ADVANCED BGP POLICY

Tawfiq Khan
TCOM610 BGP

Advanced Topics

- Route redundancy
- Load balancing
- Routing Symmetry

Route Optimization Issues

- Redundancy
  - provide multiple alternate paths
  - usually multiple connections for a particular AS
- Symmetry
  - traffic leaving an AS from particular exit point returns through the same connection to AS
- Load Balancing
  - divide traffic in reasonable way over multiple links

Redundancy

- Device and provider failure are inevitable
  - Physical error: router, interfaces, cabling, access link
  - Software error: bugs
  - Human and administrators errors
  - Natural disaster
- Redundancy design to provide failure protection for mission critical applications
  - Redundant router
  - Redundant links
  - Redundant providers
  - Redundant route from different sources
- Redundancy and symmetry are conflicting goals
  - more redundancy implies more choice and more difficult route management

Symmetry

- May want traffic to return at the same location as it exits the network
- But due to peering agreement, asymmetric routing is a given fact in today's Internet routing
- Both side Hot-potato/Cold potato routing generate asymmetrical routing
- Cold-potato routing adds cost to the network – hold onto bits longer (long-haul circuit)
- One side (provider) cold-potato routing, other side (customer) hot-potato routing scheme usually produce more symmetrical routing

Load-balancing

- Load-balancing: equal amount of traffic over multiple links;
  - Load-sharing: multiple links share some load of total traffic, not necessary balanced
- Balance both inbound and outbound traffic
- Apply inbound to control outbound traffic or vice versa
  - Route announcement direction is opposite to the direction of traffic
- Typical BGP attributes to manipulate: Local-Pref, AS_PATH, MED, BGP Community, AND routes!

Routing Options

- Static route vs dynamical route
  - Static: simple; small memory; fast processing
  - Dynamical: accommodate network condition dynamically
- Default, partial vs full routes
  - Default only: simple and small routing table, but limit the capability to apply policy and load-balancing
  - Partial plus default: moderate routing table size, flexibility for routing policy management
  - Full table: large routing table, memory requirement, poor performance for low-end router; most flexible for routing policy management; a must-be if you are a transit provider

Routing Options

- Single Homed vs Multi-homed
  - SH: good for small, not critical network; no redundancy for failure; no traffic management; no need to run BGP; usually get address space from provider
  - MH: may be the result of geographical pressure or failure protection; may require traffic load-balancing
- Single Provider vs Multi-providers
  - SP: get consistent service from one source; one contact for troubleshooting
  - MP: provide carrier failure protection; may play with competitive pricing pressure; harder for traffic management

Default Routing

- **Default route: 0.0.0.0/0 – match of last resort**
- **Static Default**
  - Point to next-hop IP address
    - ip route 0.0.0.0 0.0.0.0 1.1.1.1
  - Point to specific router interface
    - ip route 0.0.0.0 0.0.0.0 serial0
  - Point to a big aggregate network number
    - ip route 0.0.0.0 0.0.0.0 192.213.0.0/16
- **Dynamical default:**
  - Learnt via BGP session
  - Neighbor x.x.x.x default-information
- **Default routing**
  - Place default route as far upstream as possible
  - Use multiple static default with caution: use local-pref or admin distance to distinguish them, otherwise may get routing loop
  - Use static default to interface and then originate default in IGP (OSPF: default-information originate)

Issues with Default Routing

- Need defaults to disappear when entity pointed to (next hop or network) disappears
- Cisco implementation
  - statically defined default follows existence of entity pointed to
  - removes default from IP routing table if entity disappears
  - allows failover from primary to backups
- Default networks should not be specific subnets
  - stability of default will depend on stability of subnet
  - point 0/0 to major aggregate or a provider supernet

Route Optimization Issues

- Methodology
  - no single answer
  - multiple administrative entities involved
  - need effective coordination
  - limited capabilities / controls
- Design choices depend on specifics
  - goals
  - configurations

Single Home: default only

- Single-home: default only
- Single connection to a service provider
- Customer points defaults towards provider
  - might choose to receive no Internet routes
  - save memory and processing
  - No BGP needed
- Provider might use static routing towards customer
  - ip route 192.138.3.0 255.255.255.0 serial2/1 tag 100
  - redistribute static route-map FR-STATIC
- No redundancy or symmetry issues
- Typically get address space from the provider

Multi-homed: default with primary and backup

- Multiple connections to one provider, BGP can used with private AS number
- May get your own address or get them from the provider
- Customer configures defaults toward provider: one as primary (higher local-pref), and the other as backup
- Not accepting full or partial routes
- Customer selects NY link as primary, SF link as backup
- Outbound traffic
  - single router: use distance values to select preferred route
  - might learn 0/0 or some aggregates from provider to set default, then use local preference
  - no load balancing, all traffic sent on primary unless the primary fails

Multi-homed: single provider with full or partial route

- Get partial routes or full route from providers
  - If partial, typically provider's customer routes
  - Or customer can control and filter routes
  - If partial route, must also get default route from provider
- For outbound traffic load-balancing
  - Set local preference, based on AS_PATH match or prefix list
  - Raise/lower local-pref on one side, and leave other side as default
  - Better to manipulate routes with route's geographical considerations
  - May also listen to provider's MED so that traffic will get to provider close to destinations

Multi-homed to Multiple Providers

- Motivation
  - Redundancy and geographical restrictions
- Options
  - Default only: primary and backup
  - Default with primary and backup, plus partial or full routes
  - Default-free full routing
- Requirements
  - Primary or backup (premium provider vs cheap provider?)
  - Load-sharing or strict failure protection?
  - Best performance or least cost?
- Difference with Single Provider
  - If manipulate routes by local-pref, outbound traffic is very similar except MED is not meaningful across providers
  - Major difference in inbound traffic behaviors
  - It is much harder to load-share inbound traffic for multi-providers environments

General Principle

- If multiple links are provisioned
  - Want multi-links for primary/backup or load-sharing?
  - If for load-sharing, do you want to control both inbound or outbound traffic (or traffic ratio)
- Route announcement
  - From provider: default, partial or full route. More routes, more flexible to manipulate
  - To provider: one aggregate is usually not good enough for load-sharing, may need to leak some specific with no-export community for load-balancing

Customer of Same Provider with backup

- Customers may provide mutual internal and backup internet connectivity
- Might have same provider, and a private link between customers
- Scenarios
  - private link as strict backup only: get Internet routes from other network but with lower Local-Pref
  - private link as primary for local traffic, and backup for Internet access in case of failure of provider: for other network's routes, lower LP from provider, raise LP from neighbor networks
  - Customer's network provider transit service to other customer in case of failure

Customer of different provider with backup

- Customers may provide mutual internal and Internet backup connectivity
- Might have different providers, and private links
- Difficult to achieve this; Require cooperation among customers and providers – ask one ISP to set lower local-pref to routes learnt from other providers
- Might make it work easier if both providers have community-based route management policy
- Cover more later in ISP section

Inbound Traffic Management

- Inbound traffic
  - customer can advertise its routes over multiple links
  - usually default behavior is for traffic to go to customer's AS depending on which of provider's exit points is closest to the destination
  - customer can advertise with different MEDs to influence provider behavior - lower MED selected if all else equal – not possible with a single aggregate!
  - MED is limited to affect your direct provider only, may need to consider AS_PATH prepending to influence other networks
  - May also leak more specifics along with aggregate to achieve load-balancing

Inbound Traffic Management

- If multi-homed to different providers
  - MED does not work across providers
  - AS_PATH prepend works to certain extent
  - If ISP1 is better connected than ISP2 – most inbound traffic will come in from ISP1 until you AS prepends multiple times for routes announced to ISP1
  - It might still not work after AS prepending, since ISP1 may always place higher LOCAL_PREF for customer routes than peering routes
  - You may call your ISP1 and ask them to lower LOCAL_PREF!
  - You can only fully control your outbound traffic, and provide hints to your neighbors to influence your inbound traffic from them – you can never have full control over your inbound unless you pay $$$$!

Load-balancing

- Load balancing versus load sharing
  - Load balancing is on paths between two routers
  - Load sharing is on paths between two ASs
- In both cases
  - Approach for ingress differs from egress
  - Requires a knowledge of the traffic mix and both topologies

Load-balancing

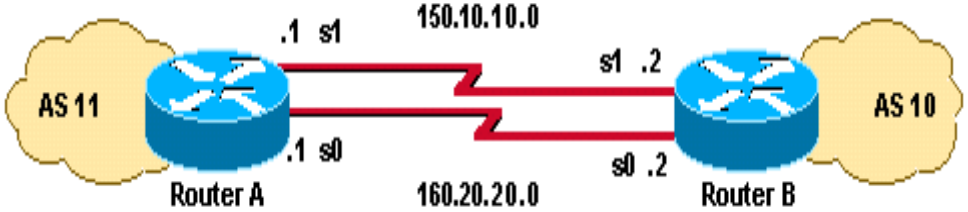
- Forwarding table can maintain multiple interfaces for a single prefix
  - Multiple interfaces for a single prefix
  - Host to interface mapping is consistent
  - Different hosts within the same prefix space may map to different interfaces
  - Use a hash of (src, dest) to pick which paths
- Effectiveness of balancing
  - Customer egress is reasonable
  - Customer ingress more difficult
  - Backbone links have a larger mix of addresses and this balances them nicely

Load-balancing Modes

- Per-destination
  - If more than one paths exist, packets to the same destination will take the same paths
  - Keep the packet order
  - But may result in uneven link usage sine certain host (web servers) may have more traffic than other hosts (desktop)
  - Balanced over per destination, not per prefix
  - The default mode for Cisco IOS
- Per-packet
  - Evenly distribute each packets over parallel links
  - Guarantee even utilization over multi-links
  - But packet may arrive out of order

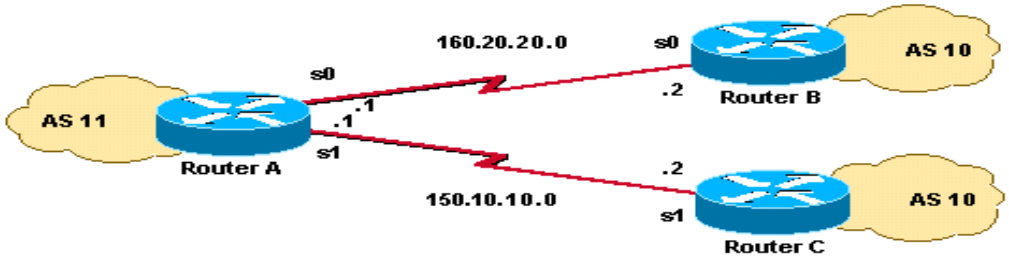
Two Parallel Links Between Two Routers

- BGP peering over virtual interface: router loopback0
- Set static route next hop for virtual interface to two interfaces
- Recursive lookup over static routes will automatically balance over two external links
- Basically use IGP to load-balance; BGP still picks one best route



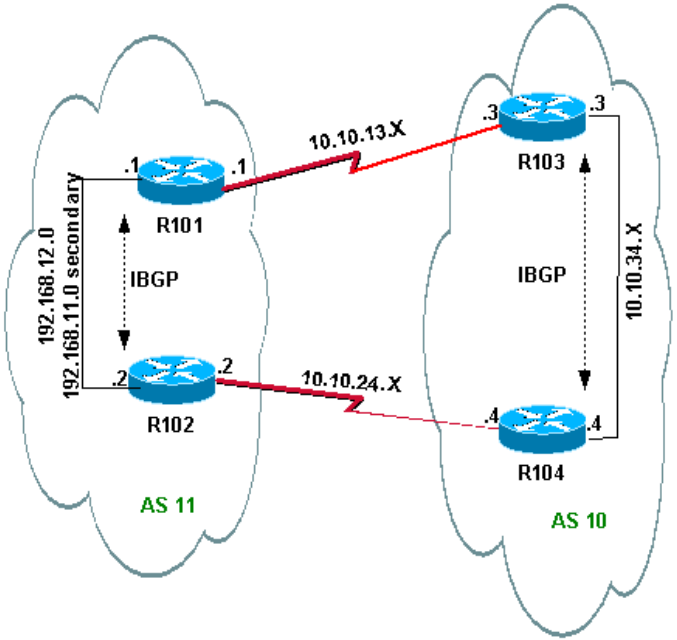
Two links over different routers

- EBGP will select a single best path to use
- maximum-paths command
  - router bgp 65001
  - maximum-paths 2
- Or you can set LOCAL_PREF or MED to all routes received based on certain odd-even scheme



Two Links Over Two Routers

- Impossible to load-balancing – load-sharing
- By default, eBGP routes are preferred over iBGP routes
- Need to apply inbound routing policy to set LOCAL-PREF to achieve load-balancing
- The two links backup each other under failure



Example

```
router bgp 100
neighbor 204.70.4.77 remote-as 200
neighbor 204.70.4.77 ebgp-multihop 4
neighbor 204.70.4.77 update-source Loopback0
neighbor 204.70.4.77 version 4
neighbor 204.70.4.77 prefix-list filter-in in
neighbor 204.70.4.77 distribute-list 111 out
neighbor 204.70.4.77 route-map no-transit- in in
neighbor 204.70.4.77 route-map no-transit-out out

ip route 204.70.4.77 255.255.255.255 serial1/0
ip route 204.70.4.77 255.255.255.255 serial2/0
```