BGP Scaling Techniques

ISP/IXP Workshops

BGP Scaling Techniques

- How to scale iBGP mesh beyond a few peers?
- How to implement new policy without causing flaps and route churning?
- How to reduce the overhead on the routers?

BGP Scaling Techniques

- Soft reconfiguration/Route Refresh
- Peer groups
- Route flap dampening
- Route reflectors
- (Confederations)

Dynamic Reconfiguration

Soft Reconfiguration and Route Refresh

Soft Reconfiguration

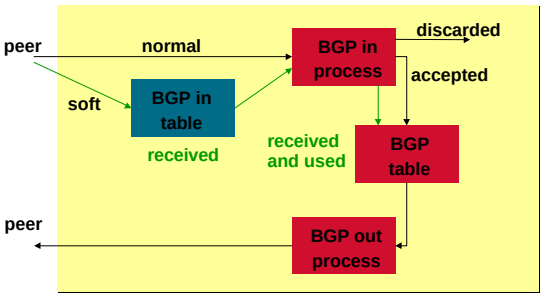
Problem:

- Hard BGP peer clear required after every policy change because the router does not store prefixes that are denied by a filter
- Hard BGP peer clearing consumes CPU and affects connectivity for all networks

Solution:

- Soft-reconfiguration

Soft Reconfiguration



Soft Reconfiguration

- New policy is activated without tearing down and restarting the peering session
- Per-neighbour basis
- Use more memory to keep prefixes whose attributes have been changed or have not been accepted

Configuring Soft reconfiguration

```
router bgp 100
 neighbor 1.1.1.1 remote-as 101
 neighbor 1.1.1.1 route-map infilter in
 neighbor 1.1.1.1 soft-reconfiguration inbound
```

! *Outbound does not need to be configured* !

Then when we change the policy, we issue an exec command

```
clear ip bgp 1.1.1.1 soft [in | out]
```


Managing Policy Changes

- `clear ip bgp <addr> [soft] [in|out]`

  <addr> may be any of the following

x.x.x.x	IP address of a peer
*	all peers
ASN	all peers in an AS
external	all external peers
peer-group <name>	all peers in a peer-group

Route Refresh Capability

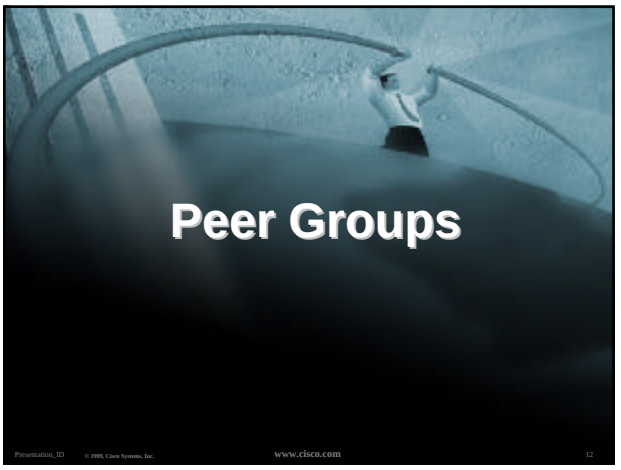
- Facilitates non-disruptive policy changes
- No configuration is needed
- No additional memory is used
- Requires peering routers to support "route refresh capability" - RFC2842
- clear ip bgp x.x.x.x in tells peer to resend full BGP announcement

Soft Reconfiguration vs Route Refresh

- Use Route Refresh capability if supported

  find out from "show ip bgp neighbor"

  uses much less memory
- Otherwise use Soft Reconfiguration



Peer Groups

Without peer groups

- iBGP neighbours receive same update
- Large iBGP mesh slow to build
- Router CPU wasted on repeat calculations

Solution - peer groups!

- Group peers with same outbound policy
- Updates are generated once per group

Peer Groups - Advantages

- Makes configuration easier
- Makes configuration less prone to error
- Makes configuration more readable
- Lower router CPU load
- iBGP mesh builds more quickly
- Members can have different inbound policy
- Can be used for eBGP neighbours too!

Configuring Peer Group

```
router bgp 100
 neighbor ibgp-peer peer-group
 neighbor ibgp-peer remote-as 100
 neighbor ibgp-peer update-source loopback 0
 neighbor ibgp-peer send-community
 neighbor ibgp-peer route-map outfilter out
 neighbor 1.1.1.1 peer-group ibgp-peer
 neighbor 2.2.2.2 peer-group ibgp-peer
 neighbor 2.2.2.2 route-map  infilter in
 neighbor 3.3.3.3 peer-group ibgp-peer
```

! *note how 2.2.2.2 has different inbound filter from peer-group* !

Configuring Peer Group

```
router bgp 109
 neighbor external-peer peer-group
 neighbor external-peer send-community
 neighbor external-peer route-map set-metric out
 neighbor 160.89.1.2 remote-as 200
 neighbor 160.89.1.2 peer-group external-peer
 neighbor 160.89.1.4 remote-as 300
 neighbor 160.89.1.4 peer-group external-peer
 neighbor 160.89.1.6 remote-as 400
 neighbor 160.89.1.6 peer-group external-peer
 neighbor 160.89.1.6 filter-list infilter in
```


Route Flap Dampening

- Route flap
  - Going up and down of path
  - Change in attribute
  - Ripples through the entire Internet
  - Wastes CPU
- Dampening aims to reduce scope of route flap propagation

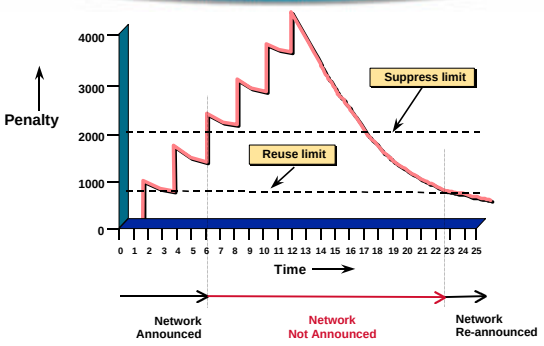
Route Flap Dampening (Continued)

- Requirements
  - Fast convergence for normal route changes
  - History predicts future behaviour
  - Suppress oscillating routes
  - Advertise stable routes
- Described in RFC2439

Route Flap Dampening - Operation

- Add penalty (1000) for each flap
- Exponentially decay penalty
  - half life determines decay rate
- Penalty above suppress-limit
  - do not advertise route to BGP peers
- Penalty decayed below reuse-limit
  - re-advertise route to BGP peers

Route Flap Dampening



Route Flap Dampening - Operation

- Only applied to inbound announcements from eBGP peers
- Alternate paths still usable
- Controlled by:
  - Half-life (default 15 minutes)
  - reuse-limit (default 750)
  - suppress-limit (default 2000)
  - maximum suppress time (default 30 minutes)

Flap Dampening: Enhancements

- Selective dampening based on
  - AS-path, Community, Prefix
- Variable dampening
  - recommendations for ISPs
  - http://www.ripe.net/docs/ripe-210.html
- Flap statistics

```
show ip bgp neighbor <x.x.x.x> [dampened-routes
| flap-statistics]
```


Configuring Route Flap Dampening

Fixed dampening

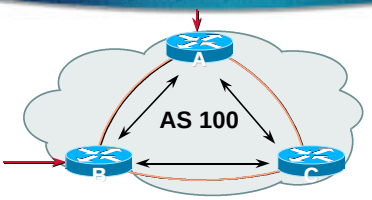
```
router bgp 100
 bgp dampening [<half-life> <reuse-value> <suppress-
  penalty> <maximum suppress time>]
```

Selective and variable dampening

```
 bgp dampening [route-map <name>]
  route-map <name> permit 10
   match ip address prefix-list FLAP-LIST
   set dampening [<half-life> <reuse-value>
  <suppress-penalty> <maximum suppress time>]
 ip prefix-list FLAP-LIST permit 192.0.2.0/24 le 32
```

Route Reflectors

Scaling iBGP mesh

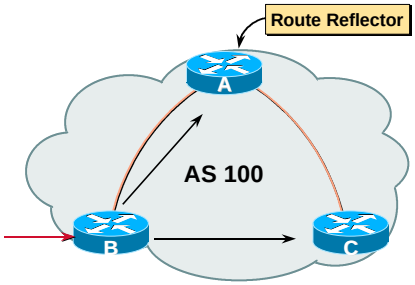


- Need to avoid routing information loop
- Solution should not change the current behaviour
- Two solutions

  Route reflector - simpler to deploy and run

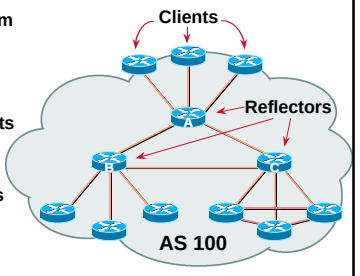
  Confederation - more complex to manage, corner case benefits

Route Reflector: Principle



Route Reflector

- Reflector receives path from clients and non-clients
- Selects best path
- If best path is from client, reflect to other clients and non-clients
- If best path is from non-client, reflect to clients only
- Non-meshed clients
- Described in RFC2796



Route Reflector Topology

- Divide the backbone into multiple clusters
- At least one route reflector and few clients per cluster
- Route reflectors are fully meshed
- Clients in a cluster could be fully meshed
- Single IGP to carry next hop and local routes

Route Reflectors: Loop Avoidance

- Originator_ID attribute

  Carries the RID of the originator of the route in the local AS (created by the RR)

- Cluster_list attribute

  The local cluster-id is added when the update is sent to (added by the RR)

  bgp cluster-id x.x.x.x

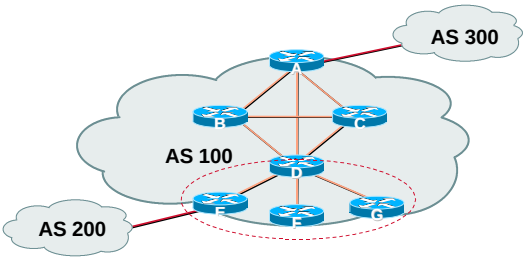
Route Reflector: Benefits

- Solves iBGP mesh problem
- Packet forwarding is not affected
- Normal BGP speakers co-exist
- Multiple reflectors for redundancy
- Easy migration
- Multiple levels of route reflectors

Route Reflectors: Migration

- Where to place the route reflectors?
  - Follow the physical topology!
  - This will guarantee that the packet forwarding won't be affected
- Configure one RR at a time
  - Eliminate redundant iBGP sessions
  - Place one RR per cluster

Route Reflector: Migration



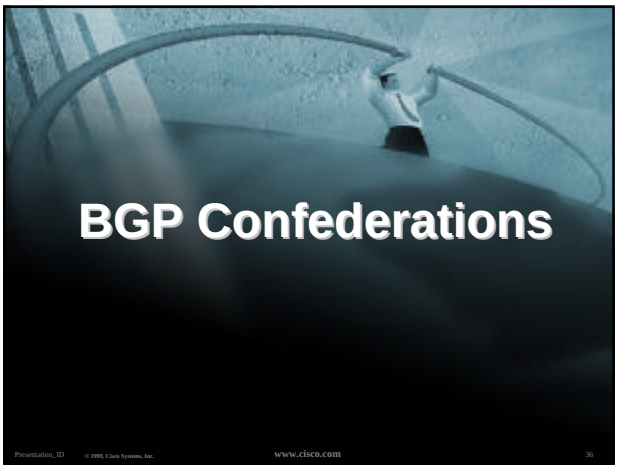
- Migrate small parts of the network, one part at a time.

Configuring a Route Reflector

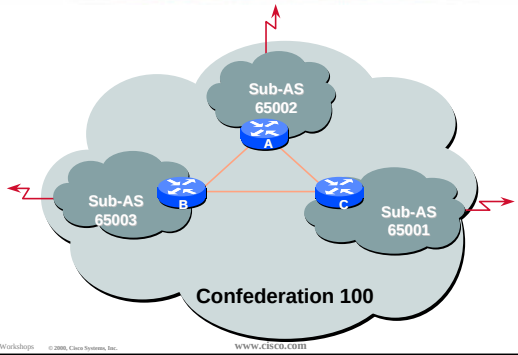
```
router bgp 100
 neighbor 1.1.1.1 remote-as 100
 neighbor 1.1.1.1 route-reflector-client
 neighbor 2.2.2.2 remote-as 100
 neighbor 2.2.2.2 route-reflector-client
 neighbor 3.3.3.3 remote-as 100
 neighbor 3.3.3.3 route-reflector-client
```


BGP Scaling Techniques

- These 4 techniques should be core requirements on all ISP networks
  - Soft reconfiguration/Route Refresh
  - Peer groups
  - Route flap dampening
  - Route reflectors



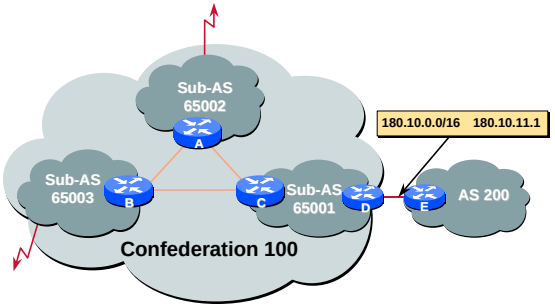
Confederations



Confederations: Principle

- Best path sent to neighbour sub-AS
- Packet forwarding depends on next hop
- IGP carries next hops and local networks
- Preserve next hop across sub-AS eBGP

Confederations: Next Hop



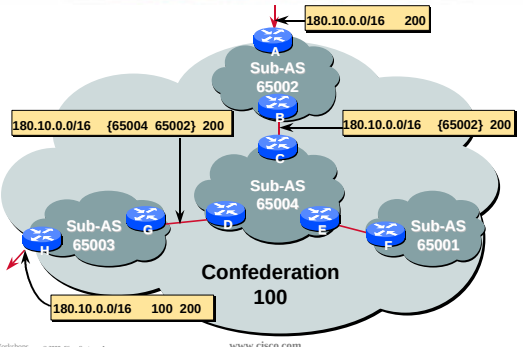
Confederation: Principle

- Local preference and MED influence path selection
- Preserve local preference and MED across sub-AS boundary
- Sub-AS eBGP path administrative distance

Confederations: Loop Avoidance

- Sub-AS traversed are carried as part of AS-path
- AS-sequence and AS path length
- Confederation boundary
- AS-sequence should be skipped during MED comparison

Confederations: AS-Sequence



Confederations: Benefits

- Solves iBGP mesh problem
- Packet forwarding not affected
- Can be used with route reflectors
- Policies could be applied to route traffic between sub-AS's

Confederations: Caveats

- Minimal number of sub-AS
- Sub-AS hierarchy
- Minimal inter-connectivity between sub-AS's
- Path diversity
- Difficult migration
  - BGP reconfigured into sub-AS
  - must be applied across the network