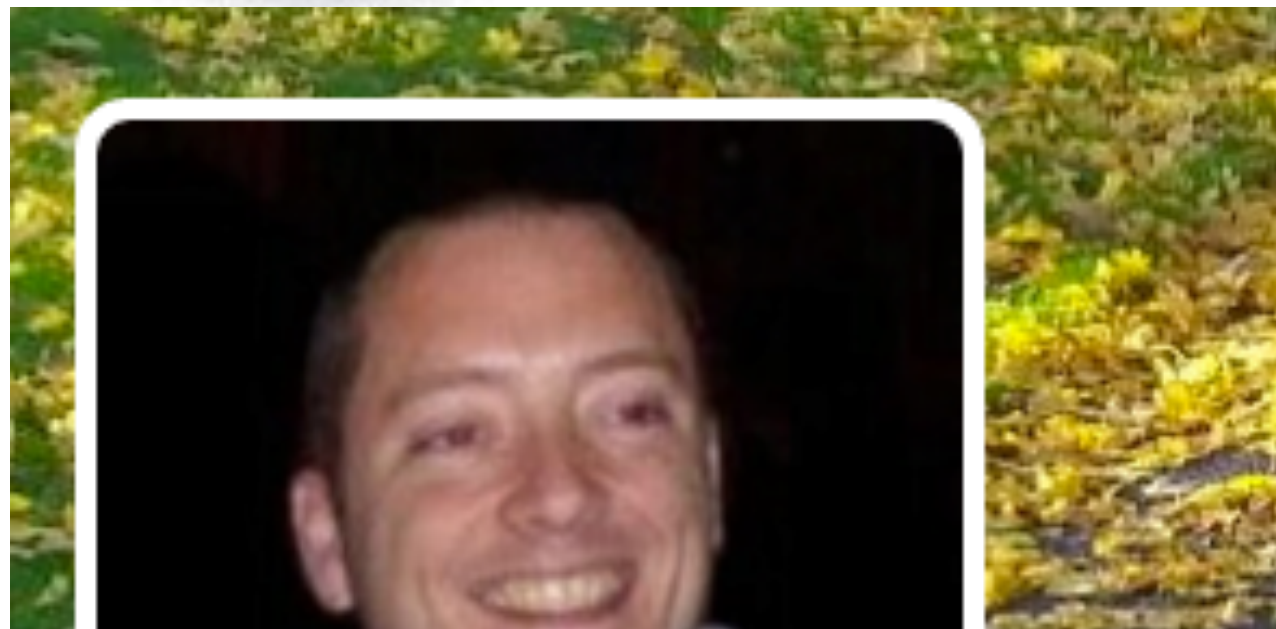# BGP 101
or
"Some Good BGP Practices"
(especially at IXPs)

Barry O'Donovan

INEX Operations

UK & Ireland Peering Forum 1.0

September 16th, 2015 - Sheffield, UK

- Member owned association, not for profit
- Island of Ireland exchange
- Six PoPs around Dublin, two infrastructures
  - LAN1 - recently fork-lift upgraded to Extreme x670-G2's and x460's
  - LAN2 - Brocade TI24X and Netiron CES, upgrade planned in 2016
  - 3 x Telecity, 2 x Interxion, 1 x Vodafone (C&W)
- Dark fibre rings connecting all PoPs (dual-hub and spoke)
  - MRV LD800 WDM active transmission equipment
  - New core links being installed as passive with coloured SFP+
- ~80 peering members
- Reseller program

- Just covering the basics
- Examples based on Cisco IOS and IPv4 only
- Setting up a BGP session
  - Securing a BGP session
  - Examples with route-map's and template's
- Simple traffic engineering
- Community tagging / next-hop filtering

- (e)BGP is a routing protocol that allows one network to signal to other networks what destinations can be reached through it.
- These networks are Autonomous Systems (AS)
  - identified by a 32-bit number called an AS Number / ASN
- These relationships are called peers / neighbors:
  - Transit - your upstream ISP
  - Peerings - typically settlement free via IXPs and PNIs
  - Customers - downstream networks of whom you are the ISP
  - Default route - gateway of last resort
- Default Free Zone (DFZ) - full internet routing table (~.5 million IPv4 routes)

Generally, BGP tries to select the best path by traversing the least number of networks (AS').

```
*   2.20.208.0/20      193.242.111.55  20940 i
*   2.20.240.0/23      193.242.111.55  20940 i
*   4.53.201.0/24      193.242.111.69  6939 18779 26481 i
*   5.2.16.0/23        193.242.111.80  5552 21472 i
*   5.2.16.0/21        193.242.111.80  5552 21472 i
```

```
barryo@rtr02> show route all 8.8.8.8

8.8.8.0/24
   *[BGP/170] 2w6d 03:32:56, MED 0, localpref 100, from 193.242.111.8
                   AS path: 15169 I
                 > to 193.242.111.57 via ge-0/0/1.0
    [BGP/170] 2w0d 18:58:53, MED 0, localpref 100, from 193.242.111.9
                   AS path: 15169 I
                 > to 193.242.111.57 via ge-0/0/1.0
    [BGP/170] 6d 10:25:25, MED 0, localpref 100
                   AS path: 3356 15169 I
                 > to 213.242.106.181 via ge-6/0/0.0
```

```
barryo@rtr02> show route all 8.8.8.8

8.8.8.0/24
    *[BGP/170] 2w6d 03:32:56, MED 0, localpref 100, from 193.242.111.8
                      AS path: 15169 I
                    > to 193.242.111.57 via ge-0/0/1.0
     [BGP/170] 2w0d 18:58:53, MED 0, localpref 100, from 193.242.111.9
                      AS path: 15169 I
                    > to 193.242.111.57 via ge-0/0/1.0
     [BGP/170] 6d 10:25:25, MED 0, localpref 100
                      AS path: 3356 15169 I
                    > to 213.242.106.181 via ge-6/0/0.0
```

```
barryo@rtr02> show route all 8.8.8.8

8.8.8.0/24
   *[BGP/170] 2w6d 03:32:56, MED 0, localpref 100, from 193.242.111.8
                    AS path: 15169 I
                  > to 193.242.111.57 via ge-0/0/1.0
    [BGP/170] 2w0d 18:58:53, MED 0, localpref 100, from 193.242.111.9
                    AS path: 15169 I
                  > to 193.242.111.57 via ge-0/0/1.0
    [BGP/170] 6d 10:25:25, MED 0, localpref 100
                    AS path: 3356 15169 I
                  > to 213.242.106.181 via ge-6/0/0.0
```

```
barryo@rtr02> show route all 8.8.8.8

8.8.8.0/24
   *[BGP/170] 2w6d 03:32:56, MED 0, localpref 100, from 193.242.111.8
                   AS path: 15169 I
                   > to 193.242.111.57 via ge-0/0/1.0
    [BGP/170] 2w0d 18:58:53, MED 0, localpref 100, from 193.242.111.9
                   AS path: 15169 I
                   > to 193.242.111.57 via ge-0/0/1.0
    [BGP/170] 6d 10:25:25, MED 0, localpref 100
                   AS path: 3356 15169 I
                   > to 213.242.106.181 via ge-6/0/0.0
```

- Layer 2 connectivity between routers
- Layer 3 subnet for IP communication
  - Typically /30 or /31 for single router transit peering
  - Typically a /25 - /22 for IXP peering LANs
- Routable address space (e.g. an allocation from RIPE)
- AS number

```
interface GigabitEthernet0/0
 description Link to INEX Peering LAN 1
 ip address 193.242.111.111 255.255.255.128
 no ip redirects
 no ip proxy-arp
 no mop enabled
```

- Our ASN is: 64510
- Advertising two prefixes:

```
ip route 192.0.2.0   255.255.255.0 Null0 254
ip route 203.0.113.0 255.255.255.0 Null0 254

interface Loopback0
  description Router handle (loopback address)
  ip address 192.0.2.0 255.255.255.255
```

```
router bgp 64510
  bgp router-id 192.0.2.0
  no bgp enforce-first-as
  bgp maxas-limit 50
  no bgp default ipv4-unicast
  bgp always-compare-med
  bgp deterministic-med
```

```
router bgp 64510
  ...
  address-family ipv4 [unicast]
    distance bgp 200 200 200
    network 192.0.2.0   mask 255.255.255.0
    network 203.0.113.0 mask 255.255.255.0
  exit-address-family
```

```
router bgp 64510
 ...
  neighbor 193.242.111.126 remote-as 2128
  neighbor 193.242.111.126 description INEX RC
  neighbor 193.242.111.126 password soopersecret

  address-family ipv4
    neighbor 193.242.111.126 activate
 exit-address-family
```

```
router bgp 64510
 ...
  neighbor 193.242.111.8 remote-as 43760
  neighbor 193.242.111.8 description INEX RS1
  neighbor 193.242.111.8 password soopersecret

  address-family ipv4
    neighbor 193.242.111.8 activate
exit-address-family
```

```
gw1#show bgp ipv4 unicast summary

Neighbor           V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down   State/PfxRcd
...
193.242.111.6     4          112    53470    58319 87880824    0    0 2w4d           2
193.242.111.8     4        43760  1846575   563990 87880824    0    0 25w4d      29931
193.242.111.9     4        43760   170109    41117 87880824    0    0 1w6d       29931
193.242.111.126 4           2128  4777695  8197048 87880852    0    0 31w3d          0
...
```

- Inbound Prefix Filters
- Outbound prefix filters
- AS path filters (CPU hog)
- Maximum Prefixes
- MD5 shared secret
- Next hop verification

- Inbound Prefix Filters
- Outbound prefix filters
- AS path filters (CPU hog)
- Maximum Prefixes
- MD5 shared secret
- Next hop verification

- Inbound Prefix Filters
  - Block your own routes (and subnets of)
  - Block the default route
  - Block martians (RFC1918, 5735, 6598)

```
ip prefix-list pl-v4-as64510 description My addresses
ip prefix-list pl-v4-as64510 seq 10 permit 192.0.2.0/24 le 32
ip prefix-list pl-v4-as64510 seq 15 permit 203.0.113.0/24 le 32

ip prefix-list pl-v4-martians description IPv4 Martians
ip prefix-list pl-v4-martians seq 10 permit 10.0.0.0/8 le 32
ip prefix-list pl-v4-martians seq 15 permit 127.0.0.0/8 le 32
...

ip prefix-list pl-v4-default description IPv4 Default
ip prefix-list pl-v4-default seq 10 permit 0.0.0.0/0
```

```
route-map rm-v4-ebgp-inex-in deny 10
  match ip address prefix-list pl-v4-as64510 pl-v4-martians pl-v4-default


route-map rm-v4-ebgp-inex-in permit 20


router bgp 64510
  address-family ipv4
    neighbor 193.242.111.8   route-map rm-v4-ebgp-inex-in in
    neighbor 193.242.111.126 route-map rm-v4-ebgp-inex-in in
```

Outbound Prefix Filters

- Only advertise your own prefixes
- BGP will share all learnt routes by default!

```
ip prefix-list pl-v4-ebgp-as64510-out description My addresses
ip prefix-list pl-v4-ebgp-as64510-out seq 10 permit 192.0.2.0/24
ip prefix-list pl-v4-ebgp-as64510-out seq 15 permit 203.0.113.0/24
```

```
route-map rm-v4-ebgp-inex-out permit 10
  match ip address prefix-list pl-v4-ebgp-as64510-out

route-map rm-v4-ebgp-inex-out deny 20


router bgp 64510
  neighbor 193.242.111.8   route-map rm-v4-ebgp-inex-out out
  neighbor 193.242.111.126 route-map rm-v4-ebgp-inex-out out
```

- Sets maximum number of prefixes accepted
- Simple **but prevents many problems (DFZ leak)**

```
router bgp 64510
 address-family ipv4
  neighbor 193.242.111.8 maximum-prefix 110000 restart 30
  neighbor 193.242.111.9 maximum-prefix 110000 restart 30
  neighbor 193.242.111.x maximum-prefix 20 restart 30
  neighbor 193.242.111.126 maximum-prefix 1 restart 30
```

- Neighbor configuration is repetitive
  - Repetition leads to error
- We'd traditionally use peer-group's
  - Was more than syntactic sugar - update efficiency
- Replaced in modern IOS with templates
  - Clean and consistent configuration; ease of maintenance; grouping of neighbour types

```
router bgp 64510
  template peer-session ps-v4-ebgp-inex
    description eBGP IXP (INEX) session policy
    timers 10 30
  exit-peer-session
```

```
router bgp 64510
  template peer-session ps-v4-ebgp-inex
    description eBGP IXP (INEX) session policy
    timers 10 30
  exit-peer-session
  template peer-policy pp-v4-ebgp-inex
    route-map rm-v4-ebgp-inex-in in
    route-map rm-v4-ebgp-inex-out out
    soft-reconfiguration-inbound
    next-hop-self
  exit-peer-policy
```

```
router bgp 64510
 template peer-session ps-v4-ebgp-inex
 template peer-policy pp-v4-ebgp-inex

 ...
 neighbor 193.242.111.126 remote-as 2128
 neighbor 193.242.111.126 description INEX RC
 neighbor 193.242.111.126 inherit peer-session ps-v4-ebgp-inex

 address-family ipv4
  neighbor 193.242.111.126 activate
  neighbor 193.242.111.126 inherit peer-policy pp-v4-ebgp-inex
 exit-address-family
```

- Example of usefulness with IXP maintenance:

```
router bgp 64510
 template peer-session ps-v4-ebgp-inex
  description eBGP IXP (INEX) session policy
  timers 10 30
  shutdown
 exit-peer-session
```

- You may want to influence traffic for reasons such as:
  - Membership at multiple IXPs
  - Membership at an IXP with multiple peering LANs
  - Congested links

- Prefer the path with the highest WEIGHT (Cisco only)
- **Prefer the path with the highest LOCAL_PREF**   **[easily manipulated]**
- Prefer the path that was locally originated via an IGP
- **Prefer the path with the shortest AS_PATH**   **[typical default decision]**
- Prefer the path with the lowest origin type
- **Prefer the path with the lowest MED**   **[easily manipulated]**
- Prefer eBGP over iBGP
- Prefer the oldest path
- Prefer the path from the router with lower router-id
- Prefer the path that comes from the lowest neighbor address

- Using local preference to prefer a particular link / peering LAN
  - Default local-preference on Cisco is 100

```
route-map rm-v4-ebgp-inex-in deny 10
  match ip address prefix-list pl-v4-as64510 pl-v4-martians pl-v4-default

route-map rm-v4-ebgp-inex-in permit 20
  set local-preference 300
```

- Using MEDs to influence inbound routing
  - Unlike local-preference, the **lower** MED is preferred

```
route-map rm-v4-ebgp-inex1-out permit 10
  match ip address prefix-list pl-v4-ebgp-as64510-out
  set metric 200

route-map rm-v4-ebgp-inex2-out permit 10
  match ip address prefix-list pl-v4-ebgp-as64510-out
  set metric 100
```

Route Servers & Communities

- Well known communities for route servers
  - Do not announce to a peer: 0:peer-as
  - Announce a route to a certain peer: 43760:peer-as
  - Don't announce to all peers: 0:43760
  - Announce a route to all peers: 43760:43760

- Well known communities for route servers
    - Do not announce to a peer: 0:peer-as
    - Announce a route to a certain peer: 43760:peer-as
    - Don't announce to all peers: 0:43760
    - Announce a route to all peers: 43760:43760
- E.g. to announce to only 64511 and 64512, tag with:
    - 0:43760 43760:64511 43760:64512

- Example: tagging communities to filter announcements:

```
route-map rm-v4-ebgp-inex-rs-out permit 10
  match ip address prefix-list pl-v4-ebgp-as64510-out
  set community 0:43760 43760:64511 43760:64512


route-map rm-v4-ebgp-inex-out deny 20
```

Note that should illustrate why using well planned template's and route-map's from the start is a Good Idea^{TM}

● Example: tagging communities to filter announcements:

```
ip bgp-community new-format

router bgp 64510
  template peer-policy pp-v4-ebgp-inex-rs
    inherit peer-policy pp-v4-ebgp-inex
    route-map rm-v4-ebgp-inex-rs-in in
    route-map rm-v4-ebgp-inex-rs-out out
    send-community
  exit-peer-policy
```

- Filtering received routes is a little different
  - A relatively easy way is on next-hop (i.e. peer address) filtering:

```
access-list 80 permit 193.242.111.x 255.255.255.255
access-list 80 permit 193.242.111.x 255.255.255.255

route-map rm-v4-ebgp-inex-rs-in deny 10
  match ip next-hop 80

…
```

Thank you!

Barry O'Donovan

INEX Operations

@barryo79 / @ComePeerWithMe

barry.odonovan@inex.ie