Part 2)

- BGP uses TCP to transmit routing information between BGP “speakers” (aka, routers).
  - Port 179 is the official BGP port
  - This relieves BGP from dealing with sequencing, retransmission, etc.
  - Note: TCP is only used to exchange routing information… not the packets to be routed!
- A BGP peer might be internal (in the same AS) or external (in a different AS)
  - Internal BGP is called IBGP.
  - External BFP is called EBGP.

Routes

- For BGP's purposes a *route* is…
  - … a pair consisting of a set of destination addresses together with a path.
  - Destination addresses are expressed as prefixes: 155.42.0.0/16, etc. A "route" might contain multiple prefixes (that all can be reached using the same associated path).
  - A path consists of a set of path attributes that describe, among other things:
    - The next-hop IP address of where packets should go to delivered to the destination addresses.

Routing Information Base

- Each BGP speaker maintains a database of routes in three groups:
  - **Adj-RIBs-In**. Routes learned from peer BGP speakers, unprocessed
  - **Loc-RIBs**. Routes used by the local BGP speaker, processed from Adj-RIBs-In. Processing entails applying policy rules (if any) and removing routing loops (if any).
  - **Adj-RIBs-Out**. Routes that will be advertised to other BGP peers. This might be a subset of what is in Loc-RIBs (BGP speakers should only advertise routes they use).
    - Note that the NEXT_HOP attribute of the paths in these routes may be modified from what is in Adj-RIBs-In. That is, a BGP speaker might or might not change them to point at itself.

Routing Table

- BGP speakers maintain a routing table (like any router) that contains…
  - … routes to directly connected networks
  - … statically defined routes
  - … routes learned by an IGP protocol (for destinations inside the AS)
  - … routes learned by BGP (address prefix + next hop router in another AS)
- Example
  - Some BGP speaker “out there” has a packet for 155.42.107.97.
  - The routing table has a BGP route for 155.42.0.0/16 and a next hop address.
  - The next hop address must be “resolved” by doing recursive lookups in the routing table until an address on a directly connected network is found.

Route Consolidation

- Example:
  - Adj-RIBs-In contains a BGP route learned from a peer for 155.42.0.0/16
  - Adj-RIBs-In contains a BGP route learned from a peer for 155.43.0.0/16
  - Both routes have the same NEXT_HOP attribute.
  - The BGP speaker assigns to Loc-RIB a single route for 155.42.0.0/15 using the common NEXT_HOP attribute.
  - This *massively* reduces the size of the routing table
  - CIDR!