Peering Workshop

BGP

Barry O'Donovan
December 12th 2013
barry.odonovan@inex.ie

What is BGP

- BGP => Border Gateway Protocol
  - BGP – 1989 (RFC1105)
  - BGP-2 – 1990 (RFC1163)
  - BGP-3 – 1991 (RFC1267)
  - BGP-4 – 1995 (RFC1654, 1771, 4271)
- AS – Autonomous System: a network managed by a single entity; uniquely identified by an AS number (ASN)
- BGP is an EGP – Exterior Gateway Protocol
  - Sets up inter-AS routing
  - IGPs are used for intra-AS routing

What is BGP / Definitions

- BGP is the routing protocol that allows one network (AS) to signal to other networks what destinations can be reached through it
- These relationships are called peers / neighbors:
  - Transit – your *upstream* ISP
  - Peerings – settlement free; IXPs and PIs
  - Customer – you are the ISP
- Default route – gateway of last resort
- Default Free Zone (DFZ) – the full internet routing table

BGP Route Propagation Example (Contrived!)

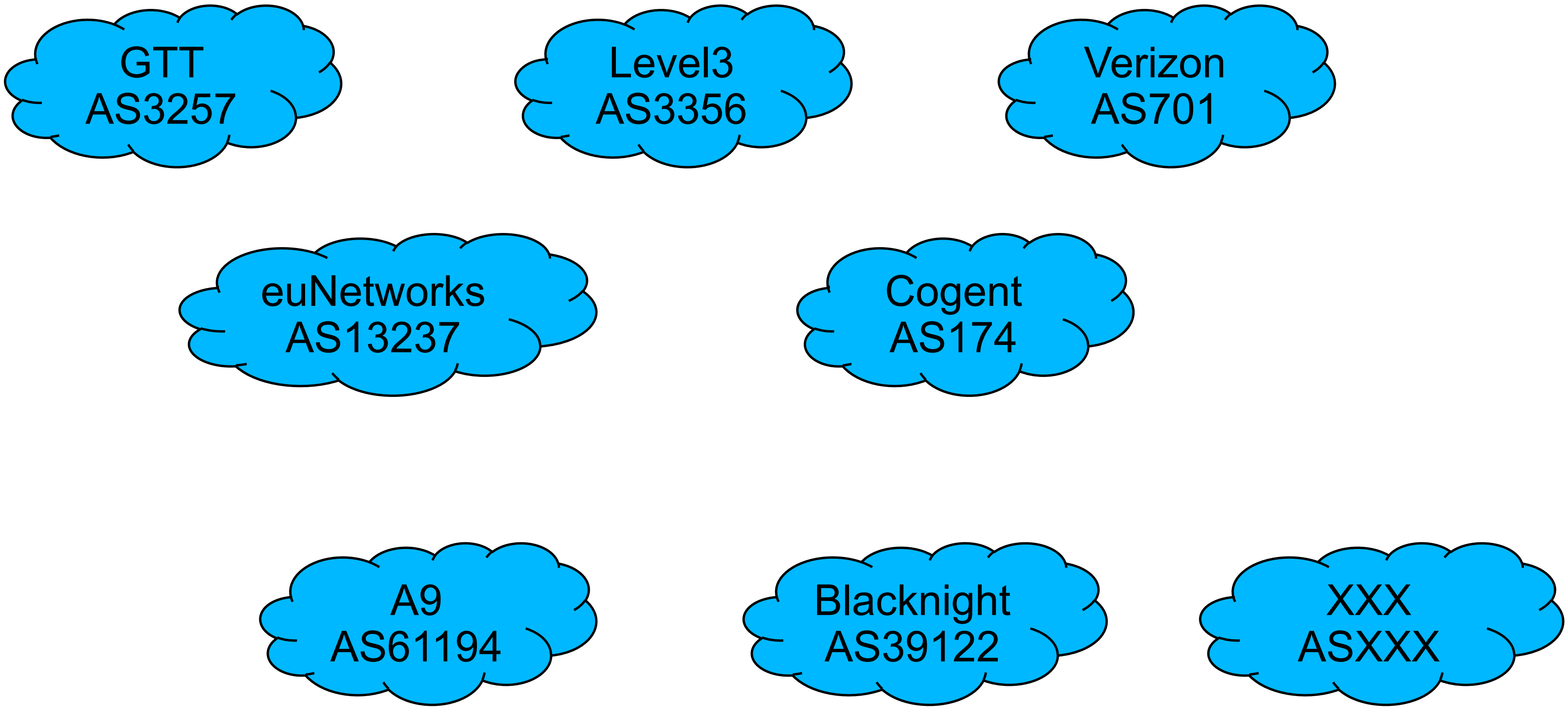
THE FOLLOWING EXAMPLE IS CONTRIVED!

FOR EXAMPLE BLACKNIGHT AND A9 HAVE MORE TRANSIT PROVIDERS THAN INDICATED.

THEY WERE SIMPLY CHOSEN AS THEY ARE INEX MEMBERS WITH A IP TRANSIT PROVIDER IN COMMON WHICH HELPS DESCRIBE HOW BGP WORKS.

ALSO - WITHOUT BEING TALKED THROUGH THESE SLIDES, THEY MAKE LITTLE SENSE…

BGP Route Propagation Example (Contrived!)



GTT
AS3257

Level3
AS3356

Verizon
AS701

euNetworks
AS13237

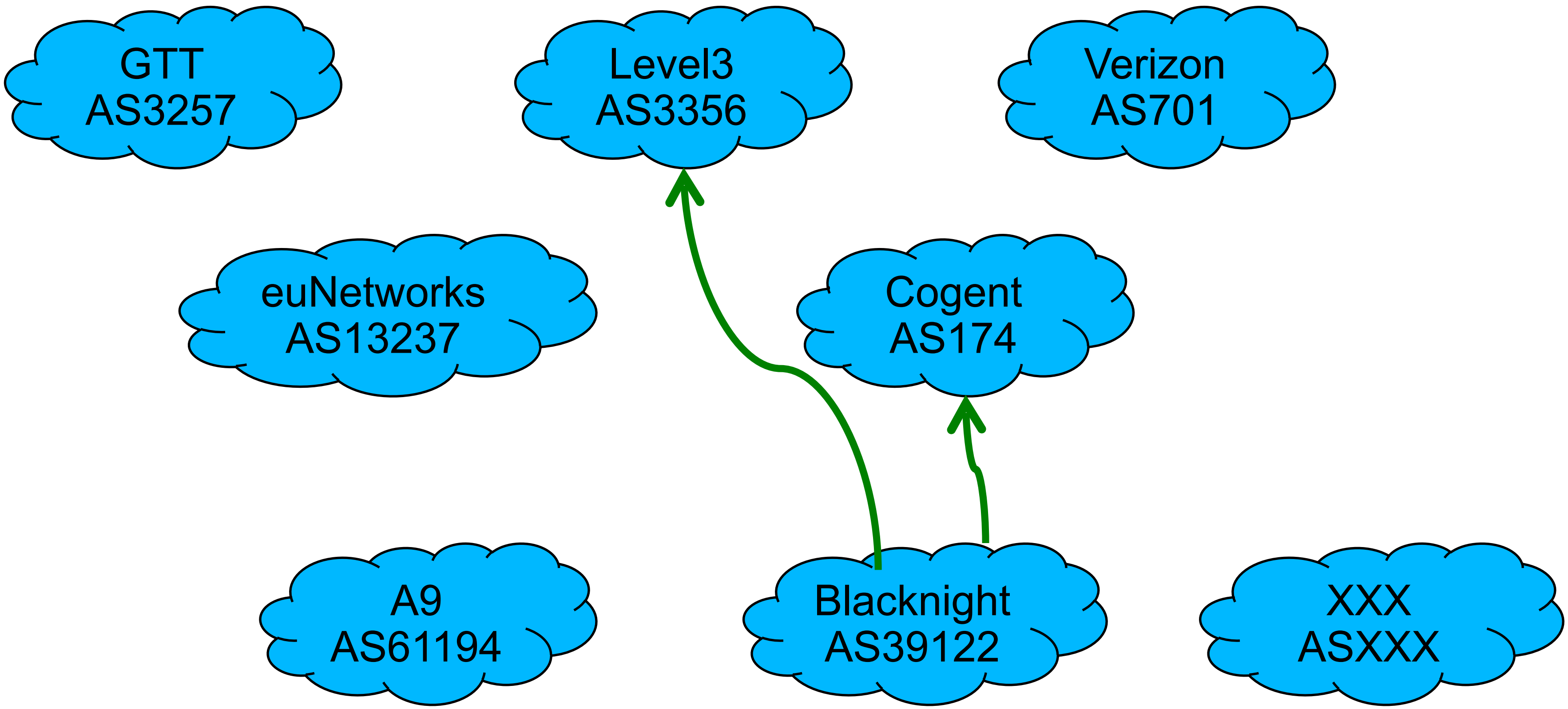
Cogent
AS174

A9
AS61194

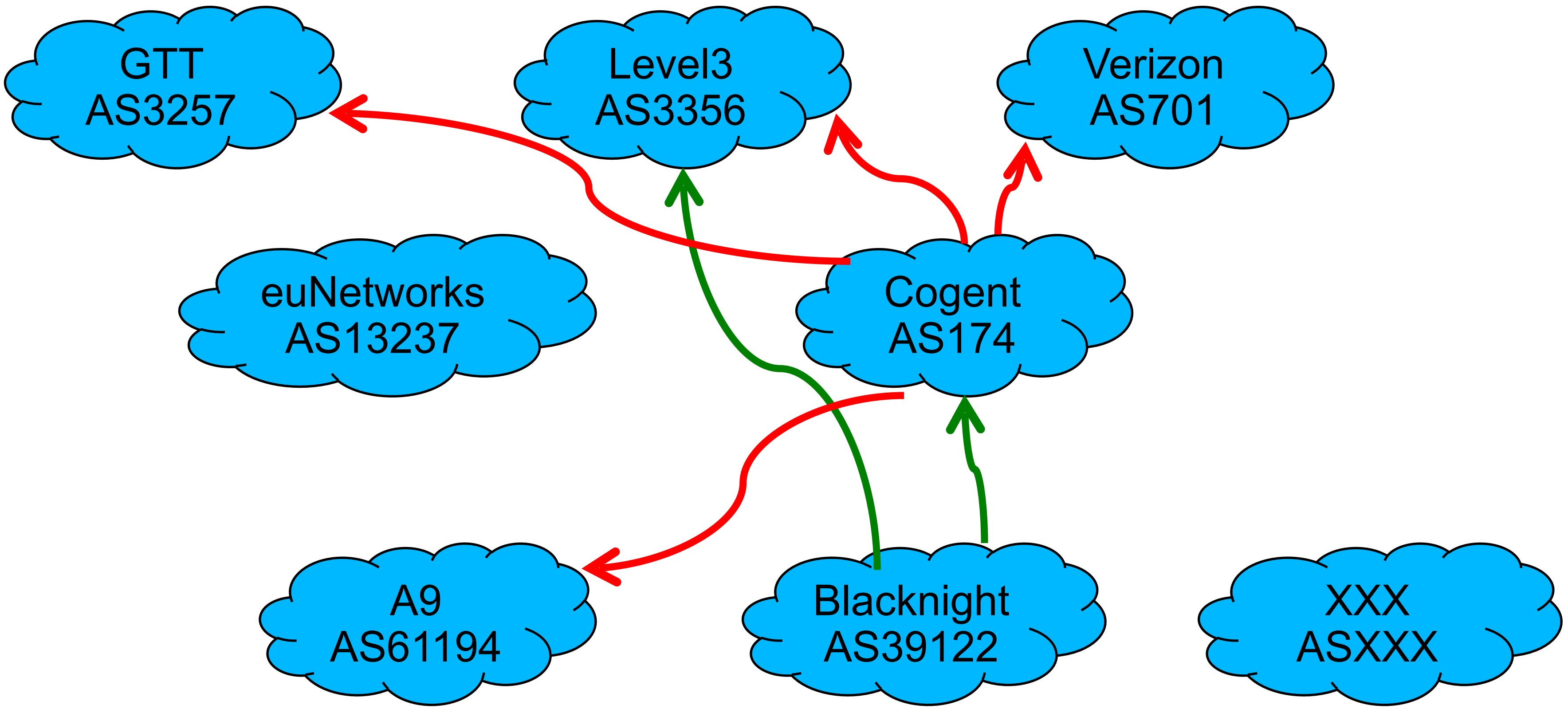
Blacknight
AS39122

XXX
ASXXX

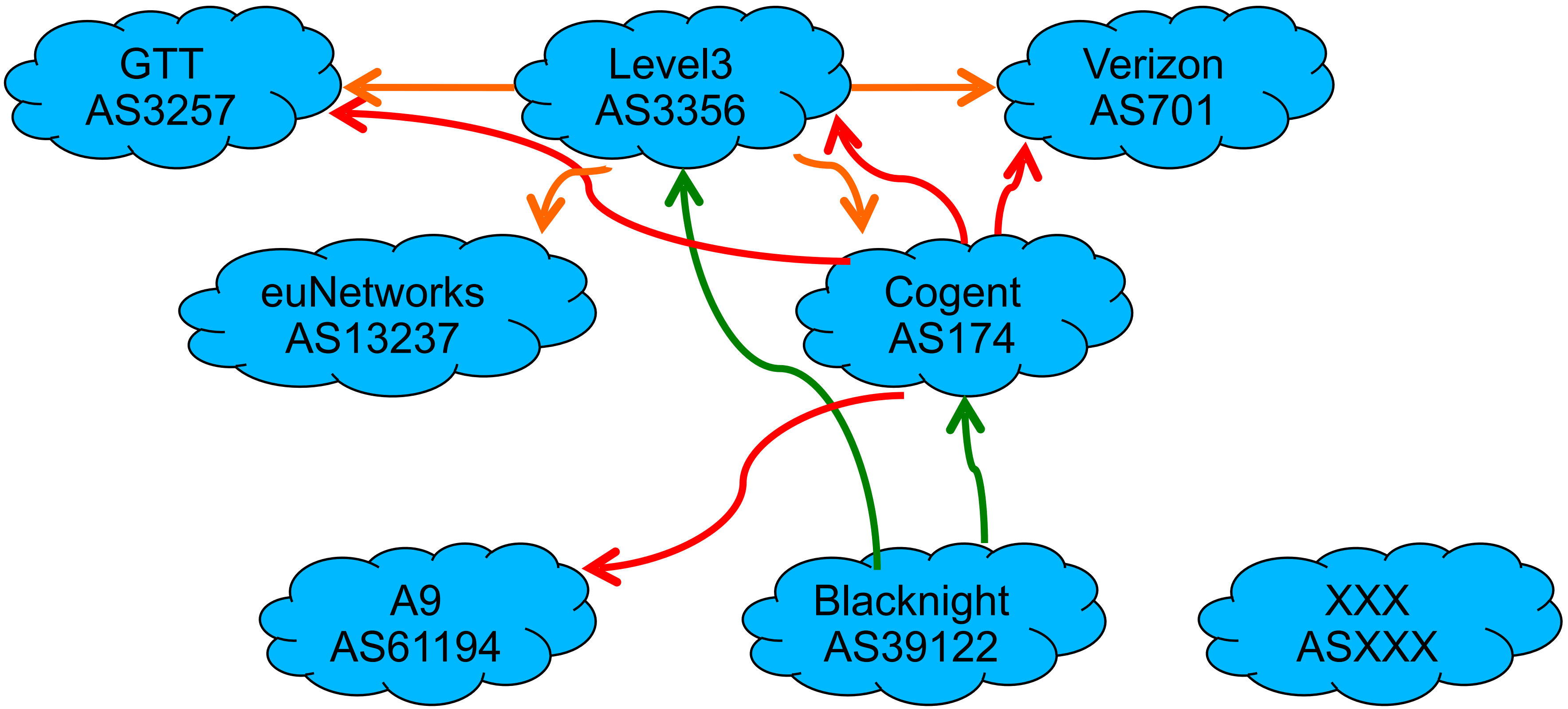
BGP Route Propagation Example (Contrived!)



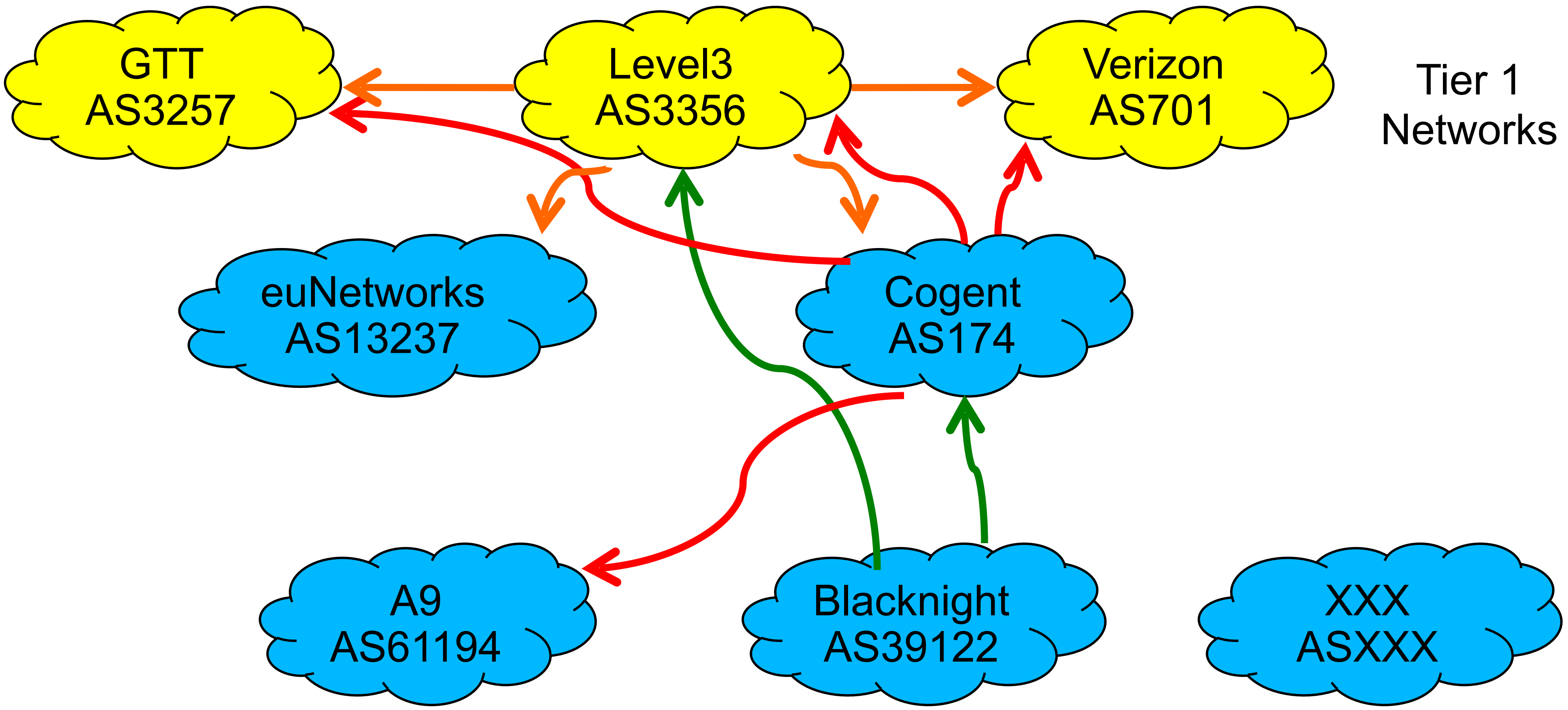
BGP Route Propagation Example (Contrived!)



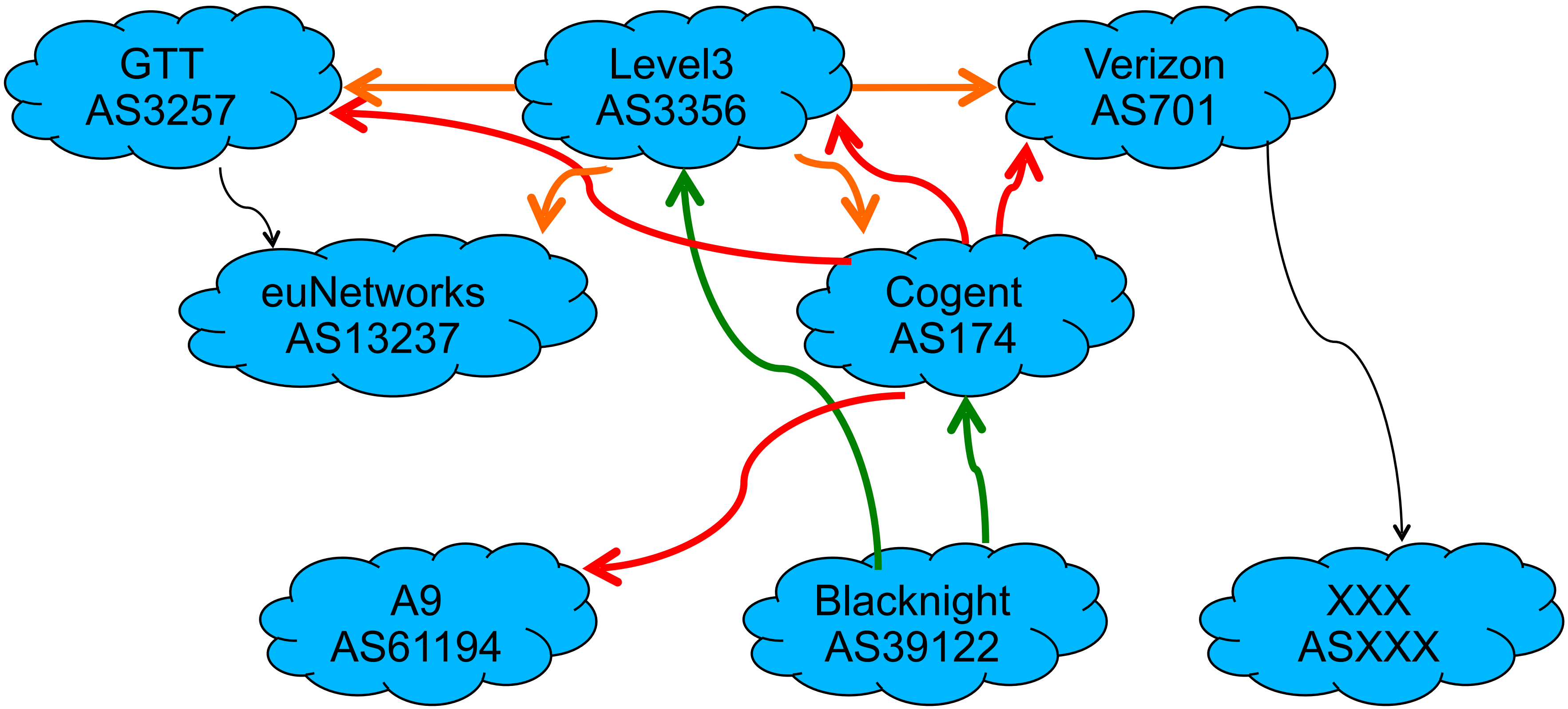
BGP Route Propagation Example (Contrived!)



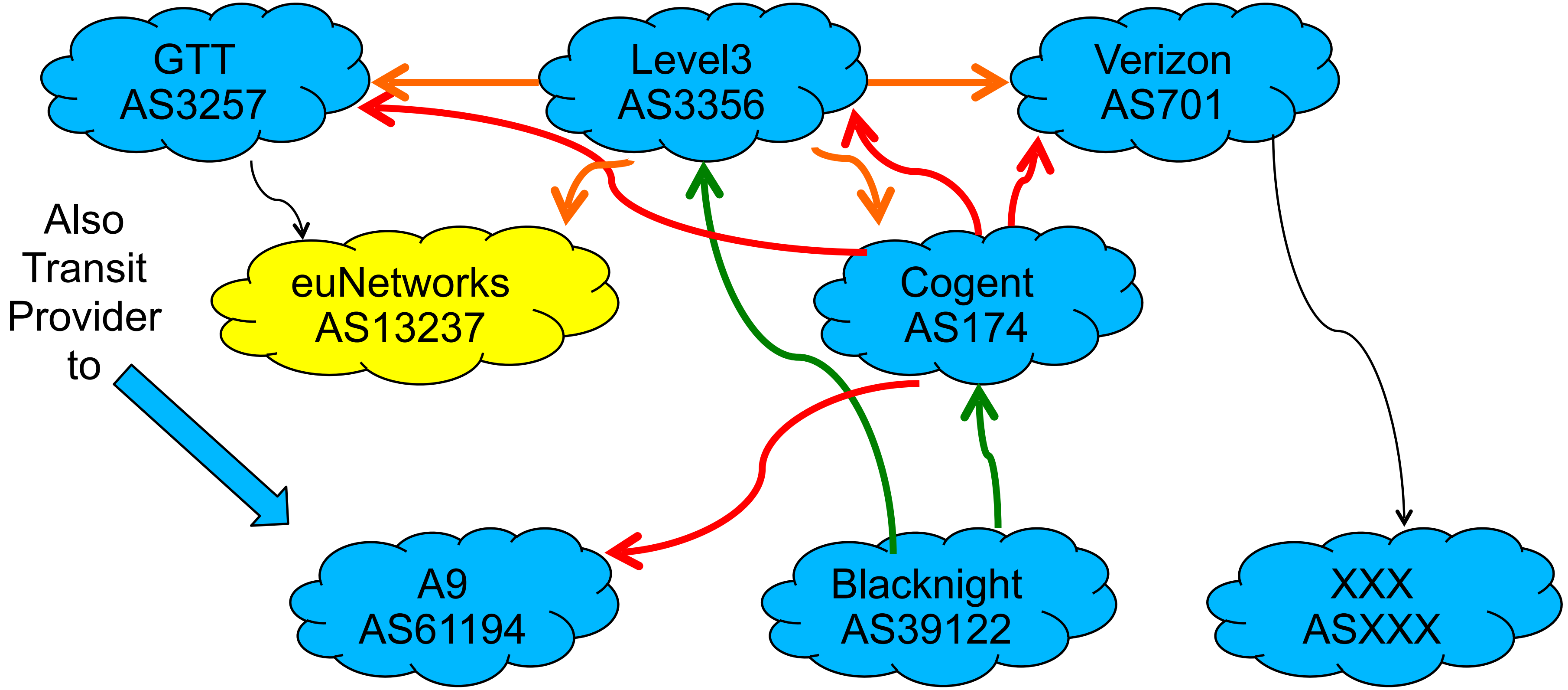
BGP Route Propagation Example (Contrived!)



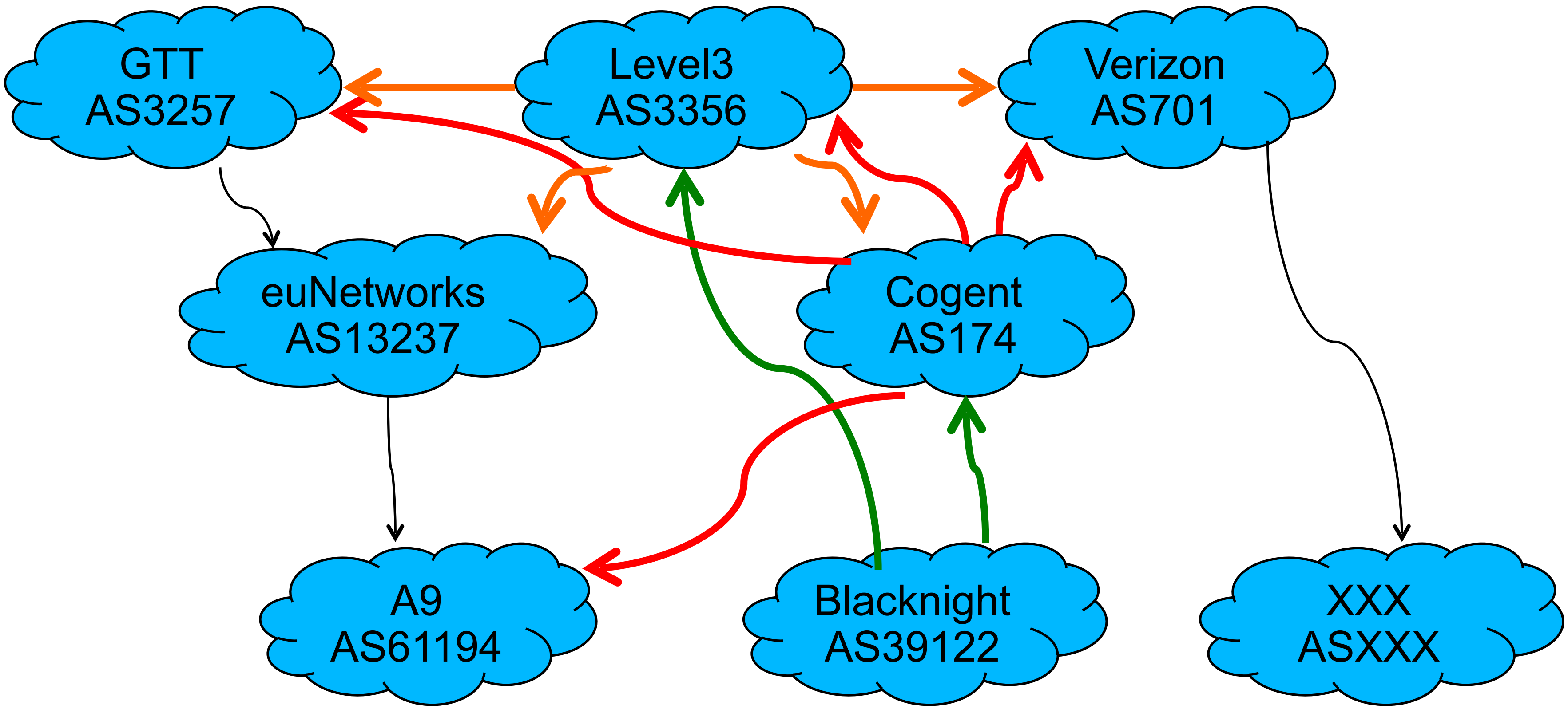
BGP Route Propagation Example (Contrived!)



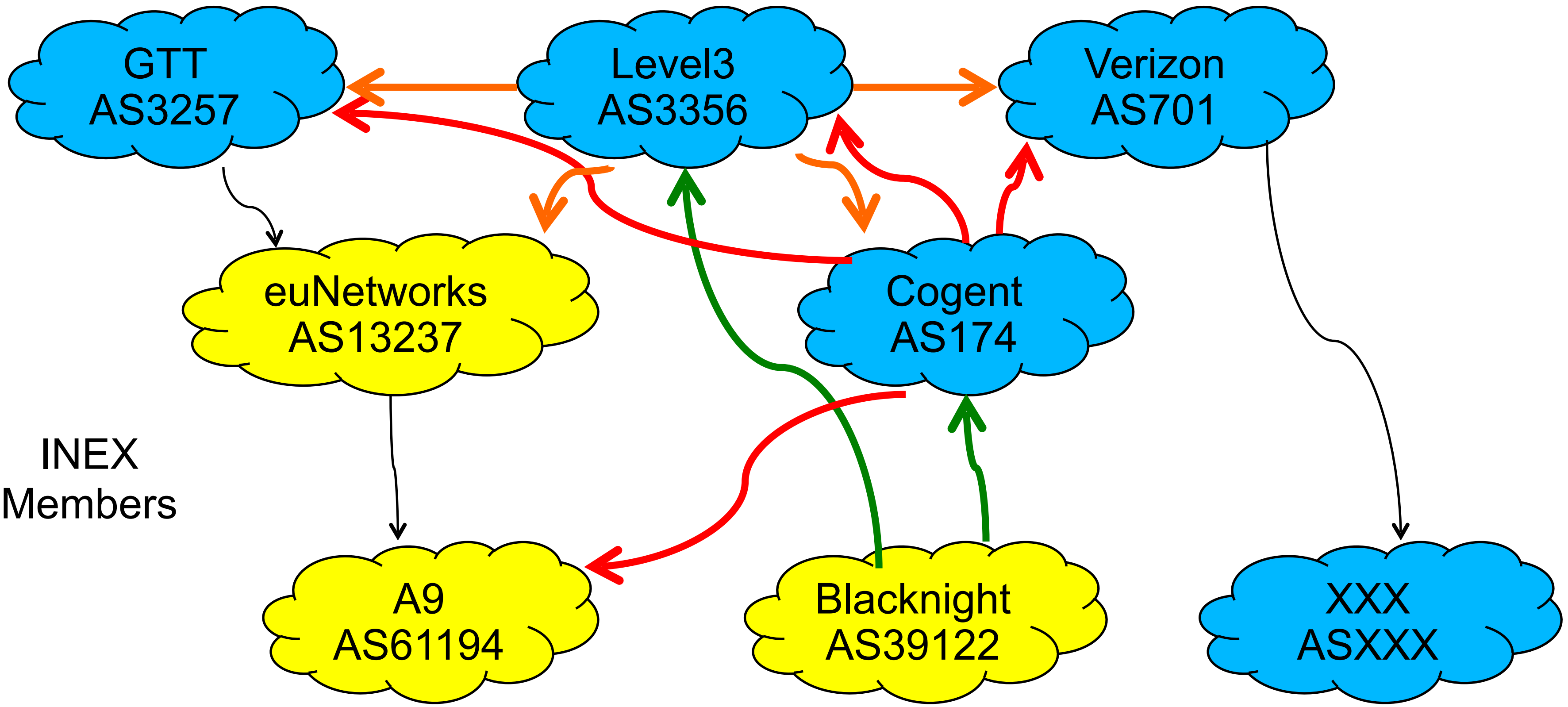
BGP Route Propagation Example (Contrived!)



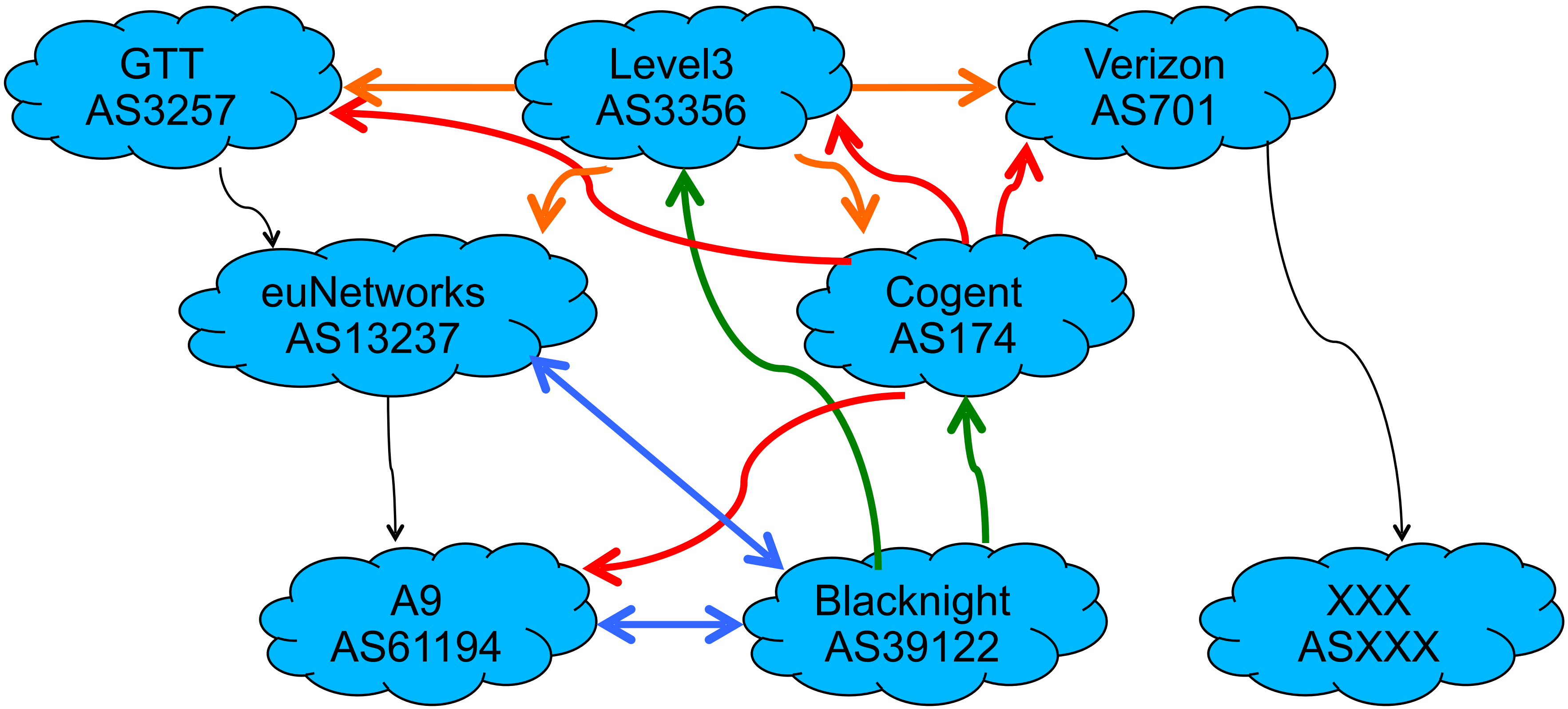
BGP Route Propagation Example (Contrived!)



BGP Route Propagation Example (Contrived!)



BGP Route Propagation Example (Contrived!)



BGP – What We’ll Look At

- Configuring a BGP session – step by step
- Securing a BGP session
- Route (Best Path) selection algorithm
- Routing examples
- Traffic shaping
  - Local preferences
  - MEDs
  - AS path prepending

BGP – What We Will NOT Look At

- iBGP
- Multihop eBGP
- IGPs and redistribution
- Protocol internals
- Route reflectors
- Communities
- Examples will be IPv4 only
- Examples will be Cisco IOS

Ingredients for a BGP Session

- Layer 2 connectivity between routers
- Layer 3 subnet for communication
  - E.g. 193.242.111.0/25
  - Typically a /30 for single router IPT
  - Or /29 for “full mesh” peering with two routers each
- Routes to advertise
- AS number

Ingredients for a BGP Session

- Security
  - Inbound prefix filters
  - Outbound prefix filters
  - AS path filters (CPU hog)
  - MD5 shared secret
  - Maximum prefixes
  - Next hop verification

Configuring: Interface

```
interface GigabitEthernet0/0
 description Link to INEX Peering LAN 1
 ip address 193.242.111.X 255.255.255.128
 no ip redirects
 no ip proxy-arp
 duplex full
 speed 1000
 ipv6 address 2001:7F8:18::X/64
 ipv6 enable
 ipv6 nd ra suppress
 no ipv6 redirects
```

Configuring: Your Routes & ASN

- Our ASN is: 65550
- We want to advertise:
  - 192.0.2.0/24
  - 203.0.113.0/24
- We need a *null* route and loopback:

```
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 Loopback address for router handles
  ip address 192.0.2.0 255.255.255.255
```

Configuring: BGP Boilerplate

```
router bgp 65550
  bgp router-id 192.0.2.0
  no bgp enforce-first-as
  bgp maxas-limit 50
  no bgp default ipv4-unicast

  address-family ipv4
    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
```

Configuring: First Peer – INEX Route Collector

```
router bgp 65550

  neighbor 193.242.111.126 remote-as 2128
  neighbor 193.242.111.126 description INEX Route Collector
  neighbor 193.242.111.126 password soopersecret

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

Configuring: Second Peer INEX Route Server #1

```
router bgp 65550

  neighbor 193.242.111.8 remote-as 43760
  neighbor 193.242.111.8 description INEX Route Server 1
  neighbor 193.242.111.8 password soopersecret

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

Sample Sessions on a INEX Member Router

# show bhp ipv4 unicast summary

```
Neighbor        Spk    AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down  St/PfxRcd
149.6.Z.ZZZ       0   174 3416949   31673 14859079    0    0     1w5d     466726
193.242.111.6     0   112   63408   63409 14859079    0    0     1w3d          1
193.242.111.8     0 43760   38844   31713 14859079    0    0     3w1d       2008
193.242.111.9     0 43760   46524   31705 14859079    0    0     2w0d       2002
193.242.111.16    0  1213  127295  126836 14859079    0    0     3w1d         21
…
193.242.111.126   0  2128   34875   31713 14859079    0    0     3w1d          2
```

Configuring: Inbound Filters

```
ip prefix-list pl-bgp-in description Routes we filter from
    BGP neighbors
ip prefix-list pl-bgp-in seq 10 deny 192.0.2.0/24 le 32
ip prefix-list pl-bgp-in seq 20 deny 203.0.113.0/24 le 32
ip prefix-list pl-bgp-in seq 30 deny 10.0.0.0/8 le 32
ip prefix-list pl-bgp-in seq 40 deny 192.168.0.0/16 le 32
ip prefix-list pl-bgp-in seq 50 deny 172.16.0.0/12 le 32
ip prefix-list pl-bgp-in seq 60 deny 127.0.0.0/8 le 32
…
ip prefix-list pl-bgp-in seq 900 deny 0.0.0.0/0
ip prefix-list pl-bgp-in seq 999 permit 0.0.0.0/0 le 32
```

Configuring: Outbound Filters

```
ip prefix-list pl-bgp-out description Routes we advertise
      over BGP
ip prefix-list pl-bgp-out seq 10 permit 192.0.2.0/24 le 32
ip prefix-list pl-bgp-out seq 20 permit 203.0.113.0/24 le 32
ip prefix-list pl-bgp-out seq 30 deny 0.0.0.0/0 le 32
```

Configuring: Applying Filters

```
router bgp 65550

  address-family ipv4

    neighbor 193.242.111.8 prefix-list pl-bgp-in in
    neighbor 193.242.111.8 prefix-list pl-bgp-out out

    neighbor 193.242.111.126 prefix-list pl-bgp-in in
    neighbor 193.242.111.126 prefix-list pl-bgp-out out

  exit-address-family
```

Configuring: Maximum Prefixes

- Sets the maximum number of prefixes accepted in a BGP session
- Simple tool but prevents many problems - particularly DFZ leeks

```
router bgp 65550
  address-family ipv4
    neighbor 193.242.111.8 maximum-prefix 20000 restart 5
    neighbor 193.242.111.126 maximum-prefix 20 restart 5
  exit-address-family
```

- INEX recommends 200 as a sane default for INEX peers
- IXP Manager will show if more is required

Configuring: Peer Groups

```
router bgp 65550
  neighbor pg-inex1 peer-group
  neighbor pg-inex1 description INEX LAN1 peer template
  neighbor pg-inex1 timers 10 30
  neighbor pg-inex2 peer-group
  neighbor pg-inex2 description INEX LAN2 peer template
  …
  address-family ipv4
    neighbor pg-inex1 maximum-prefix 200 restart 5
    neighbor pg-inex1 prefix-list pl-bgp-in in
    neighbor pg-inex1 prefix-list pl-bgp-out out
    neighbor pg-inex1 soft-reconfiguration inbound
  exit-address-family
```

Configuring: Using Peer Groups

```
router bgp 65550
  neighbor 193.242.111.8 remote-as 43760
  neighbor 193.242.111.8 description INEX Route Server 1
  neighbor 193.242.111.8 peer-group pg-inex1
  neighbor 193.242.111.9 remote-as 43760
  neighbor 193.242.111.9 description INEX Route Server 2
  neighbor 193.242.111.9 peer-group pg-inex1
  address-family ipv4
    neighbor 193.242.111.8 maximum-prefix 20000 restart 5
    neighbor 193.242.111.8 activate
    neighbor 193.242.111.9 maximum-prefix 20000 restart 5
    neighbor 193.242.111.9 activate
  exit-address-family
```

Peer Groups

- More than syntactic sugar – update processing more efficient
- Keeps your configuration clean and consistent
- Ensures you won't forget prefix-lists, etc
- Create peer-groups for IXPs, IPT providers and customers
- Also allows ease of maintenance:

```
router bgp 65550
 neighbor pg-inex1 shutdown
```

BGP Best Path Selection Algorithm - Cisco

- Prefer the path with the highest WEIGHT (Cisco only)
- Prefer the path with the highest LOCAL_PREF (def: 100)
- Prefer the path that was locally originated via an IGP
- Prefer the path with the shortest AS_PATH
- Prefer the path with the lowest origin type
- Prefer the path with the lowest MED
- 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

(some other steps omitted)

BGP Best Path Selection Algorithm - Cisco

- ~~Prefer the path with the highest WEIGHT (Cisco only)~~
- **Prefer the path with the highest LOCAL_PREF**
- ~~Prefer the path that was locally originated via an IGP~~
- **Prefer the path with the shortest AS_PATH**
- ~~Prefer the path with the lowest origin type~~
- **Prefer the path with the lowest MED**
- 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

Typical default decision. **What you can effect.**

BGP Best Path Selection Algorithm - Cisco

```
gw1#sh bgp ipv4 unicast 46.245.208.0
BGP routing table entry for 46.245.208.0/21, …
Paths: (4 available, best #3, table default)
  61194
    193.242.111.74 from 193.242.111.9 (193.242.111.9)
      Origin IGP, localpref 100, valid, external
  1213 61194
    193.242.111.74 from 193.242.111.16 (193.1.238.129)
      Origin IGP, localpref 50, valid, external
  61194
    193.242.111.74 from 193.242.111.8 (193.242.111.8)
      Origin IGP, localpref 100, valid, external, best
  61194
    193.242.111.74 from 193.242.111.126 (193.242.111.227)
      Origin IGP, metric 0, localpref 100, valid, internal
```

Traffic Shaping / Engineering / Load Balancing

- Using local pref to force a preferred route via a peer
  - Ensure all routes learnt from INEX LAN2 go via LAN2

```
route-map rm-prefer-inex2-out
  set local-preference 300

router bgp 65550
  address-family ipv4
    neighbor pg-inex2 route-map rm-prefer-inex2-out in
  exit-address-family
```

Traffic Shaping / Engineering / Load Balancing

- Using MEDs to influence inbound routing
  - Influence routes sent via INEX LAN2 to prefer LAN2
  - Remember – the lower MED wins!

```
route-map rm-deprefer-inex1-in
  set metric 200

route-map rm-prefer-inex2-in
  set metric 100
```

Traffic Shaping / Engineering / Load Balancing

```
router bgp 65550
  address-family ipv4
    neighbor pg-inex1 route-map rm-deprefer-inex1 out
    neighbor pg-inex2 route-map rm-prefer-inex2 out
  exit-address-family
```

Traffic Shaping / Engineering / Load Balancing

- Using AS Path prepending to *devalue* an IPT provider

```
route-map rm-add-two-hops
  description Increase AS path length by 2 hops
  set as-path prepend 65550 65550

router bgp 65550
  address-family ipv4
    neighbor 1.2.3.4 route-map rm-add-two-hops out
  exit-address-family
```

Enough BGP!

-

General Security

BCP38 – Defeating DOS Attacks via IP Source Address Spoofing

- http://tools.ietf.org/html/bcp38
- In a nutshell:

All traffic originating from your network should have a source address within your network.

I.e. block spoofed addresses.

In large service provider networks, typically done via uRPF

```
ip verify unicast source reachable-via {rx | any}
```

Poor Man’s uRPF at the Network Edge

```
ip access-list extended world-out
  remark Drop spoofed traffic leaving the network
  permit ip 192.0.2.0 0.0.0.255 any
  permit ip 203.0.113.0 0.0.0.255 any
  # allow peer IP ranges for BGP and ICMP
  deny ip any any log

interface GigabitEthernet0/0
  ip access-group world-out out
```

Poor Man’s uRPF at the Network Edge
Protecting Against Inbound Spoofing

```
ip access-list extended world-in
  remark Drop spoofed traffic entering the network
  deny ip 192.0.2.0 0.0.0.255 any log-input
  deny ip 203.0.113.0 0.0.0.255 any log-input
  permit ip any 192.0.2.0 0.0.0.255
  permit ip any 203.0.113.0 0.0.0.255
  # allow peer IP ranges for BGP and ICMP
  deny ip any any log-input

interface GigabitEthernet0/0
  ip access-group world-in in
```

RIPE Objects

RIPE Objects

- RIPE will have assigned you an ASN object:

```
aut-num:        AS39122
as-name:        BLACKNIGHT-AS
descr:          Blacknight Internet Solutions Ltd
org:            ORG-BISL2-RIPE
…
```

RIPE Objects

- If you plan to offer IPT to your customers, create an AS-SET:

```
as-set:          AS-BLACKNIGHT
descr:           Blacknight Solutions AS
members:         AS39122 #Blacknight
members:         AS42909 #Community DNS
members:         AS48410 #Protocol
members:         AS49567 #Aptus
tech-c:          BK1905-RIPE
admin-c:         BK1906-RIPE
mnt-by:          MNT-BLACKNIGHT
source:          RIPE # Filtered
```

w

- **If you want the route servers to accept your prefixes – create route[6]: objects:**

```
route:          81.17.240.0/20
descr:          IE-BLACKNIGHT-PA
origin:         AS39122
mnt-by:         MNT-BLACKNIGHT
source:         RIPE # Filtered

route6:         2a01:a8::/32
descr:          IE-BLACKNIGHT-PA-IPV6
origin:         AS39122
mnt-by:         MNT-BLACKNIGHT
source:         RIPE # Filtered
```