BGP Best Current Practices

AFNOG 2013 AR-E Workshop

Configuring BGP

Where do we start?

IOS Good Practices

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

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

Replace with public ASN

Make ebgp and ibgp distance the same

- If supporting more than just IPv4 unicast neighbours

```
no bgp default ipv4-unicast
```
  - is also very important and required

3

Cisco IOS Good Practices

- BGP in Cisco IOS is permissive by default
- 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
- Good practice is to ensure that each eBGP neighbour has inbound and outbound filter applied:

```
router bgp 64511
  neighbor 1.2.3.4 remote-as 64510
  neighbor 1.2.3.4 prefix-list as64510-in in
  neighbor 1.2.3.4 prefix-list as64510-out out
```

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
  - Note: Same is happening for /48s with IPv6

Configuring Aggregation – Cisco IOS

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

```
router bgp 64511
  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 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
 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
```

13

Announcing an Aggregate

- ISPs who don't and won't aggregate are held in poor regard by community
- Registries publish their minimum allocation size
  - Now ranging from a /20 to a /24 depending on RIR
  - Different sizes for different address blocks
  - (APNIC changed its minimum allocation to /24 in October 2010)
- Until recently there was no real reason to see anything longer than a /22 prefix in the Internet
  - BUT there are currently (Feb 2013) >232000 /24s!
  - IPv4 run-out is starting to have an impact

Aggregation – Example

100.10.10.0/23
100.10.0.0/24
100.10.4.0/22
…

Internet

AS100

customer

100.10.10.0/23
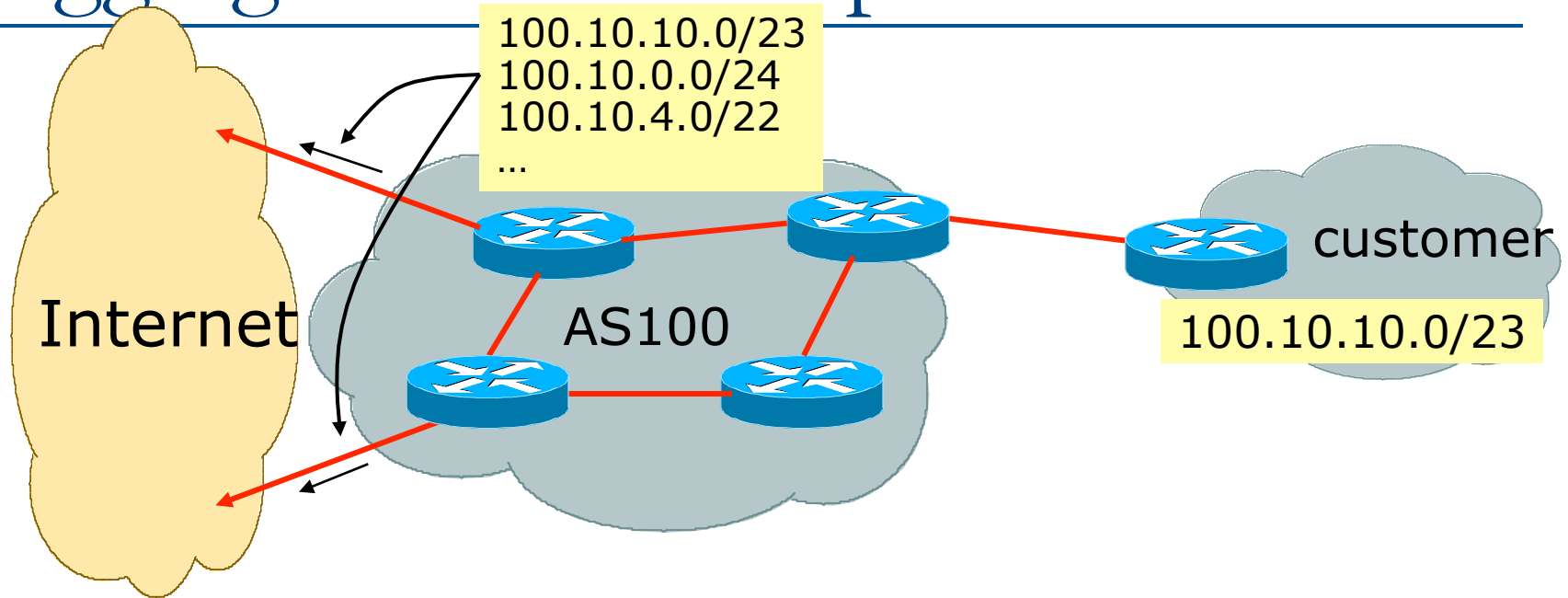
- Customer has /23 network assigned from AS100's /19 address block
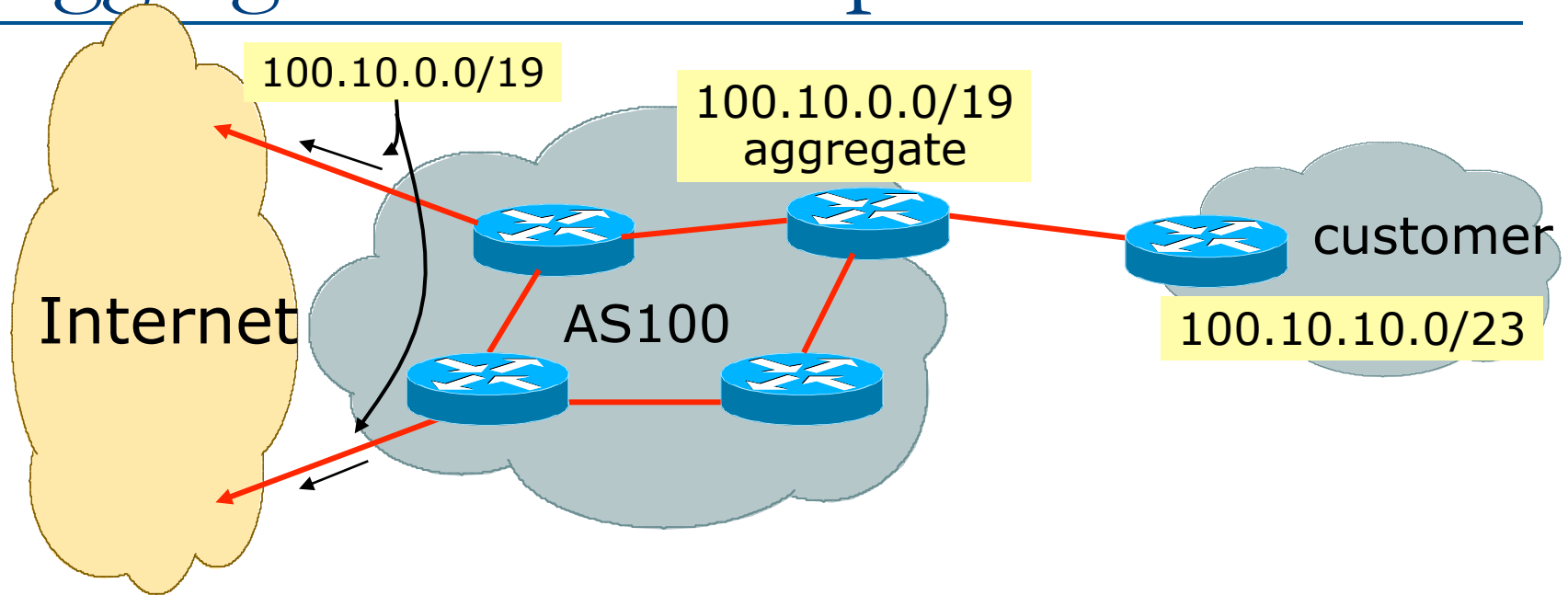- AS100 announces customers' individual networks to the Internet

15

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???

16

Aggregation – Example

100.10.0.0/19

100.10.0.0/19
aggregate

Internet

AS100

customer

100.10.10.0/23

- 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

18

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

The Internet Today (February 2013)

- Current Internet Routing Table Statistics
    - BGP Routing Table Entries   442978
    - Prefixes after maximum aggregation   182249
    - Unique prefixes in Internet   217374
    - Prefixes smaller than registry alloc   156337
    - /24s announced   232304
    - ASes in use   43294

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 recommendation
  - RIPE-399 — 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

Status Summary

Table History

Date	Prefixes	CIDR Aggregated
10-02-13	445138	254258
11-02-13	445138	254536
12-02-13	445341	254468
13-02-13	444932	254971
14-02-13	445625	254890
15-02-13	445440	255148
16-02-13	445491	255598
17-02-13	445709	255589

Plot: BGP Table Size
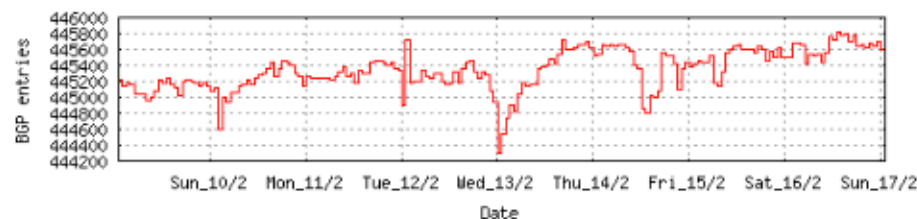
AS Summary

|	
43411	Number of ASes in routing system
18008	Number of ASes announcing only one prefix
3064	Largest number of prefixes announced by an AS
	AS6389: BELLSOUTH-NET-BLK - BellSouth.net Inc.
116912864	Largest address span announced by an AS (/32s)
	AS4134: CHINANET-BACKBONE No.31,Jin-rong Street

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)
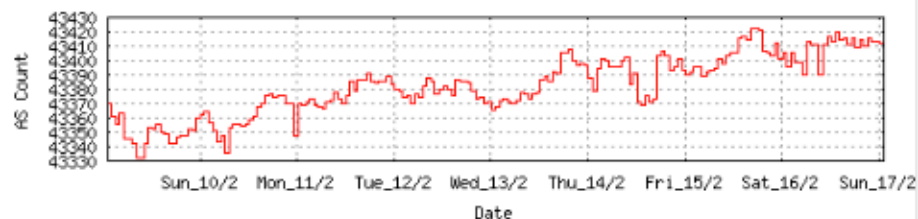
Announced Prefixes

```
Rank  AS        Type     Originate Addr Space  (pfx)   Transit Addr space  (pfx)  Description
15    AS6389             ORG+TRN Originate:    29723648 /7.17   Transit:      966656 /12.12 BELLSOUTH-NET-BLK - BellSouth.net Inc.
```

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       %
   2 AS6389          BELLSOUTH-NET-BLK - BellSouth.net Inc.      3064   2970    10    104   2960   96.61%


Prefix               AS Path                           Aggregation Suggestion
12.81.90.0/23        4777 2516 3356 7018 6389
12.81.120.0/24       4777 2516 3356 7018 6389
12.83.3.0/24         4777 2516 3356 7018 6389
12.83.5.0/24         4777 2516 3356 7018 6389
12.83.7.0/24         4777 2516 3356 7018 6389
65.0.0.0/12          4777 2516 3356 7018 6389
65.0.0.0/18          4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.40.0/22         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.50.0/23         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.64.0/18         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.128.0/18        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.192.0/19        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.224.0/19        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.0.0/19          4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.32.0/19         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.64.0/19         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.128.0/18        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.224.0/20        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.240.0/20        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.2.0.0/16          4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.2.0.0/17          4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.2.128.0/17        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.3.224.0/19        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.4.64.0/18         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.4.192.0/18        4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.1.0/24          4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.12.0/22         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.16.0/22         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.20.0/23         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.21.0/24         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.22.0/23         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.24.0/22         4777 2516 3356 7018 6389 - Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
```

Announced Prefixes

```
Rank  AS         Type      Originate Addr Space  (pfx)   Transit Addr space  (pfx)  Description
181   AS18566    ORG+TRN Originate:    2797568 /10.58  Transit:        1280 /21.68 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       %
   7 AS18566         COVAD - Covad Communications Co.       2080   1853    200    427   1653  79.47%


Prefix                  AS Path                          Aggregation Suggestion
64.81.16.0/22           4777 2516 3356 18566
64.81.20.0/22           4777 2516 4565 18566
64.81.22.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.20.0/22 4777 2516 4565 18566
64.81.24.0/21           4777 2516 3356 18566 + Announce - aggregate of 64.81.24.0/22 (4777 2516 3356 18566) and 64.81.28.0/22 (4777 25
64.81.24.0/22           4777 2516 3356 18566 - Withdrawn - aggregated with 64.81.28.0/22 (4777 2516 3356 18566)
64.81.28.0/22           4777 2516 3356 18566 - Withdrawn - aggregated with 64.81.24.0/22 (4777 2516 3356 18566)
64.81.32.0/20           4777 2516 4565 18566
64.81.32.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.33.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.34.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.35.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.36.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.37.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.38.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.39.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.40.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.44.0/24           4777 2516 4565 18566 - Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.48.0/20           4777 2516 3356 18566
64.81.48.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.49.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.50.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.51.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.52.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.53.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.54.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.55.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.56.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.57.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.58.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.59.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.60.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.61.0/24           4777 2516 3356 18566 - Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
```

Importance of Aggregation

- Size of routing table
  - Router Memory is not so much of a problem as it was in the 1990s
  - 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
  - bgpupdates.potaroo.net/instability/bgpupd.html

The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 16 February 2013 06:20 (UTC+1000)

50 Most active ASes for the past 7 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	9498	241243	7.80%	1071	225.25	BBIL-AP BHARTI Airtel Ltd.
2	24560	185739	6.00%	1044	177.91	AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
3	18207	92322	2.98%	548	168.47	YOU-INDIA-AP YOU Broadband & Cable India Ltd.
4	45609	56960	1.84%	264	215.76	BHARTI-MOBILITY-AS-AP Bharti Airtel Ltd. AS for GPRS Service
5	18002	55390	1.79%	210	263.76	WORLDPHONE-IN AS Number for Interdomain Routing
6	8402	52369	1.69%	2092	25.03	CORBINA-AS OJSC "Vimpelcom"
7	45514	44050	1.42%	306	143.95	TELEMEDIA-SMB-AS-AP Bharti Airtel Ltd., TELEMEDIA Services, for SMB customers
8	3909	43008	1.39%	34	1264.94	QWEST-AS-3908 - Qwest Communications Company, LLC
9	7029	38595	1.25%	2566	15.04	WINDSTREAM - Windstream Communications Inc
10	45528	34120	1.10%	602	56.68	TDN Tikona Digital Networks Pvt Ltd.
11	8151	30868	1.00%	1561	19.77	Uninet S.A. de C.V.
12	17488	29045	0.94%	672	43.22	HATHWAY-NET-AP Hathway IP Over Cable Internet
13	9829	26393	0.85%	1431	18.44	BSNL-NIB National Internet Backbone
14	29049	25326	0.82%	334	75.83	DELTA-TELECOM-AS Delta Telecom LTD.
15	2708	20135	0.65%	140	143.82	Universidad de Guanajuato
16	17447	19554	0.63%	129	151.58	NET4INDIA Net4India Ltd.
17	4	19416	0.63%	20	970.80	ISI-AS - University of Southern California
18	23682	19323	0.62%	147	131.45	PACENET-AS Broadband Pacenet Pvt. Ltd
19	45769	19250	0.62%	249	77.31	DVOIS-IN D-Vois Broadband Pvt Ltd
20	17917	18911	0.61%	118	160.26	QTLTELECOM-AS-AP Quadrant Televentures Limited
21	17439	16947	0.55%	208	81.48	NETMAGIC-AP Netmagic Datacenter Mumbai
22	45271	16863	0.54%	337	50.04	ICLNET-AS-AP 5th Floor, Windsor Building, Off: CST Road
23	36998	15469	0.50%	1286	12.03	SDN-MOBITEL
24	6648	13840	0.45%	222	62.34	BAYAN-TELECOMMUNICATIONS Bayan Telecommunications, Inc.

50 Most active Prefixes for the past 7 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	151.118.18.0/24	14307	0.44%	3909 -- QWEST-AS-3908 - Qwest Communications Company, LLC
2	151.118.255.0/24	14297	0.44%	3909 -- QWEST-AS-3908 - Qwest Communications Company, LLC
3	151.118.254.0/24	14297	0.44%	3909 -- QWEST-AS-3908 - Qwest Communications Company, LLC
4	202.41.70.0/24	9225	0.29%	2697 -- ERX-ERNET-AS Education and Research Network
5	37.9.248.0/21	9076	0.28%	57816 -- SAMEN Samen Ertebat Asr Co. (P.J.S.)
6	208.92.131.0/24	8585	0.27%	40946 -- PROCON - Sat Track
7	196.1.167.0/24	8131	0.25%	11139 -- CWRIN CW BARBADOS
8	192.58.232.0/24	7498	0.23%	6629 -- NOAA-AS - NOAA
9	208.14.186.0/24	6980	0.22%	22140 -- T-MOBILE-AS22140 - T-Mobile USA, Inc.
10	12.139.133.0/24	5049	0.16%	14680 -- REALE-6 - Auction.com
11	58.184.229.0/24	4825	0.15%	9950 -- PUBNETPLUS2-AS-KR DACOM
12	194.63.9.0/24	4673	0.14%	1273 -- CW Cable and Wireless Worldwide plc
13	69.38.178.0/24	4187	0.13%	19406 -- TWRS-MA - Towerstream I, Inc.
14	84.205.66.0/24	3216	0.10%	12654 -- RIPE-NCC-RIS-AS Reseaux IP Europeens Network Coordination Centre (RIPE NCC)
15	206.105.75.0/24	3009	0.09%	6174 -- SPRINTLINK8 - Sprint
16	208.16.110.0/24	3009	0.09%	6174 -- SPRINTLINK8 - Sprint
17	192.207.8.0/22	2963	0.09%	701 -- UUNET - MCI Communications Services, Inc. d/b/a Verizon Business
18	192.207.4.0/22	2961	0.09%	701 -- UUNET - MCI Communications Services, Inc. d/b/a Verizon Business
19	167.235.247.0/24	2818	0.09%	25806 -- RALEYNET - Raleys
20	211.240.105.0/24	2740	0.08%	4663 -- ELIMNET-AS ELIMNET, INC
21	41.188.128.0/21	2590	0.08%	37084 -- simbanet-tz
22	115.170.128.0/17	2536	0.08%	4847 -- CNIX-AP China Networks Inter-Exchange
23	66.55.234.0/24	2450	0.08%	7029 -- WINDSTREAM - Windstream Communications Inc
24	8.12.92.0/22	2364	0.07%	32777 -- SMART-CITY-FORT-WORTH - Smart City Networks, L.P.
25	66.78.242.0/23	2364	0.07%	32777 -- SMART-CITY-FORT-WORTH - Smart City Networks, L.P.
26	67.110.75.0/24	2364	0.07%	32777 -- SMART-CITY-FORT-WORTH - Smart City Networks, L.P.
27	65.172.212.0/22	2364	0.07%	32777 -- SMART-CITY-FORT-WORTH - Smart City Networks, L.P.
28	97.65.228.0/22	2364	0.07%	32777 -- SMART-CITY-FORT-WORTH - Smart City Networks, L.P.
29	2.93.235.0/24	2327	0.07%	8402 -- CORBINA-AS OJSC "Vimpelcom"

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
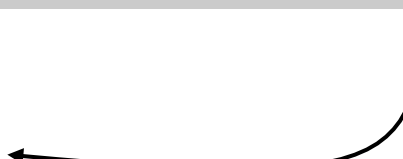 - The tool: whois –h jwhois.apnic.net x.x.x.0/24
    - (jwhois queries all RIR databases)

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:        202.12.28.0 - 202.12.29.255
netname:        APNIC-AP
descr:          Asia Pacific Network Information Centre
descr:          Regional Internet Registry for the Asia-Pacific
descr:          6 Cordelia Street
descr:          South Brisbane, QLD 4101
descr:          Australia
country:        AU
admin-c:        AIC1-AP
tech-c:         NO4-AP
mnt-by:         APNIC-HM
mnt-irt:        IRT-APNIC-AP
changed:        hm-changed@apnic.net
status:         ASSIGNED PORTABLE
changed:        hm-changed@apnic.net 20110309
source:         APNIC
```

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

33

Receiving Prefixes: From Customers

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

```
$ whois -h whois.ripe.net 193.128.0.0
inetnum:        193.128.0.0 - 193.133.255.255
netname:        UK-PIPEX-193-128-133
descr:          Verizon UK Limited
country:        GB
org:            ORG-UA24-RIPE
admin-c:        WERT1-RIPE
tech-c:         UPHM1-RIPE
status:         ALLOCATED UNSPECIFIED
remarks:        Please send abuse notification to abuse@uk.uu.net
mnt-by:         RIPE-NCC-HM-MNT
mnt-lower:      AS1849-MNT
mnt-routes:     AS1849-MNT
mnt-routes:     WCOM-EMEA-RICE-MNT
mnt-irt:        IRT-MCI-GB
source:         RIPE # Filtered
```

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

34

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
  - 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
  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 default (unless you need it)
  - Don't accept your own prefixes
- For IPv4:
  - Don't accept private (RFC1918) and certain special use prefixes:
    http://www.rfc-editor.org/rfc/rfc5735.txt
  - Don't accept prefixes longer than /24 (?)
- For IPv6:
  - Don't accept certain special use prefixes:
    http://www.rfc-editor.org/rfc/rfc5156.txt
  - Don't accept prefixes longer than /48 (?)

Receiving Prefixes:
From Upstream/Transit Provider

- Check Team Cymru's list of "bogons"

  www.team-cymru.org/Services/Bogons/http.html

- For IPv4 also consult:

  www.rfc-editor.org/rfc/rfc6441.txt

- For IPv6 also consult:

  www.space.net/~gert/RIPE/ipv6-filters.html

- Bogon Route Server:

  www.team-cymru.org/Services/Bogons/routeserver.html

  - 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 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                  ! Default
ip prefix-list in-filter deny 0.0.0.0/8 le 32            ! Network Zero
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 addr
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 192.0.2.0/24 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
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
```

44

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 2001::/32          ! Teredo
ipv6 prefix-list v6in-filter deny 2001::/32 le 128     ! Teredo subnets
ipv6 prefix-list v6in-filter deny 2001:db8::/32 le 128  ! Documentation
ipv6 prefix-list v6in-filter permit 2002::/16          ! 6to4
ipv6 prefix-list v6in-filter deny 2002::/16 le 128     ! 6to4 subnets
ipv6 prefix-list v6in-filter deny 2020:3030::/32 le 128 ! Local Prefix
ipv6 prefix-list v6in-filter deny 3ffe::/16 le 128     ! Old 6bone
ipv6 prefix-list v6in-filter permit 2000::/3 le 48     ! Global Unicast
ipv6 prefix-list v6in-filter deny ::/0 le 128
```

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

Summary

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

BGP Best Current Practices

AFNOG 2013 AR-E Workshop