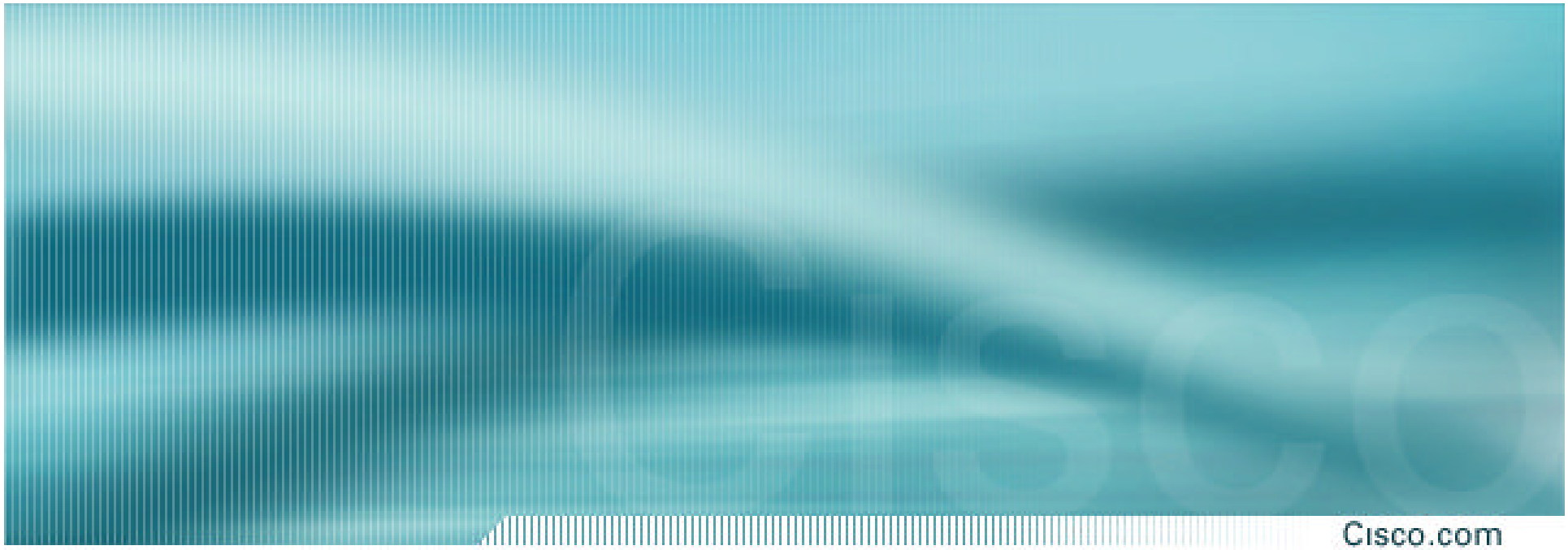


BGP Tutorial
Part 2 – Deployment Techniques

Philip Smith <pfs@cisco.com>

APRICOT 2003, Taipei

February 2003

Presentation Slides

- **Slides are available at**

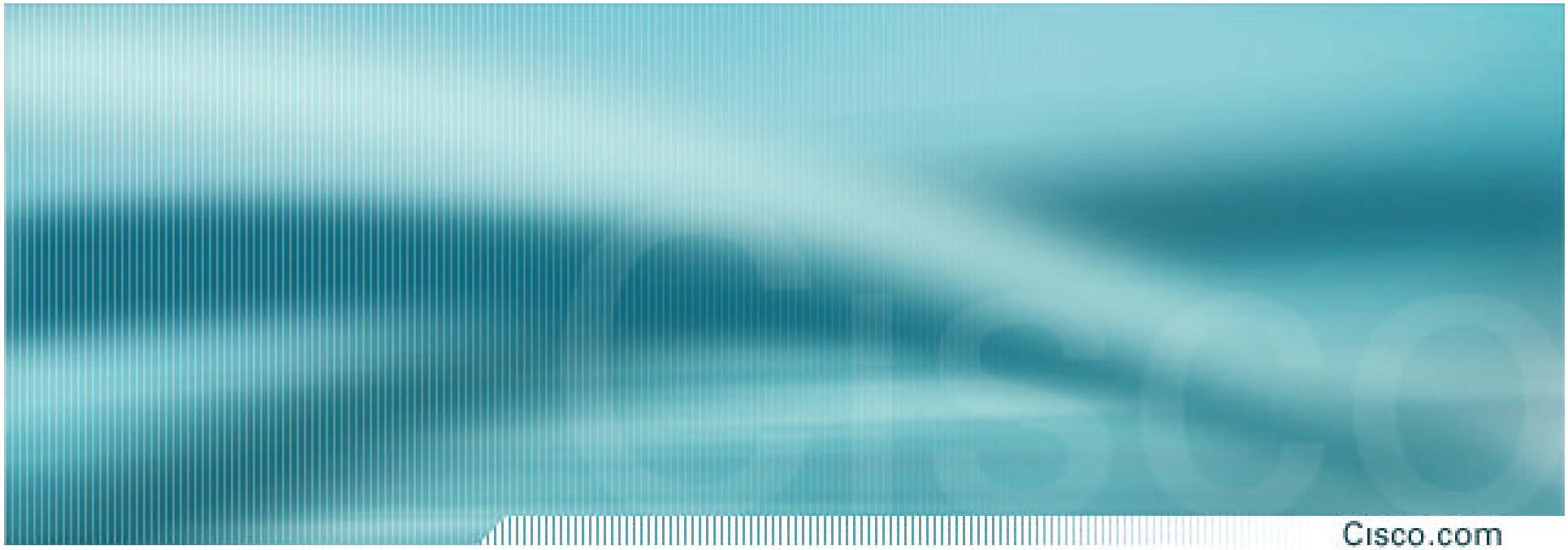
  ftp://ftp-eng.cisco.com/pfs/seminars/APRICOT02-BGP01.pdf

- **Feel free to ask questions any time**

BGP for Internet Service Providers

Cisco.com

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



BGP versus IGP

BGP versus OSPF/ISIS

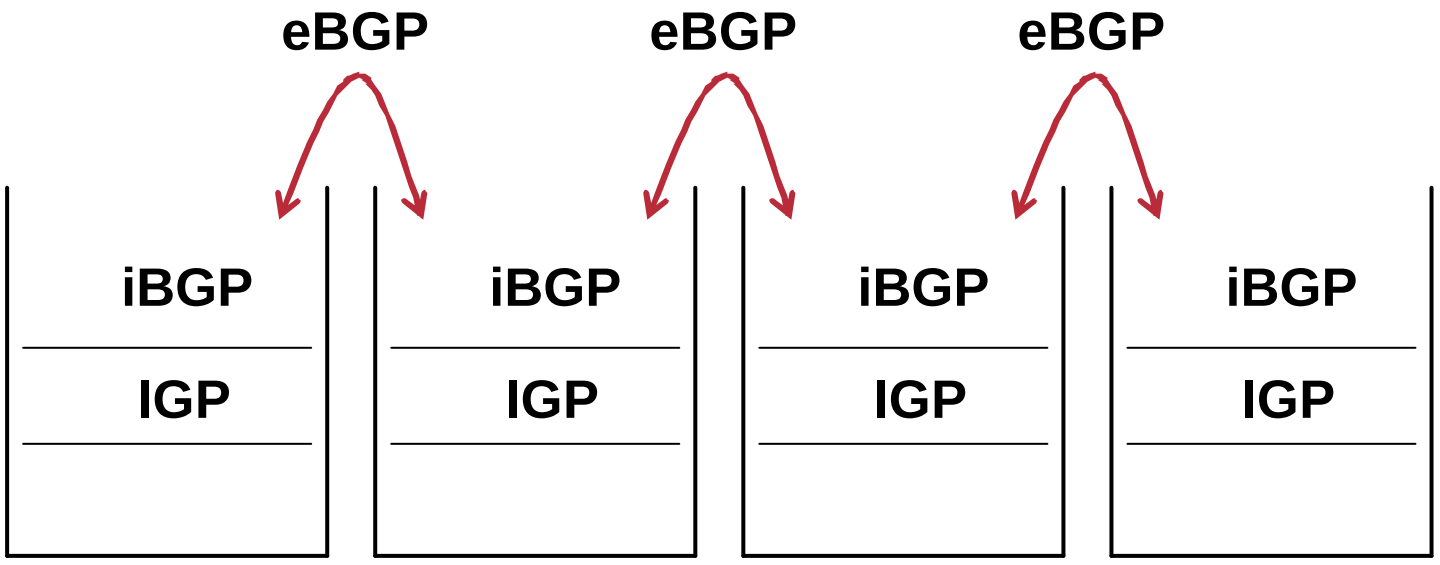
- **Internal Routing Protocols (IGPs)**
  - **examples are ISIS and OSPF**
  - **used for carrying infrastructure addresses**
  - **NOT used for carrying Internet prefixes or customer prefixes**
  - **design goal is to minimise number of prefixes in IGP to aid scalability and rapid convergence**

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 Configuration Example

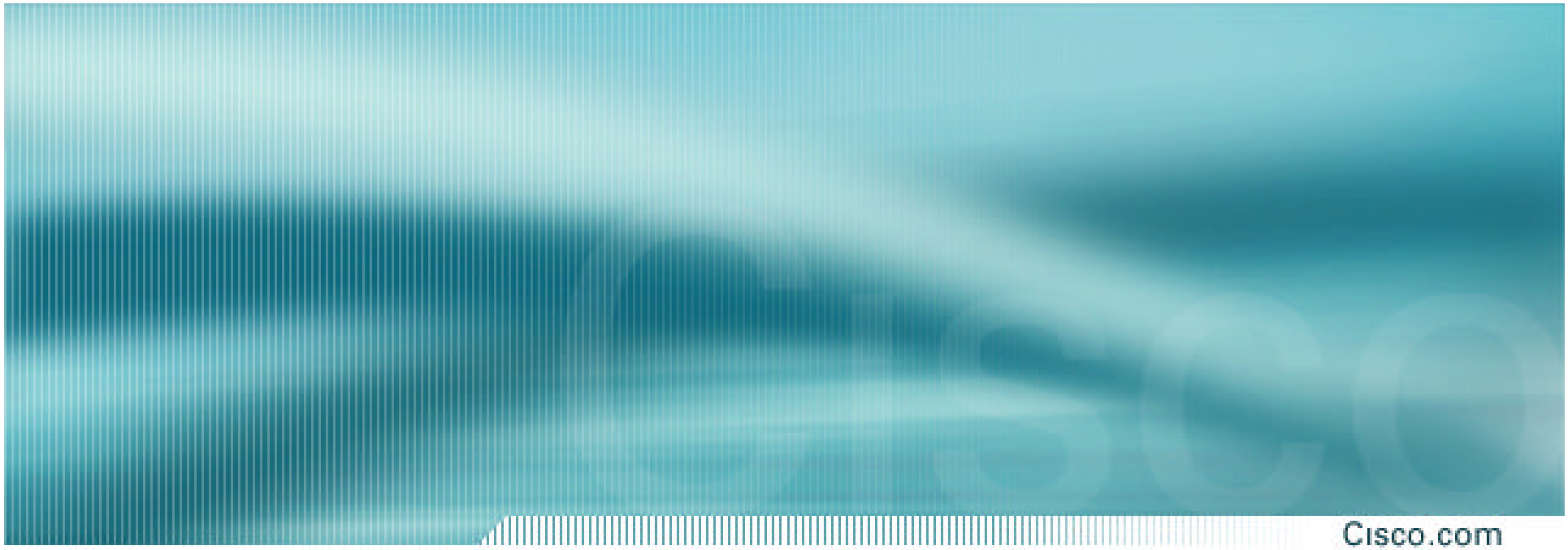

```
router bgp 34567
 neighbor core-ibgp peer-group
 neighbor core-ibgp remote-as 34567
 neighbor core-ibgp update-source Loopback0
 neighbor core-ibgp send-community
 neighbor core-ibgp-partial peer-group
 neighbor core-ibgp-partial remote-as 34567
 neighbor core-ibgp-partial update-source Loopback0
 neighbor core-ibgp-partial send-community
 neighbor core-ibgp-partial prefix-list network-ibgp out
 neighbor 222.1.9.10 peer-group core-ibgp
 neighbor 222.1.9.13 peer-group core-ibgp-partial
 neighbor 222.1.9.14 peer-group core-ibgp-partial
```

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**

BGP for Internet Service Providers

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



Prefixes into iBGP

Injecting prefixes into iBGP

- **Use iBGP to carry customer prefixes**

  don't ever use an 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 re-announce**

  use “ip route…permanent”

  Static route always exists, even if interface is down ® prefix announced in iBGP

- **many ISPs use redistribute static rather than network statement**

  only use this if you understand why

Inserting prefixes into BGP: redistribute static

- **Care required with redistribute!**

  redistribute <routing-protocol> means everything in the <routing-protocol> will be transferred into the current routing protocol

  Does not scale if uncontrolled

  Best avoided if at all possible

  redistribute normally used with “route-maps” and under tight administrative control

Router Configuration: redistribute static

- **Example:**

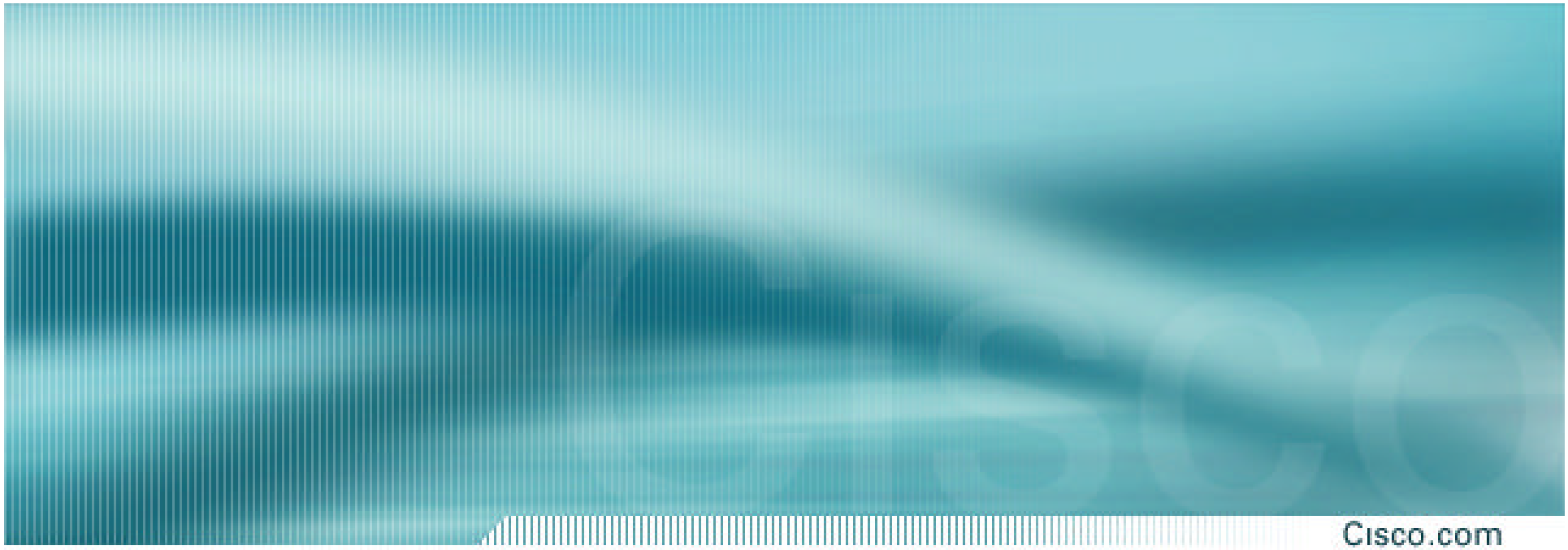
```
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 redistribute static route-map static-to-bgp
<snip>
!
route-map static-to-bgp permit 10
 match ip address prefix-list ISP-block
 set origin igp
<snip>
!
ip prefix-list ISP-block permit 215.34.10.0/22 le 30
!
```

Injecting prefixes into iBGP

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

BGP for Internet Service Providers

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



Aggregation

Quality or Quantity?

Aggregation

- **ISPs receive address block from Regional Registry or upstream provider**
- **Aggregation means announcing the address block only, not subprefixes**
  - **Subprefixes should only be announced in special cases – see later.**
- **Aggregate should be generated internally**
  - **Not on the network borders!**

Configuring Aggregation

- ISP has 221.10.0.0/19 address block
- To put into BGP as an aggregate:

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
ip route 221.10.0.0 255.255.224.0 null0
```

- The static route is a “pull up” route

  more specific prefixes within this address block ensure connectivity to ISP’s customers

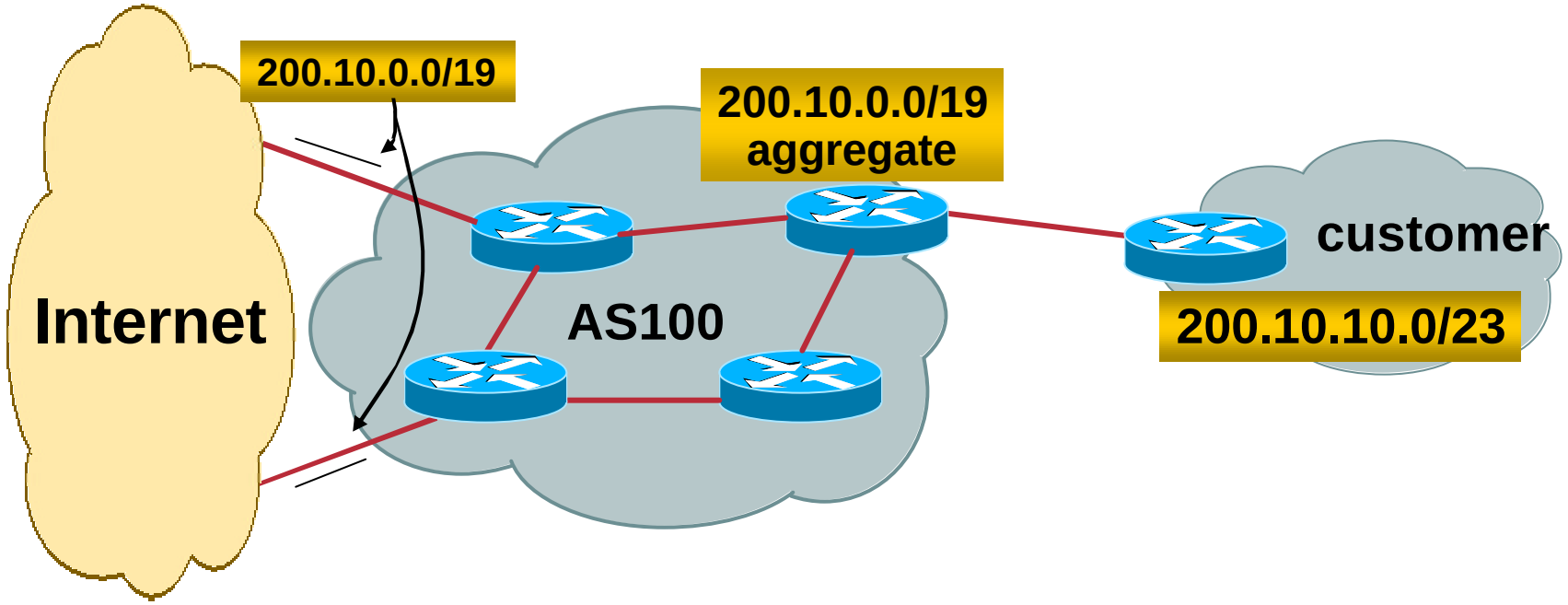
  “longest match lookup”

Announcing Aggregate – Cisco IOS

- **Configuration Example**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 101
 neighbor 222.222.10.1 prefix-list out-filter out
!
ip route 221.10.0.0 255.255.224.0 null0
!
ip prefix-list out-filter permit 221.10.0.0/19
```

Aggregation – Example



- **Customer has /23 network assigned from AS100's /19 address block**
- **AS100 announced /19 aggregate to the Internet**

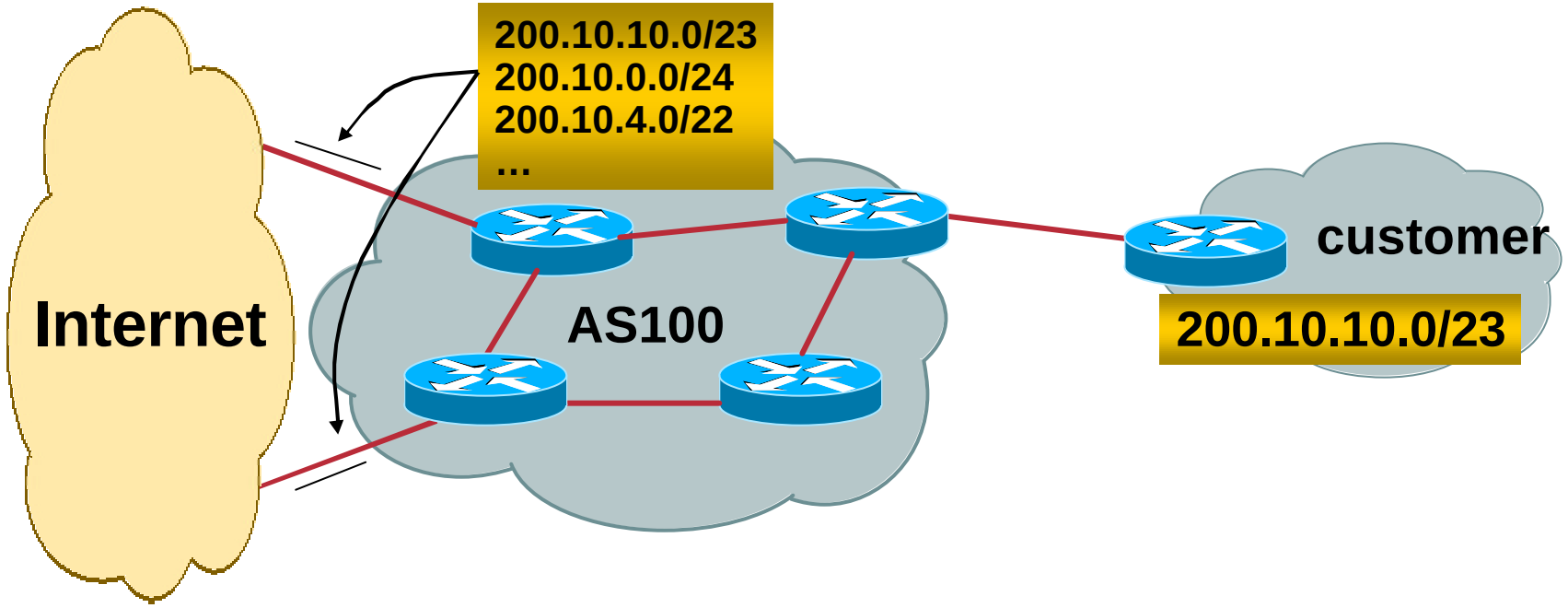
Aggregation – Good Example

- **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**

Aggregation – Good Example

- 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 – 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**

Aggregation – Bad Example

- **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 – 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!**
  - **Laziness? Lack of knowledge?**

Announcing an Aggregate

- **ISPs who don't and won't aggregate are held in poor regard by community**
- **Registries' minimum allocation size is now a /20**

  no real reason to see subprefixes of allocated blocks in the Internet

  BUT there are currently >65000 /24s!

The Internet Today (February 2003)

- **Current Internet Routing Table Statistics**

BGP Routing Table Entries	121374
Prefixes after maximum aggregation	77228
Unique prefixes in Internet	57860
Prefixes smaller than registry alloc	56731
/24s announced	66665
only 5281 /24s are from 192.0.0.0/8	
ASes in use	14677

“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**

“The New Swamp” July 2000

- **192/3 space contributes 69000 networks – rest of Internet contributes 16000 networks**

Block	Networks	Block	Networks	Block	Networks	Block	Networks
192/8	6352	204/8	4694	216/8	4177	64/8	1439
193/8	2746	205/8	3210	217/8	0	65/8	0
194/8	2963	206/8	4206	218/8	0	66/8	0
195/8	1689	207/8	3943	219/8	0	67/8	0
196/8	525	208/8	4804	220/8	0	68/8	0
198/8	4481	209/8	4755	221/8	0	69/8	0
199/8	4084	210/8	1375	24/8	1122	80/8	0
200/8	2436	211/8	532	61/8	80	81/8	0
202/8	3712	212/8	1859	62/8	428		
203/8	5494	213/8	635	63/8	2198		

“The New Swamp” February 2003

- **192/3 space contributes 83000 networks – rest of Internet contributes 38000 networks**

Block	Networks	Block	Networks	Block	Networks	Block	Networks
192/8	6478	204/8	4269	216/8	5967	64/8	3512
193/8	3761	205/8	2839	217/8	1379	65/8	3172
194/8	3110	206/8	3858	218/8	584	66/8	4395
195/8	2057	207/8	3769	219/8	419	67/8	899
196/8	678	208/8	4274	220/8	245	68/8	1805
198/8	4653	209/8	4623	221/8	12	69/8	234
199/8	4187	210/8	2277	24/8	2238	80/8	984
200/8	4639	211/8	1548	61/8	1377	81/8	392
202/8	5789	212/8	2500	62/8	1327		
203/8	7162	213/8	1976	63/8	2955		

“The New Swamp” Summary

- 192/3 space shows creeping increase in bad aggregation
  - 192/8, 193/8, 200/8, 202/7 and 216/8 show major changes not consistent with fresh RIR allocations
- Rest of address space is showing similar increase too
  - New RIR blocks in former A space are showing deaggregation
  - Other nets in former A and B space are also being deaggregated
- Why??
  - Excuses usually are traffic engineering
  - Real reason tends to be lack of knowledge and laziness

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**

Efforts to improve aggregation
The CIDR Report

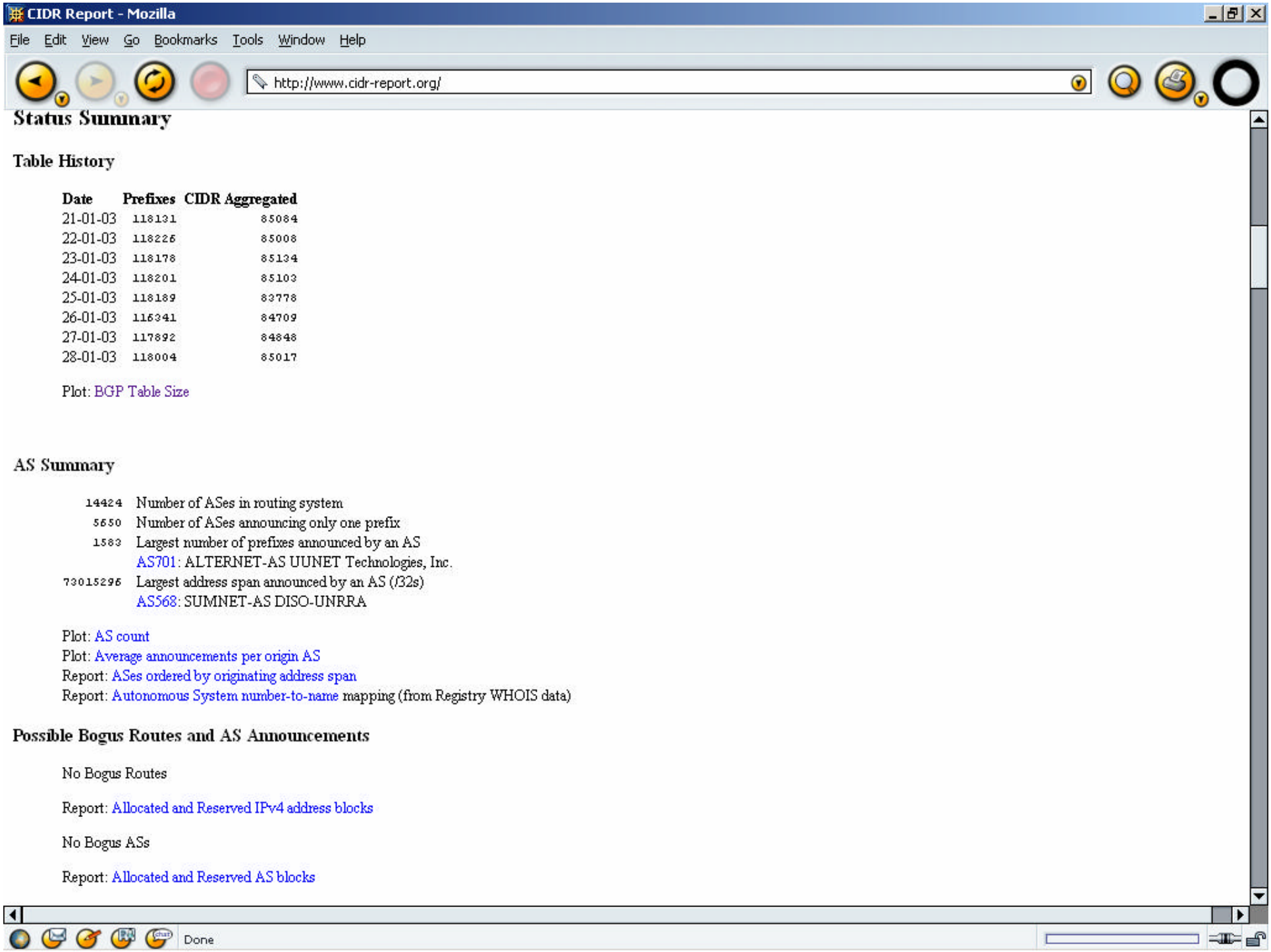
- Also computes the size of the routing table assuming ISPs performed optimal aggregation
- Website allows searches and computations of aggregation to be made on a per AS basis

  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
21-01-03	118131	85084
22-01-03	118226	85008
23-01-03	118178	85134
24-01-03	118201	85103
25-01-03	118189	83778
26-01-03	116341	84709
27-01-03	117892	84848
28-01-03	118004	85017

Plot: BGP Table Size

AS Summary

14424	Number of ASes in routing system
5650	Number of ASes announcing only one prefix
1583	Largest number of prefixes announced by an AS
	AS701: ALTERNET-AS UUNET Technologies, Inc.
73015296	Largest address span announced by an AS (/32s)
	AS568: SUMNET-AS DISO-UNRRA

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

Possible Bogus Routes and AS Announcements

No Bogus Routes

Report: Allocated and Reserved IPv4 address blocks

No Bogus ASs

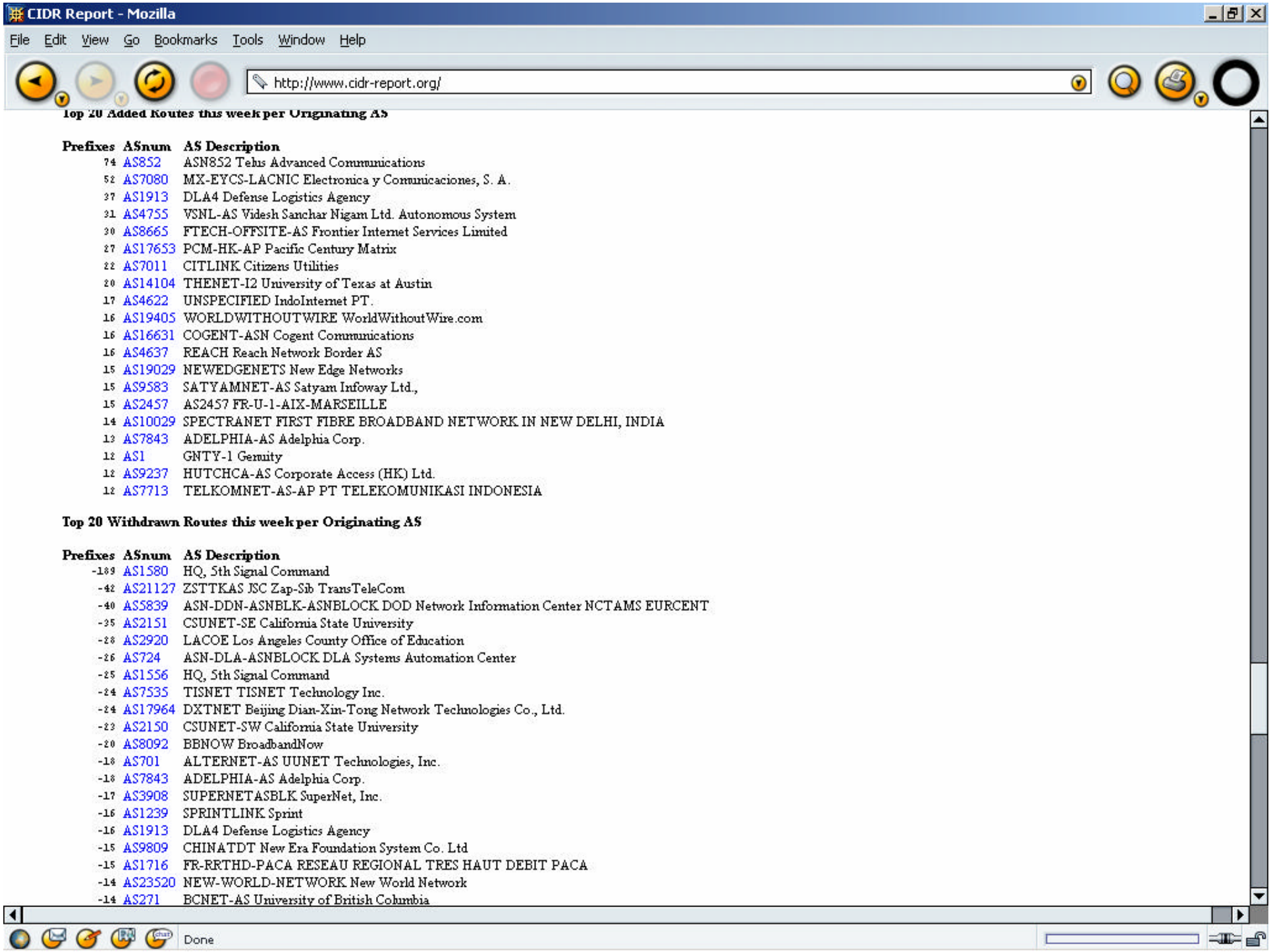
Report: Allocated and Reserved AS blocks

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').

--- **28Jan03** ---

ASnum	NetsNow	NetsAggr	NetGain	% Gain	Description
Table	118104	85000	33104	28.0%	All ASes
AS3908	1180	689	491	41.6%	SUPERNETASBLK SuperNet, Inc.
AS18566	445	5	440	98.9%	COVAD Covad Communications
AS701	1583	1172	411	26.0%	ALTERNET-AS UUNET Technologies, Inc.
AS7018	1435	1034	401	27.9%	ATT-INTERNET4 AT&T WorldNet Services
AS7843	591	250	341	57.7%	ADELPHIA-AS Adelphia Corp.
AS4323	527	188	339	64.3%	TW-COMM Time Warner Communications, Inc.
AS6197	464	154	310	66.8%	BATI-ATL BellSouth Network Solutions, Inc
AS1221	1128	833	295	26.2%	ASN-TELSTRA Telstra Pty Ltd
AS6347	372	85	287	77.2%	DIAMOND SAVVIS Communications Corporation
AS1239	957	674	283	29.6%	SPRINTLINK Sprint
AS4355	409	133	276	67.5%	ERMS-EARTHLNK EARTHLINK, INC
AS7046	552	280	272	49.3%	UUNET-CUSTOMER UUNET Technologies, Inc.
AS4151	329	58	271	82.4%	USDA-1 USDA
AS22927	289	22	267	92.4%	AR-TEAR2-LACNIC TELEFONICA DE ARGENTINA
AS4814	261	15	246	94.3%	CHINANET-BEIJING-AP China Telecom (Group)
AS705	424	181	243	57.3%	ASN-ALTERNET UUNET Technologies, Inc.
AS852	680	446	234	34.4%	ASN852 Telus Advanced Communications
AS1	663	433	230	34.7%	GNTY-1 Genuity
AS6198	423	202	221	52.2%	BATI-MIA BellSouth Network Solutions, Inc
AS17676	227	28	199	87.7%	GIGAINFRA XTAGE CORPORATION
AS22291	228	29	199	87.3%	CHARTER-LA Charter Communications
AS690	521	326	195	37.4%	MERIT-AS-27 Merit Network Inc.
AS209	522	334	188	36.0%	ASN-QWEST Qwest
AS4134	298	114	184	61.7%	ERX-CHINALINK Data Communications Bureau
AS6140	305	126	179	58.7%	IMPSAT-USA ImpSat
AS2048	259	86	173	66.8%	LANET-1 State of Louisiana
AS2386	421	249	172	40.9%	INS-AS AT&T Data Communications Services
AS6327	187	36	151	80.7%	SHAWFIBER Shaw Fiberlink Limited
AS17557	329	182	147	44.7%	PKTELECOM-AS-AP Pakistan Telecom
AS11492	299	157	142	47.5%	CABLEONE CABLE ONE
Total	16308	8521	7787	47.7%	Top 30 total

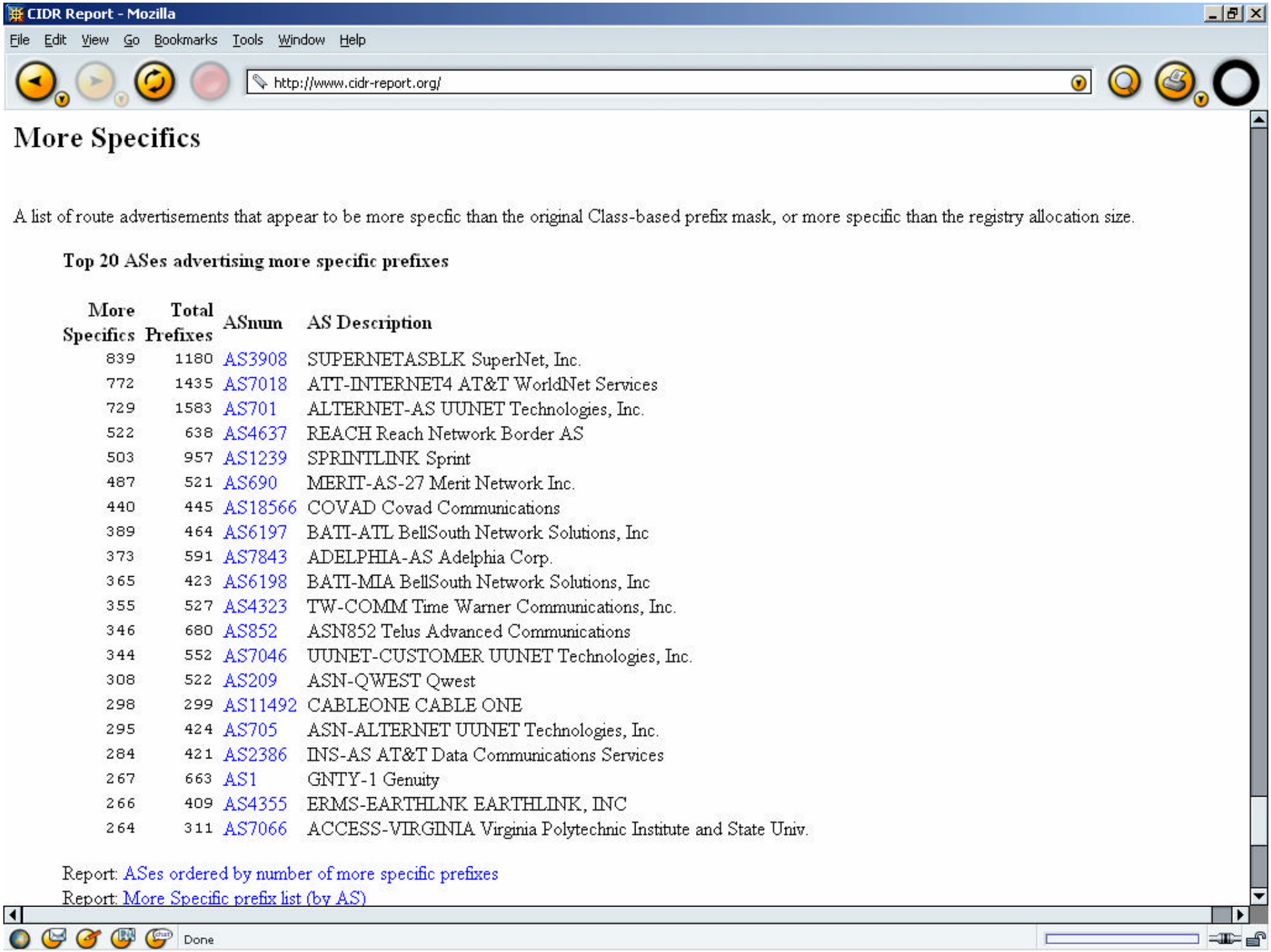


Top 20 Added Routes this week per Originating AS

Prefixes	ASnum	AS Description
74	AS852	ASN852 Telus Advanced Communications
52	AS7080	MX-EYCS-LACNIC Electronica y Comunicaciones, S. A.
37	AS1913	DLA4 Defense Logistics Agency
31	AS4755	VSNL-AS Videsh Sanchar Nigam Ltd. Autonomous System
30	AS8665	FTECH-OFFSITE-AS Frontier Internet Services Limited
27	AS17653	PCM-HK-AP Pacific Century Matrix
22	AS7011	CITLINK Citizens Utilities
20	AS14104	THENET-I2 University of Texas at Austin
17	AS4622	UNSPECIFIED IndoInternet PT.
16	AS19405	WORLDWITHOUTWIRE WorldWithoutWire.com
16	AS16631	COGENT-ASN Cogent Communications
16	AS4637	REACH Reach Network Border AS
15	AS19029	NEWEDGENETS New Edge Networks
15	AS9583	SATYAMNET-AS Satyam Infoway Ltd.,
15	AS2457	AS2457 FR-U-1-AIX-MARSEILLE
14	AS10029	SPECTRANET FIRST FIBRE BROADBAND NETWORK IN NEW DELHI, INDIA
13	AS7843	ADELPHIA-AS Adelphia Corp.
12	AS1	GNTY-1 Genuity
12	AS9237	HUTCHCA-AS Corporate Access (HK) Ltd.
12	AS7713	TELKOMNET-AS-AP PT TELEKOMUNIKASI INDONESIA

Top 20 Withdrawn Routes this week per Originating AS

Prefixes	ASnum	AS Description
-189	AS1580	HQ, 5th Signal Command
-42	AS21127	ZSTTKAS JSC Zap-Sib TransTeleCom
-40	AS5839	ASN-DDN-ASNBLK-ASNBLOCK DOD Network Information Center NCTAMS EURCENT
-35	AS2151	CSUNET-SE California State University
-28	AS2920	LACOE Los Angeles County Office of Education
-26	AS724	ASN-DLA-ASNBLOCK DLA Systems Automation Center
-25	AS1556	HQ, 5th Signal Command
-24	AS7535	TISNET TISNET Technology Inc.
-24	AS17964	DXTNET Beijing Dian-Xin-Tong Network Technologies Co., Ltd.
-23	AS2150	CSUNET-SW California State University
-20	AS8092	BBNOW BroadbandNow
-18	AS701	ALTERNET-AS UUNET Technologies, Inc.
-18	AS7843	ADELPHIA-AS Adelphia Corp.
-17	AS3908	SUPERNETASBLK SuperNet, Inc.
-16	AS1239	SPRINTLINK Sprint
-16	AS1913	DLA4 Defense Logistics Agency
-15	AS9809	CHINATDT New Era Foundation System Co. Ltd
-15	AS1716	FR-RRTHD-PACA RESEAU REGIONAL TRES HAUT DEBIT PACA
-14	AS23520	NEW-WORLD-NETWORK New World Network
-14	AS271	BCNET-AS University of British Columbia



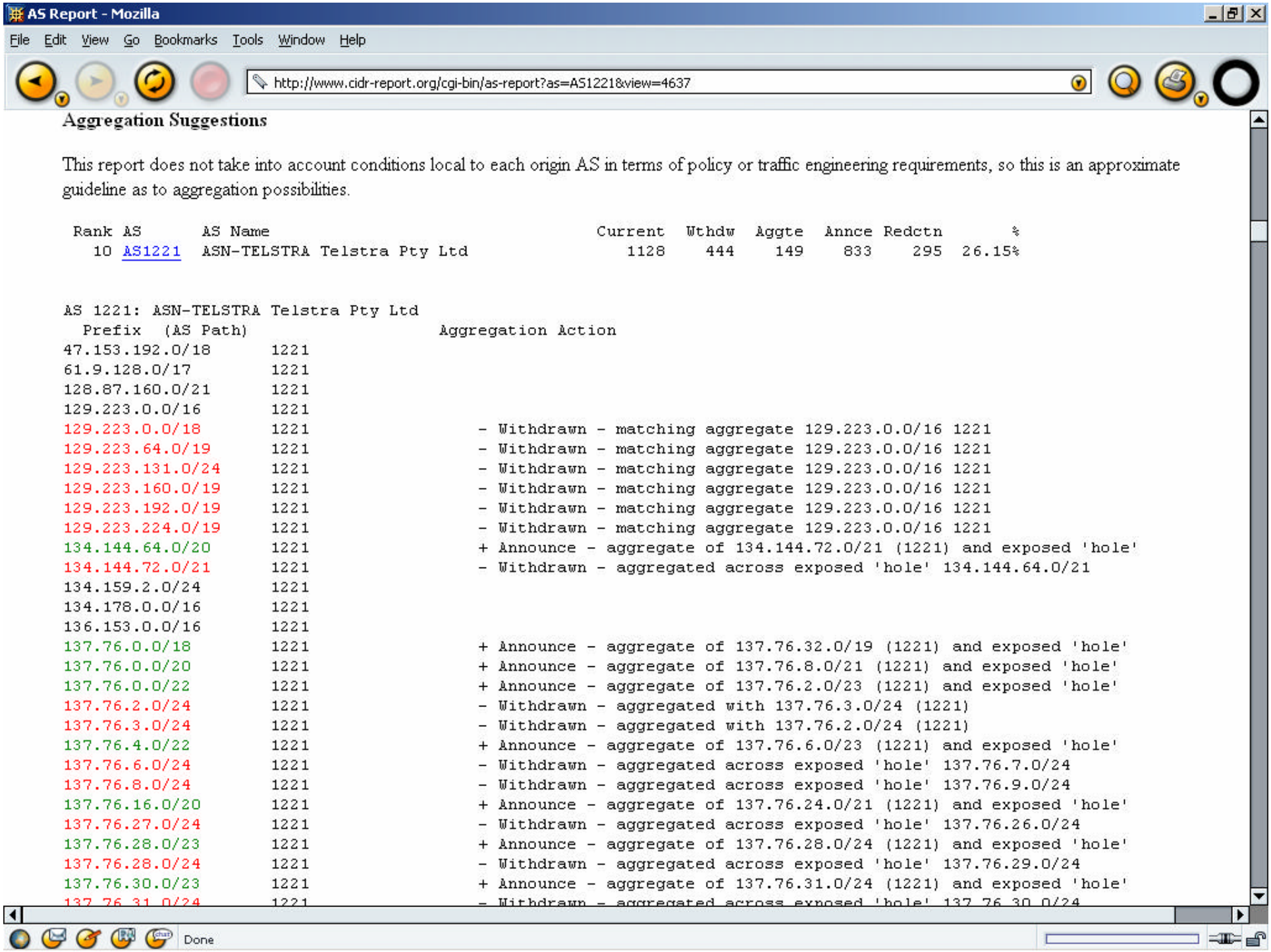
More Specifics

A list of route advertisements that appear to be more specific 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
839	1180	AS3908	SUPERNETASBLK SuperNet, Inc.
772	1435	AS7018	ATT-INTERNET4 AT&T WorldNet Services
729	1583	AS701	ALTERNET-AS UUNET Technologies, Inc.
522	638	AS4637	REACH Reach Network Border AS
503	957	AS1239	SPRINTLINK Sprint
487	521	AS690	MERIT-AS-27 Merit Network Inc.
440	445	AS18566	COVAD Covad Communications
389	464	AS6197	BATI-ATL BellSouth Network Solutions, Inc
373	591	AS7843	ADELPHIA-AS Adelphia Corp.
365	423	AS6198	BATI-MIA BellSouth Network Solutions, Inc
355	527	AS4323	TW-COMM Time Warner Communications, Inc.
346	680	AS852	ASN852 Telus Advanced Communications
344	552	AS7046	UUNET-CUSTOMER UUNET Technologies, Inc.
308	522	AS209	ASN-QWEST Qwest
298	299	AS11492	CABLEONE CABLE ONE
295	424	AS705	ASN-ALTERNET UUNET Technologies, Inc.
284	421	AS2386	INS-AS AT&T Data Communications Services
267	663	AS1	GNTY-1 Genuity
266	409	AS4355	ERMS-EARTHLNK EARTHLINK, INC
264	311	AS7066	ACCESS-VIRGINIA Virginia Polytechnic Institute and State Univ.

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



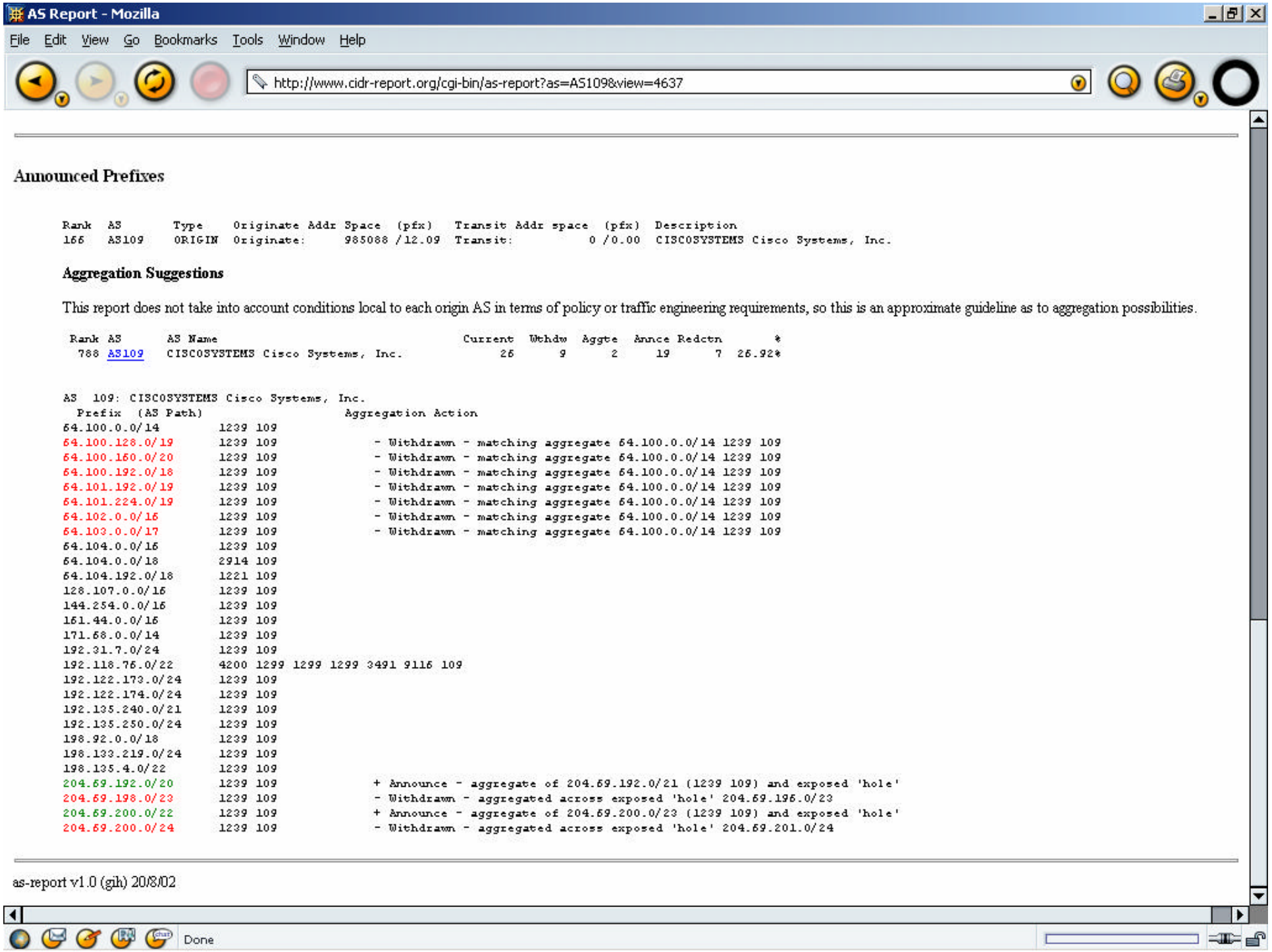
Aggregation Suggestions

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	%
10	AS1221	ASN-TELSTRA Telstra Pty Ltd	1128	444	149	833	295	26.15%

AS 1221: ASN-TELSTRA Telstra Pty Ltd

Prefix	(AS Path)	Aggregation Action
47.153.192.0/18	1221	
61.9.128.0/17	1221	
128.87.160.0/21	1221	
129.223.0.0/16	1221	
129.223.0.0/18	1221	- Withdrawn - matching aggregate 129.223.0.0/16 1221
129.223.64.0/19	1221	- Withdrawn - matching aggregate 129.223.0.0/16 1221
129.223.131.0/24	1221	- Withdrawn - matching aggregate 129.223.0.0/16 1221
129.223.160.0/19	1221	- Withdrawn - matching aggregate 129.223.0.0/16 1221
129.223.192.0/19	1221	- Withdrawn - matching aggregate 129.223.0.0/16 1221
129.223.224.0/19	1221	- Withdrawn - matching aggregate 129.223.0.0/16 1221
134.144.64.0/20	1221	+ Announce - aggregate of 134.144.72.0/21 (1221) and exposed 'hole'
134.144.72.0/21	1221	- Withdrawn - aggregated across exposed 'hole' 134.144.64.0/21
134.159.2.0/24	1221	
134.178.0.0/16	1221	
136.153.0.0/16	1221	
137.76.0.0/18	1221	+ Announce - aggregate of 137.76.32.0/19 (1221) and exposed 'hole'
137.76.0.0/20	1221	+ Announce - aggregate of 137.76.8.0/21 (1221) and exposed 'hole'
137.76.0.0/22	1221	+ Announce - aggregate of 137.76.2.0/23 (1221) and exposed 'hole'
137.76.2.0/24	1221	- Withdrawn - aggregated with 137.76.3.0/24 (1221)
137.76.3.0/24	1221	- Withdrawn - aggregated with 137.76.2.0/24 (1221)
137.76.4.0/22	1221	+ Announce - aggregate of 137.76.6.0/23 (1221) and exposed 'hole'
137.76.6.0/24	1221	- Withdrawn - aggregated across exposed 'hole' 137.76.7.0/24
137.76.8.0/24	1221	- Withdrawn - aggregated across exposed 'hole' 137.76.9.0/24
137.76.16.0/20	1221	+ Announce - aggregate of 137.76.24.0/21 (1221) and exposed 'hole'
137.76.27.0/24	1221	- Withdrawn - aggregated across exposed 'hole' 137.76.26.0/24
137.76.28.0/23	1221	+ Announce - aggregate of 137.76.28.0/24 (1221) and exposed 'hole'
137.76.28.0/24	1221	- Withdrawn - aggregated across exposed 'hole' 137.76.29.0/24
137.76.30.0/23	1221	+ Announce - aggregate of 137.76.31.0/24 (1221) and exposed 'hole'
137.76.31.0/24	1221	- Withdrawn - aggregated across exposed 'hole' 137.76.30.0/24



Announced Prefixes

Rank	AS	Type	Originate Addr Space	(pfx)	Transit Addr space	(pfx)	Description
166	AS109	ORIGIN	Originate:	985088 /12.09	Transit:	0 /0.00	CISCOSYSTEMS Cisco Systems, 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.

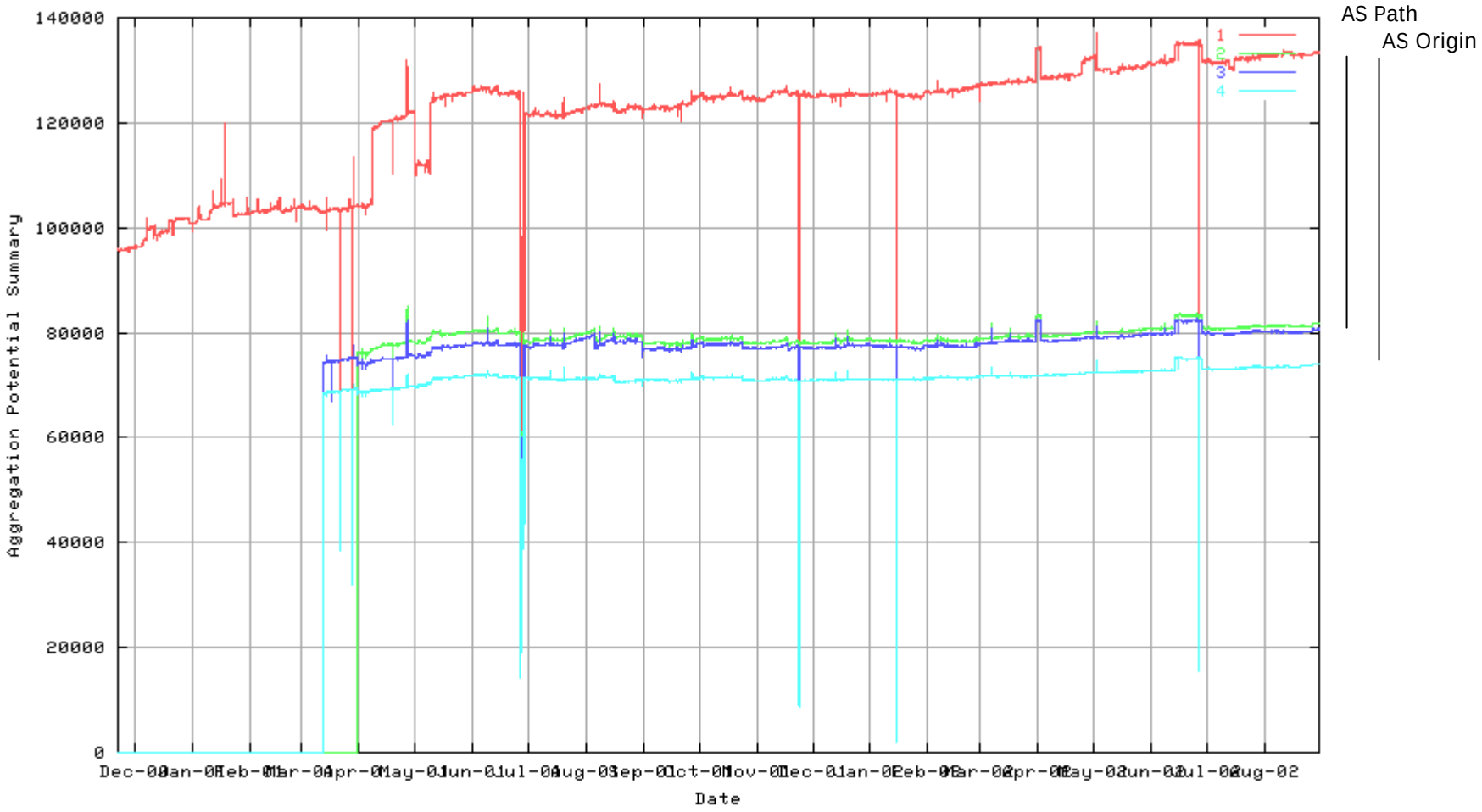
Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
788	AS109	CISCOSYSTEMS Cisco Systems, Inc.	26	9	2	19	7	26.92%

AS 109: CISCOSYSTEMS Cisco Systems, Inc.

Prefix	(AS Path)	Aggregation Action
64.100.0.0/14	1239 109	
64.100.128.0/19	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.100.160.0/20	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.100.192.0/18	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.101.192.0/19	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.101.224.0/19	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.102.0.0/16	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.103.0.0/17	1239 109	- Withdrawn - matching aggregate 64.100.0.0/14 1239 109
64.104.0.0/16	1239 109	
64.104.0.0/18	2914 109	
64.104.192.0/18	1221 109	
128.107.0.0/16	1239 109	
144.254.0.0/16	1239 109	
161.44.0.0/16	1239 109	
171.68.0.0/14	1239 109	
192.31.7.0/24	1239 109	
192.118.76.0/22	4200 1299 1299 1299 3491 9116 109	
192.122.173.0/24	1239 109	
192.122.174.0/24	1239 109	
192.135.240.0/21	1239 109	
192.135.250.0/24	1239 109	
198.92.0.0/18	1239 109	
198.133.219.0/24	1239 109	
198.135.4.0/22	1239 109	
204.69.192.0/20	1239 109	+ Announce - aggregate of 204.69.192.0/21 (1239 109) and exposed 'hole'
204.69.198.0/23	1239 109	- Withdrawn - aggregated across exposed 'hole' 204.69.196.0/23
204.69.200.0/22	1239 109	+ Announce - aggregate of 204.69.200.0/23 (1239 109) and exposed 'hole'
204.69.200.0/24	1239 109	- Withdrawn - aggregated across exposed 'hole' 204.69.201.0/24

as-report v1.0 (gih) 20/8/02

Aggregation Potential

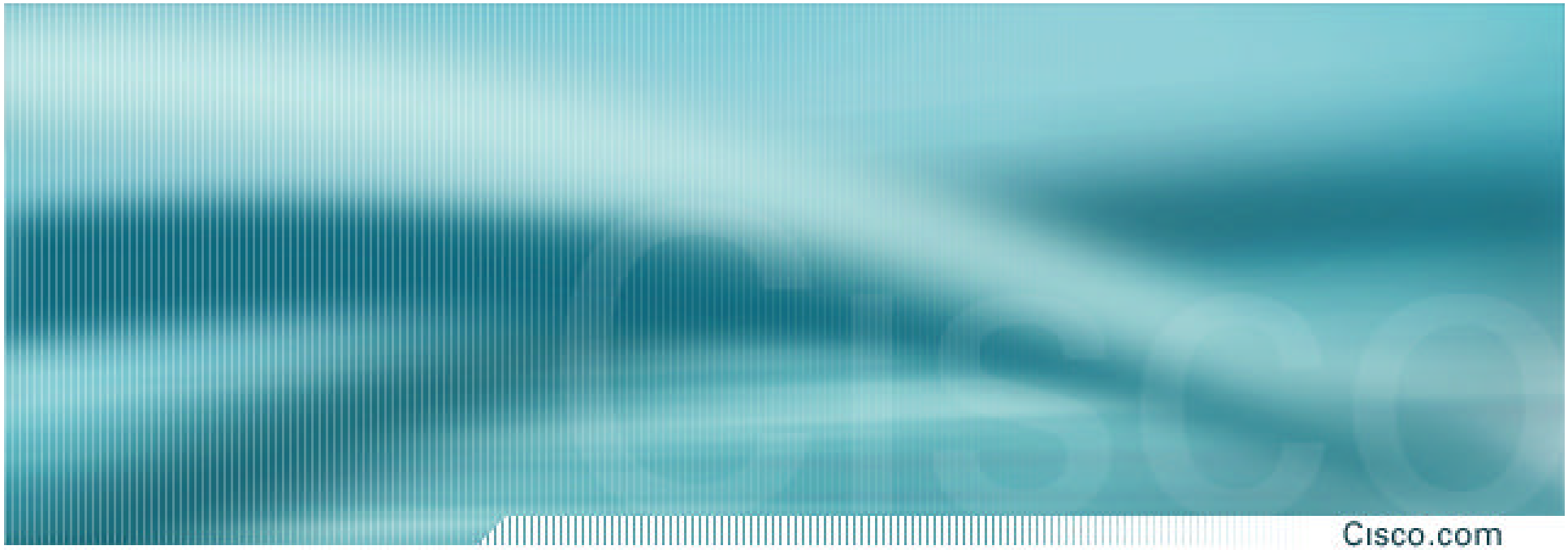


Aggregation Summary

- **Aggregation on the Internet could be MUCH better**
  - **35% saving on Internet routing table size is quite feasible**
  - **Tools are available**
    - **Commands on the router are not hard**
    - **CIDR-Report webpage**

BGP for Internet Service Providers

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



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 four 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:**

```
pfs-pc$ 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 220.50.0.0/20 block**
  - **should only announce this to upstreams**
  - **upstreams should only accept this from them**
- **Configuration on upstream**

```
router bgp 100
 neighbor 222.222.10.1 remote-as 101
 neighbor 222.222.10.1 prefix-list customer in
!
ip prefix-list customer permit 220.50.0.0/20
```

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.ripe.net/ripencc/pub-services/db/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 222.222.10.1 remote-as 101
 neighbor 222.222.10.1 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 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 101
 neighbor 221.5.7.1 prefix-list infilter in
 neighbor 221.5.7.1 prefix-list outfilter out
!
ip prefix-list infilter permit 0.0.0.0/0
!
ip prefix-list outfilter permit 221.10.0.0/19
```

Receiving Prefixes: From Upstream/Transit Provider

- **Upstream Router Configuration**

```
router bgp 101
 neighbor 221.5.7.2 remote-as 100
 neighbor 221.5.7.2 default-originate
 neighbor 221.5.7.2 prefix-list cust-in in
 neighbor 221.5.7.2 prefix-list cust-out out
!
ip prefix-list cust-in permit 221.10.0.0/19
!
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**
    - http://www.ietf.org/internet-drafts/draft-manning-dsua-08.txt
    - 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 Rob Thomas' list of "bogons"**
  - http://www.cymru.org/Documents/bogon-list.html

Receiving Prefixes

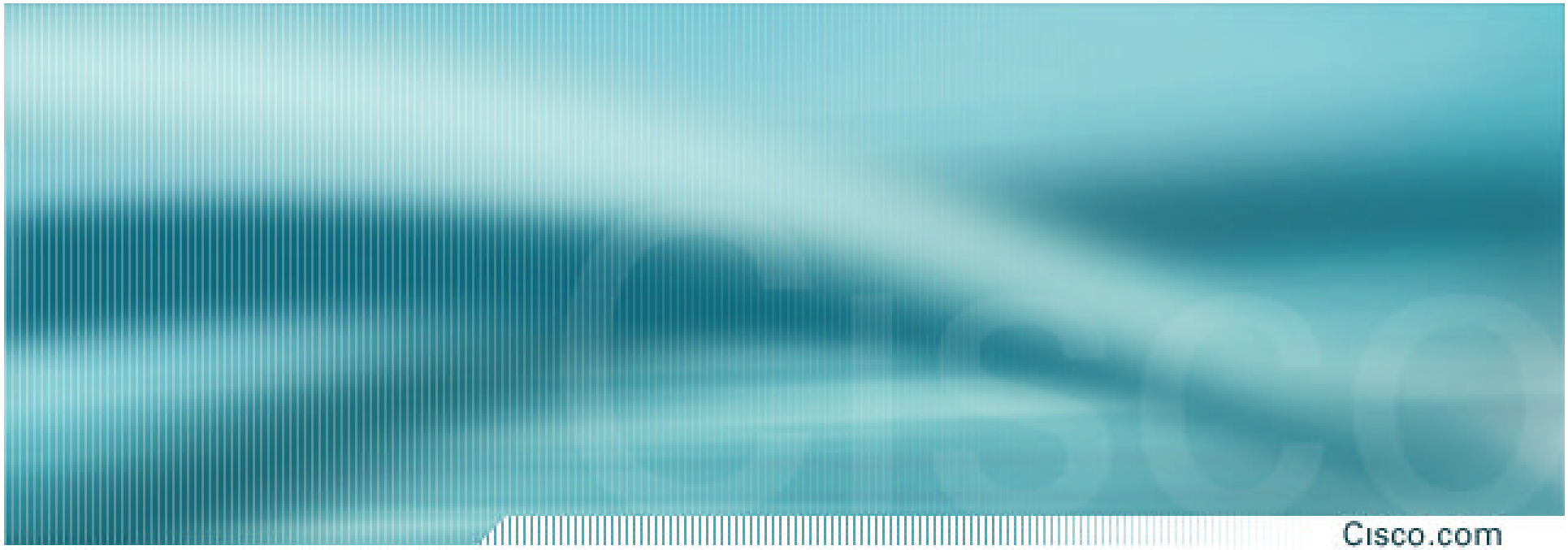

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 101
 neighbor 221.5.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0           ! Block default
ip prefix-list in-filter deny 0.0.0.0/8 le 32
ip prefix-list in-filter deny 10.0.0.0/8 le 32
ip prefix-list in-filter deny 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 221.10.0.0/19 le 32 ! Block local prefix
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**

BGP for Internet Service Providers

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



Configuration Tips

iBGP and IGPs

- **Make sure loopback is configured on router**
  - **iBGP between loopbacks, NOT real interfaces**
- **Make sure IGP carries loopback /32 address**
- **Make sure IGP carries DMZ nets**
  - **Use ip-unnumbered where possible**
  - **Or use next-hop-self on iBGP neighbours**
  - **neighbor x.x.x.x next-hop-self**

Next-hop-self

- **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 ip unnumbered**
  - **Helps scale network**
  - **BGP speaker announces external network using local address (loopback) as next-hop**

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**
  - **Sample iBGP and eBGP templates follow for Cisco IOS**

BGP Template – iBGP peers

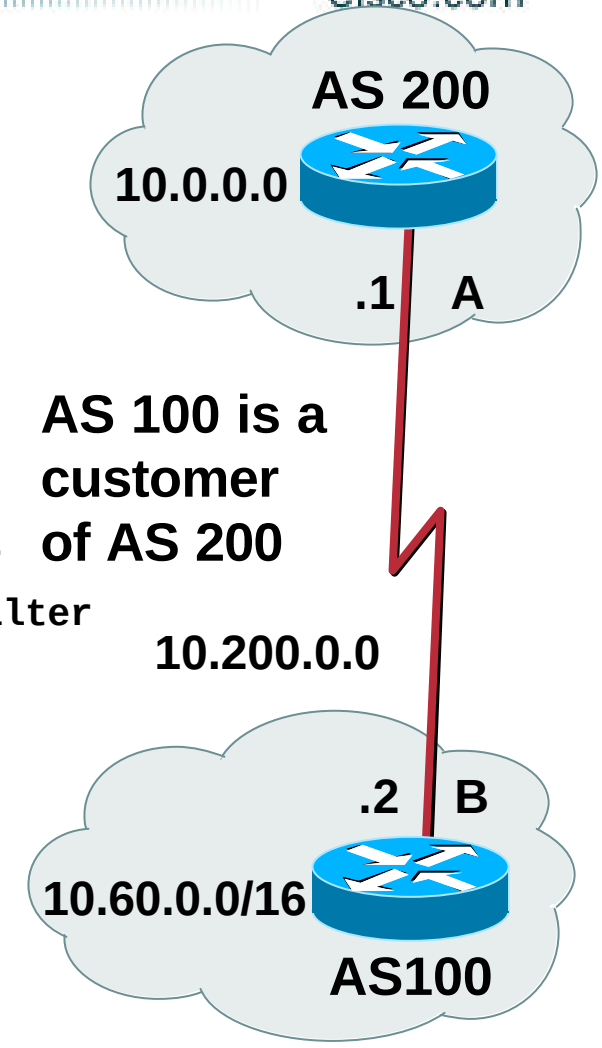

```
router bgp 100
neighbor internal peer-group
neighbor internal description ibgp peers
neighbor internal remote-as 100
neighbor internal update-source Loopback0
neighbor internal next-hop-self
neighbor internal send-community
neighbor internal version 4
neighbor internal password 7 03085A09
neighbor 1.0.0.1 peer-group internal
neighbor 1.0.0.2 peer-group internal
```

BGP Template – iBGP peers

- Use peer-groups
- iBGP between loopbacks!
- Next-hop-self
  - Keep DMZ and 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!
- Use passwords on iBGP session
  - Not being paranoid, VERY necessary

BGP Template – eBGP peers


```
Router B:
router bgp 100
bgp dampening route-map RIPE229-flap
network 10.60.0.0 mask 255.255.0.0
neighbor external peer-group
neighbor external remote-as 200
neighbor external description ISP connection
neighbor external remove-private-AS
neighbor external version 4
neighbor external prefix-list ispout out ! “real” filter
neighbor external filter-list 1 out      ! “accident” filter
neighbor external route-map ispout out
neighbor external prefix-list ispin in
neighbor external filter-list 2 in
neighbor external route-map ispin in
neighbor external password 7 020A0559
neighbor external maximum-prefix 130000 [warning-only]
neighbor 10.200.0.1 peer-group external
!
ip route 10.60.0.0 255.255.0.0 null0 254
```



BGP Template – eBGP peers

- BGP damping – use RIPE-229 parameters
- Remove private ASes from announcements
  - Common omission today
- Use extensive filters, with “backup”
  - Use as-path filters to backup prefix-lists
  - Use route-maps for policy
- Use password agreed between you and peer on eBGP session
- Use maximum-prefix tracking
  - Router will warn you if there are sudden increases in BGP table size, bringing down eBGP if desired

More BGP “defaults”

- **Log neighbour changes**

  bgp log-neighbor-changes

- **Enable deterministic MED**

  bgp deterministic-med

  Otherwise bestpath could be different every time BGP session is reset

- **Make BGP admin distance higher than any IGP**

  distance bgp 200 200 200

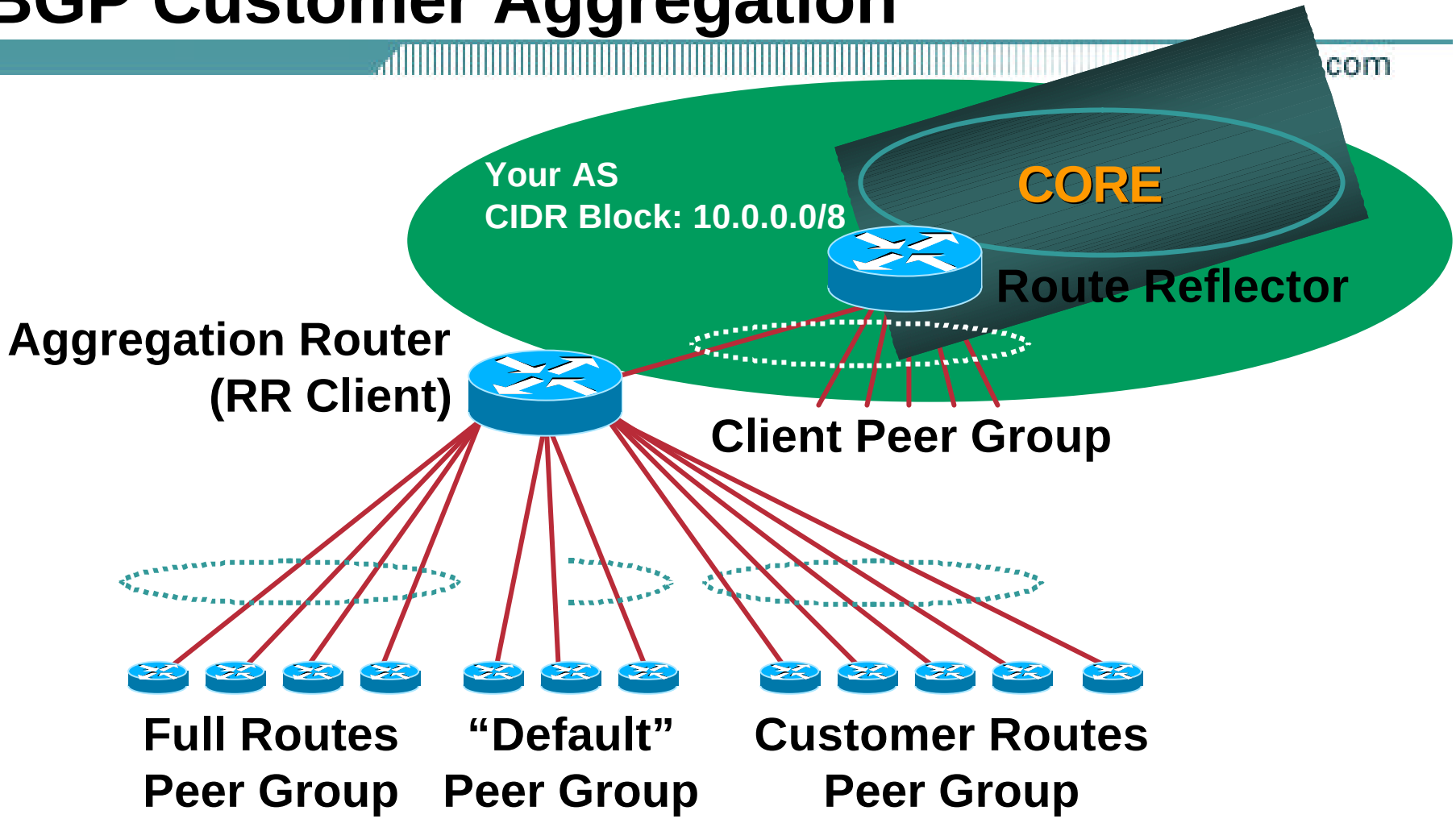
Customer Aggregation

- **BGP customers**
  - **Offer max 3 types of feeds (easier than custom configuration per peer)**
  - **Use communities**
- **Static customers**
  - **Use communities**
- **Differentiate between different types of prefixes**
  - **Makes eBGP filtering easy**

BGP Customer Aggregation Guidelines

- **Define at least three peer groups:**
  - **cust-default—send default route only**
  - **cust-cust—send customer routes only**
  - **cust-full —send full Internet routes**
- **Identify routes via communities e.g.**
  - **100:4100=customers; 100:4500=peers**
- **Apply passwords per neighbour**
- **Apply inbound & outbound prefix-list per neighbour**

BGP Customer Aggregation



Apply passwords and in/outbound prefix-list directly to each neighbour

Static Customer Aggregation Guidelines

- Identify routes via communities, e.g.
  - 100:4000 = my address blocks
  - 100:4100 = “specials” from my blocks
  - 100:4200 = customers from my blocks
  - 100:4300 = customers outside my blocks
  - Helps with aggregation, iBGP, filtering
- BGP network statements on aggregation routers set correct community

Sample core configuration

- **eBGP peers and upstreams**

  Send communities 100:4000, 100:4100 and 100:4300, receive everything

- **iBGP full routes**

  Send everything (only to network core)

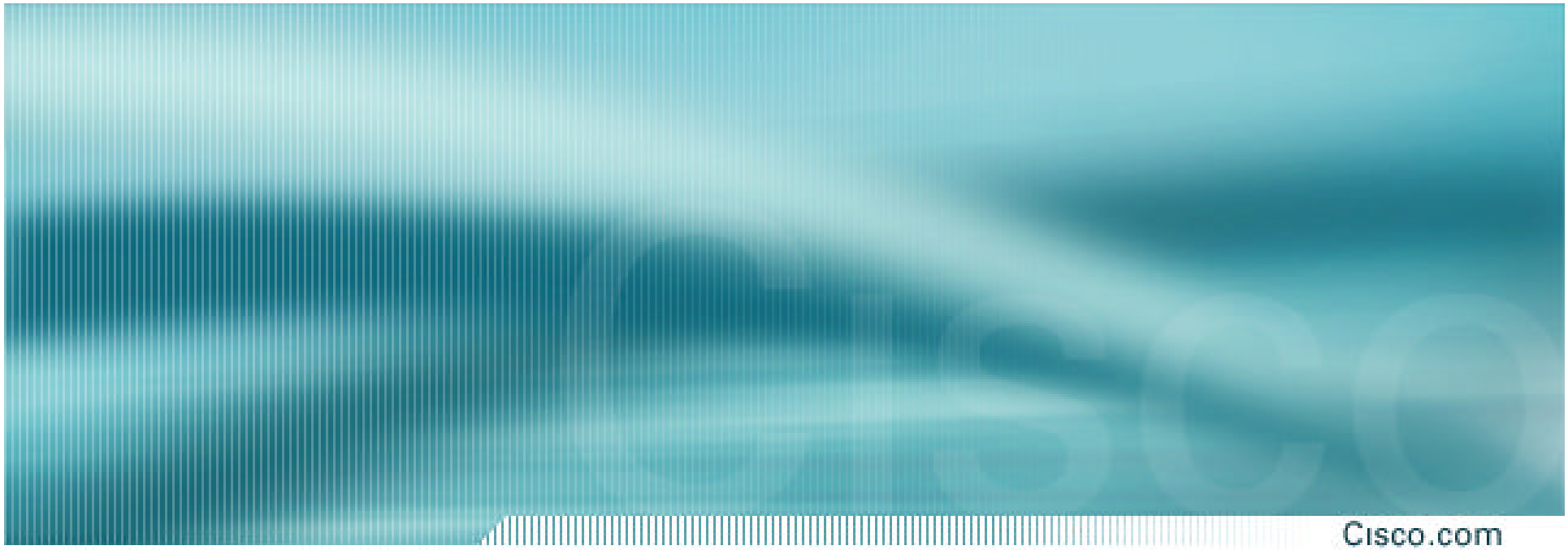
- **iBGP partial routes**

  Send communities 100:4000, 100:4100, 100:4200, 100:4300 and 100:4500 (to edge routers, peering routers, IXP routers)

- **Simple configuration with peer-groups and route-maps**

Summary

- Use configuration templates
- Standardise the configuration
- 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



More Configuration Tips

Hot off the Press!

Deterministic-MED

- MED is multi-exit discriminator
- Part of BGP path selection process
  - Prefix paths compared when heard from the same neighbouring AS
  - Applicable when multihomed to the same AS
  - RFC1771 does not specify how paths should be compared in this case
  - So IOS compares newest path with next newest *etc*
  - Which results in non-deterministic path selection when neighbour relationships are reset

Deterministic-MED

- **Solution is deterministic-MED**
  - **Paths are sorted in order of increasing ASN value**
  - **This is immune to neighbour resets**
- **MUST be configured on all BGP speakers in network**
  - **Existing deployments need to be careful in migration – whole network needs to be moved**
  - **New deployments should configure from scratch**

```
router bgp 100
 bgp deterministic-med
```

Acquisitions!

- **Your ISP has just bought another ISP**

  How to merge networks?

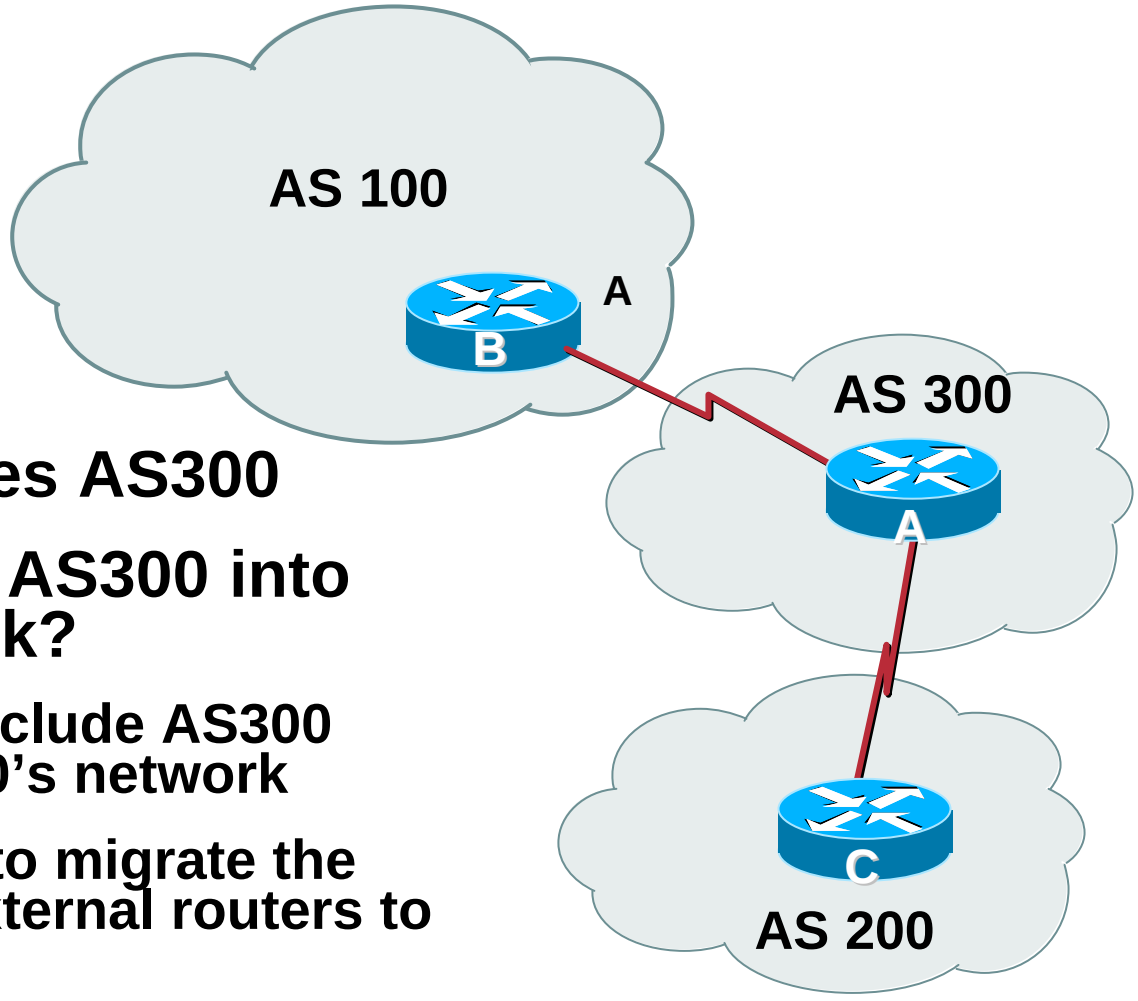
- **Options:**

  use confederations – make their AS a sub-AS (only useful if you are using confederations already)

  use the BGP local-as feature to implement a gradual transition – overrides BGP process ID

  neighbor *x.x.x.x* local-as *as-number* [no-prepend]

local-AS – Application



- AS100 purchases AS300
- How to migrate AS300 into AS100's network?
  - One task is to include AS300 routers in AS100's network
  - Another task is to migrate the peerings with external routers to the new AS

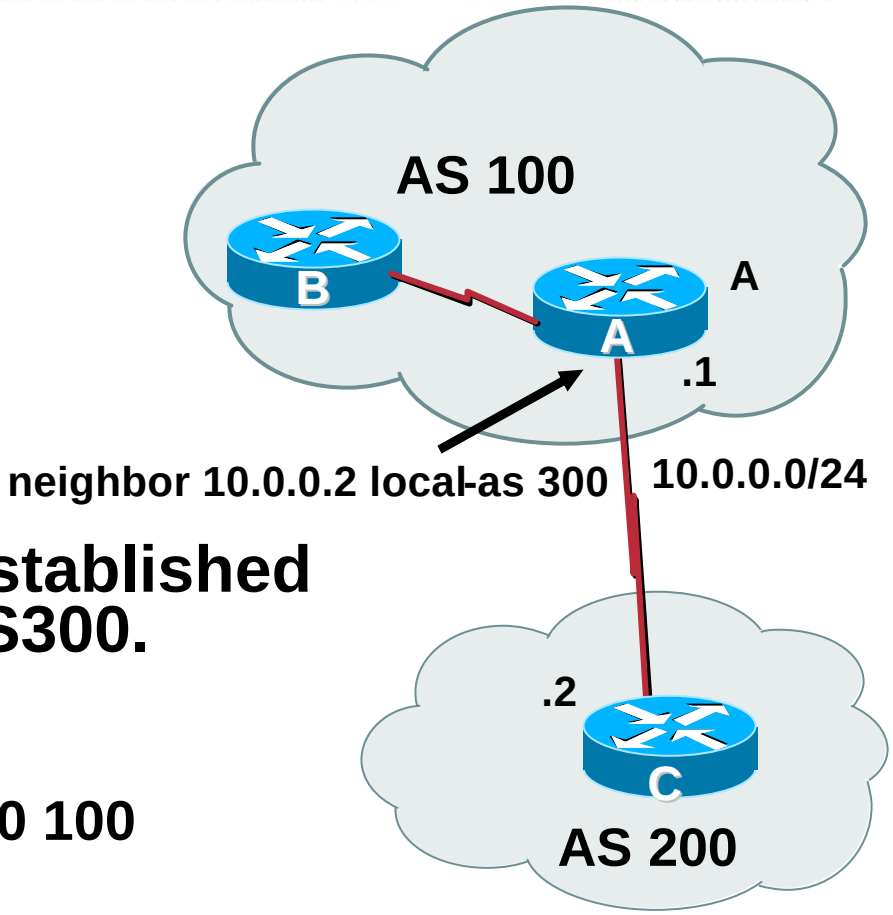
local-AS – Application

- **Migrating internal network can be done during ISP's maintenance periods**
- **During this work, the eBGP sessions need to be migrated to the new AS**

  But peers or customers or upstreams may not be available during ISP maintenance period

  local-AS comes to the rescue
- **Local-AS configured on specific eBGP peerings so that router in new AS appears as though it is still in its original AS**

local-AS – Application

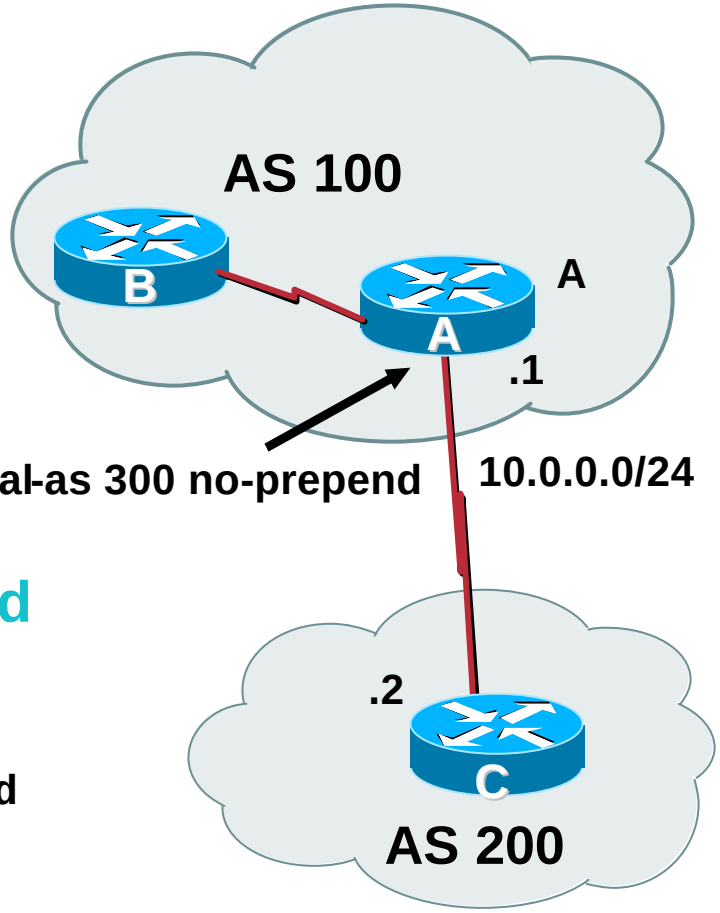


- **Router A is in AS100**
- **The peering with AS200 is established as if router A belonged to AS300.**
- **AS-PATH on Router C**

  routes originated in AS100 = 300 100
- **AS-PATH on Router A**

  routes received from AS200 = 300 200

local-AS – Application



- **Router A sees the old AS300 in the path**

  If this is not desired, the no-prepend option can be used

```
router bgp 100
 neigh 10.0.0.2 local-as 300 no-prepend
```

- **Routes received now appear as though they come directly from AS200 and not through AS300**

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   3FFE:C00:8023:5::2   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

Use bgp maxas-limit to ignore this bogus announcement

```
router bgp 100
 bgp maxas-limit 15
```

Limits the AS-path length to 15 ASNs only

Limiting the prefixes received

- **Maximum-prefix was introduced earlier in the eBGP template**

  The feature is actually quite powerful and has many options:

```
neighbor {neighbor} maximum-prefix {max} [thresh-int]
[warning-only][restart interval]
```

max	**the maximum number of prefixes before the peering is shutdown**
thresh-int	**percentage threshold before warnings are issued**
warning-only	**only issue warnings, never tear down BGP session**
restart *interval*	**how long to wait before attempting to restart a shutdown BGP session**

Maximum Prefix Example


```
neighbor 1.1.1.6 maximum-prefix 1000
```

- **When we receive the route number 1001 from neighbor 1.1.1.6**

```
%BGP-3-MAXPFXEXCEED: No. of unicast prefix
received from 1.1.1.6: 1001 exceed limit 1000
```

- **And the BGP session is closed**

  With warning-only option, message is logged, but BGP session is not closed

Maximum Prefix
Closing The Session

- The TCP/BGP session is closed:

  Update Malformed NOTIFICATION sent

  draft-chen-bgp-cease-subcod-00 adds specific codes for this condition (not implemented yet)

- The peering goes into ADMIN IDLE (equivalent) state

  Other side will try to reestablish (unsuccessfully) the BGP session

Maximum Prefix
Closing The Session


```
#show ip bgp summary
  Neighbor    (…)      State/PfxRcd
  1.1.1.6     (…)      Idle (PfxCt)
```

```
#show ip bgp neighbor
  Last reset 00:01:32, due to BGP Notification sent,
  update malformed
  Peer had exceeded the max. no. of prefixes configured
  Reduce the no. of prefix and clear ip bgp 1.1.1.6 to
  restore peering
```

Maximum Prefix
Closing The Session

- **To re-establish the peering, operators need to do a hard reset on the BGP peering:**

  `clear ip bgp {`*`neighbor`*`}`

  neighbor **must be an ip address**

- **The other side cannot do anything to re-establish the peering**

Maximum Prefix Threshold Option

- **Router will also log a warning if we exceed *thres-int***

  thres-int* is a percentage of *max

```
neighbor 1.1.1.6 maximum-prefix 1000 60
```

  After receiving the prefix # 601

```
%BGP-4-MAXPFX: No. of unicast prefix received from
1.1.1.6 reaches 601, max 1000
```

- **The default value for *thres-int* is 75%**

Maximum Prefix restart Option

- If the restart option is configured, the router will try to restart the session automatically each interval

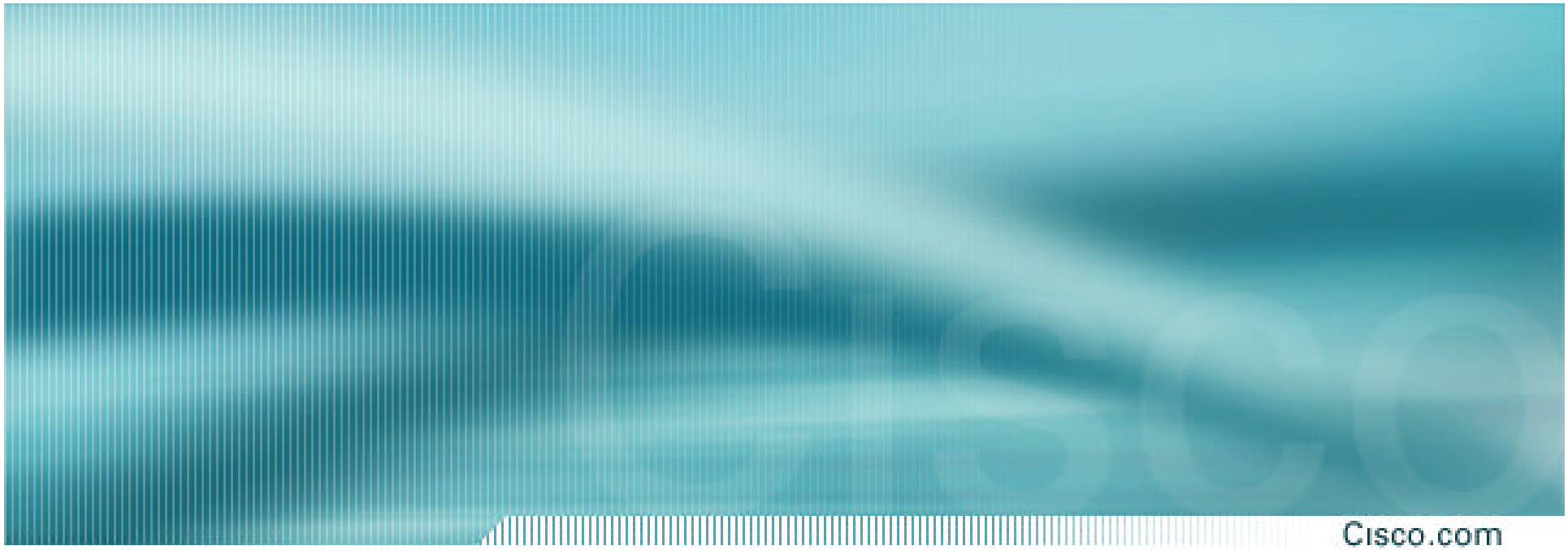
  Interval: 1-65535 minutes

  If not configured, restarting will need to be done by hand

```
#show ip bgp neighbor
  Threshold for warning message 75%, restart interval 60 min
  Reduce the no. of prefix from 145.10.10.2, will restart in
  00:00:40
```

BGP for Internet Service Providers

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



IP Addressing

How to do addressing within an ISP Network with a view to optimising the IGP and iBGP

IP Addressing

- **IPv4 Address space is a resource shared amongst all Internet users**
  - **Regional Internet Registries delegated allocation responsibility by the IANA**
  - **APNIC, ARIN, RIPE NCC and LACNIC are the four RIRs**
  - **RIRs allocate address space to ISPs and Local Internet Registries**
  - **ISPs/LIRs assign address space to end customers**
- **56% of available IPv4 address space used**

Definitions

- **Non-portable – ‘provider aggregatable’ (PA)**
  - **Customer uses RIR member’s address space while connected to Internet**
  - **Customer renumbers when changing ISP**
  - **Helps control of size of Internet routing table**
  - **May fragment provider block when multihoming**
- **PA space is allocated to the RIR member with the requirement that all assignments are announced as an aggregate**

Definitions

- **Portable – ‘provider independent’ (PI)**
  - **Customer gets or has address space independent of ISP**
  - **Customer keeps addresses when changing ISP**
  - **Bad for size of Internet routing table**
  - **PI space is rarely distributed by the RIRs**

Address Space

- **Approach upstream ISP or consider RIR membership for address space**
- **Supply addressing plan when requested**
  - **remember Internet is classless**
  - **addresses assigned according to need not want**
- **Assign addresses to backbone and other network layers – remember scalability!**
- **Some examples follow…**

Principles of Addressing

- **Separate customer & infrastructure address pools**
  - **Manageability**
    - **Different personnel manage infrastructure and assignments to customers**
  - **Scalability**
    - **Easier renumbering – customer renumbering is harder, infrastructure is easy**

Principles of Addressing

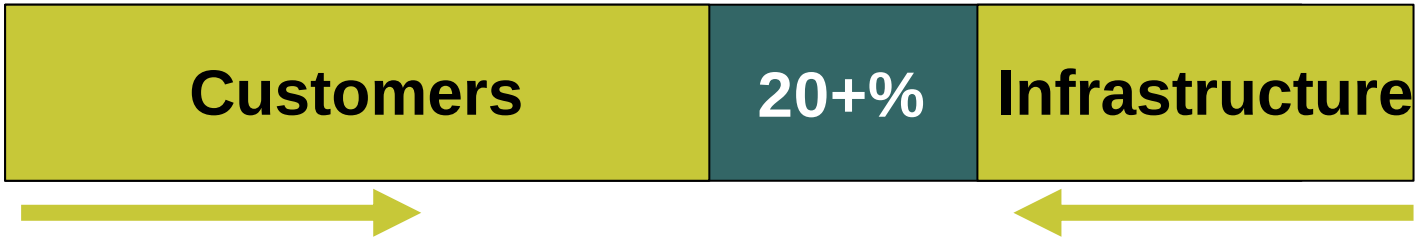
- **Further separate infrastructure**
  - **In the IGP:**
    - **p2p addresses of backbone connections**
    - **router loopback addresses**
  - **Not in the IGP:**
    - **RAS server address pools**
    - **Virtual web and content hosting LANs**
    - **Mail, DNS server system LANs**

Principles of Addressing

- Customer networks
  - Carry in iBGP
  - Do not put in IGP – ever!
- Do not need to aggregate address space assigned to customers
  - iBGP can carry in excess of 200,000 prefixes, no IGP is designed to do this

Management – Simple Network

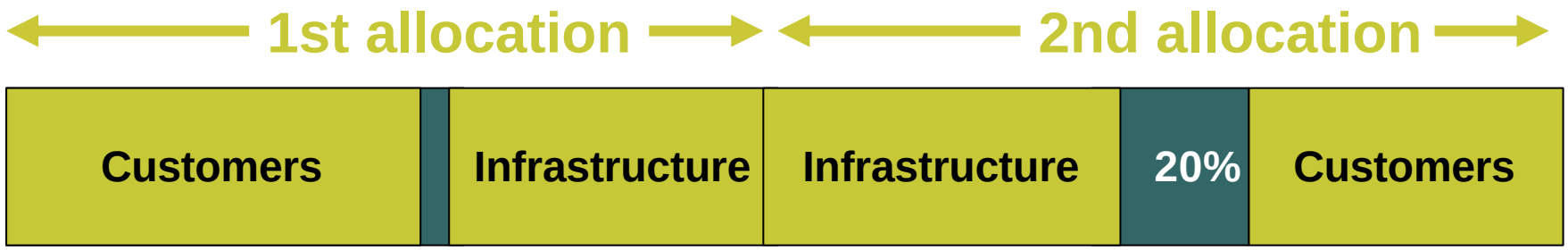
- **First allocation from APNIC**
  - **Infrastructure is known, customers are not**
  - **20% free is trigger for next request**



- **Grow usage of blocks from edges**
- **Assign customers sequentially**

Management – Simple Network

- **If second allocation is contiguous**



Reverse order of division of first block

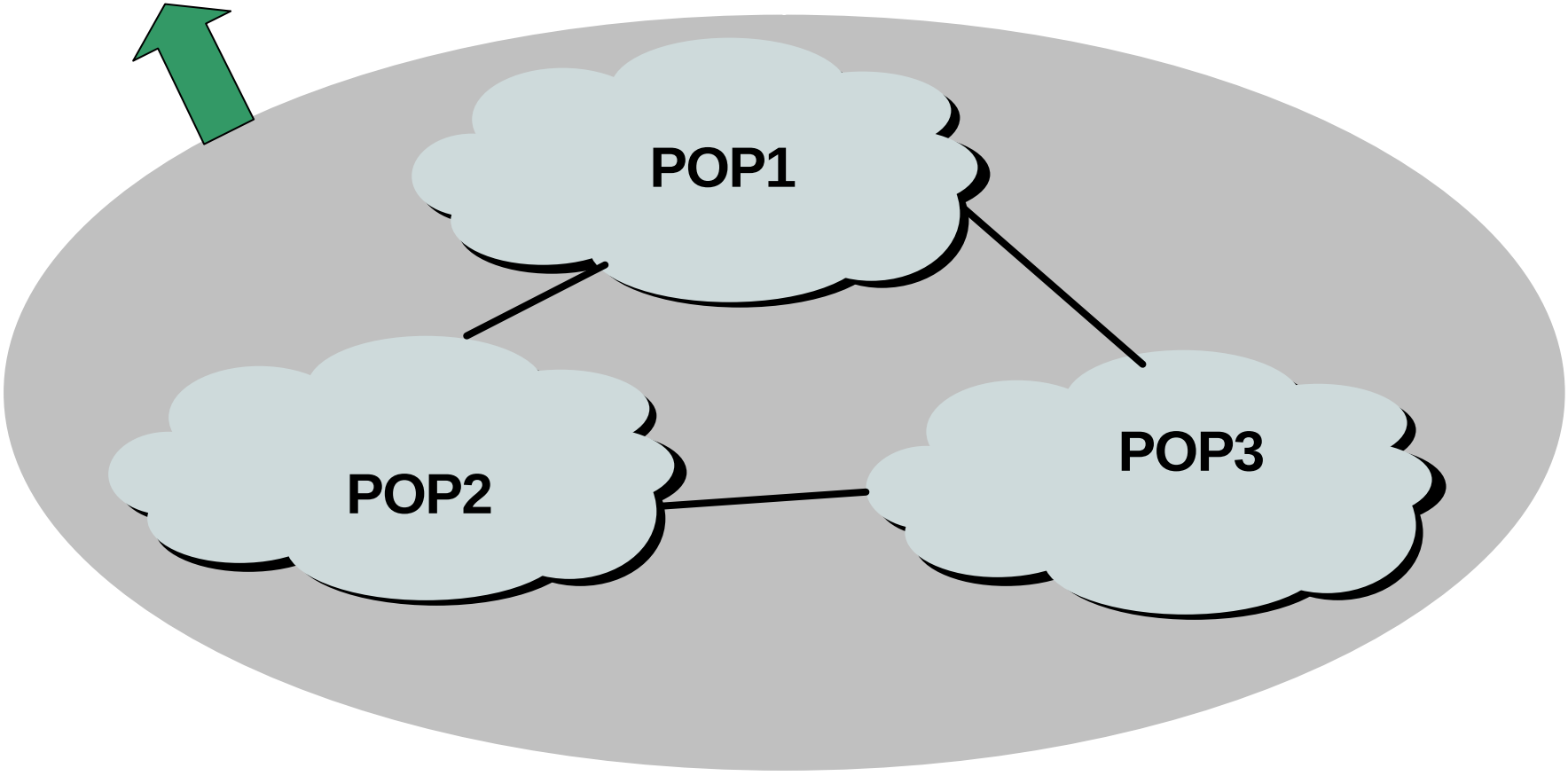
Maximise contiguous space for infrastructure

Easier for debugging

Customer networks can be discontiguous

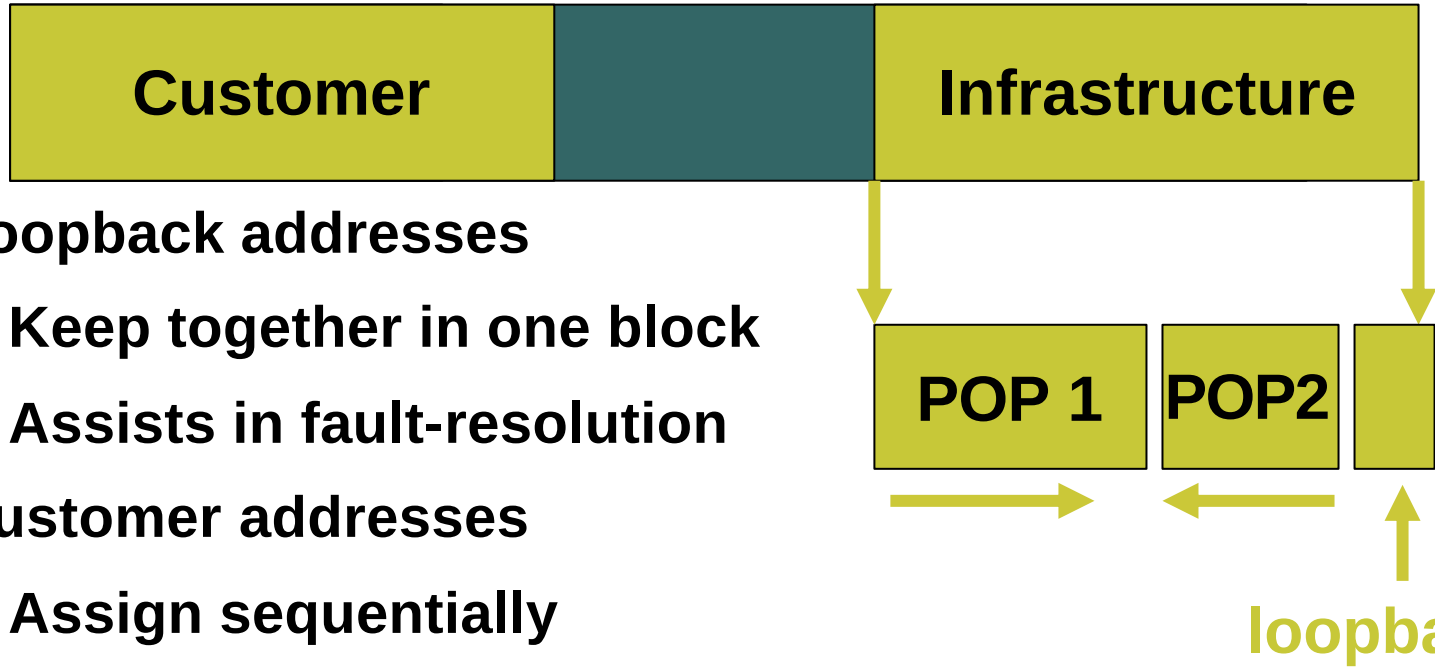
Management – Many POPs

WAN link to single transit ISP



Management – Many POPs

- **POP sizes**
  - **Choose address pool for each POP according to need**



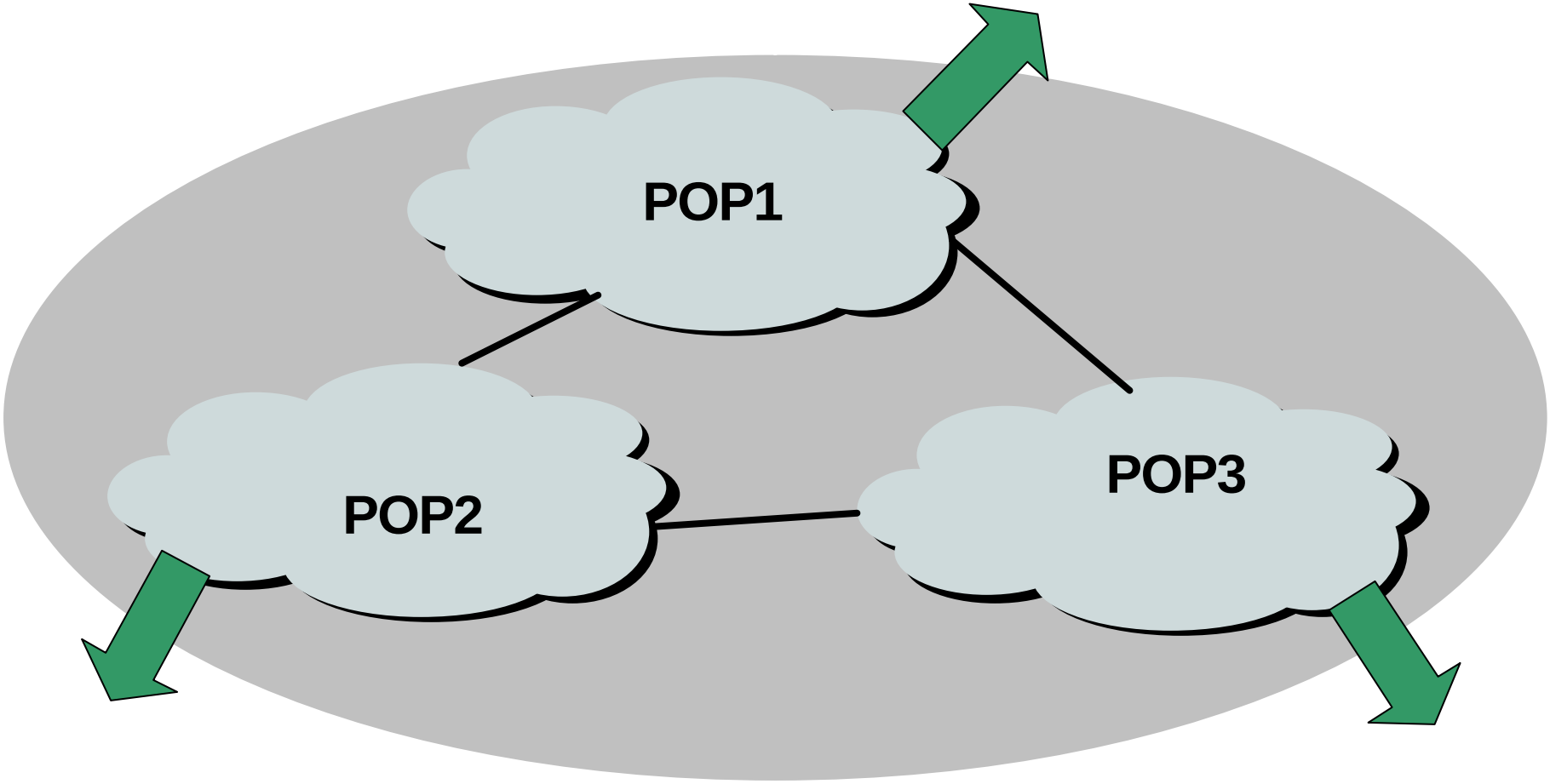
  - **Loopback addresses**
    - **Keep together in one block**
    - **Assists in fault-resolution**
  - **Customer addresses**
    - **Assign sequentially**

Management – Many POPs

- **/20 minimum allocation is not enough for all your POPs?**
  - **Deploy addresses on infrastructure first**
- **Common mistake:**
  - **Reserving customer addresses on a per POP basis**
- **Do not constrain network plans due to lack of address space**
  - **Re-apply once address space has been used**
  - **There is plenty of it!**

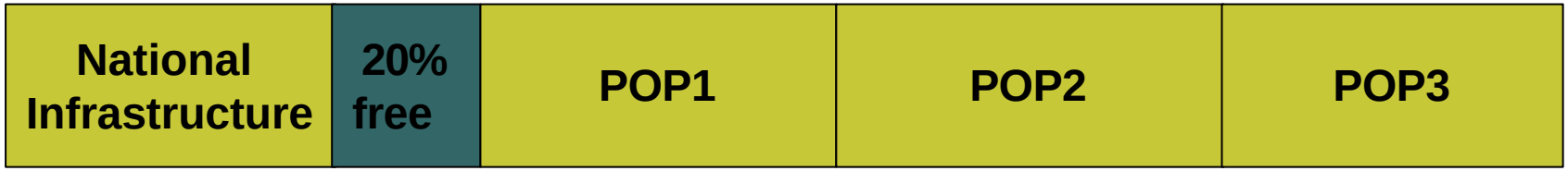
Management – Multiple Exits

- **WAN links to different ISPs**



Management – Multiple Exits

- **Create a 'national' infrastructure pool**



Carry in IGP

E.g. loopbacks, p2p links, infrastructure connecting routers and hosts which are multiply connected

On a per POP basis

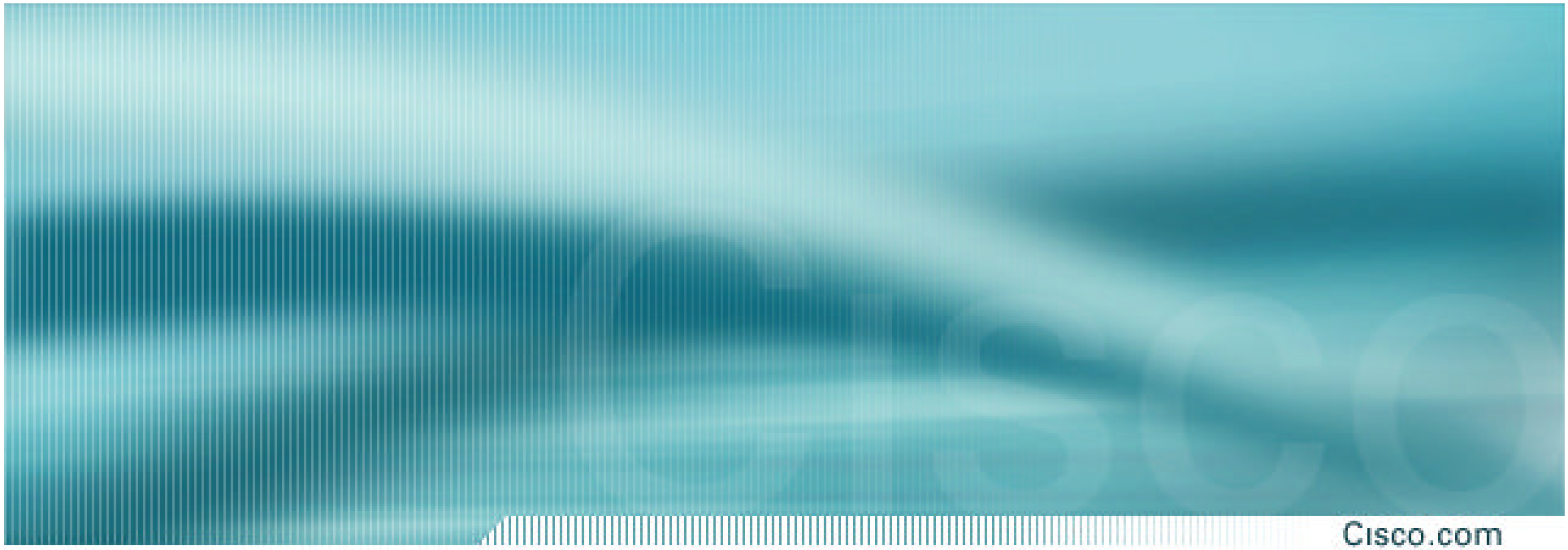
Consider separate memberships if requirement for each POP is very large from day one

Summary

- Set up an addressing plan which is sensitive to your backbone needs
- IGP carries only infrastructure addresses
- iBGP can carry the rest
  - Aggregation of customer assignments within the iBGP is usually not necessary
  - Aggregation of external announcements is VERY necessary

BGP for Internet Service Providers

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



Service Providers use of Communities

Some examples of how ISPs make life easier for themselves

BGP Communities

- Another ISP “scaling technique”
- Prefixes are grouped into different “classes” or communities within the ISP network
- Each community means a different thing, has a different result in the ISP network

BGP Communities

- **Communities are generally set at the edge of the ISP network**

  Customer edge: customer prefixes belong to different communities depending on the services they have purchased

  Internet edge: transit provider prefixes belong to difference communities, depending on the loadsharing or traffic engineering requirements of the local ISP, or what the demands from its BGP customers might be

- **Two simple examples follow to explain the concept**

Community Example – Customer Edge

- This demonstrates how communities might be used at the customer edge of an ISP network
- ISP has three connections to the Internet:
  - IXP connection, for local peers
  - Private peering with a competing ISP in the region
  - Transit provider, who provides visibility to the entire Internet
- Customers have the option of purchasing combinations of the above connections

Community Example – Customer Edge

- **Community assignments:**

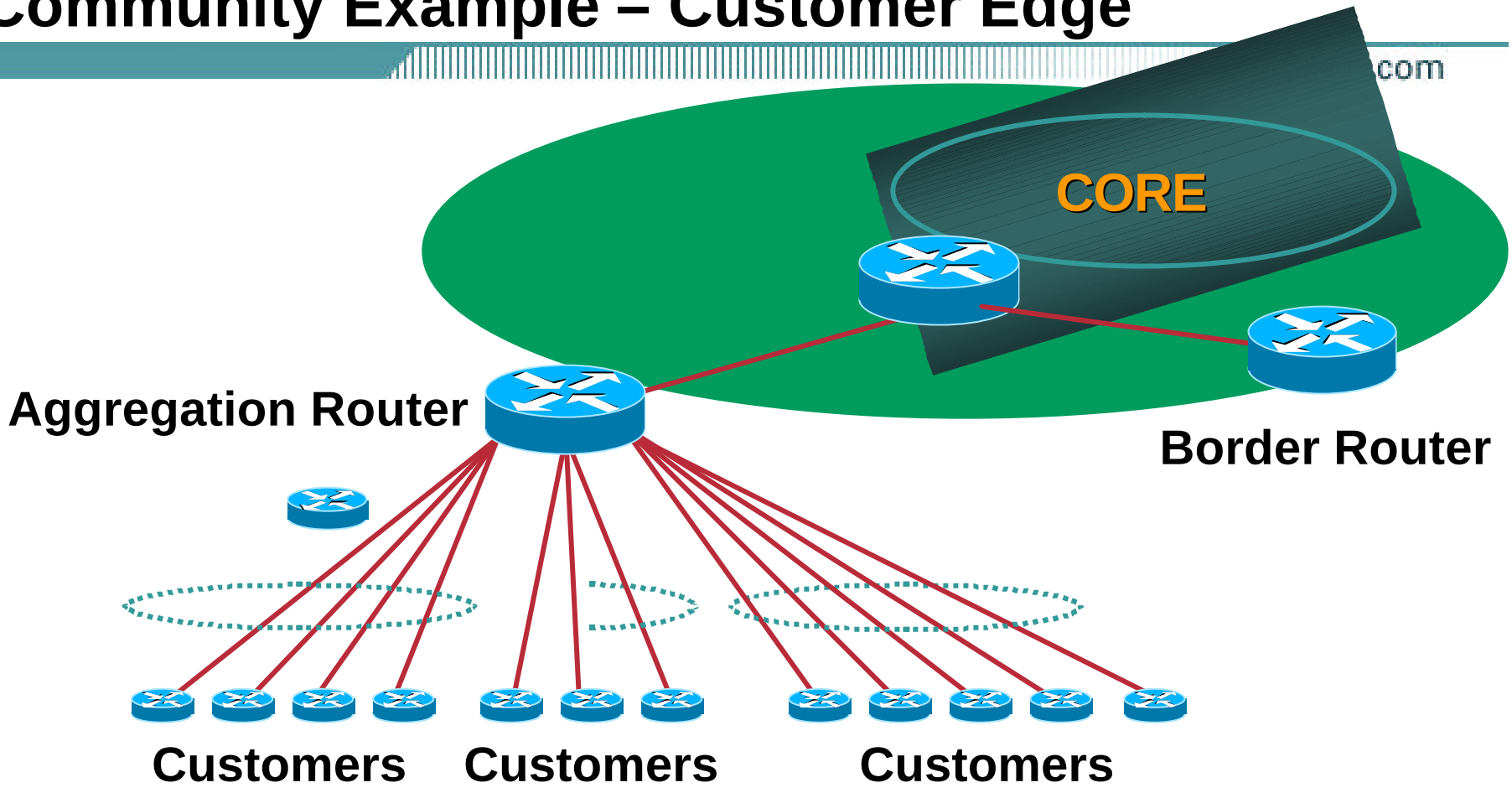
  IXP connection: community 100:2100

  Private peer: community 100:2200

- **Customer who buys local connectivity (via IXP) is put in community 100:2100**
- **Customer who buys peer connectivity is put in community 100:2200**
- **Customer who wants both IXP and peer connectivity is put in 100:2100 and 100:2200**
- **Customer who wants “the Internet” has no community set**

  We are going to announce his prefix everywhere

Community Example – Customer Edge

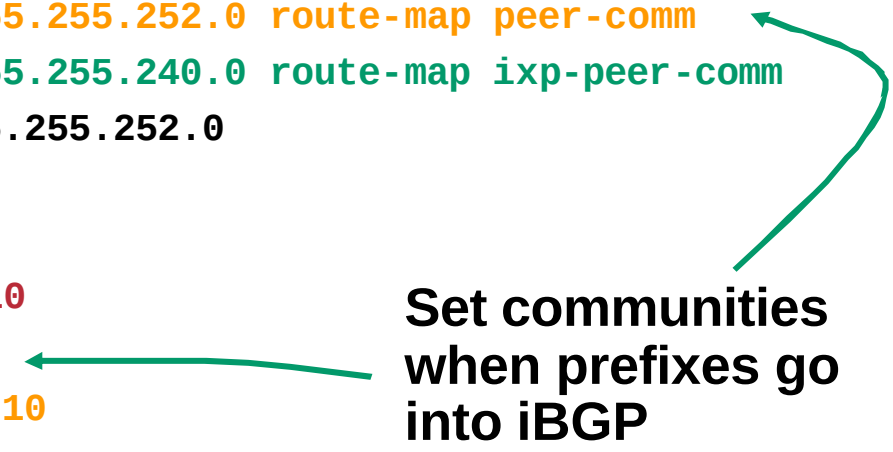


Communities set at the aggregation router where the prefix is injected into the ISP's iBGP

Community Example – Customer Edge

Aggregation Router configuration

```
ip route 222.1.20.0 255.255.255.0 serial 0 ! IXP only
ip route 222.1.28.0 255.255.252.0 serial 1 ! Peer only
ip route 222.1.64.0 255.255.240.0 serial 3 ! IXP+Peer
ip route 222.1.0.0  255.255.252.0 serial 4 ! everything
!
router bgp 100
 network 222.1.20.0 mask 255.255.255.0 route-map ixp-comm
 network 222.1.28.0 mask 255.255.252.0 route-map peer-comm
 network 222.1.64.0 mask 255.255.240.0 route-map ixp-peer-comm
 network 222.1.0.0 mask 255.255.252.0
 neighbor ...
!
route-map ixp-comm permit 10
 set community 100:2100
route-map peer-comm permit 10
 set community 100:2200
route-map ixp-peer-comm permit 10
 set community 100:2100 100:2200
```

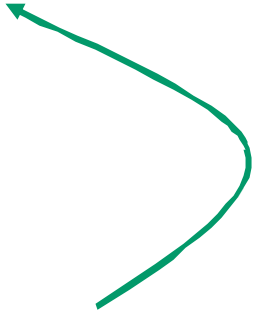


Set communities when prefixes go into iBGP

Community Example – Customer Edge

Border Router configuration

```
router bgp 100
 network 221.1.0.0 mask 255.255.0.0
 neighbor ixp-peer peer-group
 neighbor ixp-peer route-map ixp-out out
 neighbor private-peer peer-group
 neighbor private-peer route-map ppeer-out out
 neighbor upstream peer-group
 neighbor upstream prefix-list aggregate out
 neighbor ...
!
route-map ixp-out permit 10
 match community 11
route-map ppeer-out permit 10
 match community 12
!
ip community-list 11 permit 100:2100
ip community-list 12 permit 100:2200
ip prefix-list aggregate permit 221.1.0.0/16
```



Filter outgoing announcements based on communities set

Community Example – Customer Edge

- **No need to alter filters at the network border when adding a new customer**
- **New customer simply is added to the appropriate community**
  - **Border filters already in place take care of announcements**
    - **Ease of operation!**

Community Example – Internet Edge

- This demonstrates how communities might be used at the peering edge of an ISP network
- ISP has four types of BGP peers:
  - Customer
  - IXP peer
  - Private peer
  - Transit provider
- The prefixes received from each can be classified using communities
- Customers can opt to receive any or all of the above

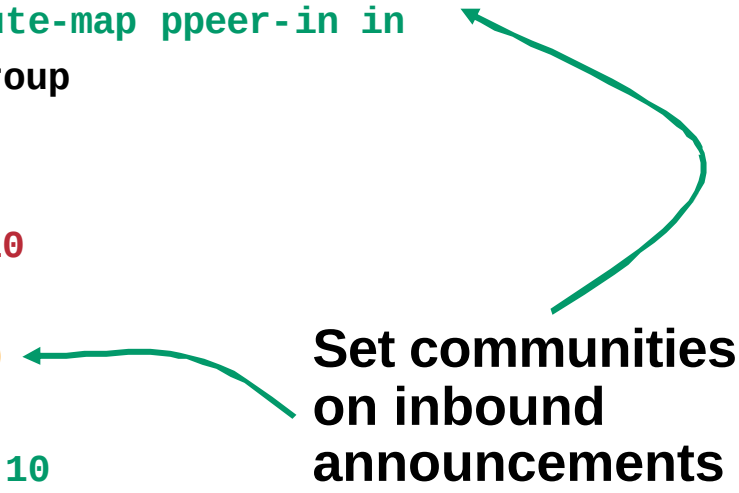
Community Example – Internet Edge

- **Community assignments:**
  - **Customer prefix: community 100:3000**
  - **IXP prefix: community 100:3100**
  - **Private peer prefix: community 100:3200**
- **BGP customer who buys local connectivity gets 100:3000**
- **BGP customer who buys local and IXP connectivity receives community 100:3000 and 100:3100**
- **BGP customer who buys full peer connectivity receives community 100:3000, 100:3100, and 100:3200**
- **Customer who wants “the Internet” gets everything**
  - **Gets default route via “default-originate”**
  - **Or pays money to get all 120k prefixes**

Community Example – Internet Edge

Border Router configuration

```
router bgp 100
 neighbor customer peer-group
 neighbor customer route-map cust-in in
 neighbor ixp-peer peer-group
 neighbor ixp-peer route-map ixp-in in
 neighbor private-peer peer-group
 neighbor private-peer route-map ppeer-in in
 neighbor upstream peer-group
 neighbor ...
!
route-map cust-in permit 10
 set community 100:3000
route-map ixp-in permit 10
 set community 100:3100
route-map ppeer-in permit 10
 set community 100:3200
!
```

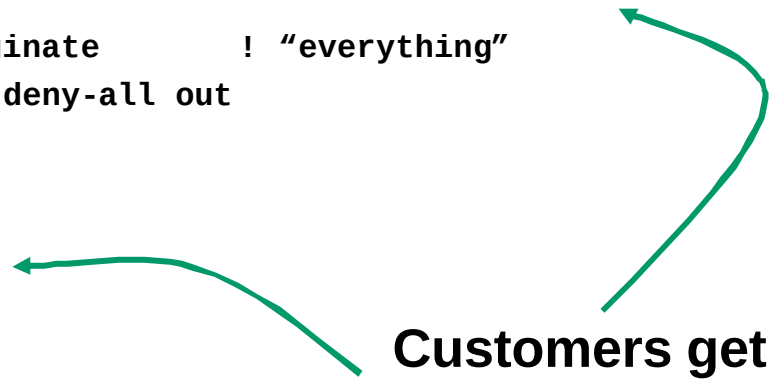


Set communities on inbound announcements

Community Example – Internet Edge

Aggregation Router configuration

```
router bgp 100
 neighbor customer1 peer-group
 neighbor customer1 route-map cust1-out     ! local routes
 neighbor customer2 peer-group
 neighbor customer2 route-map cust2-out     ! local+IXP routes
 neighbor customer3 peer-group
 neighbor customer3 route-map cust3-out     ! all routes except internet
 neighbor customer4 peer-group
 neighbor customer4 default-originate       ! “everything”
 neighbor customer4 prefix-list deny-all out
!
route-map cust1-out permit 10
 match community 23
route-map cust2-out permit 10
 match community 24
route-map cust3-out permit 10
 match community 25
!
ip community-list 23 permit 100:3000
ip community-list 24 permit 100:3000
ip community-list 24 permit 100:3100
...
```



Customers get prefixes according to community matches

Community Example – Internet Edge

- **No need to create customised filters when adding customers**
  - **Border router already sets communities**
  - **Installation engineers pick the appropriate community set when establishing the customer BGP session**
    - **Ease of operation!**

Community Example – Summary

- **Two examples of customer edge and internet edge can be combined to form a simple community solution for ISP prefix policy control**
- **More experienced operators tend to have more sophisticated options available**
  - **Advice is to start with the easy examples given, and then proceed onwards as experience is gained**

Some ISP Examples

- ISPs also create communities to give customers bigger routing policy control
- Public policy is usually listed in the IRR
  - Following examples are all in the IRR
  - Examples build on the configuration concepts from the introductory example
- Consider creating communities to give policy control to customers
  - Reduces technical support burden
  - Reduces the amount of router reconfiguration, and the chance of mistakes

Some ISP Examples
Connect.com.au

- Australian ISP
- Run their own Routing Registry

  Whois.connect.com.au

- Permit customers to send up 8 types of communities to allow traffic engineering

Some ISP
Conne

```
aut-num:        AS2764
as-name:        ASN-CONNECT-NET
descr:          connect.com.au pty ltd
admin-c:        CC89
tech-c:         MP151
remarks:        Community  Definition
remarks:        ----------------------------------------
remarks:        2764:1 Announce to "domestic" rate ASes only
remarks:        2764:2 Don't announce outside local POP
remarks:        2764:3 Lower local preference by 25
remarks:        2764:4 Lower local preference by 15
remarks:        2764:5 Lower local preference by 5
remarks:        2764:6 Announce to non customers with "no-export"
remarks:        2764:7 Only announce route to customers
remarks:        2764:8 Announce route over satellite link
notify:         routing@connect.com.au
mnt-by:         CONNECT-AU
changed:        mrp@connect.com.au 19990506
source:         CCAIR
```

Some ISP Examples
UUNET Europe

- UUNET's European operation
- Permits customers to send communities which determine
  - local preferences within UUNET's network
  - Reachability of the prefix
  - How the prefix is announced outside of UUNET's network

Some IS
UUN

```
aut-num: AS702
as-name: AS702
descr:   UUNET - Commercial IP service provider in Europe
remarks: ------------------------------------------------------------
remarks: UUNET uses the following communities with its customers:
remarks: 702:80  Set Local Pref 80 within AS702
remarks: 702:120 Set Local Pref 120 within AS702
remarks: 702:20  Announce only to UUNET AS'es and UUNET customers
remarks: 702:30  Keep within Europe, don't announce to other UUNET AS's
remarks: 702:1   Prepend AS702 once at edges of UUNET to Peers
remarks: 702:2   Prepend AS702 twice at edges of UUNET to Peers
remarks: 702:3   Prepend AS702 thrice at edges of UUNET to Peers
remarks: Details of UUNET's peering policy and how to get in touch with
remarks: UUNET regarding peering policy matters can be found at:
remarks: http://www.uu.net/peering/
remarks: ------------------------------------------------------------
mnt-by:  UUNET-MNT
changed: eric-apps@eu.uu.net 20010928
source:  RIPE
```

Some ISP Examples
BT Ignite

- Formerly Concert's European network
- One of the most comprehensive community lists around
  - Seems to be based on definitions originally used in Tiscali's network
  - whois –h whois.ripe.net AS5400 reveals all
- Extensive community definitions allow sophisticated traffic engineering by customers

Some ISP
BT Ig

```
aut-num:        AS5400
as-name:        CIPCORE
descr:          BT Ignite European Backbone
remarks:        The following BGP communities can be set by BT Ignite
remarks:        BGP customers to affect announcements to major peers.
remarks:
remarks:        Community to                                Community to
remarks:        Not announce      To peer:                  AS prepend 5400
remarks:
remarks:        5400:1000 European peers                    5400:2000
remarks:        5400:1001 Sprint (AS1239)                   5400:2001
remarks:        5400:1003 Unisource (AS3300)                5400:2003
remarks:        5400:1005 UUnet (AS702)                     5400:2005
remarks:        5400:1006 Carrier1 (AS8918)                 5400:2006
remarks:        5400:1007 SupportNet (8582)                 5400:2007
remarks:        5400:1008 AT&T (AS2686)                     5400:2008
remarks:        5400:1009 Level 3 (AS9057)                  5400:2009
remarks:        5400:1010 RIPE (AS3333)                     5400:2010
<snip>
remarks:        5400:1100         US peers                  5400:2100
notify:         notify@eu.ignite.net
mnt-by:         CIP-MNT
source:         RIPE
```

And many many more!

Some ISP Examples
Carrier1

- European ISP
- Another very comprehensive list of community definitions

  whois –h whois.ripe.net AS8918 reveals all

Some ISP
Carrier

```
aut-num:        AS8918
descr:          Carrier1 Autonomous System
<snip>
remarks:        Community Support Definitions:
remarks:        Communities that determine the geographic
remarks:        entry point of routes into the Carrier1 network:
remarks:        *
remarks:        Community     Entry Point
remarks:        --------------------------------
remarks:        8918:10       London
remarks:        8918:15       Hamburg
remarks:        8918:18       Chicago
remarks:        8918:20       Amsterdam
remarks:        8918:25       Milan
remarks:        8918:28       Berlin
remarks:        8918:30       Frankfurt
remarks:        8918:35       Zurich
remarks:        8918:40       Geneva
remarks:        8918:45       Stockholm
<snip>
notify:         inoc@carrier1.net
mnt-by:         CARRIER1-MNT
source:         RIPE
```

Some ISP Examples
Level 3

- **Highly detailed AS object held on the RIPE Routing Registry**
- **Also a very comprehensive list of community definitions**

  whois –h whois.ripe.net AS3356 reveals all

Some IS
Leve

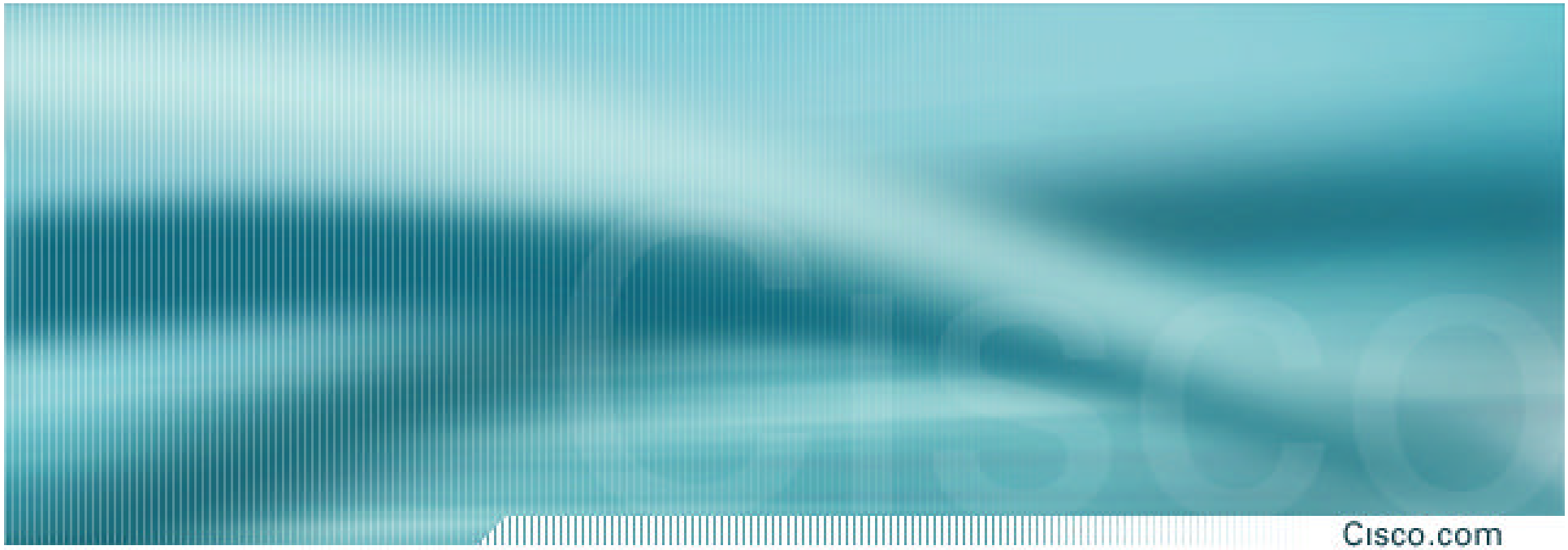
```
aut-num:      AS3356
descr:        Level 3 Communications
<snip>
remarks:      ------------------------------------------------------
remarks:      customer traffic engineering communities - Suppression
remarks:      ------------------------------------------------------
remarks:      64960:XXX - announce to AS XXX if 65000:0
remarks:      65000:0   - announce to customers but not to peers
remarks:      65000:XXX - do not announce at peerings to AS XXX
remarks:      ------------------------------------------------------
remarks:      customer traffic engineering communities - Prepending
remarks:      ------------------------------------------------------
remarks:      65001:0   - prepend once  to all peers
remarks:      65001:XXX - prepend once  at peerings to AS XXX
remarks:      65002:0   - prepend twice to all peers
remarks:      65002:XXX - prepend twice at peerings to AS XXX
remarks:      65003:0   - prepend 3x to all peers
remarks:      65003:XXX - prepend 3x    at peerings to AS XXX
remarks:      65004:0   - prepend 4x to all peers
remarks:      65004:XXX - prepend 4x    at peerings to AS XXX
<snip>
mnt-by:       LEVEL3-MNT
source:       RIPE
```

And many many more!

BGP for Internet Service Providers

Cisco.com

- IGP versus BGP
- Injecting Prefixes into iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips
- Addressing Planning
- Service Provider use of Communities



BGP Tutorial

End of Part 2 – Deployment Techniques

Part 3 – Multihoming is next