BGP Best Current Practices

ISP/IXP Workshops

Last updated 4th April 2025

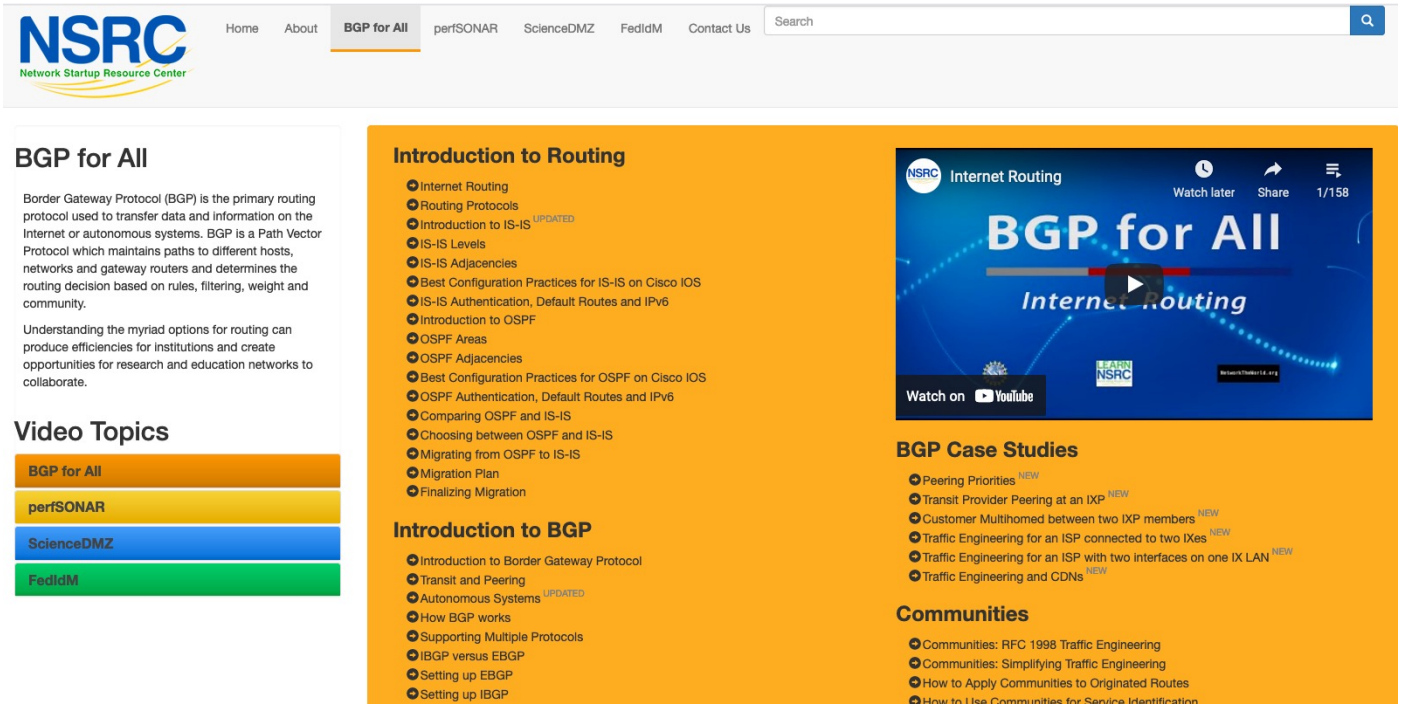
Acknowledgements

- This material originated from the Cisco ISP/IXP Workshop Programme developed by Philip Smith & Barry Greene
- Use of these materials is encouraged as long as the source is fully acknowledged and this notice remains in place
- Bug fixes and improvements are welcomed
  - Please email *workshop (at) bgp4all.com*

Philip Smith

BGP Videos

- NSRC has made a video recording of this presentation, as part of a library of BGP videos for the whole community to use:
  - https://learn.nsrc.org/bgp#bgp_best_practices



Configuring BGP

Where do we start?

Cisco IOS Good Practices

- Network Operators should start off with the following BGP commands as a basic template:

```
router bgp 64511
 bgp deterministic-med
 no bgp default ipv4-unicast
 distance bgp 200 200 200
 no synchronization
 no auto-summary
```

Replace with public ASN

Turn off IOS assumption that all neighbours will exchange IPv4 prefixes

Make EBGP and IBGP distance the same & more than any IGP

EBGP Default Behaviour

- Industry standard is described in RFC8212
  - https://tools.ietf.org/html/rfc8212
  - External BGP (EBGP) Route Propagation Behaviour without Policies
- NB: BGP in Cisco IOS is permissive by default
  - This is contrary to industry standard and RFC8212
- Configuring BGP peering without using filters means:
  - All best paths on the local router are passed to the neighbour
  - All routes announced by the neighbour are received by the local router
  - Can have disastrous consequences (see RFC8212)

EBGP Default Behaviour

- Best practice is to ensure that each EBGP neighbour has inbound and outbound filter applied:

```
router bgp 64511
 address-family ipv4
  neighbor 100.64.0.1 remote-as 64510
  neighbor 100.64.0.1 prefix-list as64510-in in
  neighbor 100.64.0.1 prefix-list as64510-out out
  neighbor 100.64.0.1 activate
```

EBGP Default Behaviour

- FRR turns on RFC8212 support by default:
  - https://frrouting.org/

```
frr.pfs.lab(config)# router bgp 64512 view LAB
frr.pfs.lab(config-router)# bgp ?
<snip>
ebgp-requires-policy          Require in and out policy for eBGP peers (RFC8212)
<snip>
```

- No prefixes will be sent or received to external peers in the absence of inbound and outbound policy

What is BGP for??

What is an IGP not for?

BGP versus OSPF/ISIS

- Internal Routing Protocols (IGPs)
  - Examples are IS-IS and OSPF
  - Used for carrying **infrastructure** addresses
  - NOT used for carrying Internet prefixes or customer prefixes
  - Design goal is to **minimise** number of prefixes in IGP to aid **scalability** and **rapid convergence**

BGP versus OSPF/IS-IS

- BGP is used
  - Internally (IBGP)
  - Externally (EBGP)
- IBGP is used to carry:
  - Some/all Internet prefixes across backbone
  - Customer prefixes
- EBGP is used to:
  - Exchange prefixes with other ASes
  - Implement routing policy

BGP versus OSPF/IS-IS

- DO NOT:
  - Distribute BGP prefixes into an IGP
  - Distribute IGP routes into BGP
  - Use an IGP to carry customer prefixes
- **YOUR NETWORK WILL NOT SCALE**

Aggregation

Aggregation

- Aggregation means announcing the address block received from the RIR to the other ASes connected to your network
- Subprefixes of this aggregate may be:
  - Used internally in the provider network
  - Announced to other ASes to aid with multihoming
- Too many operators are still thinking about class Cs, resulting in a proliferation of /24s in the Internet routing table
  - April 2025: 602052 /24s in IPv4 table of 979075 prefixes
- The same is happening for /48s with IPv6
  - April 2025: 97816 /48s in IPv6 table of 208948 prefixes

Configuring Aggregation – Cisco IOS

- Service Provider has 100.66.0.0/19 address block
- To put into BGP as an aggregate:

```
router bgp 64511
 address-family ipv4
  network 100.66.0.0 mask 255.255.224.0
ip route 100.66.0.0 255.255.224.0 null0
```

- The static route is a "pull up" route
  - More specific prefixes within this address block ensure connectivity to Service Provider's customers
  - "Longest match" lookup

Aggregation

- Address block should be announced to the Internet as an aggregate
- Subprefixes of address block should NOT be announced to Internet unless for traffic engineering
  - See BGP Multihoming presentations
- Aggregate should be generated internally
  - Not on the network borders!

Announcing Aggregate – Cisco IOS

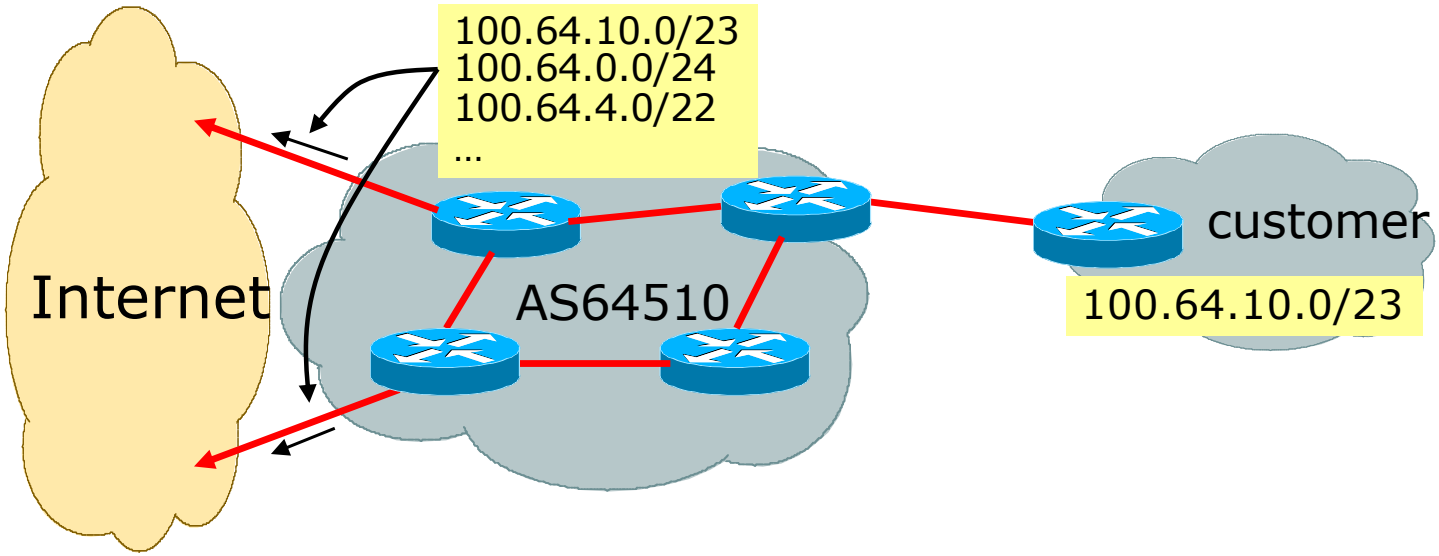
- Configuration Example

```
router bgp 64511
 address-family ipv4
  network 100.66.0.0 mask 255.255.224.0
  neighbor 100.67.10.1 remote-as 64510
  neighbor 100.67.10.1 prefix-list out-filter out
  neighbor 100.67.10.1 prefix-list default in
  neighbor 100.67.10.1 activate
!
ip route 100.66.0.0 255.255.224.0 null0
!
ip prefix-list out-filter permit 100.66.0.0/19
ip prefix-list out-filter deny 0.0.0.0/0 le 32
!
ip prefix-list default permit 0.0.0.0/0
```

Announcing an Aggregate

- Network Operators who don't and won't aggregate are held in poor regard by community
- Registries publish their minimum allocation size
  - For IPv4:
    - /24
  - For IPv6:
    - /48 for assignment, /32 for allocation
- Until 2010, there was no real reason to see anything longer than a /22 IPv4 prefix on the Internet. But now?
  - IPv4 run-out is having an impact
  - It is expected that eventually the global IPv4 table will be mostly /24s

Aggregation – Example

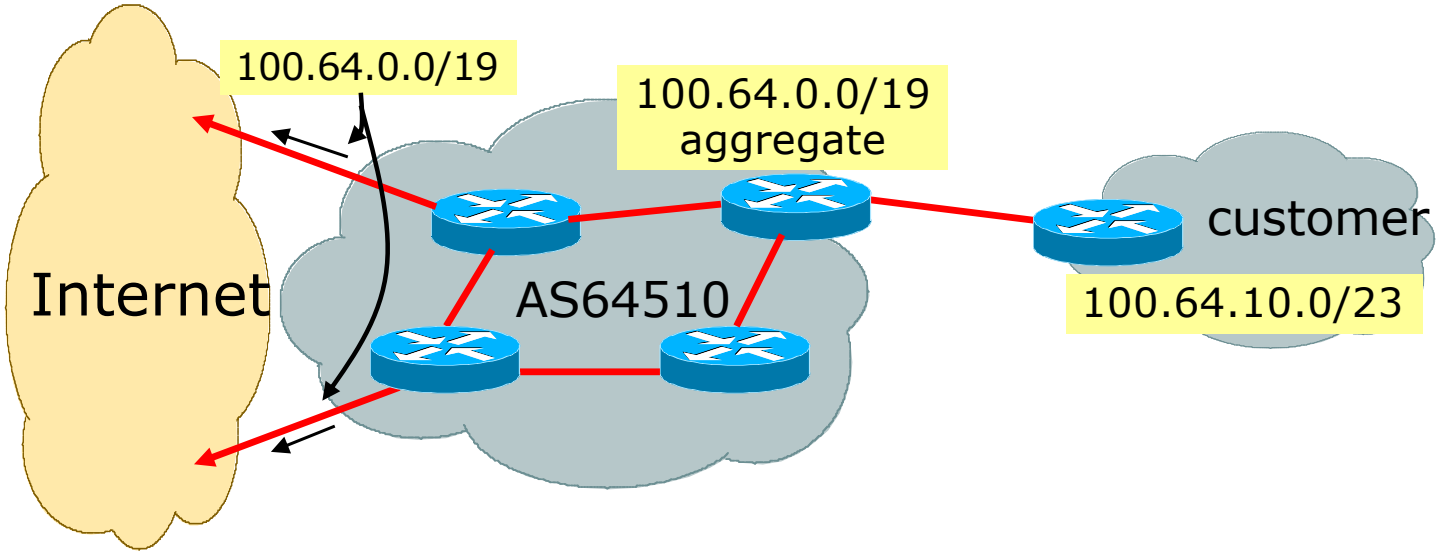


- Customer has /23 network assigned from AS64510's /19 address block
- AS64510 announces customers' individual networks to the Internet

Aggregation – Bad Example

- Customer link goes down
  - Their /23 network becomes unreachable
  - /23 is withdrawn from AS64510's IBGP
- Their service provider doesn't aggregate its /19 network block
  - /23 network withdrawal announced to peers
  - Starts rippling through the Internet
  - Added load on all Internet backbone routers as network is removed from routing table
- Customer link returns
  - Their /23 network is now visible to their provider
  - Their /23 network is re-advertised to peers
  - Starts rippling through Internet
  - Load on Internet backbone routers as network is reinserted into routing table
  - Some service providers suppress the flaps
  - Internet may take 10-20 min or longer to be visible
  - Where is the Quality of Service???

Aggregation – Example



- Customer has /23 network assigned from AS64510's /19 address block
- AS64510 announced /19 aggregate to the Internet

Aggregation – Good Example

- Customer link goes down
  - Their /23 network becomes unreachable
  - /23 is withdrawn from AS64510's IBGP
- /19 aggregate is still being announced
  - No BGP hold down problems
  - No BGP propagation delays
  - No damping by other network operators
- Customer link returns
- Their /23 network is visible again
  - The /23 is re-injected into AS64510's IBGP
- The whole Internet becomes visible immediately
- Customer has Quality of Service perception

Aggregation – Summary

- Good example is what everyone should do!
  - Adds to Internet stability
  - Reduces size of routing table
  - Reduces routing churn
  - Improves Internet QoS for **everyone**
- Bad example is what too many still do!
  - Why? Lack of knowledge?
  - Laziness?

Separation of IBGP and EBGP

- Many network operators do not understand the importance of separating IBGP and EBGP
  - IBGP is where all customer prefixes are carried
  - EBGP is used for announcing aggregate to Internet and for Traffic Engineering
- Do NOT do traffic engineering with customer originated IBGP prefixes
  - Leads to instability similar to that mentioned in the earlier bad example
  - Even though aggregate is announced, a flapping subprefix will lead to instability for the customer concerned
- Generate traffic engineering prefixes on the Border Router

The Internet Today (April 2025)

- Current IPv4 Internet Routing Table Statistics

BGP Routing Table Entries	979075
Prefixes after maximum aggregation	375657
Unique prefixes in Internet	476317
/24s announced	602052
ASNs in use	76783

  - (maximum aggregation is calculated by Origin AS)
  - (unique prefixes > max aggregation means that operators are announcing prefixes from their blocks without a covering aggregate)

The Internet Today (April 2025)

- Current IPv6 Internet Routing Table Statistics

BGP Routing Table Entries	208948
/48s announced	97816
ASNs in use	34240

Efforts to improve aggregation

- The CIDR Report
  - Initiated and operated for many years by Tony Bates
  - Now combined with Geoff Huston's routing analysis
    - www.cidr-report.org
    - (covers both IPv4 and IPv6 BGP tables)
  - Results e-mailed on a weekly basis to most operations lists around the world
  - Lists the top 30 service providers who could do better at aggregating
- RIPE Routing WG aggregation recommendations
  - IPv4: RIPE-399 — www.ripe.net/ripe/docs/ripe-399.html
  - IPv6: RIPE-532 — www.ripe.net/ripe/docs/ripe-532.html

Efforts to Improve Aggregation
The CIDR Report

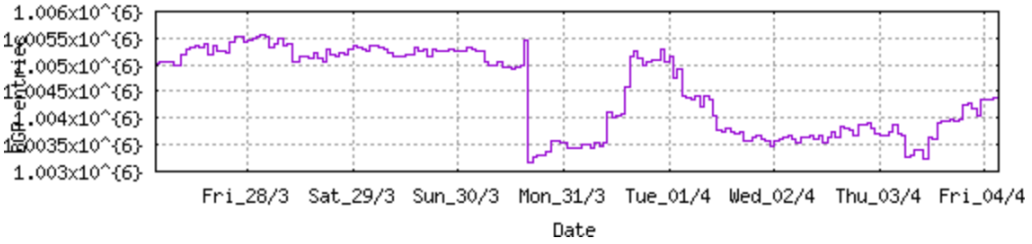
- Also computes the size of the routing table assuming network operators performed optimal aggregation
- Website allows searches and computations of aggregation to be made on a per AS basis
  - Flexible and powerful tool to aid network operators
  - Intended to show how greater efficiency in terms of BGP table size can be obtained without loss of routing and policy information
  - Shows what forms of origin AS aggregation could be performed and the potential benefit of such actions to the total table size
  - Very effectively challenges the traffic engineering excuse

Status Summary

Table History

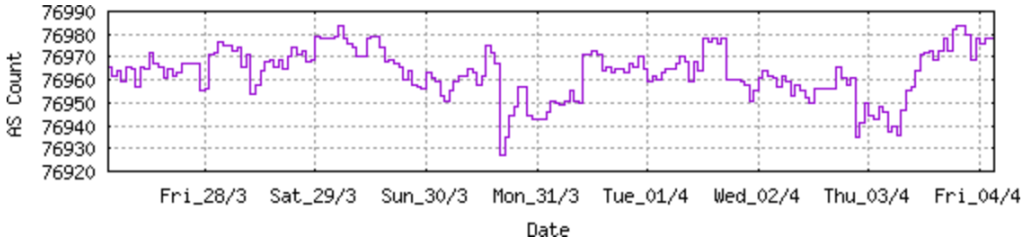
Date	Prefixes	CIDR Aggregated
28-03-25	1005425	567716
29-03-25	1005291	554547
30-03-25	1005242	554966
31-03-25	1003523	555763
01-04-25	1005069	556277
02-04-25	1003465	557202
03-04-25	1003718	558403
04-04-25	1004358	558356

Plot: BGP Table Size



AS Summary

76980	Number of ASes in routing system
26760	Number of ASes announcing only one prefix
13011	Largest number of prefixes announced by an AS AS16509: AMAZON-02, US
228349696	Largest address span announced by an AS (/32s) AS749: DNIC-AS-00749, US



Plot: AS count
Plot: Average announcements per origin AS
Report: ASes ordered by originating address span
Report: ASes ordered by transit address span
Report: Autonomous System number-to-name mapping (from Registry WHOIS data)

Announced Prefixes

```
Rank  AS       Type    Originate Addr Space  (pfx)   Transit Addr space  (pfx)  Description
109   AS6389         ORG+TRN Originate:    6923264 /9.28   Transit:      16640 /17.98 BELLSOUTH-NET-BLK, US
```

Long term deaggregator – BellSouth in the US

Aggregation Suggestions

Filter: Aggregates, Specifics

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation p

```
Rank AS           AS Name                                  Current  Wthdw  Aggte  Annce Redctn       %
 109 AS6389       BELLSOUTH-NET-BLK, US                        547    504     10     53    494  90.31%
```

Prefix	AS Path	Aggregation Suggestion
12.81.120.0/24	4608 7575 2914 7018 6389	
12.130.209.0/24	4608 7575 2914 7018 6389 6389 6389 6389	
65.5.0.0/16	4608 7575 2914 7018 6389	
65.5.160.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.164.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.172.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.200.0/21	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.228.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.232.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.236.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.240.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.244.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.248.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.5.252.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.5.0.0/16 4608 7575 2914 7018 6389
65.7.0.0/16	4608 7575 2914 7018 6389	
65.7.64.0/18	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.7.0.0/16 4608 7575 2914 7018 6389
65.12.0.0/14	4608 7575 2914 7018 6389	
65.13.84.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.92.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.120.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.124.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.136.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.176.0/21	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.184.0/21	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389
65.13.192.0/22	4608 7575 2914 7018 6389	Withdrawn - matching aggregate 65.12.0.0/14 4608 7575 2914 7018 6389

Announced Prefixes

Rank	AS	Type	Originate Addr Space	(pfx)	Transit Addr space	(pfx)	Description
119	AS18403	ORG+TRN	Originate:	6080256 /9.46	Transit:	509184 /13.04	FPT-AS-AP FPT Telecom Company, VN

Long term deaggregator – FPT in Vietnam

Aggregation Suggestions

Filter: Aggregates, Specifics

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation p

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
9	AS18403	FPT-AS-AP FPT Telecom Company, VN	4226	3754	129	601	3625	85.78%

Prefix	AS Path	Aggregation Suggestion
1.52.0.0/14	4608 4635 18403 18403 18403	
1.52.0.0/18	4608 4635 18403 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/14 4608 4635 18403 18403 18403
1.52.0.0/14	4608 4826 18403 18403	+ Announce - aggregate of 1.52.0.0/15 (4608 4826 18403 18403) and 1.54.0.0/15 (4608 4826 18403 18403)
1.52.0.0/20	4608 4826 18403 18403	- Withdrawn - aggregated with 1.52.16.0/20 (4608 4826 18403 18403)
1.52.0.0/23	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.2.0/23	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.4.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.5.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.6.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.7.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.8.0/23	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.10.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.11.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.12.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.13.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.14.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.15.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.0.0/20 4608 4826 18403 18403
1.52.16.0/20	4608 4826 18403 18403	- Withdrawn - aggregated with 1.52.0.0/20 (4608 4826 18403 18403)
1.52.16.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403
1.52.17.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403
1.52.18.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403
1.52.19.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403
1.52.20.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403
1.52.21.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403
1.52.22.0/24	4608 4826 18403 18403	- Withdrawn - matching aggregate 1.52.16.0/20 4608 4826 18403 18403

Announced Prefixes

Rank	AS	Type	Originate Addr Space	(pfx)	Transit Addr space	(pfx)	Description
126	AS7545	ORG+TRN	Originate:	5745152 /9.55	Transit:	6014720 /9.48	TPG-INTERNET-AP TPG Telecom Limited, AU

Long term deaggregator – TPG in Australia

Aggregation Suggestions

Filter: Aggregates, Specifics

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation p

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
4	AS7545	TPG-INTERNET-AP TPG Telecom Limited, AU	6214	5716	170	668	5546	89.25%

Prefix	AS Path	Aggregation Suggestion
14.2.0.0/17	4608 7575 7545	+ Announce - aggregate of 14.2.0.0/18 (4608 7575 7545) and 14.2.64.0/18 (4608 7575 7545)
14.2.0.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.32.0/19 (4608 7575 7545)
14.2.32.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.0.0/19 (4608 7575 7545)
14.2.32.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.40.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.48.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.56.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.64.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.96.0/19 (4608 7575 7545)
14.2.96.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.64.0/19 (4608 7575 7545)
14.2.128.0/18	4608 7575 7545	
14.2.192.0/20	4608 7575 7545	
14.200.0.0/14	4608 7575 7545	
14.200.0.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.1.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.2.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.3.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.4.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.5.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.6.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.7.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.8.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.9.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.10.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.11.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.12.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545

Announced Prefixes

```
Rank  AS       Type    Originate Addr Space  (pfx)   Transit Addr space  (pfx)  Description
50    AS12479        ORG+TRN Originate:   14689536 /8.19   Transit:     381184 /13.46 UNI2-AS, ES
```

Long term deaggregator – Orange in Spain

Aggregation Suggestions

Filter: Aggregates, Specifics

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation p

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
11	AS12479	UNI2-AS, ES	7711	4003	947	4655	3056	39.63%

Prefix	AS Path	Aggregation Suggestion
1.178.224.0/19	4608 1221 4637 5511 12479	
1.178.224.0/20	4608 1221 4637 5511 12479	- Withdrawn - matching aggregate 1.178.224.0/19 4608 1221 4637 5511 12479
1.178.240.0/20	4777 2516 1299 5511 12479	
37.11.0.0/16	4608 1221 4637 5511 12479	
37.11.0.0/22	4777 2516 1299 5511 12479	
37.11.8.0/22	4608 7575 6461 5511 12479	
37.11.12.0/22	4777 2516 1299 5511 12479	
37.11.16.0/22	4608 7575 6461 5511 12479	
37.11.20.0/22	4608 1221 4637 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 1221 4637 5511 12479
37.11.24.0/21	4777 2516 1299 5511 12479	+ Announce - aggregate of 37.11.24.0/22 (4777 2516 1299 5511 12479) and 37.11.28.0/22 (4777 2516 129
37.11.24.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.28.0/22 (4777 2516 1299 5511 12479)
37.11.28.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.24.0/22 (4777 2516 1299 5511 12479)
37.11.32.0/21	4608 7575 6461 5511 12479	+ Announce - aggregate of 37.11.32.0/22 (4608 7575 6461 5511 12479) and 37.11.36.0/22 (4608 7575 646
37.11.32.0/22	4608 7575 6461 5511 12479	- Withdrawn - aggregated with 37.11.36.0/22 (4608 7575 6461 5511 12479)
37.11.36.0/22	4608 7575 6461 5511 12479	- Withdrawn - aggregated with 37.11.32.0/22 (4608 7575 6461 5511 12479)
37.11.40.0/22	4777 2516 1299 5511 12479	
37.11.44.0/24	4608 7575 6461 5511 12479	
37.11.45.0/24	4777 2516 1299 5511 12479	
37.11.48.0/22	4777 2516 1299 5511 12479	
37.11.52.0/22	4608 7575 6461 5511 12479	
37.11.56.0/21	4608 7575 6461 5511 12479	+ Announce - aggregate of 37.11.56.0/22 (4608 7575 6461 5511 12479) and 37.11.60.0/22 (4608 7575 646
37.11.56.0/23	4608 7575 6461 5511 12479	- Withdrawn - aggregated with 37.11.58.0/23 (4608 7575 6461 5511 12479)
37.11.58.0/23	4608 7575 6461 5511 12479	- Withdrawn - aggregated with 37.11.56.0/23 (4608 7575 6461 5511 12479)
37.11.60.0/22	4608 7575 6461 5511 12479	- Withdrawn - aggregated with 37.11.56.0/22 (4608 7575 6461 5511 12479)
37.11.64.0/22	4777 2516 1299 5511 12479	

Importance of Aggregation

- Size of routing table
  - Router Memory is not so much of a problem as it was in the 1990s
  - Routers routinely carry over 2 million prefixes
- Convergence of the Routing System
  - This is a problem
  - Bigger table takes longer for CPU to process
  - BGP updates take longer to deal with
  - BGP Instability Report tracks routing system update activity
  - bgpupdates.potaroo.net/instability/bgpupd.html

The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 03 April 2025 06:26 (UTC+1000)

50 Most active ASes for the past 14 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	16509	382427	3.57%	13195	28.98	AMAZON-02, US
2	8151	300673	2.80%	11972	25.11	UNINET, MX
3	7552	220403	2.06%	3916	56.28	VIETEL-AS-AP Viettel Group, VN
4	39891	192259	1.79%	4797	40.08	ALJAWWALSTC-AS, SA
5	209181	94201	0.88%	44	2140.93	ZENEX5IVE-NL, GB
6	9009	90289	0.84%	3801	23.75	M247, RO
7	9829	83243	0.78%	2043	40.75	BSNL-NIB National Internet Backbone, IN
8	45899	72765	0.68%	3288	22.13	VNPT-AS-VN VNPT Corp, VN
9	4155	72733	0.68%	2596	28.02	USDA-1, US
10	9155	53011	0.49%	435	121.86	QNET Kuwait, KW
11	25048	50770	0.47%	5	10154.00	DSNET DSNet, GB
12	9583	50590	0.47%	1916	26.40	SIFY-AS-IN Sify Limited, IN
13	51375	50140	0.47%	421	119.10	VIVA, BH
14	36914	48264	0.45%	569	84.82	KENET-AS, KE
15	6057	46757	0.44%	575	81.32	Administracion Nacional de Telecomunicaciones, UY
16	4268	46742	0.44%	7	6677.43	CERNET-ASN-BLOCK, US
17	28526	46583	0.43%	74	629.50	Universidad Autonoma del Estado de Mexico, MX
18	17794	41681	0.39%	13	3206.23	HTCL-ORANGE-HK-AP Hutchison Telephone Company Limited, HK
19	52752	40749	0.38%	27	1509.22	Vegas Telecom Informatica Ltda., BR
20	7545	40073	0.37%	6222	6.44	TPG-INTERNET-AP TPG Telecom Limited, AU
21	36903	39187	0.37%	1249	31.37	MT-MPLS, MA
22	5371	38946	0.36%	281	138.60	DNIC-ASBLK-05120-05376, US
23	647	38890	0.36%	422	92.16	DNIC-ASBLK-00616-00665, US

50 Most active Prefixes for the past 14 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	140.174.37.0/24	30959	0.28%	63376 -- TOUCHTONE-CUSTS, US
2	38.224.214.0/24	17700	0.16%	174 -- COGENT-174, US
3	81.90.132.0/24	16930	0.15%	25048 -- DSNET DSNet, GB
4	81.90.139.0/24	16921	0.15%	25048 -- DSNET DSNet, GB
5	81.90.133.0/24	16915	0.15%	25048 -- DSNET DSNet, GB
6	189.76.64.0/23	16548	0.15%	52752 -- Vegas Telecom Informatica Ltda., BR
7	189.76.66.0/23	16535	0.15%	52752 -- Vegas Telecom Informatica Ltda., BR
8	103.223.2.0/24	15901	0.14%	135445 -- IDNIC-AIRPAY-AS-ID PT. Airpay International Indonesia, ID
9	168.0.128.0/22	15651	0.14%	263069 -- BRT Comercio de Produtos de Informatica LTDA, BR
10	124.195.190.0/24	15606	0.14%	38684 -- CMBDAEJEON-AS-KR CMB Daejeon Broadcasting Co,.Ltd, KR
11	223.196.5.0/24	14134	0.13%	55644 -- VIL-AS-AP Vodafone Idea Ltd, IN
12	201.71.207.0/24	14051	0.13%	265626 -- SITE TELECOM, MX
13	203.145.78.0/24	13897	0.13%	17794 -- HTCL-ORANGE-HK-AP Hutchison Telephone Company Limited, HK
14	203.145.74.0/24	13892	0.13%	17794 -- HTCL-ORANGE-HK-AP Hutchison Telephone Company Limited, HK
15	202.45.88.0/24	13892	0.13%	17794 -- HTCL-ORANGE-HK-AP Hutchison Telephone Company Limited, HK
16	202.181.232.0/23	13629	0.12%	7540 -- HKCIX-AS-AP HongKong Commercial Internet Exchange, HK
17	207.167.116.0/22	13407	0.12%	7954 -- IMMENSE-NETWORKS, US
18	130.137.231.0/24	12633	0.11%	16509 -- AMAZON-02, US
19	72.43.207.0/24	12187	0.11%	12220 -- I-EVOLVE-TECHNOLOGY-SERVICES, US
20	67.96.211.0/24	11958	0.11%	3356 -- LEVEL3, US
21	130.137.12.0/24	10520	0.10%	16509 -- AMAZON-02, US
22	138.99.97.0/24	10238	0.09%	28657 -- MD Brasil - Tecnologia da Informacao Ltda, BR
23	130.137.108.0/24	9932	0.09%	16509 -- AMAZON-02, US
24	160.238.104.0/22	9509	0.09%	266151 -- Mania NET, BR
25	84.205.66.0/24	9140	0.08%	12654 -- RIPE-NCC-RIS-AS Reseaux IP Europeens Network Coordination Centre RIPE NCC, NL
26	130.137.124.0/24	9098	0.08%	16509 -- AMAZON-02, US
27	130.137.86.0/24	8987	0.08%	16509 -- AMAZON-02, US
28	130.137.147.0/24	8853	0.08%	16509 -- AMAZON-02, US

The BGP IPv6 Instability Report

This report is updated daily. The current report was generated on 4 April 2025 01:20 (UTC+1000)

50 Most active ASes for the past 14 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	151194	653542	15.71%	138	4735.81	STELIGHT-AS-AP Zhu Yucheng, CN
2	210842	294252	7.07%	99	2972.24	RKZED-AS, ID
3	51559	276494	6.65%	113	2446.85	NETINTERNET Netinternet Bilisim Teknolojileri AS, TR
4	22616	129875	3.12%	33	3935.61	ZSCALER-SJC1, US
5	262505	114594	2.75%	15	7639.60	N4 Telecomunicacoes LTDA - ME, BR
6	11014	108611	2.61%	28	3878.96	CPS, AR
7	16509	94442	2.27%	5634	16.76	AMAZON-02, US
8	202256	88163	2.12%	832	105.97	LAWLIETNET, CN
9	11172	80524	1.94%	2719	29.62	Alestra, S. de R.L. de C.V., MX
10	59449	74613	1.79%	4	18653.25	MITJA KonzolaNET, SI
11	264987	71121	1.71%	7	10160.14	RBR TELECOM SERVICOS DE TELEFONIA LIMITADA, BR
12	2472	68919	1.66%	1	68919.00	FR-DOM-GUYANE Guyane Francaise, EU
13	39572	66807	1.61%	44	1518.34	ADVANCEDHOSTERS-AS, NL
14	2199	65247	1.57%	1	65247.00	FR-DOM-MARTINIQUE Iles de la Martinique, EU
15	263390	54110	1.30%	23	2352.61	FNT Telecomunicacoes e Acesso a Redes de Internet, BR
16	263028	52860	1.27%	7	7551.43	Roveri Opcao Provedor de Acesso a Internet Ltda ME, BR
17	30083	50243	1.21%	3	16747.67	AS-30083-US-VELIA-NET, DE
18	40138	44187	1.06%	42	1052.07	MDNET, US
19	53667	41659	1.00%	996	41.83	PONYNET, US
20	28598	33626	0.81%	15	2241.73	DB3 SERVICOS DE TELECOMUNICACOES S.A, BR
21	28458	32536	0.78%	2	16268.00	IENTC S DE RL DE CV, MX
22	45271	31401	0.75%	1615	19.44	VIL-AS-AP Vodafone Idea Ltd, IN
23	211759	31035	0.75%	68	456.40	MIKU-NETWORK-TECHNOLOGY-LIMITED, HK
24	42298	28598	0.69%	645	44.34	GCC-MPLS-PEERING GCC MPLS peering, QA

50 Most active Prefixes for the past 14 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	2a03:eec0:3212::/48	129774	2.94%	22616 -- ZSCALER-SJC1, US
2	2a14:6781:ff04::/48	73165	1.66%	59449 -- MITJA KonzolaNET, SI
3	2001:661:4000::/35	68919	1.56%	2472 -- FR-DOM-GUYANE Guyane Francaise, EU
4	2001:661:2000::/35	65247	1.48%	2199 -- FR-DOM-MARTINIQUE Iles de la Martinique, EU
5	2605:de00:bb::/48	50227	1.14%	30083 -- AS-30083-US-VELIA-NET, DE
6	2804:248:100::/40	33583	0.76%	28598 -- DB3 SERVICOS DE TELECOMUNICACOES S.A, BR
7	2806:202::/32	32530	0.74%	28458 -- IENTC S DE RL DE CV, MX
8	2a02:5420:11::/48	27010	0.61%	43541 -- VSHOSTING, CZ
9	2403:e240::/32	21798	0.49%	4767 -- AIT-CS-ASN Computer Science, TH
10	2606:ab40:100::/48	19522	0.44%	400339 -- TRINITY-CYBER-01, US
11	2606:6e00:8000::/35	17034	0.39%	7296 -- AS-DYNASCALE-LAX, US
12	2a00:e200:105::/48	16970	0.38%	41313 -- NOVATEL-AS 2 Kukush str., floor 1, office 102, BG
13	2804:fdc::/32	15731	0.36%	263608 -- WSNET TELECOM LTDA ME, BR
14	2a0c:b641:302::/47	14001	0.32%	204210 -- ZEUSPACKAGINGLTD, IE
15	2804:2058::/32	13957	0.32%	264487 -- AVATO TECNOLOGIA S.A, BR
16	240d:c010:165::/48	13938	0.32%	139341 -- ACE-AS-AP ACE, SG
17	2407:5440::/48	13862	0.31%	141145 -- GIGANET-AS-ID PT Giga Digital Nusantara, ID
18	2a02:b48:83f3::/48	13538	0.31%	39572 -- ADVANCEDHOSTERS-AS, NL
19	2804:3104:4000::/34	13162	0.30%	264987 -- RBR TELECOM SERVICOS DE TELEFONIA LIMITADA, BR
20	2804:3104:c000::/34	13162	0.30%	264987 -- RBR TELECOM SERVICOS DE TELEFONIA LIMITADA, BR
21	2804:3104:8000::/34	13161	0.30%	264987 -- RBR TELECOM SERVICOS DE TELEFONIA LIMITADA, BR
22	2804:3104::/34	13154	0.30%	264987 -- RBR TELECOM SERVICOS DE TELEFONIA LIMITADA, BR
23	2804:544::/34	13128	0.30%	262505 -- N4 Telecomunicacoes LTDA - ME, BR
24	2804:544:4000::/34	13128	0.30%	262505 -- N4 Telecomunicacoes LTDA - ME, BR
25	2804:544:c000::/34	13128	0.30%	262505 -- N4 Telecomunicacoes LTDA - ME, BR
26	2804:544:8000::/34	13121	0.30%	262505 -- N4 Telecomunicacoes LTDA - ME, BR
27	2a02:b48:84ff::/48	12900	0.29%	39572 -- ADVANCEDHOSTERS-AS, NL
28	2a02:b48:83f4::/48	12164	0.28%	39572 -- ADVANCEDHOSTERS-AS, NL

Aggregation: Summary

- Aggregation on the Internet could be MUCH better
  - 50% saving on Internet routing table size is quite feasible
  - Tools are available
  - Commands on the routers are not hard
  - CIDR-Report webpage

Receiving Prefixes

Receiving Prefixes

- There are three scenarios for receiving prefixes from other ASes
  - Customer talking BGP
  - Peer talking BGP
  - Upstream/Transit talking BGP
- Each has different filtering requirements and need to be considered separately

Receiving Prefixes: From Customers

- Network Operators must only accept prefixes which have been assigned or allocated to their downstream customer
- If the Network Operator has assigned address space to its customer, then the customer IS entitled to announce it back to their provider
- If the Network Operator has NOT assigned address space to its customer, then:
  - Check in the five RIR databases to see if this address space really has been assigned to the customer
  - The tool: whois –h jwhois.apnic.net x.x.x.0/24
    - (jwhois is "joint whois" and queries the 5 RIR databases)

Receiving Prefixes: From Customers

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h jwhois.apnic.net 202.12.29.0

inetnum:        202.12.29.0 - 202.12.29.255
netname:        APNIC-SERVICES-AU
descr:          Asia Pacific Network Information Centre
descr:          Regional Internet Registry for the Asia-Pacific Region
descr:          6 Cordelia Street
descr:          South Brisbane
country:        AU
geofeed:        https://geofeed.apnic.net/geofeed.csv
org:            ORG-APNI1-AP
admin-c:        AIC1-AP
tech-c:         AIC1-AP
abuse-c:        AA1589-AP
status:         ASSIGNED PORTABLE
mnt-by:         APNIC-HM
mnt-routes:     MAINT-APNIC-IS-AP
mnt-irt:        IRT-APNIC-IS-AP
last-modified:  2024-03-01T05:17:46Z
source:         APNIC
```

inetnum – means it is an address delegation to an entity

Portable – means it is an assignment to the customer, the customer can announce it to you

Receiving Prefixes: From Customers

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h jwhois.apnic.net 194.15.141.0

inetnum:        194.15.141.0 - 194.15.141.255
netname:        INETTECH
country:        SE
org:            ORG-ITAS2-RIPE
admin-c:        KEL5-RIPE
tech-c:         KEL5-RIPE
status:         ASSIGNED PI
mnt-by:         RIPE-NCC-END-MNT
mnt-by:         KURTIS-PP-MNT
mnt-routes:     KURTIS-PP-MNT
mnt-domains:    KURTIS-PP-MNT
created:        2003-12-04T09:33:09Z
last-modified:  2016-04-14T08:21:55Z
source:         RIPE
sponsoring-org: ORG-NIE1-RIPE
```

inetnum – means it is an address delegation to an entity

Assigned PI – means its an assignment to the customer, the customer can announce it to you

Receiving Prefixes: From Customers

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h jwhois.apnic.net 193.128.0.0/22

inetnum:        193.128.0.0 - 193.128.6.255
netname:        UK-PIPEX-19931014
country:        GB
org:            ORG-UA24-RIPE
admin-c:        WERT1-RIPE
tech-c:         UPHM1-RIPE
status:         ALLOCATED PA
remarks:        Please send abuse notification to abuse@uk.uu.net
mnt-by:         RIPE-NCC-HM-MNT
mnt-by:         AS1849-MNT
mnt-routes:     AS1849-MNT
mnt-routes:     WCOM-EMEA-RICE-MNT
mnt-irt:        IRT-MCI-GB
created:        2018-07-30T09:42:04Z
last-modified:  2018-07-30T09:42:04Z
source:         RIPE # Filtered
```

inetnum – means it is an address delegation to an entity

ALLOCATED – means that this is Provider Aggregatable address space and can only be announced by the service provider holding the allocation (in this case Verizon UK)

Receiving Prefixes from customer: Cisco IOS

- For Example:
  - Downstream has 100.69.0.0/20 block
  - Should only announce this to upstreams
  - Upstreams should only accept this from them
- Configuration on upstream

```
router bgp 100
 address-family ipv4
  neighbor 100.67.10.1 remote-as 101
  neighbor 100.67.10.1 prefix-list customer in
  neighbor 100.67.10.1 prefix-list default out
  neighbor 100.67.10.1 activate
!
ip prefix-list customer permit 100.69.0.0/20
!
ip prefix-list default permit 0.0.0.0/0
```

Receiving Prefixes: From Peers

- A peer is a Network Operator with whom you agree to exchange prefixes you originate into the Internet routing table
  - Prefixes you accept from a peer are only those they have indicated they will announce
  - Prefixes you announce to your peer are only those you have indicated you will announce

Receiving Prefixes: From Peers

- Agreeing what each will announce to the other:
  - Exchange of e-mail documentation as part of the peering agreement, and then ongoing updates

    OR
  - Use of the Internet Routing Registry and configuration tools such as:
    - IRRToolSet: https://github.com/irrtoolset/irrtoolset
    - bgpq4: https://github.com/bgp/bgpq4
      (uses NTT's IRR database by default)

Receiving Prefixes from peer: Cisco IOS

- For Example:
  - Peer has 220.50.0.0/16, 61.237.64.0/18 and 81.250.128.0/17 address blocks
- Configuration on local router

```
router bgp 100
 address-family ipv4
  neighbor 100.67.10.1 remote-as 101
  neighbor 100.67.10.1 prefix-list my-peer in
  neighbor 100.67.10.1 prefix-list my-prefix out
  neighbor 100.67.10.1 activate
!
ip prefix-list my-peer permit 220.50.0.0/16
ip prefix-list my-peer permit 61.237.64.0/18
ip prefix-list my-peer permit 81.250.128.0/17
ip prefix-list my-peer deny 0.0.0.0/0 le 32
!
ip prefix-list my-prefix permit 100.67.16.0/20
```

Receiving Prefixes: From Upstream/Transit Provider

- Upstream/Transit Provider is a Network Operator who you pay to give you transit to the WHOLE Internet
- Receiving prefixes from them is not desirable unless really necessary
  - Traffic Engineering – see BGP Multihoming presentations
- Ask upstream/transit provider to either:
  - originate a default-route

    OR
  - announce one prefix you can use as default

Receiving Prefixes: From Upstream/Transit Provider

- Downstream Router Configuration

```
router bgp 100
 address-family ipv4
  network 100.66.0.0 mask 255.255.224.0
  neighbor 100.65.7.1 remote-as 101
  neighbor 100.65.7.1 prefix-list infilter in
  neighbor 100.65.7.1 prefix-list outfilter out
  neighbor 100.65.7.1 activate
!
ip prefix-list infilter permit 0.0.0.0/0
!
ip prefix-list outfilter permit 100.66.0.0/19
```

Receiving Prefixes: From Upstream/Transit Provider

- Upstream Router Configuration

```
router bgp 101
 address-family ipv4
  neighbor 100.65.7.2 remote-as 100
  neighbor 100.65.7.2 default-originate
  neighbor 100.65.7.2 prefix-list cust-in in
  neighbor 100.65.7.2 prefix-list cust-out out
  neighbor 100.65.7.2 activate
!
ip prefix-list cust-in permit 100.66.0.0/19
!
ip prefix-list cust-out permit 0.0.0.0/0
```

Receiving Prefixes: From Upstream/Transit Provider

- If it is necessary to receive prefixes from any provider, care is required.
  - Don't accept default (unless you need it)
  - Don't accept your own prefixes
- Special use prefixes for IPv4 and IPv6:
  - http://www.rfc-editor.org/rfc/rfc6890.txt
- For IPv4:
  - Don't accept prefixes longer than /24 (?)
    - /24 was the historical class C
- For IPv6:
  - Don't accept prefixes longer than /48 (?)
    - /48 is the design minimum delegated to a site

Receiving Prefixes: From Upstream/Transit Provider

- Check Team Cymru's list of "bogons"
  - https://www.team-cymru.com/bogon-reference-http
- For IPv4 also consult:
  - https://www.rfc-editor.org/rfc/rfc6441.txt (BCP171)
- Bogon Route Server:
  - https://www.team-cymru.com/bogon-reference-bgp
  - Supplies a BGP feed (IPv4 and/or IPv6) of address blocks which should not appear in the BGP table

Receiving IPv4 Prefixes

```
router bgp 100
 network 101.10.0.0 mask 255.255.224.0
 neighbor 100.65.7.1 remote-as 101
 neighbor 100.65.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0              ! Default
ip prefix-list in-filter deny 0.0.0.0/8 le 32        ! RFC1122 local host
ip prefix-list in-filter deny 10.0.0.0/8 le 32       ! RFC1918
ip prefix-list in-filter deny 100.64.0.0/10 le 32    ! RFC6598 shared address
ip prefix-list in-filter deny 101.10.0.0/19 le 32    ! Local prefix
ip prefix-list in-filter deny 127.0.0.0/8 le 32      ! Loopback
ip prefix-list in-filter deny 169.254.0.0/16 le 32   ! Auto-config
ip prefix-list in-filter deny 172.16.0.0/12 le 32    ! RFC1918
ip prefix-list in-filter deny 192.0.0.0/24 le 32     ! RFC6598 IETF protocol
ip prefix-list in-filter deny 192.0.2.0/24 le 32     ! TEST1
ip prefix-list in-filter deny 192.168.0.0/16 le 32   ! RFC1918
ip prefix-list in-filter deny 198.18.0.0/15 le 32    ! Benchmarking
ip prefix-list in-filter deny 198.51.100.0/24 le 32  ! TEST2
ip prefix-list in-filter deny 203.0.113.0/24 le 32   ! TEST3
ip prefix-list in-filter deny 224.0.0.0/3 le 32      ! Multicast & Experimental
ip prefix-list in-filter deny 0.0.0.0/0 ge 25        ! Prefixes >/24
ip prefix-list in-filter permit 0.0.0.0/0 le 32
```

Receiving IPv6 Prefixes

```
router bgp 100
 network 2020:3030::/32
 neighbor 2020:3030::1 remote-as 101
 neighbor 2020:3030::1 prefix-list v6in-filter in
!
ipv6 prefix-list v6in-filter permit 64:ff9b::/96          ! RFC6052 v4v6trans
ipv6 prefix-list v6in-filter deny 2001::/23 le 128        ! RFC2928 IETF prot
ipv6 prefix-list v6in-filter deny 2001:2::/48 le 128      ! Benchmarking (RFC5180)
ipv6 prefix-list v6in-filter deny 2001:10::/28 le 128     ! ORCHID
ipv6 prefix-list v6in-filter deny 2001:db8::/32 le 128    ! Documentation (RFC3849)
ipv6 prefix-list v6in-filter deny 2002::/16 le 128        ! Deny all 6to4
ipv6 prefix-list v6in-filter deny 2020:3030::/32 le 128   ! Local Prefix
ipv6 prefix-list v6in-filter deny 3ffe::/16 le 128        ! Formerly 6bone
ipv6 prefix-list v6in-filter deny 3fff::/20 le 128        ! Documentation (new)
ipv6 prefix-list v6in-filter permit 2000::/3 le 48        ! Global Unicast
ipv6 prefix-list v6in-filter deny ::/0 le 128
```

Note: These filters block Teredo (serious security risk) and 6to4 (deprecated by RFC7526)

Receiving Prefixes

- Paying attention to prefixes received from customers, peers and transit providers assists with:
  - The integrity of the local network
  - The integrity of the Internet
- Responsibility of all Network Operators to be good Internet citizens

Receiving BGP attributes

Receiving BGP attributes

- BGP attributes are sent as part of the BGP updates for each prefix
- Common attributes operators need to be aware of, for routing best practice, are:
  - MED
  - AS numbers (only public ASNs are routable)
  - BGP Communities

Receiving Prefixes: MEDs?

- MEDs are used by EBGP neighbours to indicate preferred entry point into their network over two or more links with their neighbour
  - Allows the operator to determine entry path into their network
    - Might have unintended consequences within their peer's network
  - Many operators will override MEDs attached to BGP announcements by setting their own local-preference values

Receiving Prefixes: Bogon ASNs?

- What about prefixes originated by bogon AS numbers?
  - Public ranges are 1-64495 (excluding 23456) and 131072-458751
    - IANA is distributing AS blocks to the RIRs from the latter range
  - All other ASNs are either for documentation, or for private use, or are unassigned
    - And any prefixes originating from those need to be dropped
    - Configuration error? Malicious intent?
- What would the AS_PATH filter look like?
  - Challenging with regular expression (as per IOS)
  - Easier with AS ranges (as per Bird or JunOS)

Filtering bogon ASNs – BIRD

- Here is a function showing how to filter bogon ASNs, as described previously:

```
function as_path_contains_bogons()
int set invalid_asns;
{
    invalid_asns = [
        0,                      # Reserved
        23456,                  # Transition AS
        64496..64511,           # Documentation ASNs
        64512..65534,           # Private ASNs
        65535,                  # Reserved
        65536..65551,           # Documentation ASNs
        65552..131071,          # Reserved
        458752..4199999999,     # IANA Reserved
        4200000000..4294967294, # Private ASNs
        4294967295              # Reserved
    ];
    return bgp_path ~ invalid_asns;
}
```

Filtering bogon ASNs – FRR

- Here is an AS-PATH regexp showing how to filter bogon ASNs:

```
bgp as-path access-list Bogon_ASNs deny _0_
bgp as-path access-list Bogon_ASNs deny _23456_
bgp as-path access-list Bogon_ASNs deny _6449[6-9]_
bgp as-path access-list Bogon_ASNs deny _64[5-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _6[5-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _[7-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _1[0-2][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _130[0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _1310[0-6][0-9]_
bgp as-path access-list Bogon_ASNs deny _13107[0-1]_
bgp as-path access-list Bogon_ASNs deny _45875[2-9]_
bgp as-path access-list Bogon_ASNs deny _4587[6-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _458[8-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _459[0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _4[6-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _[5-9][0-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]_
bgp as-path access-list Bogon_ASNs permit .*
```

Receiving Prefixes: BGP Communities?

- BGP communities are attached to BGP announcements to indicate:
  - Internal policy within an AS
  - External policy supported by a peer, for:
    - Onward routing policy/traffic engineering
    - Filtering (eg Remotely Triggered Blackhole Filtering)
    - Traffic engineering between the two networks
- Different BGP implementations have different default BGP community behaviours – consult:
  - Vendor documentation
  - https://www.rfc-editor.org/rfc/rfc8642.txt for discussion of some of the issues operators need to be aware of

Receiving Prefixes: BGP Communities

- Do **NOT** accept community values that are not expected
  - Match expected values
  - Overwrite received community values with your own default value

```
ip community-list standard lp-250 permit 65534:250
!
route-map ebgp-import permit 5
 description Set high preference
 match community lp-250
 set local-preference 250
 set community 65534:100
!
route-map ebgp-import permit 10
 description Set our default community
 set community 65534:100
!
```

We only expect this community from our EBGP neighbour

Cisco IOS: this overwrites all incoming community values

Receiving Prefixes: BGP Communities

- Do **NOT** send community values that are not needed by the peer
  - This avoids propagating your internal communities to other networks
    - Propagating your internal communities leaves you open to DoS or worse!

```
route-map ebgp-export permit 5
 description Tell upstream to set local-pref 250
 set community 65534:250
!
```

Cisco IOS: this overwrites all other community values

- Propagate all communities within the AS (by IBGP)
  - This may need changes to your equipment's default!

Receiving BGP attributes

- Care is needed when receiving prefixes, to be aware of some of the optional BGP attributes that may be attached
  - BGP communities are only intended for policy decisions within an AS or between two peering ASes
  - MEDs may have unexpected consequences for traffic flows on the peer's network
  - Bogon ASNs, like bogon address space, must never be used or announced to the global Internet

Prefixes into IBGP

Injecting prefixes into IBGP

- Use IBGP to carry customer prefixes
  - Don’t use IGP
- Point static route to customer interface
- Use BGP network statement
- As long as static route exists (interface active), prefix will be in BGP

Router Configuration: network statement

- Example:

```
interface loopback 0
 ip address 100.64.3.1 255.255.255.255
!
interface Serial 5/0
 ip unnumbered loopback 0
 ip verify unicast reverse-path
!
ip route 100.71.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 address-family ipv4
  network 100.71.10.0 mask 255.255.252.0
!
```

Injecting prefixes into IBGP

- Interface flap will result in prefix withdraw and reannounce
  - use "**`ip route . . . permanent`**"
- Many network operators redistribute static routes into BGP rather than using the network statement
  - Only do this if you understand why

Router Configuration: redistribute static

- Example:

```
ip route 100.71.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 address-family ipv4
  redistribute static route-map static-to-bgp
<snip>
!
route-map static-to-bgp permit 10
 match ip address prefix-list ISP-block
 set origin igp
 set community 100:1000
<snip>
!
ip prefix-list ISP-block permit 100.71.10.0/22 le 30
```

Injecting prefixes into IBGP

- Route-map **`static-to-bgp`** can be used for many things:
  - Setting communities and other attributes
  - Setting origin code to IGP, etc
- Be careful with prefix-lists and route-maps
  - Absence of either/both means all statically routed prefixes go into IBGP

Summary

- Best Practices Covered:
  - When to use BGP
  - When to use ISIS/OSPF
  - Aggregation
  - Receiving Prefixes
  - Prefixes into BGP

Interconnection Best Practices

PeeringDB and the Internet Routing Registry

Interconnection Best Practices

- Types of Peering
- Using the PeeringDB and IXPDB
- Using the Internet Routing Registry

Types of Peering (1)

- Private Peering
  - Where two network operators agree to interconnect their networks, and exchange their respective routes, for the purpose of ensuring their customers can reach each other directly over the peering link
- Settlement Free Peering
  - No traffic charges
  - **The most common form of peering**
- Paid Peering
  - Where two operators agree to exchange traffic charges for a peering relationship

Types of Peering (2)

- Bi-lateral Peering
  - Very similar to Private Peering, but usually takes place at a public peering point (IXP)
- Multilateral Peering
  - Takes place at Internet Exchange Points, where operators all peer with each other via a Route Server
- Mandatory Multilateral Peering
  - Where operators are forced to peer with each other as condition of IXP membership
  - **Strongly discouraged: Has no record of success**

Types of Peering (3)

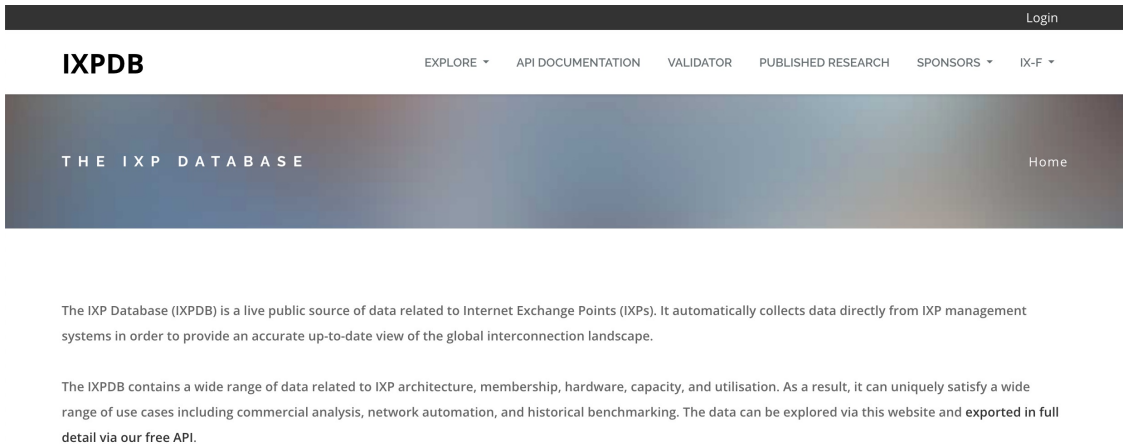
- Open Peering
  - Where a network operator publicly states that they will peer with all parties who approach them for peering
  - Commonly found at IXPs where the network operator participates via the Route Server (RS)
- Selective Peering
  - Where a network operator's peering policy depends on the nature of the operator who requests peering with them
  - At IXPs, the operator will not peer with RS but will only peer bilaterally
- Restrictive Peering
  - Where a network operator decides who its peering partners are, and is generally not approachable to considering peering opportunities

Types of Peering (4)

- The Peering Database documents network operator peering policies
  - https://www.peeringdb.com
- All operators with an AS are recommended to register in the PeeringDB
  - All operators who are considering peering or are peering must be in the PeeringDB to enhance their peering opportunities
- Participation in peering fora is encouraged too
  - Global Peering Forum (GPF) – (for North American peering)
  - Regional Peering Fora (Europe, Middle East, Africa, Asia, Caribbean, Latin America)
  - Many countries now have their own Peering Fora

Types of Peering (5)

- The IXPDB documents IXPs and their participants around the world
  - https://ixpdb.euro-ix.net/en/
- All Internet Exchange Point operators should register their IXP in the database
  - IXPs using IXP Manager will have this happen as part of the IXP Manager set up
  - Provides the LAN IP addresses of each member to facilitate automation



PeeringDB

Search here for a network, IX, or facility.

Advanced Search Legacy Search

pfsinoz

English (English)

HKIX Gold Sponsor

Peers	Connections	Open Peers	Total Speed	% with IPv6
276	353	170	16.5T	84

EXPORT

Organization	Hong Kong Internet eXchange Limited
Also Known As	
Long Name	Hong Kong Internet Exchange
City	Hong Kong
Country	HK
Continental Region	Asia Pacific
Service Level	Not Disclosed
Terms	Not Disclosed
Last Updated	2020-01-22T04:24:06Z
Notes	

Contact Information

Company Website	https://www.hkix.net/
Traffic Stats Website	https://www.hkix.net/hkix/stat/aggt/hkix-aggregate.html
Technical Email	noc@hkix.net
Technical Phone	+85239439900
Policy Email	info@hkix.net
Policy Phone	+85239438800
Sales Email	
Sales Phone	
Health Check	

Peers at this Exchange Point

Filter

Peer Name AŻ IPv4	ASN IPv6	Speed Port Location	Policy
2012 Limited 123.255.90.135	4658 2001:7fa:0:1::ca28:a087	10G	Selective
2012 Limited 123.255.90.122	4658 2001:7fa:0:1::ca28:a07a	1G	Selective
ACE CDN 123.255.92.38	139341 2001:7fa:0:1::ca28:a226	200G	Open
ACE CDN 123.255.91.67	139341 2001:7fa:0:1::ca28:a143	200G	Open
ACE CDN 123.255.91.194	139341 2001:7fa:0:1::ca28:a1c2	200G	Open
ACE CDN 123.255.91.79	139341 2001:7fa:0:1::ca28:a14f	200G	Open
ACME Universal 123.255.91.24	56190	1G	Open
ADVANCED HOSTING 123.255.91.178	39572 2001:7fa:0:1::ca28:a1b2	100G	Selective
Advanced Wireless Network Co. Ltd.(IIG) 123.255.92.80	45430 2001:7fa:0:1::ca28:a250	100G	Selective
AgotoZ HK 123.255.90.175	141167 2001:7fa:0:1::ca28:a0af	10G	Open

PeeringDB

Search here for a network, IX, or facility.

Advanced Search Legacy Search

pfsinoz

English (English)

Amazon.com Diamond Sponsor

EXPORT

Organization	Amazon.com, Inc.
Also Known As	Amazon Web Services
Long Name	
Company Website	https://www.amazon.com
ASN	16509
IRR as-set/route-set	AS16509:AS-AMAZON
Route Server URL	
Looking Glass URL	
Network Types	Enterprise
IPv4 Prefixes	16000
IPv6 Prefixes	8000
Traffic Levels	Not Disclosed
Traffic Ratios	Balanced
Geographic Scope	Global
Protocols Supported	✓ Unicast IPv4 ○ Multicast ✓ IPv6 ✓ Never via route servers
Last Updated	2025-03-28T07:45:16Z
Public Peering Info Updated	2025-03-28T07:45:02Z
Peering Facility Info Updated	2025-02-18T16:21:02Z
Contact Info Updated	2024-12-04T09:20:48Z
Notes	**AWS Peering - https://peering.aws/** **Peering requests** **IX** • Please use our peering portal to request IX peering sessions:

Public Peering Exchange Points

Filter

Exchange IPv4	ASN IPv6	Speed Port Location	RS Peer	BFD Support
1-IX EU 185.1.254.91	16509 2001:7f8:115:1::91	100G	○	○
ACT-IX 218.100.54.135	16509 2001:7fa:11:5:0:407d:0:2	10G	○	○
ACT-IX 218.100.54.134	16509 2001:7fa:11:5:0:407d:0:1	10G	○	○
AKL-IX (Auckland NZ) 43.243.21.113	16509 2001:7fa:11:6:0:407d:0:2	100G	○	○
AKL-IX (Auckland NZ) 43.243.21.112	16509 2001:7fa:11:6:0:407d:0:1	100G	○	○
AMS-IX 80.249.210.100	16509 2001:7f8:1::a501:6509:1	600G	○	○
AMS-IX 80.249.210.217	16509 2001:7f8:1::a501:6509:2	600G	○	○
AMS-IX Bangkok 103.100.140.204	16509 2402:b740:15:388:0:a501:6509:2	100G	○	○
AMS-IX Bangkok 103.100.140.201	16509 2402:b740:15:388:0:a501:6509:1	100G	○	○
AMS-IX Chicago 206.108.115.36	16509 2001:504:38:1:0:a501:6509:1	100G	○	○
AMS-IX Mumbai	16509	10G	○	○

PeeringDB

Search here for a network, IX, or facility.

Advanced Search Legacy Search

pfsinoz

English (English)

Arelion (Twelve99)

EXPORT

Organization	Arelion
Also Known As	f/k/a Telia Carrier
Long Name	
Company Website	https://www.arelion.com/
ASN	1299
IRR as-set/route-set	RIPE::AS1299:AS-TWELVE99
Route Server URL	
Looking Glass URL	https://lg.twelve99.net/
Network Types	NSP
IPv4 Prefixes	800000
IPv6 Prefixes	175000
Traffic Levels	100+Tbps
Traffic Ratios	Balanced
Geographic Scope	Global
Protocols Supported	✓ Unicast IPv4 ○ Multicast ✓ IPv6 ✓ Never via route servers
Last Updated	2024-01-02T16:09:45Z
Public Peering Info Updated	
Peering Facility Info Updated	2024-11-14T23:00:16Z
Contact Info Updated	2023-06-20T13:36:16Z
Notes	**All trouble ticket requests or support related emails should be sent to support@arelion.com.** AS1299 is matching RPKI validation state and reject invalid prefixes from peers and customers. Our looking-glass marks validation state for all prefixes. Please review your registered

Public Peering Exchange Points

Filter

Exchange AŻ IPv4	ASN IPv6	Speed Port Location	RS Peer	BFD Support

No filter matches.
You may filter by **Exchange**, **ASN** or **Speed.**

Interconnection Facilities

Filter

Facility AŻ ASN	Country City
123.NET - DC1 - 24700 Northwestern Hwy. 1299	United States of America Southfield
1530 SWIFT - NOCIX 1299	United States of America North Kansas City
1623 Farnam 1299	United States of America Omaha
165 Halsey Meet-Me Room 1299	United States of America Newark
365 Data Centers Buffalo (BU1) 1299	United States of America Buffalo
365 Data Centers Detroit (DT1) 1299	United States of America Southfield
365 Data Centers Nashville (NA1) 1299	United States of America Nashville
365 Data Centers Tampa (TA1) 1299	United States of America Tampa
3U Rechenzentrum Berlin 1299	Germany Berlin

Internet Routing Registry

- Many major transit providers and several content providers pay attention to what is contained in the Internet Routing Registry
  - There are many IRRs operating, the most commonly used being those hosted by the Regional Internet Registries, RADB, and some transit providers
- Best practice for any AS holder is to document their routing policy in the IRR
  - A route-object is the absolute minimum requirement

Internet Routing Registry

- IRR objects can be created via the database web-interfaces or submitted via email
- Policy language used to be known as RPSL
- Problems:
  - IRR contains a lot of outdated information
  - Network operators not following best practices
- Some network operators now using RPKI and ROAs to securely indicate the origin AS of their routes
  - Takes priority over IRR entries
  - RPKI and ROAs covered in other presentations

Which Internet Routing Registry database to use?

- Members of a Regional Internet Registry are strongly encouraged to use their RIR's Internet Routing Registry instance
  - Usually managed via the RIR's member portal giving easy access for creation and update of objects
  - Provided as part of the RIR's services to its members
- Operators who do not belong to any RIR generally use:
  - Their upstream transit provider's Routing Registry (if provided)
  - The RADB (https://www.radb.net)
    - Placing objects in the RADB requires an annual subscription fee
    - RADB now uses IRRDv4 – objects with RPKI **Invalid** cannot be created; existing RPKI **Invalid** objects will NOT be visible in a query, nor can they be modified

Route Object: Purpose

- Documents which Autonomous System number is originating the route listed
- Required by many major transit providers
  - They build their customer and peer filter based on the route-objects listed in the IRR
  - Referring to at least the 5 RIR routing registries and the RADB
  - Some operators run their own Routing Registry
    - May require their customers to place a Route Object there (if not using the 5 RIR or RADB versions of the IRR)

Route Object: Examples

```
route:          202.144.128.0/20
descr:          DRUKNET-BLOCK-A1
country:        BT
notify:         ioc@bt.bt
mnt-by:         MAINT-BT-DRUKNET
origin:         AS18024
last-modified:  2018-09-18T09:37:40Z
source:         APNIC
```

This declares that AS18024 is the origin of 202.144.128.0/20

```
route6:         2405:D000::/32
descr:          DRUKNET-IPV6-BLOCK
origin:         AS17660
notify:         netops@bt.bt
mnt-by:         MAINT-BT-DRUKNET
last-modified:  2010-07-21T03:46:02Z
source:         APNIC
```

This declares that AS17660 is the origin of 2405:D000::/32

AS Object: Purpose

- Documents peering policy with other Autonomous Systems
  - Lists network information
  - Lists contact information
  - Lists routes announced to neighbouring autonomous systems
  - Lists routes accepted from neighbouring autonomous systems
- Some operators pay close attention to what is contained in the AS Object
  - Some configure their border router BGP policy based on what is listed in the AS Object

AS Object: Example

```
aut-num:        AS17660
as-name:        DRUKNET-AS
descr:          DrukNet ISP, Bhutan Telecom, Thimphu
country:        BT
org:            ORG-BTL2-AP
import:         from AS6461   action pref=100;     accept ANY
export:         to AS6461     announce AS-DRUKNET-TRANSIT
import:         from AS2914   action pref=150;     accept ANY
export:         to AS2914     announce AS-DRUKNET-TRANSIT
<snip>
import:         from AS135666 action pref=250;   accept AS135666
export:         to AS135666   announce {0.0.0.0/0} AS-DRUKNET-TRANSIT
admin-c:        DNO1-AP
tech-c:         DNO1-AP
notify:         netops@bt.bt
mnt-irt:        IRT-BTTELECOM-BT
mnt-by:         APNIC-HM
mnt-lower:      MAINT-BT-DRUKNET
mnt-routes:     MAINT-BT-DRUKNET
last-modified:  2019-06-09T22:40:10Z
source:         APNIC
```

Examples of inbound and outbound policies – RPSL

AS-Set: Purpose

- The AS-Set is used by network operators to group AS numbers they provide transit for in an easier to manage form
  - Convenient for more complicated policy declarations
  - Used mostly by network operators who build their EBGP filters from their IRR entries
  - Commonly used at Internet Exchange Points to handle large numbers of peers

AS-Set: Example

```
as-set:          AS-DRUKNET-TRANSIT
descr:           DrukNet transit networks
members:         AS17660
members:         AS132232
members:         AS134715
members:         AS135666
members:         AS137925
members:         AS59219
members:         AS18024
members:         AS18025
members:         AS137994
members:         AS140695
members:         AS151498
members:         AS151955
members:         AS152317
members:         AS138558
admin-c:         DNO1-AP
tech-c:          DNO1-AP
notify:          netops@bt.bt
mnt-by:          MAINT-BT-DRUKNET
last-modified:   2024-09-16T04:35:58Z
source:          APNIC
```


Lists all the autonomous systems within the AS-DRUKNET-TRANSIT group

Hierarchical AS-Set

- The usage of hierarchical AS-Set (RFC2622) is strongly recommended now (and required for APNIC IRR) – this helps resolve name collisions

```
as-set:          AS-GEMNET
descr:           GEMNET LLC
country:         MN
members:         AS9934, AS9484, AS10219, AS9789,
                 AS38038, AS24496, AS24559, AS4850,
<snip>
tech-c:          GA263-AP
admin-c:         GA263-AP
mnt-by:          MAINT-GEMNET-MN
mnt-lower:       MAINT-GEMNET-MN
last-modified:   2023-09-26T01:25:15Z
source:          APNIC
```

VS

```
as-set:          AS-GEMNET
descr:           GEMNET s.r.o. ASes
members:         AS59479
members:         AS202733
tech-c:          DUMY-RIPE
admin-c:         DUMY-RIPE
mnt-by:          GEMNETCZ-MNT
created:         2013-08-19T09:49:13Z
last-modified:   2024-08-27T14:09:27Z
source:          RIPE
```

- Solution: AS-Set name changes to AS45204:AS-GEMNET
- Consult https://sanog.org/resources/sanog41/SANOG41_Conference-Recent-IRR-changes_Maz.pdf for more information and migration steps

Summary

- PeeringDB
  - An industry Best Practice so that:
    - Network operators can promote the interconnects they participate in and attract more peering partners
- IXPDB
  - An industry Best Practice so that:
    - Internet Exchange Points can show their participants and help make the interconnect more attractive for potential participants
- IRR
  - An industry Best Practice:
    - So that network operators can document which autonomous system is originating their prefixes
    - Used by network operators to filter prefixes received from their customers and peers

Route Origin Authorisation

Steps to securing the Routing System

Route Origin Authorisation

- Essential first step to secure the global routing system
- Covered in detail in separate presentation slide deck:
  - http://www.bgp4all.com.au/pfs/_media/workshops/02-rpki.pdf

- But there are some important best practices
  1. Signing ROAs
  2. Implementing ROV to drop ”invalids”

Route Origin Authorisation (ROA)

- A digital object that contains a list of address prefixes and one AS number
- It is an authority created by a prefix holder to authorise an AS Number to originate one or more specific route advertisements
- Publish a ROA using your RIR member portal
  - Consult your RIR for how to use their member portal to publish your ROAs

Route Origin Authorisation

- A typical ROA would look like this:

Prefix	10.10.0.0/16
Max-Length	/18
Origin-AS	AS65534

- There can be more than one ROA per address block
  - Allows the operator to originate prefixes from more than one AS
  - Caters for changes in routing policy or prefix origin

- NB: Only create ROAs for the aggregate and the exact subnets expected in the routing table!! (See RFC9319)

Route Origin Validation

- Route Origin Validation means checking if the prefix received has a valid ROA
  - Valid ROA means that the prefix (and subnet) is being originated from the correct origin AS
  - See the "BGP Origin Validation" presentation for more in-depth content
- Implementing ROV means checking the validation database with what is learned from BGP peers:
  - Valid – allow; Invalid – drop; NotFound – allow (at lower preference?)
- Problem: how is this implemented in routers today?

Route Origin Validation

- The ideal would be to write directly to the active BGP table
- Some implementations use existing EBGP policy handling routines
  - ADJ-RIB-IN: table of all prefixes received prior to policy being applied
  - Route Refresh (RFC2918)

- Routers which maintain the ADJ-RIB-IN:
  - Apply the ROV policy to the stored received BGP table
  - Updates are applied "automatically" to the BGP table and therefore the FIB
  - No impact on any BGP peers (Route Refresh not needed)

Route Origin Validation

- Routers which do NOT maintain the ADJ-RIB-IN:
  - Apply the ROV policy by sending a Route Refresh to peers
  - When there are a large number of ROAs (April 2025 sees over 670k), and frequent changes or updates of ROAs:
    - Routers are sending frequent Route Refresh requests to peers (typically every few minutes)
    - Peers are being “bombarded” by Route Refresh requests: significant resource burden when they send the full or a large portion of the BGP table
    - Severe control plane CPU impact on the peer router (effectively a Denial of Service on the peer router)
  - As more and more ROAs are created and altered globally, this problem becomes significantly more serious!

Route Refresh: Route Origin Validation

- JunOS implements ADJ-RIB-IN by default
  - ROA updates do not cause a problem when operating ROV
- Cisco does not implement ADJ-RIB-IN by default:
  - Applies to all versions of Cisco IOS/IOS-XE and older versions of IOS-XR
  - **MUST turn on soft-reconfiguration if running ROV on the router**
  - Soft-reconfiguration is similar concept to ADJ-RIB-IN
    - Note that Route Refresh CLI seems to be no longer accessible

Enabling Cisco's Soft Reconfiguration

```
router bgp 64510
 address-family ipv4
  neighbor 100.64.1.1 remote-as 64511
  neighbor 100.64.1.1 route-map infilter in
  neighbor 100.64.1.1 soft-reconfiguration inbound
```

- When the policy needs to be changed:

```
clear ip bgp 100.64.1.1 soft in
```

- Note:
  - When "soft-reconfiguration" is enabled, there is no access to the route-refresh capability CLI
  - `clear ip bgp 100.64.1.1 in` also does a soft refresh

Using Cisco's Soft-Reconfiguration

- Strongly recommended when deploying Route Origin Validation

- Operators will also use soft-reconfiguration when trouble-shooting EBGP peer problems
  - Soft reconfiguration enabled on an EBGP session means that the operator can see which prefixes were sent by a neighbour **before** any policy is applied
  - This helps saves arguments between operators about whose BGP filters may have configuration errors!

Configuration Tips

Of passwords, tricks and templates

IBGP and IGPs Reminder!

- Make sure loopback is configured on router
  - IBGP between loopbacks, NOT physical interfaces
- Make sure IGP carries loopback IPv4 /32 and IPv6 /128 address
- Consider the DMZ nets:
  - Use unnumbered interfaces?
  - Use next-hop-self on IBGP neighbours
  - Or carry the DMZ IPv4 /30s and IPv6 /127s in the IBGP
  - Basically, keep the DMZ nets out of the IGP!

IBGP: Next-hop-self

- BGP speaker announces external network to IBGP peers using router's local address (loopback) as next-hop
- Used by many service providers on edge routers
  - Preferable to carrying DMZ point-to-point link addresses in the IGP
  - Reduces size of IGP to just core infrastructure
  - Alternative to using unnumbered interfaces
  - Helps scale network
  - Many service providers consider this "best practice"

Limiting AS Path Length

- Some BGP implementations have problems with long AS_PATHS
  - Memory corruption
  - Memory fragmentation
- Even using AS_PATH prepends, it is not normal to see more than 20 ASNs in a typical AS_PATH in the Internet Routing Table today
  - The Internet is around 5 ASes deep on average
  - Largest AS_PATH is usually 16-20 ASNs

```
neighbor x.x.x.x maxas-limit 20
```

Limiting AS Path Length

- Some announcements have ridiculous lengths of AS-paths
  - This example is an error in one IPv6 implementation

```
*> 3FFE:1600::/24       22 11537 145 12199 10318 10566 13193 1930 2200 3425 293 5609 5430
13285 6939 14277 1849 33 15589 25336 6830 8002 2042 7610 i
```

  - This example shows 100 prepends (for no obvious reason)

```
*>i193.105.15.0         2516 3257 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 i
```

- If your implementation supports it, consider limiting the maximum AS-path length you will accept

BGP Maximum Prefix Tracking

- Allow configuration of the maximum number of prefixes a BGP router will receive from a peer
- Two level control:
  - Warning threshold: log warning message
  - Maximum: tear down the BGP peering, manual intervention required to restart

```
neighbor <x.x.x.x> maximum-prefix <max> [restart N] [<threshold>] [warning-only]
```

- *restart* is an optional keyword which will restart the BGP session N minutes after being torn down
- *threshold* is an optional parameter between 1 to 100
  - Specify the percentage of <max> that will cause a warning message to be generated. Default is 75%.
- *warning-only* is an optional keyword which allows log messages to be generated but peering session will not be torn down

Private-AS – Application

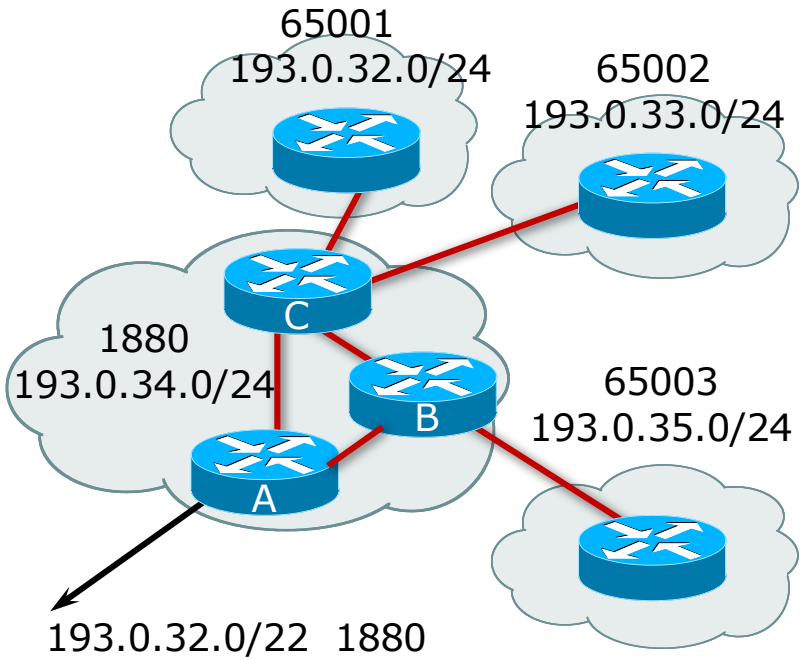
- A network operator with end-sites multihomed on their backbone (RFC2270)

  or

- A corporate network with several regions but connections to the Internet only in the core

  or

- Within a BGP Confederation



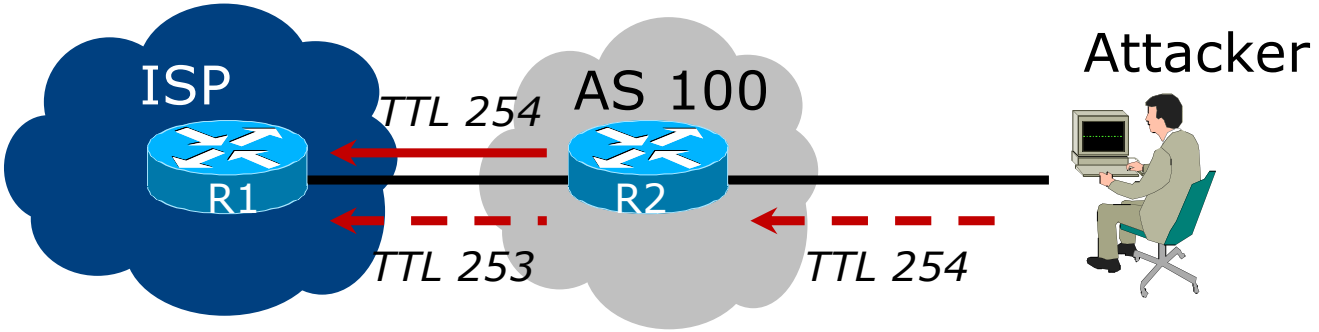
Private-AS – Removal

- Private ASNs MUST be removed from all prefixes announced to the public Internet
  - Include configuration to remove private ASNs in the EBGP template
- As with RFC1918 address space, private ASNs are intended for internal use
  - They must not be leaked to or used on the public Internet
- Cisco IOS

```
neighbor x.x.x.x remove-private-AS
```

BGP TTL "hack"

- Implement RFC5082 on BGP peerings
  - (Generalised TTL Security Mechanism)
  - Neighbour sets TTL to 255
  - Local router expects TTL of incoming BGP packets to be 254
  - No one apart from directly attached devices can send BGP packets which arrive with TTL of 254, so any possible attack by a remote miscreant is dropped due to TTL mismatch



BGP TTL "hack"

- TTL Hack:
  - Both neighbours must agree to use the feature
  - TTL check is much easier to perform than MD5
  - (Called BTSH – BGP TTL Security Hack)
- Provides "security" for BGP sessions
  - In addition to packet filters of course
  - MD5 should still be used for messages which slip through the TTL hack
  - See https://www.nanog.org/meetings/nanog27/presentations/meyer.pdf for more details

BGP TTL “hack”

- Configuration example:

```
neighbor 100.121.0.2 ttl-security hops 1
```

- BGP neighbour status:

```
Router# sh ip bgp neigh 100.121.0.2
...
Mininum incoming TTL 254, Outgoing TTL 255
Local host: 100.121.0.1, Local port: 41103
Foreign host: 100.121.0.2, Foreign port: 179
```

- The neighbour must set the same configuration
  - If they don’t, the BGP session will not come up

Templates

- Good practice to configure templates for everything
  - Vendor defaults tend not to be optimal or even very useful for service providers
  - Service providers create their own defaults by using configuration templates
- EBGP and IBGP examples follow
  - Also see Team Cymru's BGP templates
    - http://www.team-cymru.com/community-services

IBGP Template Example

- IBGP between loopbacks!
- Next-hop-self
  - Keep DMZ and external point-to-point out of IGP
- Always send communities in IBGP
  - Otherwise BGP policy accidents will happen
  - (Default on some vendor implementations, optional on others)
- Hardwire BGP to version 4
  - Yes, this is being paranoid!
  - Prevents accidental configuration of BGP version 3 which is still supported in some implementations

IBGP Template Example continued

- Use passwords on IBGP session
  - Not being paranoid, VERY necessary
  - It's a secret shared between you and your peer
  - If arriving packets don't have the correct MD5 hash, they are ignored
  - Helps defeat miscreants who wish to attack BGP sessions
- Powerful preventative tool, especially when combined with filters and the TTL "hack"

EBGP Template Example

- BGP damping
  - Do **NOT** use it unless you understand the impact
  - Do **NOT** use the vendor defaults without thinking
- Cisco's Soft Reconfiguration
  - Do **NOT** use unless troubleshooting or doing Route Origin Validation – it will consume considerable amounts of extra memory for BGP
- Remove private ASNs from announcements
  - Common omission today
- Use extensive filters, with "backup"
  - Use AS-path filters to backup prefix filters
  - Keep policy language for implementing policy, rather than basic filtering

EBGP Template Example continued

- Use password agreed between you and peer on EBGP session
- Use maximum-prefix tracking
  - Router will warn you if there are sudden increases in BGP table size, bringing down EBGP if desired
- Limit maximum as-path length inbound
- Log changes of neighbour state
  - …and monitor those logs!
- Make BGP admin distance higher than that of any IGP
  - Otherwise, prefixes heard from outside your network could override your IGP!!

Mutually Agreed Norms for Routing Security

Industry Best Practices to ensure Security of the Routing System

MANRS

Routing Security

- Implement the recommendations in https://www.manrs.org

1. Prevent propagation of incorrect routing information
   - Filter BGP peers, in & out!
2. Prevent traffic with spoofed source addresses
   - BCP38 – Unicast Reverse Path Forwarding
3. Facilitate communication between network operators
   - NOC to NOC Communication
   - Up-to-date details in Route and AS Objects, and PeeringDB
4. Facilitate validation of routing information
   - Route Origin Authorisation using RPKI

MANRS 1)

- Filtering prefixes inbound and outbound
  - RFC8212 requires all EBGP implementations to reject prefixes received and announced in the absence of any policy
- Advice: ***Never*** set up an EBGP session without inbound and outbound prefix filters
  - If full table required, block at least the bogons (see earlier)

MANRS 2)

- Implementing BCP 38
  - Unicast Reverse Path Forwarding
  - (Deny outbound traffic from customers which has spoofed source addresses)
- Advice: implement uRPF on ***all*** single-homed customer facing interfaces
  - Cheaper (CPU & RAM) than implementing packet filters

MANRS 3)

- Facilitate NOC to NOC communication
  - Know the **direct** NOC contacts for your customer Network Operators, your peer Network Operators, and your upstream Network Operators
  - This is not calling their "customer support line"
  - Make sure NOC contact info is part of any service contract
  - Up to date info in Route and AS Objects
  - Up to date AS info in PeeringDB
- Advice: NOC contact info for all connected Autonomous Networks is known to your NOC

MANRS 4)

- Facilitate validation of Routing Information
  - RPKI and Route Origin Authorisation (ROA)
  - All routes originated need to be signed to indicate that your AS is authorised to originate these routes
    - Helps secure the global routing system
- Advice: Sign ROAs for all originated routes using RPKI
  - And make sure all customer originated routes are also signed
  - Validate received routes from all peers
    - High priority for validated routes
    - Discard invalid routes
    - Low priority for unsigned routes

MANRS summary

- If your organisation supports and implements all 4 techniques in your network
  - Then join MANRS
  - https://www.manrs.org/join/
    - MANRS for Operators
    - MANRS for IXPs
    - MANRS for CDN & Cloud Providers

Summary

- Use configuration templates
- Standardise the configuration
- Be aware of standard “tricks” to avoid compromise of the BGP session
- Anything to make your life easier, network less prone to errors, network more likely to scale
- Implement the four fundamentals of MANRS
- It’s all about scaling – if your network won’t scale, then it won’t be successful

BGP Best Current Practices

ISP/IXP Workshops