External v.s. Internal BGP

Bill Cheng

http://merlot.usc.edu/cs551-712

External v.s. Internal BGP

CS551
Prefix learned from E-BGP can be advertised to I-BGP neighbor and vice versa, but the same message types, attribute types, and state machine as E-BGP.

Different rules about re-advertising prefixes:
- Prefix learned from one I-BGP neighbor cannot be advertised to another I-BGP neighbor.
  - Reason: No AS-PATH within the same AS and thus danger of looping.

R2 can only find these prefixes through a direct connection to R1.

Result: I-BGP routers must be fully connected (via TCP)!

BGP Example:
- AS1
  - Provider A
  - AS100
- AS2
  - Provider B
  - AS200
- AS3
- AS4
- AS5
- AS6

BGP Example:
- R1 advertises routes within AS1 to R2.
- R2 advertises routes within AS2 and AS3 to R1.
- R2 learns AS3 routes from I-BGP with R4.
- R4 learns AS3 routes from E-BGP with R6.
- R4 advertises routes within AS2 and AS1 to R6.
- R6 can only find these prefixes through a direct connection to R1.

Contrast with E-BGP sessions that map to physical links.

R3 can tell R1 and R2 prefixes from R4.
- R3 can tell R4 prefixes from R1 and R2.
- R3 cannot tell R2 prefixes from R1.

Failure on an E-BGP link:
- R3 cannot tell R2 prefixes from R1.
- R2 cannot tell prefixes from R1 and R2.
- R2 cannot tell R1 and R2 prefixes from R4.

Failure on an I-BGP link:
- R3 cannot tell R4 prefixes from R1.
- R2 cannot tell prefixes from R1 and R2.
- R2 cannot tell R1 and R2 prefixes from R4.

Two types of link failures:
- Failure on an I-BGP link
- Failure on an E-BGP link

Why?
- These failures are treated completely different in BGP.

Link Failures

Different approaches to advertise eBGP prefixes:
- Advertise at another iBGP neighbor
- Advertise to another eBGP neighbor
- Advertise through iBGP to advertise to eBGP neighbors.

Different rules about advertising prefixes:
- Same message types, attribute types, and state machine.
- Prefixes advertised to eBGP neighbors cannot be re-advertised to iBGP neighbors.
Failure on an E-BGP Link

If the link R1-R2 goes down, then the TCP connection breaks and so does the E-BGP connection. BGP routes are removed from the BGP table. This is the desired behavior.

AS1

R1

AS2

R2

E-BGP session

Physical link

138.39.1.1/30 138.39.1.2/30

Note that 138.39.1.1 and 138.39.1.2 are on the same network.

Failure on an I-BGP Link

R1 and R2 should, in theory, still be able to exchange traffic, i.e., the indirect path through R3 should be used. If physical link R1-R2 goes down, the 138.39.1.0/30 network becomes unreachable, connection between R1 and R2 is lost.

I-BGP session

Physical link

138.39.1.1/30 138.39.1.2/30

Note that 138.39.128.1 and 138.39.128.5 are on different networks here!

Virtual Interfaces (VIFs, a.k.a. Loop-back Interfaces)

A VIF is not associated with a physical link or hardware interface. How do routers learn of VIF addresses?

Two methods:

- Use IGP
- BGP confederations

Scaling the I-BGP Mesh

Route reflectors scale by adding hierarchical BGP routes originating.

Two methods:

- Expand the I-BGP mesh
- Use IGP

Note: 1.0-1.4 do not need a physical link with respect to TCP endpoints. This is E-BGP and I-BGP must use different conventions.

AS Confederation

AS1

R1

R2

Sub-divide a single AS into multiple, internal sub-AS's to reduce I-BGP mesh size.

BGP Confederations

1. BGP Confederations
   - Confederations allow multiple ASes to form a single AS for BGP purposes.
   - The AS-CONFED-SET and AS-CONFED-SEQUENCE attributes are used to ensure loop-free routing.

2. RR Example
   - RR1 advertises 138.39.0.0/16 learned from RRC2 into I-BGP.
   - RR1 will not advertise 128.4.0.0/16 learned from RR2.

Rules for Route Reflectors

1. RR will not advertise routes learned from non-clients.
2. RR advertises routes learned from RRC into I-BGP.
3. RR reflects advertised routes learned from clients into the I-BGP mesh.
4. RR will not re-advertise routes between non-clients.

Confederations

1. AS CONFED-SEQUENCE and AS CONFED-SET are used to avoid looped BGP sessions.
2. AS CONFED-SEQUENCE ensures that the AS numbers are in ascending order.
3. AS CONFED-SET ensures that the AS numbers are unique.

Route Reflectors

1. Route Reflector Client (RRC): A router that depends on the RR to re-advertise its routes to the entire AS.
2. Route Reflector (RR): A router whose BGP implementation allows re-advertisement of routes between I-BGP neighbors.
3. RR Example:
   - With RR, there are 7 I-BGP sessions instead of 21 (=7*6/2).

Conferences

1. AS CONFED-SEQUENCE and AS CONFED-SET ensure that the AS numbers are in ascending order.
2. AS PATH now includes AS CONFED-SET and AS CONFED-SEQUENCE to avoid loops.
3. RR sessions between sub-AS's are like regular E-BGP but with some changes.
4. Next-hop attribute traverses sub-AS boundaries (assumes single IGP running - everyone has the same route to next-hop).
5. Local-pref attribute remains meaningful within confederation (E-BGP ignores it).