BGP Techniques for Internet Service Providers

Philip Smith <pfs@cisco.com>

AfNOG 2011

Dar Es Salaam, Tanzania

5 June 2011

Presentation Slides

- Will be available on

  ftp://ftp-eng.cisco.com

  /pfs/seminars/AfNOG2011-BGP-Techniques.pdf

  And on the AfNOG2011 website

- Feel free to ask questions any time

Deploying BGP in an ISP Network

We've learned about BGP in SI-E/F and AR-E… What now?

Deploying BGP

- The role of IGPs and iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Deploying 4-byte ASNs

The role of IGP and iBGP

Ships in the night?

Or

Good foundations?

BGP versus OSPF/ISIS

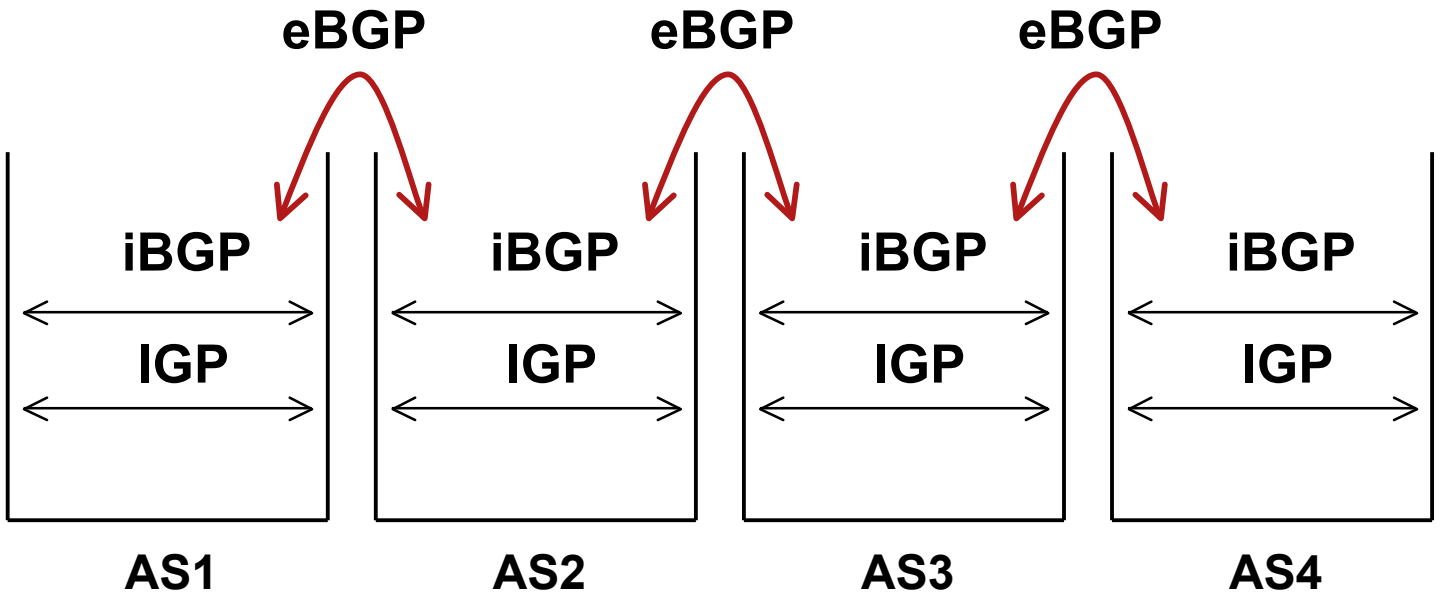
- Internal Routing Protocols (IGPs)
  - examples are ISIS and OSPF
  - used for carrying infrastructure addresses
  - **NOT** used for carrying Internet prefixes or customer prefixes
  - design goal is to minimise number of prefixes in IGP to aid scalability and rapid convergence

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/IGP model used in ISP networks

- Model representation



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

Injecting prefixes into iBGP

- Use iBGP to carry customer prefixes

  Don't ever use IGP

- Point static route to customer interface
- Enter network into BGP process

  Ensure that implementation options are used so that the prefix always remains in iBGP, regardless of state of interface

  i.e. avoid iBGP flaps caused by interface flaps

Aggregation

Quality or Quantity?

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

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 purposes

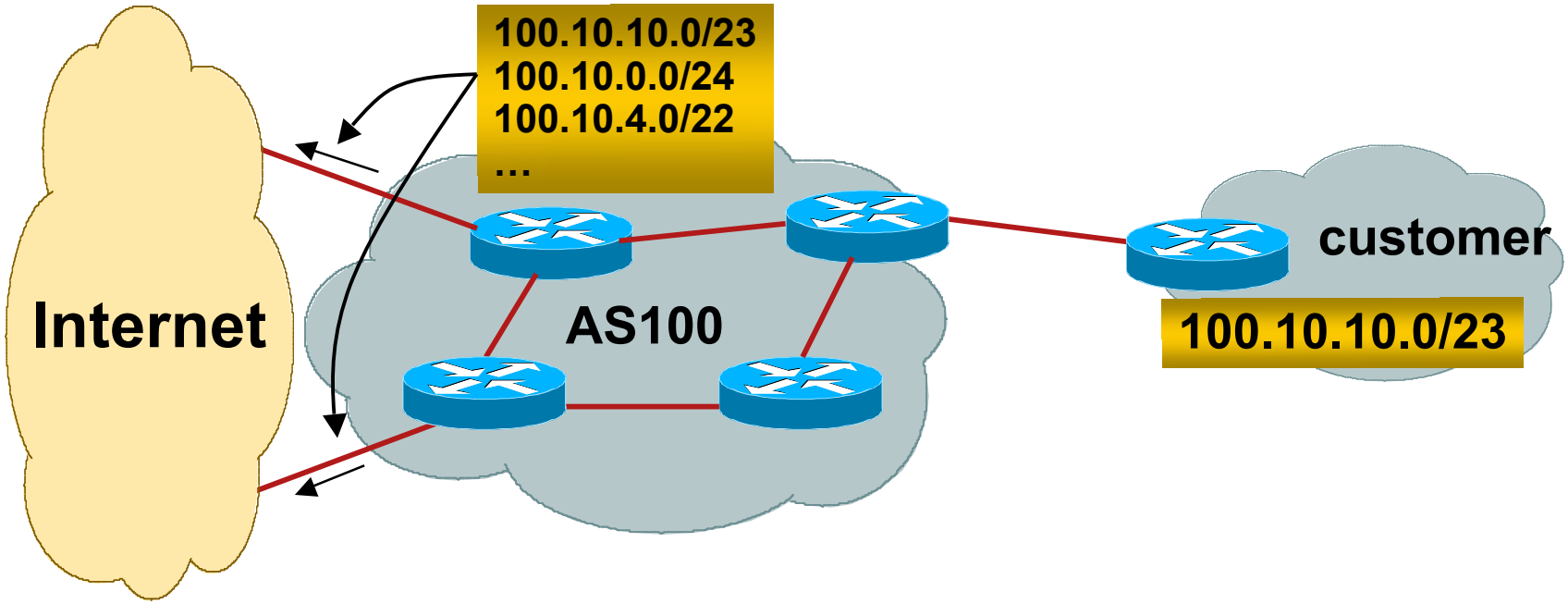
  (see BGP Multihoming Tutorial)
- Aggregate should be generated internally

  Not on the network borders!

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 /24 depending on RIR
  - Different sizes for different address blocks
  - There are currently >185000 /24s!
- APNIC changed (Oct 2010) its minimum allocation size on all blocks to /24
  - IPv4 run-out is starting to have an impact

Aggregation – Example

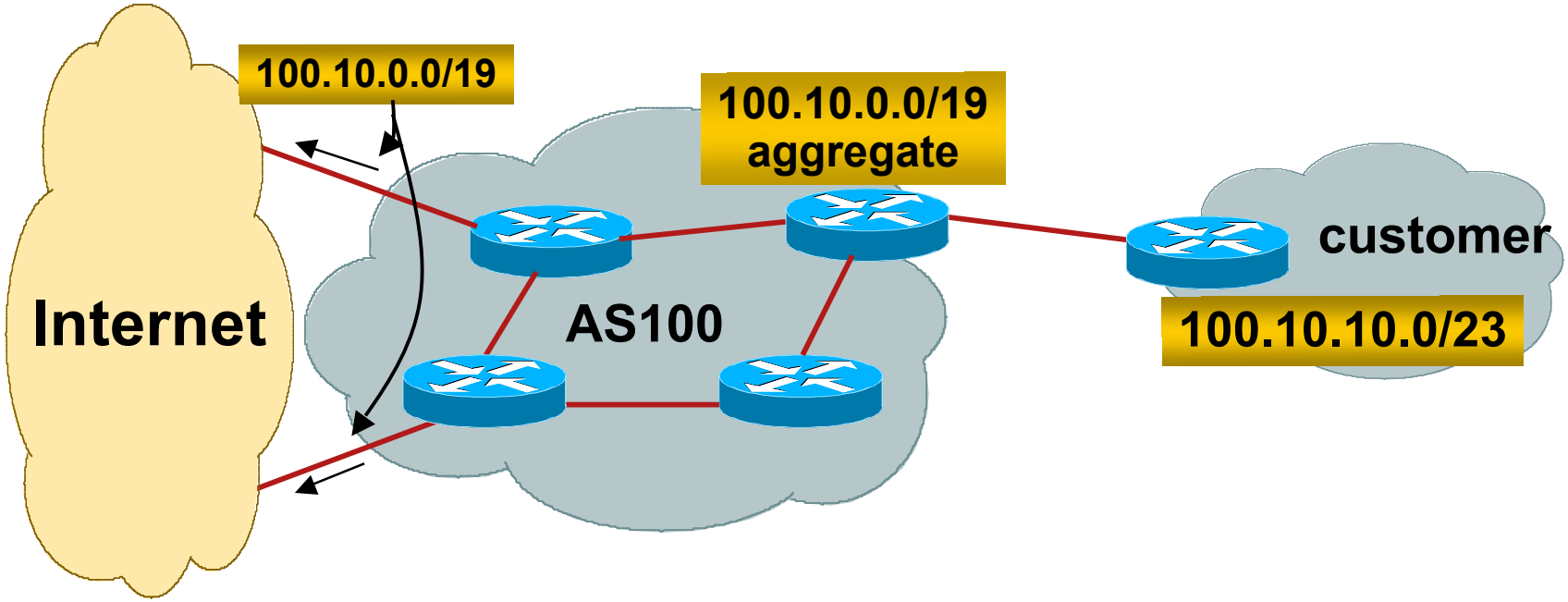


- Customer has /23 network assigned from AS100's /19 address block
- AS100 announces customers' individual networks to the Internet

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???

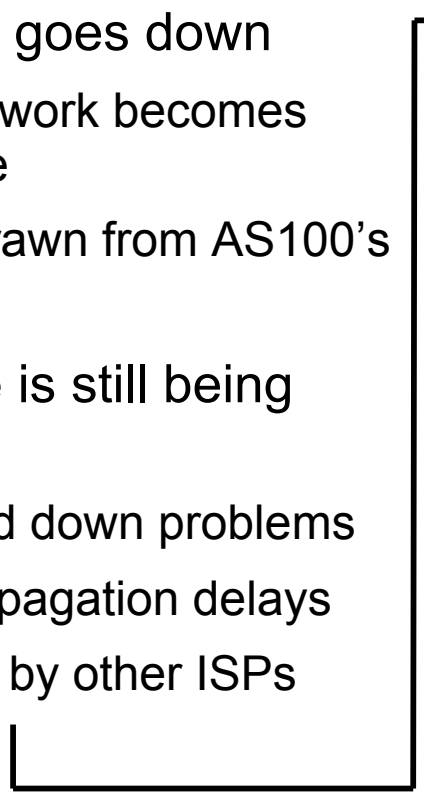
Aggregation – Example



- 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**

The Internet Today (1st June 2011)

- Current Internet Routing Table Statistics

BGP Routing Table Entries	358603
Prefixes after maximum aggregation	162337
Unique prefixes in Internet	178173
Prefixes smaller than registry alloc	149545
/24s announced	186667
ASes in use	37758

“The New Swamp”

- Swamp space is name used for areas of poor aggregation
  - The original swamp was 192.0.0.0/8 from the former class C block
    - Name given just after the deployment of CIDR
  - The new swamp is creeping across all parts of the Internet
    - Not just RIR space, but “legacy” space too

“The New Swamp”
RIR Space – February 1999

RIR blocks contribute 88% of the Internet Routing Table

Block	Networks	Block	Networks	Block	Networks	Block	Networks
24/8	**165**	*79/8*	*0*	*118/8*	*0*	*201/8*	*0*
41/8	*0*	*80/8*	*0*	*119/8*	*0*	**202/8**	**2276**
58/8	*0*	*81/8*	*0*	*120/8*	*0*	**203/8**	**3622**
59/8	*0*	*82/8*	*0*	*121/8*	*0*	**204/8**	**3792**
60/8	*0*	*83/8*	*0*	*122/8*	*0*	**205/8**	**2584**
61/8	**3**	*84/8*	*0*	*123/8*	*0*	**206/8**	**3127**
62/8	**87**	*85/8*	*0*	*124/8*	*0*	**207/8**	**2723**
63/8	**20**	*86/8*	*0*	*125/8*	*0*	**208/8**	**2817**
64/8	*0*	*87/8*	*0*	*126/8*	*0*	**209/8**	**2574**
65/8	*0*	*88/8*	*0*	*173/8*	*0*	**210/8**	**617**
66/8	*0*	*89/8*	*0*	*174/8*	*0*	*211/8*	*0*
67/8	*0*	*90/8*	*0*	*186/8*	*0*	**212/8**	**717**
68/8	*0*	*91/8*	*0*	*187/8*	*0*	**213/8**	**1**
69/8	*0*	*96/8*	*0*	*189/8*	*0*	**216/8**	**943**
70/8	*0*	*97/8*	*0*	*190/8*	*0*	*217/8*	*0*
71/8	*0*	*98/8*	*0*	**192/8**	**6275**	*218/8*	*0*
72/8	*0*	*99/8*	*0*	**193/8**	**2390**	*219/8*	*0*
73/8	*0*	*112/8*	*0*	**194/8**	**2932**	*220/8*	*0*
74/8	*0*	*113/8*	*0*	**195/8**	**1338**	*221/8*	*0*
75/8	*0*	*114/8*	*0*	**196/8**	**513**	*222/8*	*0*
76/8	*0*	*115/8*	*0*	**198/8**	**4034**		
77/8	*0*	*116/8*	*0*	**199/8**	**3495**		
78/8	*0*	*117/8*	*0*	**200/8**	**1348**		

“The New Swamp”
RIR Space – February 2010

RIR blocks contribute about 87% of the Internet Routing Table

Block	Networks	Block	Networks	Block	Networks	Block	Networks
24/8	3328	79/8	1119	118/8	1349	201/8	4136
41/8	3448	80/8	2335	119/8	1694	202/8	11354
58/8	1675	81/8	1709	120/8	531	203/8	11677
59/8	1575	82/8	1358	121/8	1756	204/8	5744
60/8	888	83/8	1357	122/8	2687	205/8	3037
61/8	2890	84/8	1341	123/8	2400	206/8	3951
62/8	2418	85/8	2492	124/8	2259	207/8	4635
63/8	3114	86/8	780	125/8	2514	208/8	6498
64/8	*6601*	87/8	1466	126/8	106	209/8	5536
65/8	3966	88/8	1068	173/8	1994	210/8	4977
66/8	*7782*	89/8	3168	174/8	1089	211/8	3130
67/8	3771	90/8	377	186/8	1223	212/8	3550
68/8	3221	91/8	4555	187/8	1501	213/8	3442
69/8	5280	96/8	778	189/8	3063	216/8	7645
70/8	2008	97/8	725	190/8	6945	217/8	3136
71/8	1327	98/8	1312	192/8	6952	218/8	1512
72/8	4050	99/8	288	193/8	6820	219/8	1303
73/8	4	112/8	883	194/8	5177	220/8	2108
74/8	*5074*	113/8	890	195/8	5325	221/8	980
75/8	1164	114/8	996	196/8	1857	222/8	1058
76/8	1034	115/8	1616	198/8	4504		
77/8	1964	116/8	1755	199/8	4372		
78/8	1397	117/8	1611	200/8	8884		

“The New Swamp” Summary

- RIR space shows creeping deaggregation

  It seems that an RIR /8 block averages around 5000 prefixes (and upwards) once fully allocated

- Food for thought:

  The 120 RIR /8s combined will cause:

  635000 prefixes with 5000 prefixes per /8 density

  762000 prefixes with 6000 prefixes per /8 density

  Plus 12% due to “non RIR space deaggregation”

  → Routing Table size of 853440 prefixes

“The New Swamp” Summary

- Rest of address space is showing similar deaggregation too ☹
- What are the reasons?
  - Main justification is traffic engineering
- Real reasons are:
  - Lack of knowledge
  - Laziness
  - Deliberate & knowing actions

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

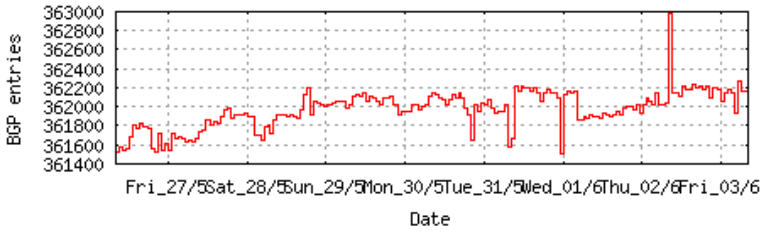
- Also computes the size of the routing table assuming ISPs performed optimal aggregation
- Website allows searches and computations of aggregation to be made on a per AS basis
  - Flexible and powerful tool to aid ISPs
  - Intended to show how greater efficiency in terms of BGP table size can be obtained 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

A list of advertisements of address blocks and Autonomous System numbers where there is no matching allocation data.

Status Summary

Table History

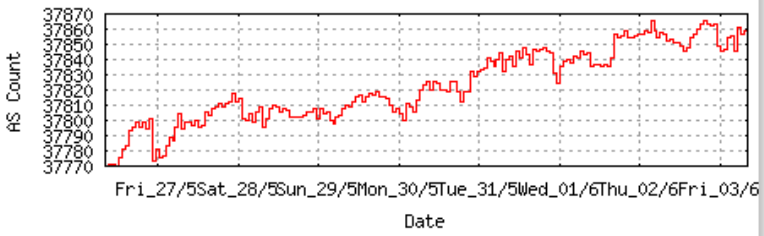
Date	Prefixes	CIDR Aggregated
27-05-11	361620	212256
28-05-11	361939	212249
29-05-11	362005	212296
30-05-11	361957	212301
31-05-11	362044	212140
01-06-11	361502	212100
02-06-11	361940	212371
03-06-11	362189	212506



Plot: BGP Table Size

AS Summary

37867	Number of ASes in routing system
15949	Number of ASes announcing only one prefix
3641	Largest number of prefixes announced by an AS AS6389: BELLSOUTH-NET-BLK - BellSouth.net Inc.
110390016	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)

Aggregation Summary

Aggregation Summary

The algorithm used in this report proposes aggregation only when there is a precise match using AS path so as to preserve traffic transit policies. Aggregation is also proposed across non-advertised address space ('holes').

--- 03Jun11 ---

ASnum	NetsNow	NetsAggr	NetGain	% Gain	Description
Table	362179	212511	149668	41.3%	All ASes
AS6389	3641	260	3381	92.9%	BELLSOUTH-NET-BLK - BellSouth.net Inc.
AS4323	1974	402	1572	79.6%	TWTC - tw telecom holdings, inc.
AS4766	2460	932	1528	62.1%	KIXS-AS-KR Korea Telecom
AS6478	1695	331	1364	80.5%	ATT-INTERNET3 - AT&T Services, Inc.
AS22773	1338	97	1241	92.8%	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications Inc.
AS19262	1492	307	1185	79.4%	VZGNI-TRANSIT - Verizon Online LLC
AS18566	1875	720	1155	61.6%	COVAD - Covad Communications Co.
AS4755	1472	386	1086	73.8%	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP
AS1785	1791	759	1032	57.6%	AS-PAETEC-NET - PaeTec Communications, Inc.
AS28573	1321	312	1009	76.4%	NET Servicos de Comunicao S.A.
AS7552	1138	131	1007	88.5%	VIETEL-AS-AP Vietel Corporation
AS10620	1512	591	921	60.9%	Telmex Colombia S.A.
AS7545	1538	746	792	51.5%	TPG-INTERNET-AP TPG Internet Pty Ltd
AS18101	933	145	788	84.5%	RELIANCE-COMMUNICATIONS-IN Reliance Communications Ltd.DAKC MUMBAI
AS24560	1155	390	765	66.2%	AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
AS4808	1094	336	758	69.3%	CHINA169-BJ CNCGROUP IP network China169 Beijing Province Network
AS8151	1380	647	733	53.1%	Uninet S.A. de C.V.
AS7303	958	292	666	69.5%	Telecom Argentina S.A.
AS3356	1113	454	659	59.2%	LEVEL3 Level 3 Communications
AS17488	930	312	618	66.5%	HATHWAY-NET-AP Hathway IP Over Cable Internet
AS17974	1541	924	617	40.0%	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
AS17676	658	70	588	89.4%	GIGAINFRA Softbank BB Corp.
AS14420	674	88	586	86.9%	CORPORACION NACIONAL DE TELECOMUNICACIONES - CNT EP
AS3549	959	403	556	58.0%	GBLX Global Crossing Ltd.
AS855	646	109	537	83.1%	CANET-ASN-4 - Bell Aliant Regional Communications, Inc.
AS4780	749	213	536	71.6%	SEEDNET Digital United Inc.
AS22047	565	30	535	94.7%	VTR BANDA ANCHA S.A.

Top 20 Added Routes this week per Originating AS

Prefixes	ASnum	AS Description
135	AS237	MERIT-AS-14 - Merit Network Inc.
66	AS5541	ADNET-TELECOM AdNet Telecom
64	AS3	MIT-GATEWAYS - Massachusetts Institute of Technology
62	AS45194	SIPL-AS Syscon Infoway Pvt. Ltd., Internet Service Provider, India.
61	AS4	ISI-AS - University of Southern California
45	AS9198	KAZTELECOM-AS JSC Kazakhtelecom
37	AS7738	Telecomunicacoes da Bahia S.A.
30	AS29571	CITelecom-AS
24	AS10695	WAL-MART - Wal-Mart Stores, Inc.
22	AS36992	ETISALAT-MISR
21	AS20299	Newcom Limited
20	AS10620	Telmex Colombia S.A.
20	AS24835	RAYA-AS
20	AS22561	DIGITAL-TELEPORT - Digital Teleport Inc.
19	AS45904	BANGLALION-WIMAX-BD Silver Tower (16 & 18th Floor)
19	AS38207	RAJESHNET-TRANSIT-AS-AP Rajesh Multi Channel Pvt Ltd.
17	AS50664	PUBGROUPE-FR PUBLICIS-TECHNOLOGY
17	AS47589	KTC3G Kuwait Telecommunication Company (Under Association)
16	AS28009	Davitel S.A.
16	AS26929	DOLLAR-PHONE-CORP-SUPERNET - Dollar Phone Corp / Supernet

Top 20 Withdrawn Routes this week per Originating AS

Prefixes	ASnum	AS Description
-320	AS17974	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
-67	AS18049	TINP-TW Taiwan Infrastructure Network Technologie
-57	AS9839	MTC-GTEPACIFICA-AS Micronesian Telecommunications Corp
-46	AS5976	DNIC-ASBLK-05800-06055 - DoD Network Information Center
-42	AS8452	TE-AS TE-AS
-38	AS29256	STE-AS Syrian Telecommunications Establishment
-32	AS24835	RAYA-AS
-24	AS32851	WAL-MART-2 - Wal-Mart Stores, Inc.
-23	AS55714	APNIC-FIBERLINK-PK Fiberlink Pvt.Ltd
-23	AS45595	PKTELECOM-AS-PK Pakistan Telecom Company Limited
-23	AS30653	EXOBIT - Exobit Networks Inc.
-21	AS27773	MILLICOM CABLE EL SALVADOR S.A. DE C.V.
-21	AS6478	ATT-INTERNET3 - AT&T Services, Inc.

More Specifics

A list of route advertisements that appear to be more specfic than the original Class-based prefix mask, or more specific than the registry allocation size.

Top 20 ASes advertising more specific prefixes

More Specifics	Total Prefixes	ASnum	AS Description
3544	3641	AS6389	BELLSOUTH-NET-BLK - BellSouth.net Inc.
2403	2460	AS4766	KIXS-AS-KR Korea Telecom
1853	1875	AS18566	COVAD - Covad Communications Co.
1775	1974	AS4323	TWTC - tw telecom holdings, inc.
1703	1791	AS1785	AS-PAETEC-NET - PaeTec Communications, Inc.
1694	1695	AS6478	ATT-INTERNET3 - AT&T Services, Inc.
1531	1586	AS20115	CHARTER-NET-HKY-NC - Charter Communications
1528	1541	AS17974	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
1510	1512	AS10620	Telmex Colombia S.A.
1482	1538	AS7545	TPG-INTERNET-AP TPG Internet Pty Ltd
1459	1472	AS4755	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP
1427	1492	AS19262	VZGNI-TRANSIT - Verizon Online LLC
1373	1380	AS8151	Uninet S.A. de C.V.
1321	1321	AS28573	NET Servicos de Comunicao S.A.
1293	1338	AS22773	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications Inc.
1265	1274	AS11492	CABLEONE - CABLE ONE, INC.
1175	1266	AS2386	INS-AS - AT&T Data Communications Services
1155	1160	AS7011	FRONTIER-AND-CITIZENS - Frontier Communications of America, Inc.
1155	1155	AS24560	AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
1135	1138	AS7552	VIETEL-AS-AP Vietel Corporation

Report: ASes ordered by number of more specific prefixes
Report: More Specific prefix list (by AS)
Report: More Specific prefix list (ordered by prefix)

Possible Bogus Routes and AS Announcements

Announced Prefixes

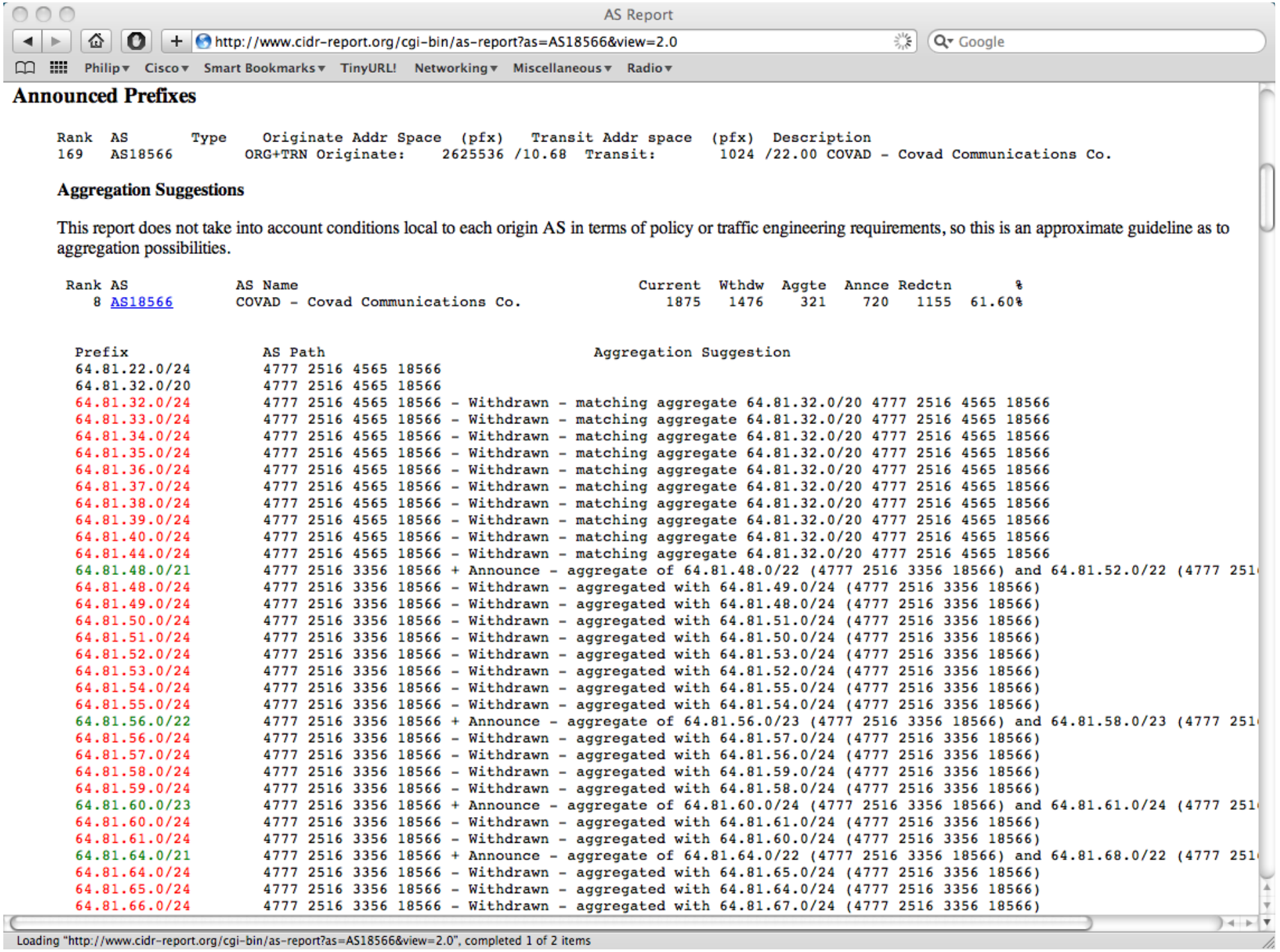
Rank	AS	Type	Originate Addr Space	(pfx)	Transit Addr space	(pfx)	Description
130	AS4755	ORG+TRN	Originate:	3627264 /10.21	Transit:	10856192 /8.63	TATACOMM-AS TATA Communications formerly VSNL is

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	%
9	AS4755	TATACOMM-AS TATA Communications formerly VSNL	1472	1139	53	386	1086	73.78%

Prefix	AS Path	Aggregation Suggestion
14.140.0.0/14	4777 2516 6453 4755	
14.140.0.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.4.0/23	4608 1221 4637 6453 4755	
14.140.6.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.16.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.20.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.24.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.32.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.40.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.48.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.56.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.64.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.72.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.80.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.82.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.84.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.88.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.96.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.104.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.112.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.254.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
49.32.0.0/12	4777 2516 6453 4755	
59.151.144.0/22	4608 1221 4637 6453 4755	
59.160.0.0/16	4777 2516 6453 4755	
59.160.0.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.4.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.5.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.8.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.11.0/24	4777 2516 6453 4755 4755	
59.160.12.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.15.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.16.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.24.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755



Announced Prefixes

Rank	AS	Type	Originate Addr Space	(pfx)	Transit Addr space	(pfx)	Description
169	AS18566	ORG+TRN	Originate:	2625536 /10.68	Transit:	1024 /22.00	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	%
8	AS18566	COVAD - Covad Communications Co.	1875	1476	321	720	1155	61.60%

Prefix	AS Path	Aggregation Suggestion
64.81.22.0/24	4777 2516 4565 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
64.81.48.0/21	4777 2516 3356 18566	+ Announce - aggregate of 64.81.48.0/22 (4777 2516 3356 18566) and 64.81.52.0/22 (4777 251
64.81.48.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.49.0/24 (4777 2516 3356 18566)
64.81.49.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.48.0/24 (4777 2516 3356 18566)
64.81.50.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.51.0/24 (4777 2516 3356 18566)
64.81.51.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.50.0/24 (4777 2516 3356 18566)
64.81.52.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.53.0/24 (4777 2516 3356 18566)
64.81.53.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.52.0/24 (4777 2516 3356 18566)
64.81.54.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.55.0/24 (4777 2516 3356 18566)
64.81.55.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.54.0/24 (4777 2516 3356 18566)
64.81.56.0/22	4777 2516 3356 18566	+ Announce - aggregate of 64.81.56.0/23 (4777 2516 3356 18566) and 64.81.58.0/23 (4777 251
64.81.56.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.57.0/24 (4777 2516 3356 18566)
64.81.57.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.56.0/24 (4777 2516 3356 18566)
64.81.58.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.59.0/24 (4777 2516 3356 18566)
64.81.59.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.58.0/24 (4777 2516 3356 18566)
64.81.60.0/23	4777 2516 3356 18566	+ Announce - aggregate of 64.81.60.0/24 (4777 2516 3356 18566) and 64.81.61.0/24 (4777 251
64.81.60.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.61.0/24 (4777 2516 3356 18566)
64.81.61.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.60.0/24 (4777 2516 3356 18566)
64.81.64.0/21	4777 2516 3356 18566	+ Announce - aggregate of 64.81.64.0/22 (4777 2516 3356 18566) and 64.81.68.0/22 (4777 251
64.81.64.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.65.0/24 (4777 2516 3356 18566)
64.81.65.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.64.0/24 (4777 2516 3356 18566)
64.81.66.0/24	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.67.0/24 (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
  - **http://bgpupdates.potaroo.net/instability/bgpupd.html**

The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 03 June 2011 06:12 (UTC+1000)

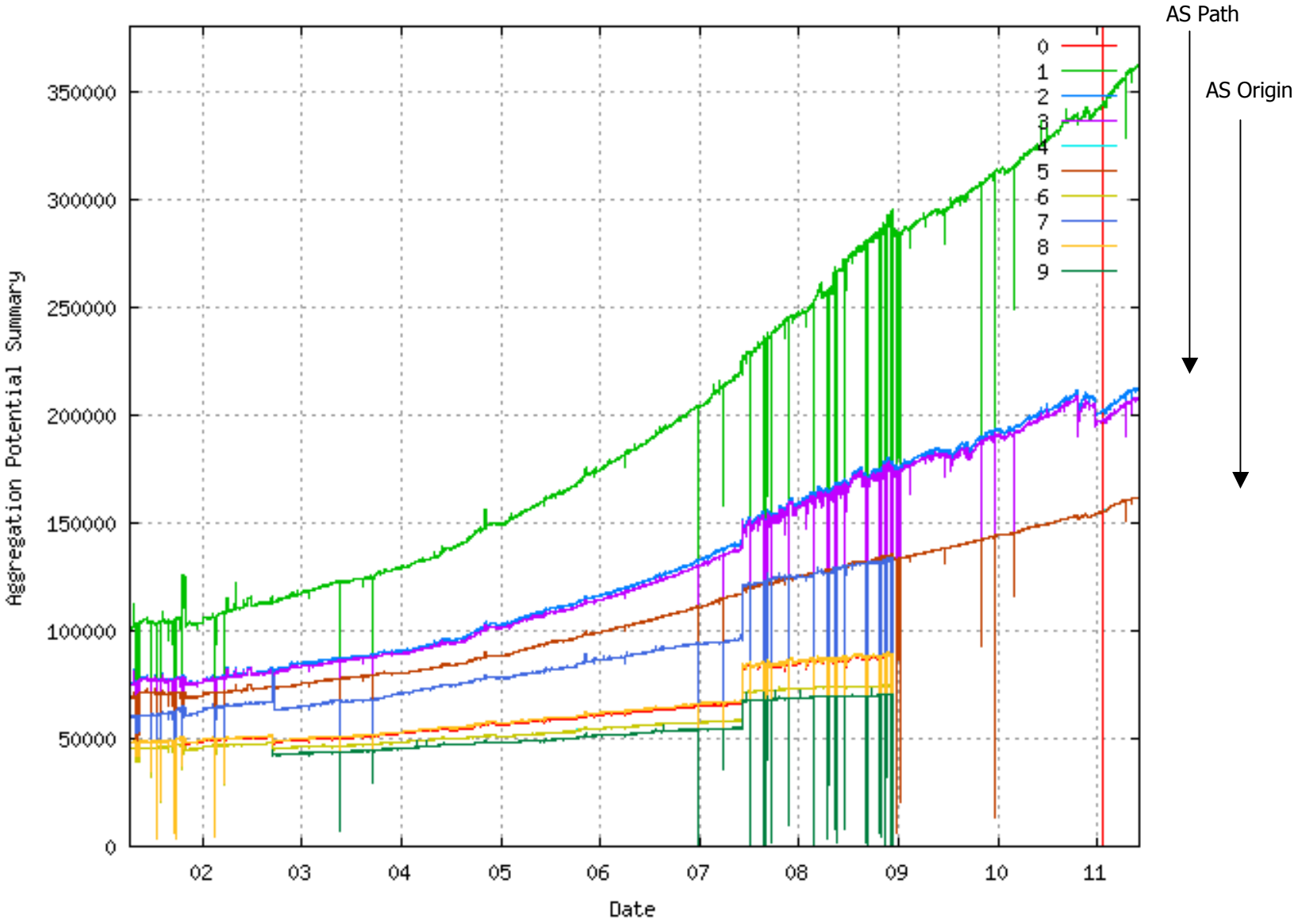
50 Most active ASes for the past 7 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	9829	57552	3.92%	1044	55.13	BSNL-NIB National Internet Backbone
2	33475	45744	3.12%	215	212.76	RSN-1 - RockSolid Network, Inc.
3	24560	33904	2.31%	1156	29.33	AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
4	19743	33903	2.31%	7	4843.29	
5	9498	27999	1.91%	828	33.82	BBIL-AP BHARTI Airtel Ltd.
6	27738	22165	1.51%	339	65.38	Ecuadortelecom S.A.
7	17974	17751	1.21%	1864	9.52	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
8	11492	17633	1.20%	1283	13.74	CABLEONE - CABLE ONE, INC.
9	32528	15576	1.06%	8	1947.00	ABBOTT Abbot Labs
10	3320	9731	0.66%	428	22.74	DTAG Deutsche Telekom AG
11	17488	9398	0.64%	934	10.06	HATHWAY-NET-AP Hathway IP Over Cable Internet
12	2697	9387	0.64%	202	46.47	ERX-ERNET-AS Education and Research Network
13	45514	8527	0.58%	304	28.05	TELEMEDIA-SMB-AS-AP Bharti Airtel Ltd., TELEMEDIA Services, for SMB customers
14	45595	8318	0.57%	364	22.85	PKTELECOM-AS-PK Pakistan Telecom Company Limited
15	27065	7747	0.53%	124	62.48	DNIC-ASBLK-27032-27159 - DoD Network Information Center
16	3454	7635	0.52%	8	954.38	Universidad Autonoma de Nuevo Leon
17	7552	7577	0.52%	1141	6.64	VIETEL-AS-AP Vietel Corporation
18	701	7050	0.48%	622	11.33	UUNET - MCI Communications Services, Inc. d/b/a Verizon Business
19	8151	6906	0.47%	1397	4.94	Uninet S.A. de C.V.
20	18002	6859	0.47%	170	40.35	WORLDPHONE-IN AS Number for Interdomain Routing
21	29049	6161	0.42%	456	13.51	DELTA-TELECOM-AS Delta Telecom LTD.
22	8402	5973	0.41%	496	12.04	CORBINA-AS Corbina Telecom
23	28573	5841	0.40%	1329	4.40	NET Servicos de Comunicao S.A.

50 Most active Prefixes for the past 7 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	202.92.235.0/24	11224	0.72%	9498 -- BBIL-AP BHARTI Airtel Ltd.
2	91.217.214.0/24	9569	0.61%	3320 -- DTAG Deutsche Telekom AG
3	130.36.35.0/24	7783	0.50%	32528 -- ABBOTT Abbot Labs
4	130.36.34.0/24	7781	0.50%	32528 -- ABBOTT Abbot Labs
5	200.23.202.0/24	7635	0.49%	3454 -- Universidad Autonoma de Nuevo Leon
6	208.54.82.0/24	6800	0.44%	701 -- UUNET - MCI Communications Services, Inc. d/b/a Verizon Business
7	65.122.196.0/24	6418	0.41%	19743 --
8	72.164.144.0/24	5504	0.35%	19743 --
9	66.238.91.0/24	5497	0.35%	19743 --
10	66.89.98.0/24	5495	0.35%	19743 --
11	65.163.182.0/24	5495	0.35%	19743 --
12	65.162.204.0/24	5494	0.35%	19743 --
13	202.153.174.0/24	3413	0.22%	17408 -- ABOVE-AS-AP AboveNet Communications Taiwan
14	205.91.160.0/20	2984	0.19%	5976 -- DNIC-ASBLK-05800-06055 - DoD Network Information Center
15	65.181.192.0/23	2041	0.13%	11492 -- CABLEONE - CABLE ONE, INC.
16	77.74.144.0/24	1536	0.10%	21429 -- SICOB Sicob S.r.l. Autonomous System 5396 -- MC-LINK MC-link Spa
17	192.80.43.0/24	1508	0.10%	1706 -- UNIV-ARIZ - University of Arizona
18	24.116.2.0/24	1497	0.10%	11492 -- CABLEONE - CABLE ONE, INC.
19	24.116.1.0/24	1388	0.09%	11492 -- CABLEONE - CABLE ONE, INC.
20	190.15.21.128/26	1360	0.09%	27817 -- Red Nacional Académica de Tecnología Avanzada - RENATA
21	72.10.56.0/21	1316	0.08%	31815 -- MEDIATEMPLE - Media Temple, Inc.
22	72.10.32.0/20	1308	0.08%	31815 -- MEDIATEMPLE - Media Temple, Inc.
23	1.231.14.0/24	1299	0.08%	38388 -- BEN-AS-KR Bukbu District Office of Education in Seoul
24	14.102.50.0/24	1268	0.08%	18002 -- WORLDPHONE-IN AS Number for Interdomain Routing
25	77.81.5.0/24	1118	0.07%	51722 -- EAD-TELECOM-AS EAD TELECOM SRL
26	203.3.121.0/24	1024	0.07%	7545 -- TPG-INTERNET-AP TPG Internet Pty Ltd
27	92.246.206.0/24	996	0.06%	48612 -- RTC-ORENBURG-AS RTC-Orenburg, affiliate of RTC CJSC.
28	204.245.102.0/24	933	0.06%	19262 -- VZGNI-TRANSIT - Verizon Online LLC

Aggregation Potential (source: bgp.potaroo.net/as2.0/)



Aggregation Summary

- Aggregation on the Internet could be MUCH better
  - 35% 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 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 the five RIR databases to see if this address space really has been assigned to the customer

  The tool: whois

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

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)

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 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 Tutorial
- 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

- 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:

  datatracker.ietf.org/doc/draft-vegoda-no-more-unallocated-slash8s

- 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 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

Configuration Tips

Of passwords, tricks and templates

iBGP and IGPs Reminder!

- Make sure loopback is configured on router
  - iBGP between loopbacks, NOT real interfaces
- Make sure IGP carries loopback /32 address
- Consider the DMZ nets:
  - Use unnumbered interfaces?
  - Use next-hop-self on iBGP neighbours
  - Or carry the DMZ /30s 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 ISPs on edge routers
  - Preferable to carrying DMZ /30 addresses in the IGP
  - Reduces size of IGP to just core infrastructure
  - Alternative to using unnumbered interfaces
  - Helps scale network
  - Many ISPs 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 ASes in a typical AS_PATH in the Internet today
  - The Internet is around 5 ASes deep on average
  - Largest AS_PATH is usually 16-20 ASNs

Limiting AS Path Length

- Some announcements have ridiculous lengths of AS-paths:

```
*> 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 is an error in one IPv6 implementation

```
*>  96.27.246.0/24       2497 1239 12026 12026 12026
12026 12026 12026 12026 12026 12026 12026 12026
12026 12026 12026 12026 12026 12026 12026 12026
12026 12026 12026 i
```

  This example shows 21 prepends (for no obvious reason)

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

BGP TTL “hack”

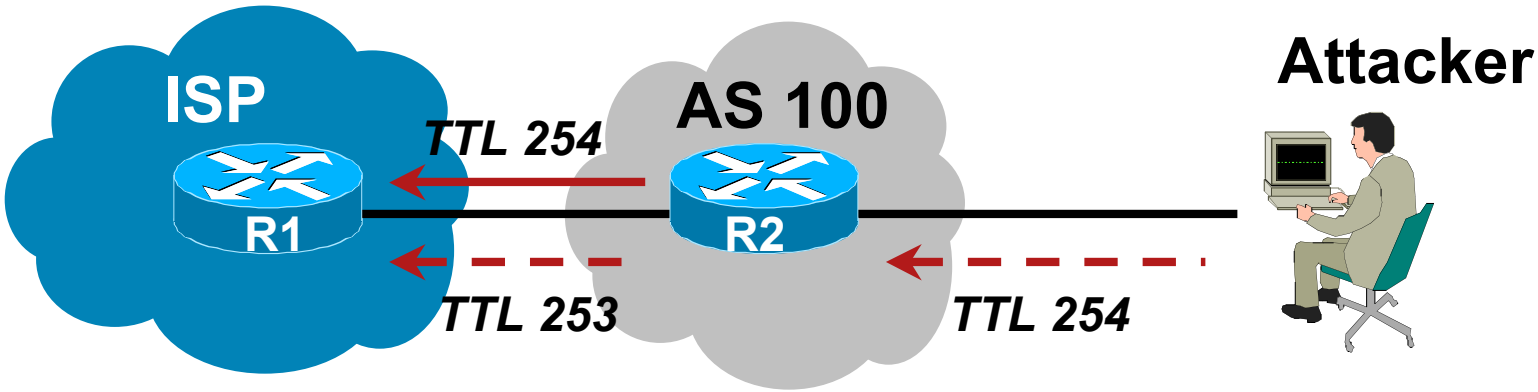
- Implement RFC5082 on BGP peerings

  (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 www.nanog.org/mtg-0302/hack.html for more details

Templates

- Good practice to configure templates for everything

  Vendor defaults tend not to be optimal or even very useful for ISPs

  ISPs create their own defaults by using configuration templates

- eBGP and iBGP examples follow

  Also see Team Cymru's BGP templates

  http://www.team-cymru.org/ReadingRoom/Documents/

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 accidents will happen
- Hardwire BGP to version 4
  - Yes, this is being paranoid!

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
- Remove private ASes 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
- Use password agreed between you and peer on eBGP session

eBGP Template Example continued

- 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!!

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
- It’s all about scaling – if your network won’t scale, then it won’t be successful

Deploying 32-bit ASNs

How to support customers using the extended ASN range

32-bit ASNs

- Standards documents
  - Description of 32-bit ASNs
    - www.rfc-editor.org/rfc/rfc4893.txt
  - Textual representation
    - www.rfc-editor.org/rfc/rfc5396.txt
  - New extended community
    - www.rfc-editor.org/rfc/rfc5668.txt
- AS 23456 is reserved as interface between 16-bit and 32-bit ASN world

32-bit ASNs – terminology

- 16-bit ASNs

  Refers to the range 0 to 65535

- 32-bit ASNs

  Refers to the range 65536 to 4294967295

  (or the extended range)

- 32-bit ASN pool

  Refers to the range 0 to 4294967295

Getting a 32-bit ASN

- Sample RIR policy

  www.apnic.net/docs/policy/asn-policy.html

- From 1st January 2007

  32-bit ASNs were available on request

- From 1st January 2009

  32-bit ASNs were assigned by default

  16-bit ASNs were only available on request

- From 1st January 2010

  No distinction – ASNs assigned from the 32-bit pool

Representation

- Representation of 0-4294967295 ASN range
  - Most operators favour traditional format (asplain)
  - A few prefer dot notation (X.Y):
    - asdot for 65536-4294967295, e.g 2.4
    - asdot+ for 0-4294967295, e.g 0.64513
  - **But regular expressions will have to be completely rewritten for asdot and asdot+ !!!**
- For example:
  - ^[0-9]+$ matches any ASN (16-bit and asplain)
  - This and equivalents extensively used in BGP multihoming configurations for traffic engineering
- Equivalent regexp for asdot is: ^([0-9]+)|([0-9]+\.[0-9]+)$
- Equivalent regexp for asdot+ is: ^[0-9]+\.[0-9]+$

Changes

- 32-bit ASNs are backward compatible with 16-bit ASNs
- **There is no flag day**
- You do NOT need to:
  - Throw out your old routers
  - Replace your 16-bit ASN with a 32-bit ASN
- You do need to be aware that:
  - Your customers will come with 32-bit ASNs
  - ASN 23456 is not a bogon!
  - You will need a router supporting 32-bit ASNs to use a 32-bit ASN locally
- If you have a proper BGP implementation, 32-bit ASNs will be transported silently across your network

How does it work?

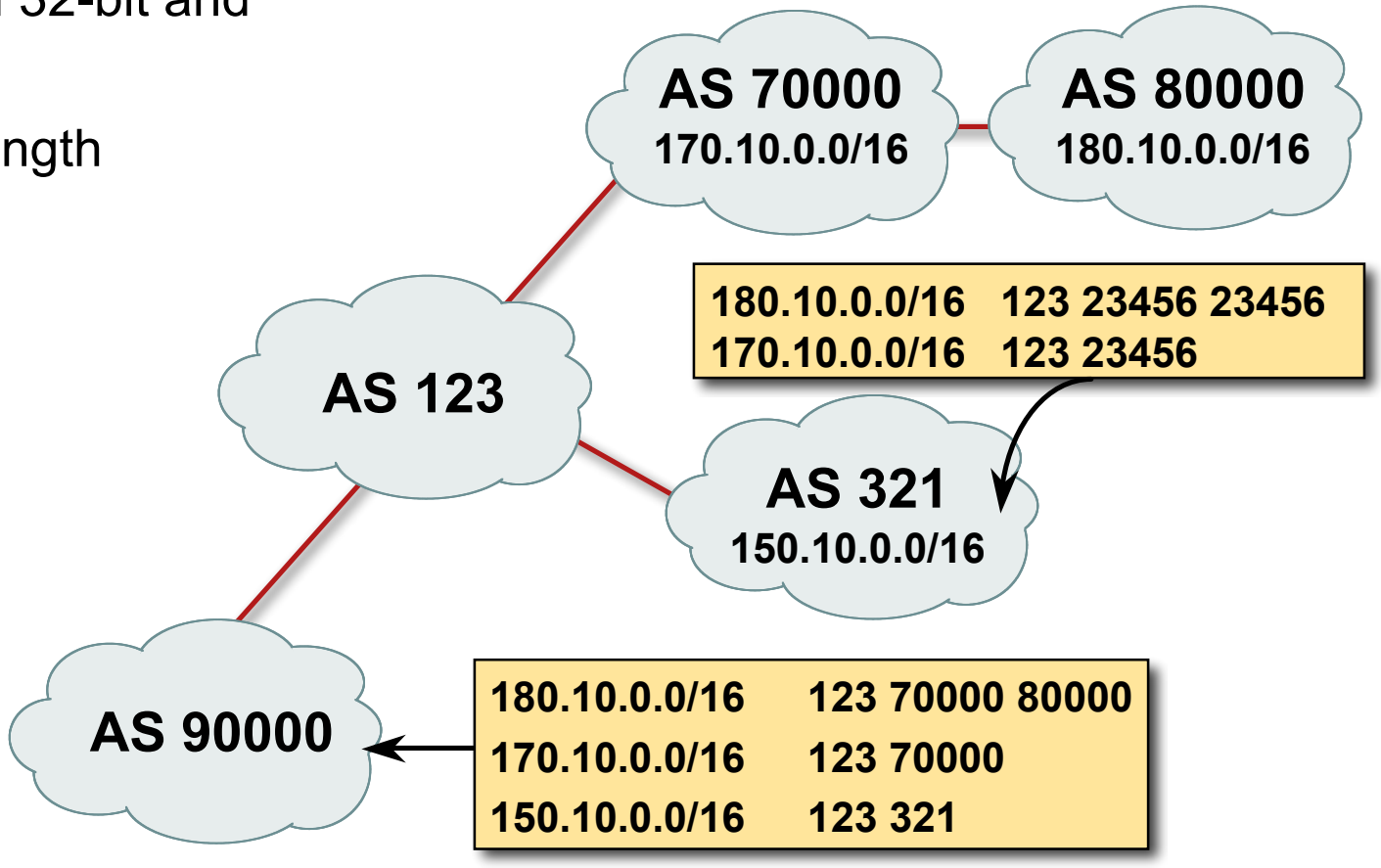
- If local router and remote router supports configuration of 32-bit ASNs
  - BGP peering is configured as normal using the 32-bit ASN
- If local router and remote router does not support configuration of 32-bit ASNs
  - BGP peering can only use a 16-bit ASN
- If local router only supports 16-bit ASN and remote router/network has a 32-bit ASN
  - Compatibility mode is initiated…

Compatibility Mode:

- Local router only supports 16-bit ASN and remote router uses 32-bit ASN
- BGP peering initiated:
  - Remote asks local if 32-bit supported (BGP capability negotiation)
  - When local says “no”, remote then presents AS23456
  - Local needs to be configured to peer with remote using AS23456
- BGP peering initiated (cont):
  - BGP session established using AS23456
  - 32-bit ASN included in a new BGP attribute called AS4_PATH
    - (as opposed to AS_PATH for 16-bit ASNs)
- Result:
  - 16-bit ASN world sees 16-bit ASNs and 23456 standing in for 32-bit ASNs
  - 32-bit ASN world sees 16 and 32-bit ASNs

Example:

- Internet with 32-bit and 16-bit ASNs
- AS-PATH length maintained



What has changed?

- Two new BGP attributes:
  - AS4_PATH
    - Carries 32-bit ASN path info
  - AS4_AGGREGATOR
    - Carries 32-bit ASN aggregator info
  - Well-behaved BGP implementations will simply pass these along if they don't understand them
- AS23456 (AS_TRANS)

What do they look like?

- IPv4 prefix originated by AS196613

asplain format

```
as4-7200#sh ip bgp 145.125.0.0/20
BGP routing table entry for 145.125.0.0/20, version 58734
Paths: (1 available, best #1, table default)
  131072 12654 196613
    204.69.200.25 from 204.69.200.25 (204.69.200.25)
      Origin IGP, localpref 100, valid, internal, best
```

- IPv4 prefix originated by AS3.5

asdot format

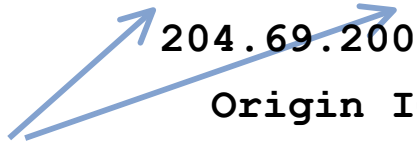
```
as4-7200#sh ip bgp 145.125.0.0/20
BGP routing table entry for 145.125.0.0/20, version 58734
Paths: (1 available, best #1, table default)
  2.0 12654 3.5
    204.69.200.25 from 204.69.200.25 (204.69.200.25)
      Origin IGP, localpref 100, valid, internal, best
```

What do they look like?

- IPv4 prefix originated by AS196613

  But 16-bit AS world view:

```
BGP-view1>sh ip bgp 145.125.0.0/20
BGP routing table entry for 145.125.0.0/20, version 113382
Paths: (1 available, best #1, table Default-IP-Routing-
Table)
23456 12654 23456
    204.69.200.25 from 204.69.200.25 (204.69.200.25)
      Origin IGP, localpref 100, valid, external, best
```



Transition AS

If 32-bit ASN not supported:

- Inability to distinguish between peer ASes using 32-bit ASNs
  - They will all be represented by AS23456
  - Could be problematic for transit provider's policy
- Inability to distinguish prefix's origin AS
  - How to tell whether origin is real or fake?
  - The real and fake both represented by AS23456
  - (There should be a better solution here!)
- Incorrect NetFlow summaries:
  - Prefixes from 32-bit ASNs will all be summarised under AS23456
  - Traffic statistics need to be measured per prefix and aggregated
  - Makes it hard to determine peerability of a neighbouring network

iBGP Deployment (1)

- Typical ISP design is thus:

  ISIS/OSPF for IGP, carrying loopback and point to point link addresses

  iBGP mesh (full/RR/Confederation) to carry customer and Internet prefixes

- All routers support 4-byte ASNs:

  Proceed with iBGP design as normal

- Not all routers support 4-byte ASNs:

  Three viable options

iBGP Deployment (2)

1. Return 4-byte ASN to the RIR and request 2-byte ASN instead:
   - Works if RIR is willing to do so
   - Works as long as there are 2-byte ASNs remaining
2. Partial iBGP mesh:
   - Routers which support 4-byte ASNs run iBGP mesh
   - Routers which do not support 4-byte ASNs either:
     - Run in private ASN (as a pseudo-customer) **or**
     - Do not run BGP at all
3. Use a BGP Confederation
   - (see AR-E Workshop)

Implementations (May 2011)

- Cisco IOS-XR 3.4 onwards
- Cisco IOS-XE 2.3 onwards
- Cisco IOS 12.0(32)S12, 12.4(24)T, 12.2SRE, 12.2(33)SXI1 onwards
- Cisco NX-OS 4.0(1) onwards
- Quagga 0.99.10 (patches for 0.99.6)
- OpenBGPd 4.2 (patches for 3.9 & 4.0)
- Juniper JunOSe 4.1.0 & JunOS 9.1 onwards
- Redback SEOS
- Force10 FTOS7.7.1 onwards

http://as4.cluepon.net/index.php/Software_Support for a complete list

Cisco Routers Supporting 4-byte ASNs

- CRS

  IOS-XR 3.4 onwards

- GSR

  IOS-XR 3.4 onwards

  IOS 12.0(32)S12, 12.0(33)S and 12.0(32)SY8 onwards

- ASR1000

  IOS-XE 2.3 onwards

- Nexus Switches

  NX-OS 4.0(1) onwards

Cisco Routers Supporting 4-byte ASNs

- Catalyst 6500

  IOS 12.2(33)SXI1 onwards

- 7600

  IOS 12.2(33)SRE1 onwards

- 7200 series

  IOS 12.0(32)S12, 12.0(33)S, 12.2(33)SRE1, 12.4(24)T, 15.0 onwards

- 7301

  IOS 12.2(33)SRE1, 12.4(24)T, 15.0 onwards

Cisco Routers Supporting 4-byte ASNs

- 3900/2900/1900 series

  IOS 15.0 onwards

- 3800/2800/1800/800 series

  IOS 12.4(24)T and IOS 15.0 onwards

- 3745/3725

  IOS 12.4(24)T

- AS5350/5400

  IOS 12.4(24)T and IOS 15.0 onwards

Cisco Routers NOT supporting 4-byte ASNs

- Routers which will never support 4-byte ASNs include:
  - 1700 series
  - 2500 series
  - 2600 series
  - 3600 series
  - AS5300
  - 7304

Summary

- Deploying 32-bit ASNs is simple
  - Your network can talk to a network which is using a 32-bit ASN
  - You have options with iBGP if not all your routers support configuration of 32-bit ASN
- Vendor support should be much better
  - Recent software support only, meaning older hardware will be problematic

Summary

Summary

- Tutorial has examined BGP deployment techniques:
  - The role of IGPs and iBGP
  - Aggregation
  - Receiving Prefixes
  - Configuration Tips
  - Deploying 4-byte ASNs

BGP Techniques for Internet Service Providers

The End! ☺