# BGP Best Current Practices

## ISP/IXP Workshops

# What is BGP for??

## What is an IGP not for?

# BGP versus OSPF/ISIS

- Internal Routing Protocols (IGPs)

  examples are ISIS 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/ISIS

- BGP used internally (iBGP) and externally (eBGP)
- iBGP used to carry
  - some/all Internet prefixes across backbone
  - customer prefixes
- eBGP used to
  - exchange prefixes with other ASes
  - implement routing policy

# BGP versus OSPF/ISIS

- 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 ISP network
  - Announced to other ASes to aid with multihoming
- Unfortunately too many people are still thinking about class Cs, resulting in a proliferation of /24s in the Internet routing table

# Configuring Aggregation – Cisco IOS

- ISP has 101.10.0.0/19 address block
- To put into BGP as an aggregate:

```
router bgp 100
 network 101.10.0.0 mask 255.255.224.0
ip route 101.10.0.0 255.255.224.0 null0
```

- The static route is a “pull up” route

  more specific prefixes within this address block ensure connectivity to ISP’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 special circumstances (more later)
- Aggregate should be generated internally
    - Not on the network borders!

# Announcing Aggregate – Cisco IOS

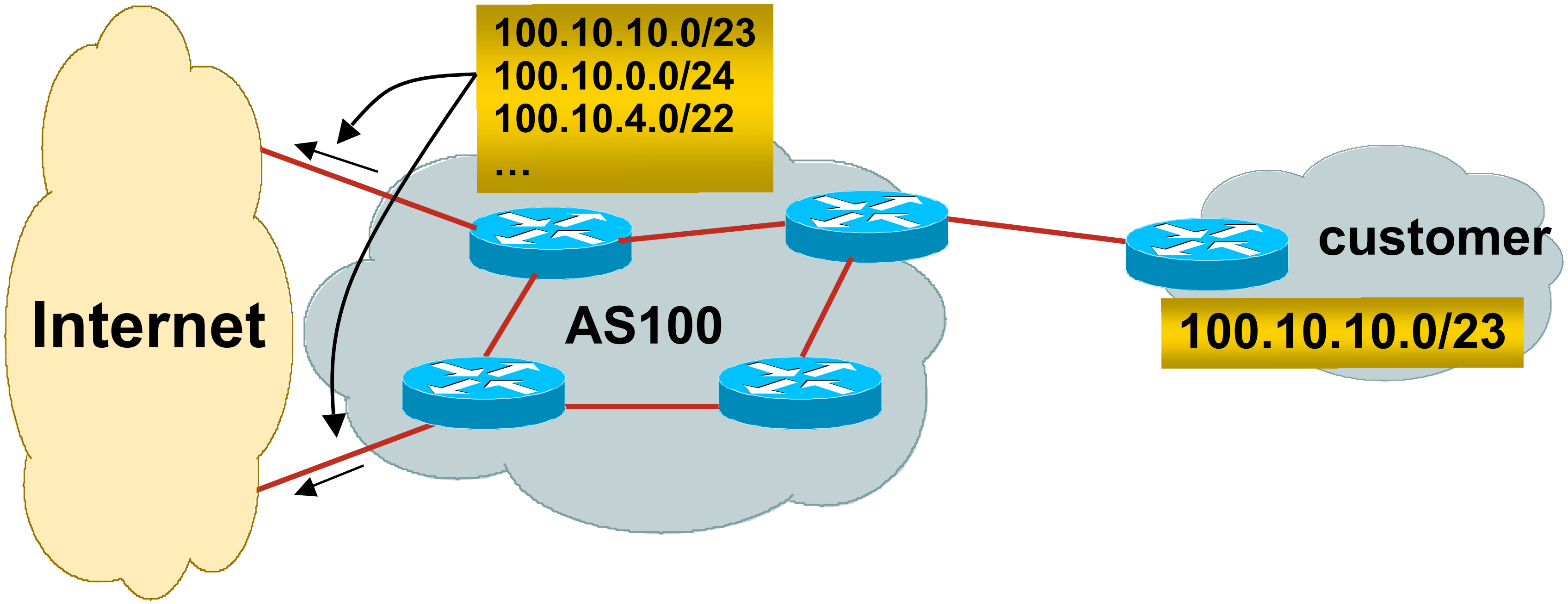
- Configuration Example

```
router bgp 100
 network 101.10.0.0 mask 255.255.224.0
 neighbor 102.102.10.1 remote-as 101
 neighbor 102.102.10.1 prefix-list out-filter out
!
ip route 101.10.0.0 255.255.224.0 null0
!
ip prefix-list out-filter permit 101.10.0.0/19
ip prefix-list out-filter deny 0.0.0.0/0 le 32
```

# Announcing an Aggregate

- ISPs who don’t and won’t aggregate are held in poor regard by community
- Registries publish their minimum allocation size
  - Anything from a /20 to a /22 depending on RIR
  - Different sizes for different address blocks
- No real reason to see anything longer than a /22 prefix in the Internet
  - BUT there are currently >146000 /24s!

# Aggregation – Example



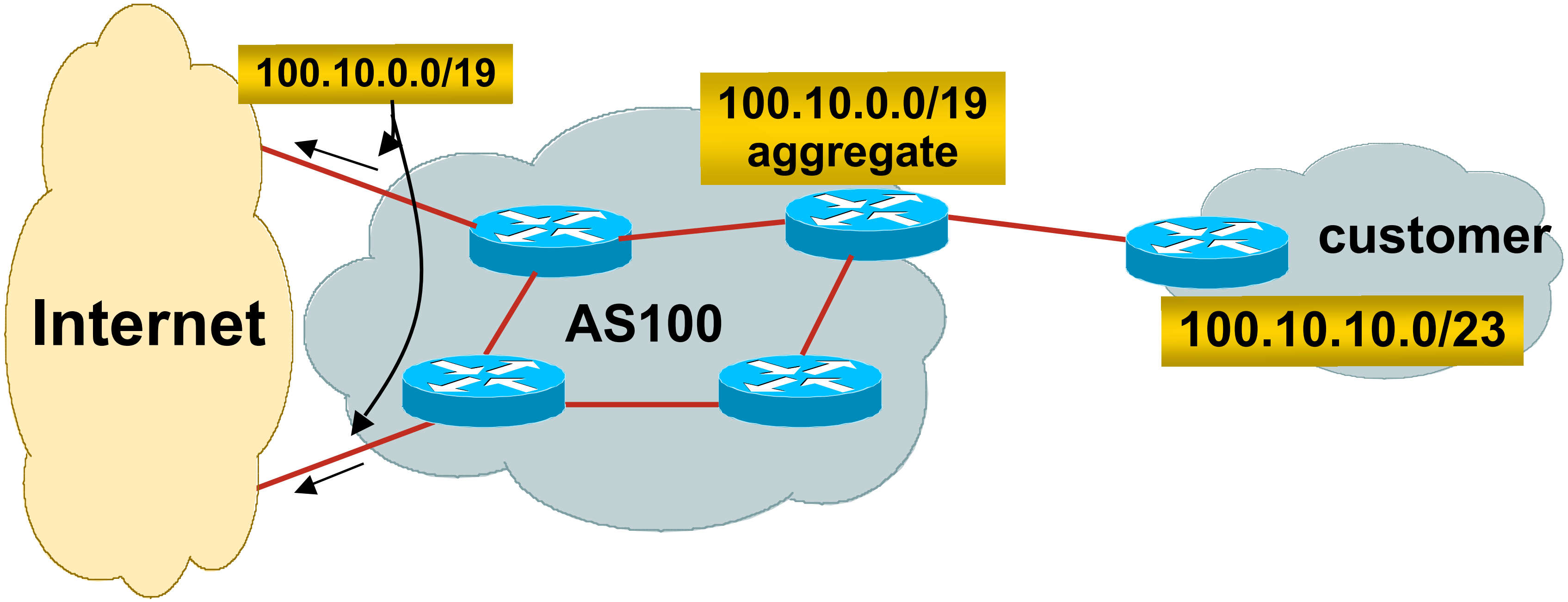

- Customer has /23 network assigned from AS100's /19 address block
- AS100 announces customers' individual networks to the Internet

# Aggregation – Bad Example

- Customer link goes down
  - Their /23 network becomes unreachable
  - /23 is withdrawn from AS100's iBGP
- Their ISP 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 ISP
  - 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 ISP's 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 AS100's /19 address block
- AS100 announced /19 aggregate to the Internet

# Aggregation – Good Example

- Customer link goes down
  - their /23 network becomes unreachable
  - /23 is withdrawn from AS100's iBGP
- /19 aggregate is still being announced
  - no BGP hold down problems
  - no BGP propagation delays
  - no damping by other ISPs
- Customer link returns
- Their /23 network is visible again
  - The /23 is re-injected into AS100's iBGP
- The whole Internet becomes visible immediately
- Customer has Quality of Service perception

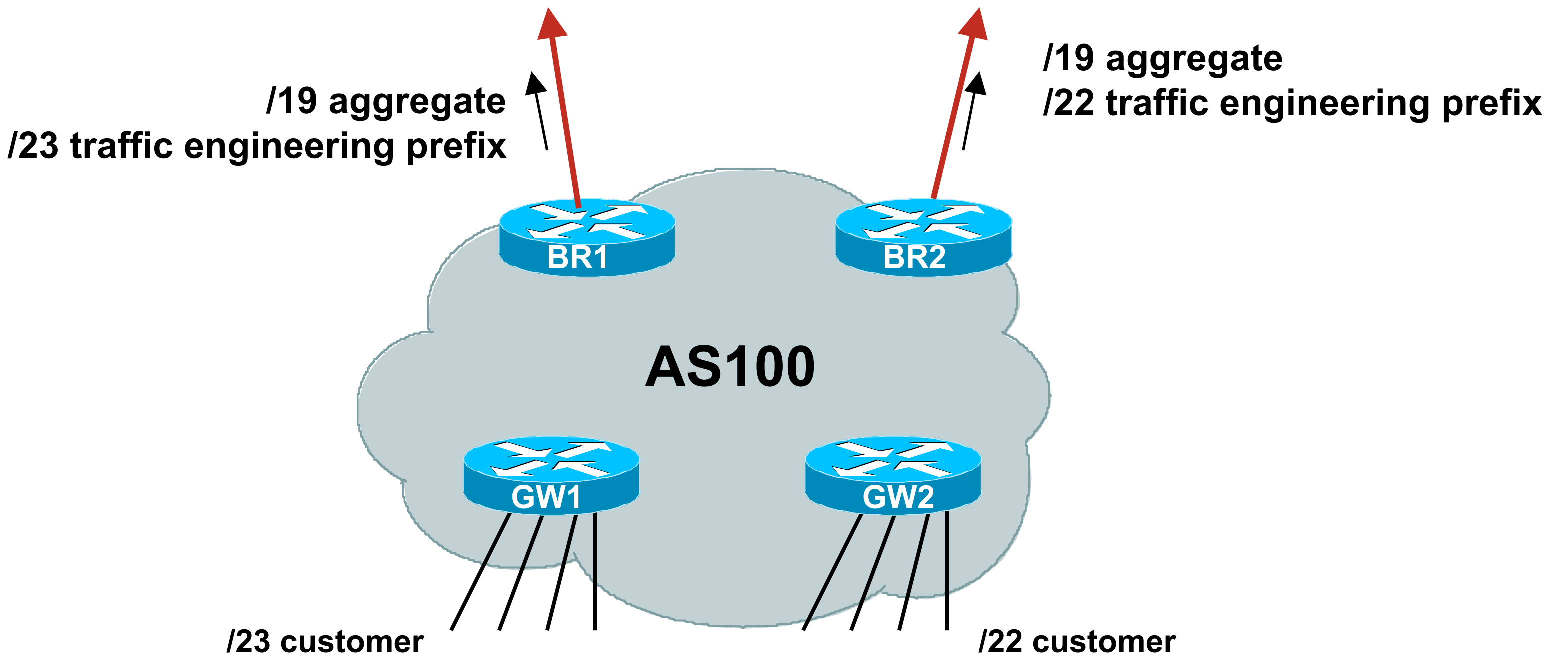
# Aggregation – Summary

- 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 ISPs 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**

# Separation of iBGP and eBGP




- /19 aggregate announced, plus /23 and /22 traffic engineering prefixes
- Even though /23 and /22 may be used by customers and visible through iBGP, they should also still be originated on the two Border Routers

# Separation of iBGP and eBGP

Priority must be greater than 200 (iBGP priority)

- Border Router configuration

```
router bgp 100
 network 100.10.0.0 mask 255.255.224.0
 network 100.10.10.0 mask 255.255.254.0
ip route 100.10.0.0 255.255.224.0 null0 254
ip route 100.10.10.0 255.255.254.0 null0 254
```

- Aggregation Router configuration

```
router bgp 100
 network 100.10.10.0 mask 255.255.254.0
ip route 100.10.10.0 255.255.254.0 serial 0/0
```

# Separation of iBGP and eBGP

- Note separate aggregation and border router configuration for the /23 being used for traffic engineering

  If customer link is up, iBGP prefix takes priority

  If customer link is down, iBGP prefix withdrawn from aggregation router, and the prefix generated on the border router “takes over”

  This ensures stable outbound announcement to the Internet even if customer link is unstable, removed, reconfigured, etc

# The Internet Today (June 2009)

- Current Internet Routing Table Statistics

| | |
|---|---|
| BGP Routing Table Entries | 288960 |
| Prefixes after maximum aggregation | 137560 |
| Unique prefixes in Internet | 143410 |
| Prefixes smaller than registry alloc | 142901 |
| /24s announced | 151247 |
| only 5819 /24s are from 192.0.0.0/8 | |
| ASes in use | 31579 |

# 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**

  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 recommendation

  **RIPE-399 — http://www.ripe.net/ripe/docs/ripe-399.html**

# Efforts to Improve Aggregation
# The CIDR Report

- Also computes the size of the routing table assuming ISPs performed optimal aggregation
- Website allows searches and computations of aggregation to be made on a per AS basis

  Flexible and powerful tool to aid ISPs

  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

CIDR Report

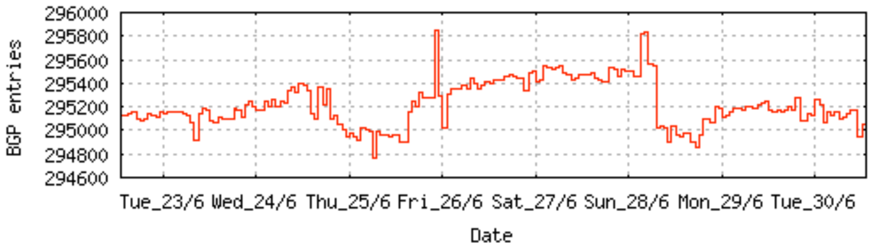
http://www.cidr-report.org/as2.0/

Radio Philip Networking Cisco Miscellaneous Smart Bookmarks TinyURL!

# Status Summary

## Table History

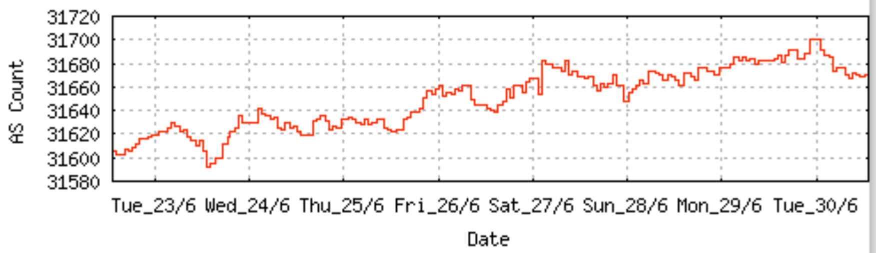
| Date | Prefixes | CIDR Aggregated |
|---|---|---|
| 23-06-09 | 295159 | 183803 |
| 24-06-09 | 295207 | 183718 |
| 25-06-09 | 294949 | 183021 |
| 26-06-09 | 295286 | 183129 |
| 27-06-09 | 295497 | 183044 |
| 28-06-09 | 295496 | 183041 |
| 29-06-09 | 295193 | 183133 |
| 30-06-09 | 295131 | 183303 |




Plot: BGP Table Size

## AS Summary

| | |
|---|---|
| 31673 | Number of ASes in routing system |
| 13449 | Number of ASes announcing only one prefix |
| 4295 | Largest number of prefixes announced by an AS<br>AS4323: TWTC - tw telecom holdings, inc. |
| 89737472 | Largest address span announced by an AS (/32s)<br>AS27064: DNIC-ASBLK-27032-27159 - DoD Network Information Center |




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)

AS Report

http://www.cidr-report.org/cgi-bin/as-report?as=AS4755&view=2.0

Radio ▾ Philip ▾ ADSL ▾ Networking ▾ Internet ▾ Cisco ▾ Miscellaneous ▾ TinyURL!

AS Report

# Announced Prefixes

| Rank | AS | Type | Originate Addr Space (pfx) | Transit Addr space (pfx) | Description |
|---|---|---|---|---|---|
| 140 | AS4755 | ORG+TRN | Originate: 2427648 /10.79 | Transit: 5044736 /9.73 | TATACOMM-AS TATA Communications formerly VSNL |

## Aggregation Suggestions

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 possibilities.

| Rank | AS | AS Name | Current | Wthdw | Aggte | Annce | Redctn | % |
|---|---|---|---|---|---|---|---|---|
| 7 | AS4755 | TATACOMM-AS TATA Communications formerly VSNL | 1200 | 1019 | 50 | 231 | 969 | 80.75% |

| Prefix | AS Path | Aggregation Suggestion |
|---|---|---|
| 59.151.144.0/22 | 4777 2516 4755 | |
| 59.160.0.0/16 | 4777 2516 4755 | |
| 59.160.0.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.4.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.5.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.8.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.12.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.15.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.16.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.24.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.24.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.28.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.32.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.38.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.40.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.44.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.46.0/23 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.48.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.48.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.56.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.64.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.71.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.72.0/21 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.73.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.81.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.82.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.83.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.88.0/22 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.88.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.89.0/24 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |
| 59.160.96.0/20 | 4777 2516 4755 | - Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 4755 |

AS Report

http://www.cidr-report.org/cgi-bin/as-report?as=AS18566&view=2.0

Radio Philip ADSL Networking Internet Cisco Miscellaneous TinyURL!

AS Report

# Announced Prefixes

| Rank | AS | Type | Originate Addr Space | (pfx) | Transit Addr space | (pfx) | Description |
|---|---|---|---|---|---|---|---|
| 147 | AS18566 | ORIGIN | Originate: | 2352896 /10.83 | Transit: | 0 /0.00 | COVAD - Covad Communications Co. |

## Aggregation Suggestions

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 possibilities.

| Rank | AS | AS Name | Current | Wthdw | Aggte | Annce | Redctn | % |
|---|---|---|---|---|---|---|---|---|
| 16 | AS18566 | COVAD - Covad Communications Co. | 1061 | 732 | 137 | 466 | 595 | 56.08% |

| Prefix | AS Path | Aggregation Suggestion |
|---|---|---|
| 64.105.0.0/16 | 4777 2497 2828 18566 | |
| 64.105.0.0/23 | 4777 2516 3356 18566 | |
| 64.105.4.0/22 | 4777 2516 3356 18566 | + Announce - aggregate of 64.105.4.0/23 (4777 2516 3356 18566) and 64.105.6.0/23 (4777 2 |
| 64.105.4.0/23 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.6.0/23 (4777 2516 3356 18566) |
| 64.105.6.0/23 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.4.0/23 (4777 2516 3356 18566) |
| 64.105.8.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.10.0/23 | 4777 2516 3356 18566 | |
| 64.105.14.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.16.0/23 | 4777 2516 3356 18566 | + Announce - aggregate of 64.105.16.0/24 (4777 2516 3356 18566) and 64.105.17.0/24 (4777 |
| 64.105.16.0/24 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.17.0/24 (4777 2516 3356 18566) |
| 64.105.17.0/24 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.16.0/24 (4777 2516 3356 18566) |
| 64.105.18.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.20.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.22.0/23 | 4777 2516 3356 18566 | |
| 64.105.24.0/21 | 4777 2516 3356 18566 | |
| 64.105.32.0/21 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.40.0/22 | 4777 2516 3356 18566 | + Announce - aggregate of 64.105.40.0/23 (4777 2516 3356 18566) and 64.105.42.0/23 (4777 |
| 64.105.40.0/23 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.42.0/23 (4777 2516 3356 18566) |
| 64.105.42.0/23 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.40.0/23 (4777 2516 3356 18566) |
| 64.105.44.0/23 | 4777 2516 3356 18566 | |
| 64.105.46.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.48.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.50.0/23 | 4777 2516 3356 18566 | |
| 64.105.52.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.54.0/23 | 4777 2516 3356 18566 | |
| 64.105.56.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |
| 64.105.58.0/23 | 4777 2516 3356 18566 | |
| 64.105.60.0/22 | 4777 2516 3356 18566 | + Announce - aggregate of 64.105.60.0/23 (4777 2516 3356 18566) and 64.105.62.0/23 (4777 |
| 64.105.60.0/23 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.62.0/23 (4777 2516 3356 18566) |
| 64.105.62.0/23 | 4777 2516 3356 18566 | - Withdrawn - aggregated with 64.105.60.0/23 (4777 2516 3356 18566) |
| 64.105.64.0/23 | 4777 2497 2828 18566 | - Withdrawn - matching aggregate 64.105.0.0/16 4777 2497 2828 18566 |

# Importance of Aggregation

- Size of routing table

  Memory is no longer a problem

  Routers can be specified to carry 1 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

  **http://bgpupdates.potaroo.net/instability/bgpupd.html**

The BGP Instability Report

http://bgpupdates.potaroo.net/instability/bgpupd.html

Radio Philip Networking Cisco Miscellaneous Smart Bookmarks TinyURL!

The BGP Instability Report | Cisco – Internal Sites WebEx Enter...

# The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 30 June 2009 06:18 (UTC+1000)

**50 Most active ASes for the past 7 days**

| RANK | ASN | UPDs | % | Prefixes | UPDs/Prefix | AS NAME |
|---|---|---|---|---|---|---|
| 1 | 9198 | 54354 | 4.87% | 287 | 189.39 | KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 2 | 21104 | 14225 | 1.27% | 14 | 1016.07 | AM-NETSYS-AS Netsys JVC Autonomous System Yerevan, Armenia |
| 3 | 22783 | 10898 | 0.98% | 6 | 1816.33 | WEBPOWER - WebPower, Inc. |
| 4 | 4812 | 9890 | 0.89% | 897 | 11.03 | CHINANET-SH-AP China Telecom (Group) |
| 5 | 6389 | 9271 | 0.83% | 4288 | 2.16 | BELLSOUTH-NET-BLK - BellSouth.net Inc. |
| 6 | 33783 | 9229 | 0.83% | 152 | 60.72 | EEPAD |
| 7 | 30969 | 8596 | 0.77% | 15 | 573.07 | TAN-NET TransAfrica Networks |
| 8 | 21491 | 8375 | 0.75% | 30 | 279.17 | UTL-ON-LINE UTL On-line is RF broadband ISP in Uganda - Africa |
| 9 | 30890 | 7989 | 0.72% | 418 | 19.11 | EVOLVA Evolva Telecom |
| 10 | 1221 | 7245 | 0.65% | 722 | 10.03 | ASN-TELSTRA Telstra Pty Ltd |
| 11 | 4134 | 7049 | 0.63% | 983 | 7.17 | CHINANET-BACKBONE No.31,Jin-rong Street |
| 12 | 47408 | 6941 | 0.62% | 22 | 315.50 | MANDARIN-AS Mandarin WIMAX Sicilia SpA |
| 13 | 3816 | 5726 | 0.51% | 392 | 14.61 | COLOMBIA TELECOMUNICACIONES S.A. ESP |
| 14 | 8452 | 5545 | 0.50% | 1008 | 5.50 | TEDATA TEDATA |
| 15 | 5668 | 5118 | 0.46% | 1040 | 4.92 | AS-5668 - CenturyTel Internet Holdings, Inc. |
| 16 | 8151 | 5023 | 0.45% | 1472 | 3.41 | Uninet S.A. de C.V. |
| 17 | 2764 | 4446 | 0.40% | 500 | 8.89 | AAPT AAPT Limited |
| 18 | 4766 | 4341 | 0.39% | 1818 | 2.39 | KIXS-AS-KR Korea Telecom |
| 19 | 4787 | 4320 | 0.39% | 170 | 25.41 | ASN-CBN Internet Service Provider |
| 20 | 17974 | 4280 | 0.38% | 622 | 6.88 | TELKOMNET-AS2-AP PT Telekomunikasi Indonesia |
| 21 | 24863 | 4210 | 0.38% | 896 | 4.70 | LINKdotNET-AS |
| 22 | 4323 | 4161 | 0.37% | 4316 | 0.96 | TWTC - tw telecom holdings, inc. |
| 23 | 25543 | 4121 | 0.37% | 49 | 84.10 | FasoNet-AS |

The BGP Instability Report

http://bgpupdates.potaroo.net/instability/bgpupd.html

Loading... Google

Radio Philip Networking Cisco Miscellaneous Smart Bookmarks TinyURL!

The BGP Instability Report | Cisco - Internal Sites WebEx Enter...

**50 Most active Prefixes for the past 7 days**

| RANK | PREFIX | UPDs | % | Origin AS -- AS NAME |
|---|---|---|---|---|
| 1 | 80.86.238.0/24 | 7165 | 0.61% | 21104 -- AM-NETSYS-AS Netsys JVC Autonomous System Yerevan, Armenia |
| 2 | 80.86.239.0/24 | 6937 | 0.59% | 21104 -- AM-NETSYS-AS Netsys JVC Autonomous System Yerevan, Armenia |
| 3 | 95.59.8.0/23 | 6866 | 0.59% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 4 | 95.59.2.0/23 | 6866 | 0.59% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 5 | 95.59.4.0/22 | 6865 | 0.59% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 6 | 92.46.244.0/23 | 6864 | 0.59% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 7 | 89.218.220.0/23 | 6860 | 0.59% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 8 | 89.218.218.0/23 | 6860 | 0.59% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 9 | 95.59.1.0/24 | 6214 | 0.53% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 10 | 88.204.221.0/24 | 6195 | 0.53% | 9198 -- KAZTELECOM-AS Kazakhtelecom Corporate Sales Administration |
| 11 | 205.246.203.0/24 | 4232 | 0.36% | 22783 -- WEBPOWER - WebPower, Inc. |
| 12 | 63.163.66.0/24 | 4230 | 0.36% | 22783 -- WEBPOWER - WebPower, Inc. |
| 13 | 72.23.246.0/24 | 3637 | 0.31% | 5050 -- PSC-EXT - Pittsburgh Supercomputing Center |
| 14 | 196.27.104.0/21 | 3562 | 0.30% | 30969 -- TAN-NET TransAfrica Networks |
| 15 | 196.27.108.0/22 | 3554 | 0.30% | 30969 -- TAN-NET TransAfrica Networks |
| 16 | 65.212.89.0/24 | 2424 | 0.21% | 22783 -- WEBPOWER - WebPower, Inc. |
| 17 | 193.201.184.0/21 | 2342 | 0.20% | 25546 -- BROOKLANDCOMP-AS Brookland Computer Services |
| 18 | 65.113.57.0/24 | 1686 | 0.14% | 36616 -- AGTLD - VeriSign Global Registry Services |
| 19 | 90.150.0.0/24 | 1585 | 0.14% | 35400 -- MFIST Interregoinal Organization Network Technologies |
| 20 | 192.12.120.0/24 | 1529 | 0.13% | 5691 -- MITRE-AS-5 - The MITRE Corporation |
| 21 | 203.178.128.0/17 | 1523 | 0.13% | 2500 -- WIDE-BB WIDE Project |
| 22 | 72.42.193.0/24 | 1423 | 0.12% | 40060 -- AAAWI - AAA Wireless, Inc. |
| 23 | 41.222.236.0/22 | 1060 | 0.09% | 37011 -- MSTARTEL-AS |
| 24 | 81.199.18.0/24 | 984 | 0.08% | 21491 -- UTL-ON-LINE UTL On-line is RF broadband ISP in Uganda - Africa |
| 25 | 81.199.22.0/24 | 931 | 0.08% | 21491 -- UTL-ON-LINE UTL On-line is RF broadband ISP in Uganda - Africa |
| 26 | 81.199.16.0/22 | 931 | 0.08% | 21491 -- UTL-ON-LINE UTL On-line is RF broadband ISP in Uganda - Africa |
| 27 | 81.199.23.0/24 | 929 | 0.08% | 21491 -- UTL-ON-LINE UTL On-line is RF broadband ISP in Uganda - Africa |
| 28 | 196.0.0.0/16 | 916 | 0.08% | 21491 -- UTL-ON-LINE UTL On-line is RF broadband ISP in Uganda - Africa |

# Receiving Prefixes

# Receiving Prefixes

- There are three scenarios for receiving prefixes from other ASNs
  - 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

- ISPs should only accept prefixes which have been assigned or allocated to their downstream customer
- If ISP has assigned address space to its customer, then the customer IS entitled to announce it back to his ISP
- If the ISP 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 whois.apnic.net x.x.x.0/24**

# Receiving Prefixes: From Customers

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

```
$ whois -h whois.apnic.net 202.12.29.0
inetnum:        202.12.29.0 - 202.12.29.255
netname:        APNIC-AP-AU-BNE
descr:          APNIC Pty Ltd - Brisbane Offices + Servers
descr:          Level 1, 33 Park Rd
descr:          PO Box 2131, Milton
descr:          Brisbane, QLD.
country:        AU
admin-c:        HM20-AP
tech-c:         NO4-AP
mnt-by:         APNIC-HM
changed:        hm-changed@apnic.net 20030108
status:         ASSIGNED PORTABLE
source:         APNIC
```

**Portable – 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 whois.ripe.net 193.128.2.0
inetnum:       193.128.2.0 - 193.128.2.15
descr:         Wood Mackenzie
country:       GB
admin-c:       DB635-RIPE
tech-c:        DB635-RIPE
status:        ASSIGNED PA
mnt-by:        AS1849-MNT
changed:       davids@uk.uu.net 20020211
source:        RIPE

route:         193.128.0.0/14
descr:         PIPEX-BLOCK1
origin:        AS1849
notify:        routing@uk.uu.net
mnt-by:        AS1849-MNT
changed:       beny@uk.uu.net 20020321
source:        RIPE
```

**ASSIGNED PA – means that it is Provider Aggregatable address space and can only be used for connecting to the ISP who assigned it**

# Receiving Prefixes from customer: Cisco IOS

- For Example:

  downstream has 100.50.0.0/20 block

  should only announce this to upstreams

  upstreams should only accept this from them

- Configuration on upstream

```
router bgp 100
 neighbor 102.102.10.1 remote-as 101
 neighbor 102.102.10.1 prefix-list customer in
!
ip prefix-list customer permit 100.50.0.0/20
```

# Receiving Prefixes: From Peers

- A peer is an ISP 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 the IRRToolSet

  **www.isc.org/sw/IRRToolSet/**

# 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
 neighbor 102.102.10.1 remote-as 101
 neighbor 102.102.10.1 prefix-list my-peer in
!
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
```

# Receiving Prefixes: From Upstream/Transit Provider

- Upstream/Transit Provider is an ISP who you pay to give you transit to the WHOLE Internet
- Receiving prefixes from them is not desirable unless really necessary

  special circumstances – see later
- 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
 network 101.10.0.0 mask 255.255.224.0
 neighbor 101.5.7.1 remote-as 101
 neighbor 101.5.7.1 prefix-list infilter in
 neighbor 101.5.7.1 prefix-list outfilter out
!
ip prefix-list infilter permit 0.0.0.0/0
!
ip prefix-list outfilter permit 101.10.0.0/19
```

# Receiving Prefixes: From Upstream/Transit Provider

- Upstream Router Configuration

```
router bgp 101
 neighbor 101.5.7.2 remote-as 100
 neighbor 101.5.7.2 default-originate
 neighbor 101.5.7.2 prefix-list cust-in in
 neighbor 101.5.7.2 prefix-list cust-out out
!
ip prefix-list cust-in permit 101.10.0.0/19
!
ip prefix-list cust-out permit 0.0.0.0/0
```

# Receiving Prefixes: From Upstream/Transit Provider

- If necessary to receive prefixes from any provider, care is required
  - don’t accept RFC1918 etc prefixes
    - **ftp://ftp.rfc-editor.org/in-notes/rfc3330.txt**
  - don’t accept your own prefixes
  - don’t accept default (unless you need it)
  - don’t accept prefixes longer than /24
- Check Team Cymru’s list of “bogons”
  - **www.cymru.com/Documents/bogon-list.html**
  - **www.team-cymru.org/Services/Bogons/routeserver.html**

# Receiving Prefixes

```
router bgp 100
 network 101.10.0.0 mask 255.255.224.0
 neighbor 101.5.7.1 remote-as 101
 neighbor 101.5.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0           ! Block default
ip prefix-list in-filter deny 0.0.0.0/8 le 32
ip prefix-list in-filter deny 10.0.0.0/8 le 32
ip prefix-list in-filter deny 101.10.0.0/19 le 32 ! Block local prefix
ip prefix-list in-filter deny 127.0.0.0/8 le 32
ip prefix-list in-filter deny 169.254.0.0/16 le 32
ip prefix-list in-filter deny 172.16.0.0/12 le 32
ip prefix-list in-filter deny 192.0.2.0/24 le 32
ip prefix-list in-filter deny 192.168.0.0/16 le 32
ip prefix-list in-filter deny 224.0.0.0/3 le 32   ! Block multicast
ip prefix-list in-filter deny 0.0.0.0/0 ge 25     ! Block prefixes >/24
ip prefix-list in-filter permit 0.0.0.0/0 le 32
```

# 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 ISPs to be good Internet citizens

# 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 215.17.3.1 255.255.255.255
!
interface Serial 5/0
 ip unnumbered loopback 0
 ip verify unicast reverse-path
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 network 215.34.10.0 mask 255.255.252.0
```

# Injecting prefixes into iBGP

- Interface flap will result in prefix withdraw and reannounce
  - use "**`ip route…permanent`**"
- Many ISPs 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 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 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
<snip>
!
ip prefix-list ISP-block permit 215.34.10.0/22 le 30
```

# Injecting prefixes into iBGP

- Route-map ISP-block 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

# Scaling the network

## How to get out of carrying all prefixes in IGP

# Why use BGP rather than IGP?

- IGP has Limitations:
  - The more routing information in the network
    - Periodic updates/flooding “overload”
    - Long convergence times
    - Affects the core first
  - Policy definition
    - Not easy to do

# Preparing the Network

- We want to deploy BGP now…
- BGP will be used therefore an ASN is required
- If multihoming to different ISPs is intended in the near future, a public ASN should be obtained:

  Either go to upstream ISP who is a registry member, or

  Apply to the RIR yourself for a one off assignment, or

  Ask an ISP who is a registry member, or

  **Join the RIR and get your own IP address allocation too (this option strongly recommended)!**

# Preparing the Network
# Initial Assumptions

- The network is not running any BGP at the moment
  - single statically routed connection to upstream ISP
- The network is not running any IGP at all
  - Static default and routes through the network to do “routing”

# Preparing the Network
# First Step: IGP

- Decide on an IGP: OSPF or ISIS ☺
- Assign loopback interfaces and /32 address to each router which will run the IGP
  - Loopback is used for OSPF and BGP router id anchor
  - Used for iBGP and route origination
- Deploy IGP (e.g. OSPF)
  - IGP can be deployed with NO IMPACT on the existing static routing
  - e.g. OSPF distance might be 110; static distance is 1
  - **Smallest distance wins**

# Preparing the Network
# IGP (cont)

- Be prudent deploying IGP – keep the Link State Database Lean!
  - Router loopbacks go in IGP
  - WAN point to point links go in IGP
  - (In fact, any link where IGP dynamic routing will be run should go into IGP)
  - Summarise on area/level boundaries (if possible) – i.e. think about your IGP address plan

# Preparing the Network IGP (cont)

- Routes which don’t go into the IGP include:

  Dynamic assignment pools (DSL/Cable/Dial)

  Customer point to point link addressing

    (using next-hop-self in iBGP ensures that these do NOT need to be in IGP)

  Static/Hosting LANs

  Customer assigned address space

  Anything else not listed in the previous slide

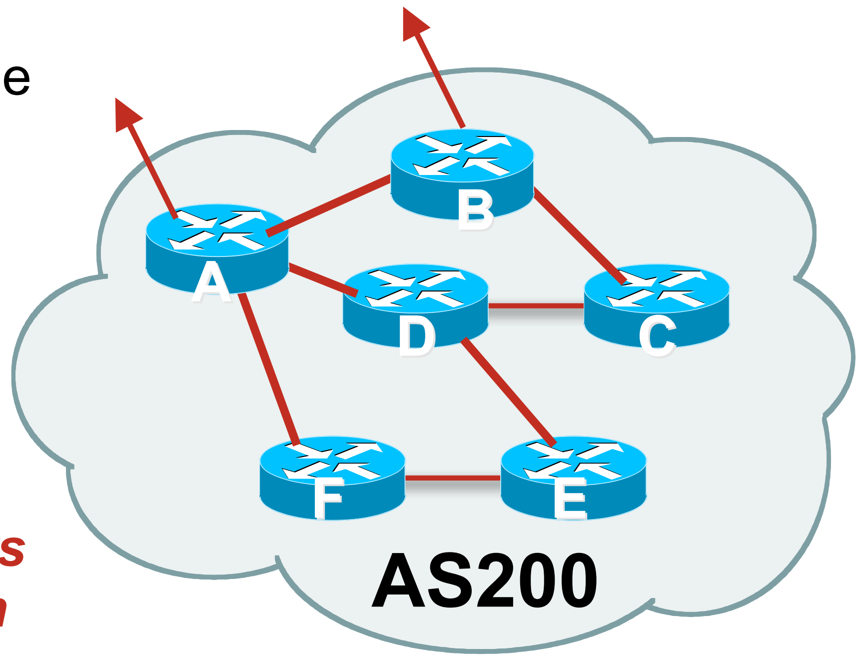
# Preparing the Network
# Second Step: iBGP

- Second step is to configure the local network to use iBGP
- iBGP can run on

  all routers, or

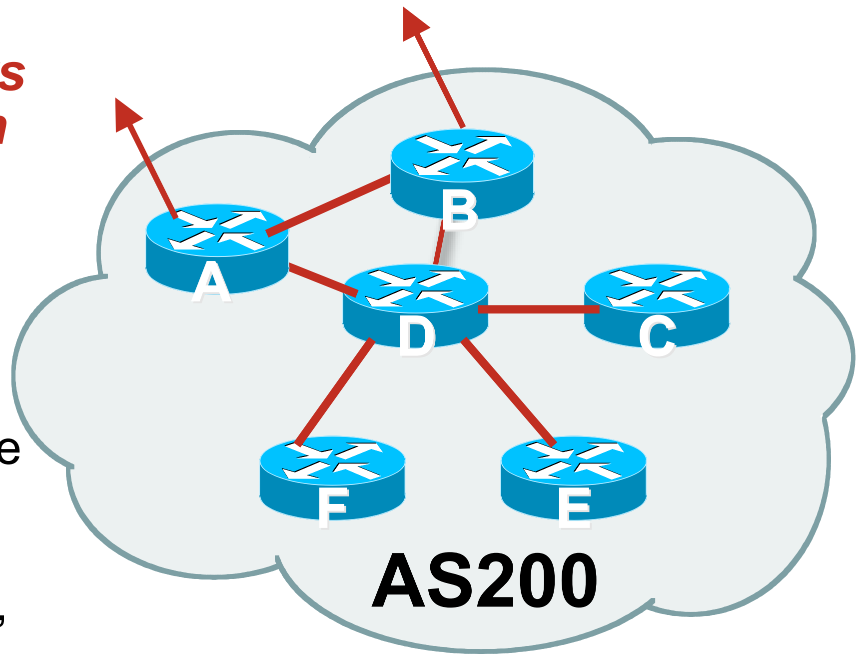
  a subset of routers, or

  just on the upstream edge
- ***iBGP must run on all routers which are in the transit path between external connections***



# Preparing the Network
# Second Step: iBGP (Transit Path)

- ***iBGP must run on all routers which are in the transit path between external connections***
- Routers C, E and F are not in the transit path
  - Static routes or IGP will suffice
- Router D is in the transit path
  - Will need to be in iBGP mesh, otherwise routing loops will result



# Preparing the Network
# Layers

- Typical SP networks have three layers:

  Core – the backbone, usually the transit path

  Distribution – the middle, PoP aggregation layer

  Aggregation – the edge, the devices connecting customers

# Preparing the Network Aggregation Layer

- iBGP is optional
  - Many ISPs run iBGP here, either partial routing (more common) or full routing (less common)
  - Full routing is not needed unless customers want full table
  - Partial routing is cheaper/easier, might usually consist of internal prefixes and, optionally, external prefixes to aid external load balancing
    - Communities and peer-groups make this administratively easy
- Many aggregation devices can't run iBGP
  - Static routes from distribution devices for address pools
  - IGP for best exit

# Preparing the Network
# Distribution Layer

- Usually runs iBGP
  - Partial or full routing (as with aggregation layer)
- But does not have to run iBGP
  - IGP is then used to carry customer prefixes (does not scale)
  - IGP is used to determine nearest exit
- Networks which plan to grow large should deploy iBGP from day one
  - Migration at a later date is extra work
  - No extra overhead in deploying iBGP, indeed IGP benefits

# Preparing the Network
# Core Layer

- Core of network is usually the transit path
- iBGP necessary between core devices
    - Full routes or partial routes:
        - Transit ISPs carry full routes in core
        - Edge ISPs carry partial routes only
- Core layer includes AS border routers

# Preparing the Network
# iBGP Implementation

Decide on:

- Best iBGP policy

  Will it be full routes everywhere, or partial, or some mix?

- iBGP scaling technique

  Community policy?

  Route-reflectors?

  Techniques such as peer groups and peer templates?

# Preparing the Network
# iBGP Implementation

- Then deploy iBGP:

  Step 1: Introduce iBGP mesh on chosen routers

  make sure that iBGP distance is greater than IGP distance (it usually is)

  Step 2: Install "customer" prefixes into iBGP

  **Check!** Does the network still work?

  Step 3: Carefully remove the static routing for the prefixes now in IGP and iBGP

  **Check!** Does the network still work?

  Step 4: Deployment of eBGP follows

# Preparing the Network
# iBGP Implementation

***Install "customer" prefixes into iBGP?***

- Customer assigned address space
    - Network statement/static route combination
    - Use unique community to identify customer assignments
- Customer facing point-to-point links
    - Redistribute connected through filters which only permit point-to-point link addresses to enter iBGP
    - Use a unique community to identify point-to-point link addresses (these are only required for your monitoring system)
- Dynamic assignment pools & local LANs
    - Simple network statement will do this
    - Use unique community to identify these networks

# Preparing the Network
# iBGP Implementation

***Carefully remove static routes?***

- Work on one router at a time:

    Check that static route for a particular destination is also learned by the iBGP

    If so, remove it

    If not, establish why and fix the problem

    (Remember to look in the RIB, not the FIB!)

- Then the next router, until the whole PoP is done
- Then the next PoP, and so on until the network is now dependent on the IGP and iBGP you have deployed

# Preparing the Network
# Completion

- Previous steps are NOT flag day steps
  - Each can be carried out during different maintenance periods, for example:
  - Step One on Week One
  - Step Two on Week Two
  - Step Three on Week Three
  - And so on
  - And with proper planning will have NO customer visible impact at all

# Preparing the Network
# Configuration Summary

- IGP essential networks are in IGP
- Customer networks are now in iBGP
  - iBGP deployed over the backbone
  - Full or Partial or Upstream Edge only
- BGP distance is greater than any IGP
- Now ready to deploy eBGP

# BGP Best Current Practices

ISP/IXP Workshops