BGP Techniques for Internet Service Providers

Philip Smith <pfs@cisco.com>

NANOG 34

Seattle, 15-17 May 2005

Presentation Slides

- **Slides are at:**

  ftp://ftp-eng.cisco.com /pfs/seminars/NANOG34-BGP-Techniques.pdf

  And on the NANOG 34 meeting website

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

BGP for Internet Service Providers

- BGP Basics
- Scaling BGP
- Using Communities
- Deploying BGP in an ISP network

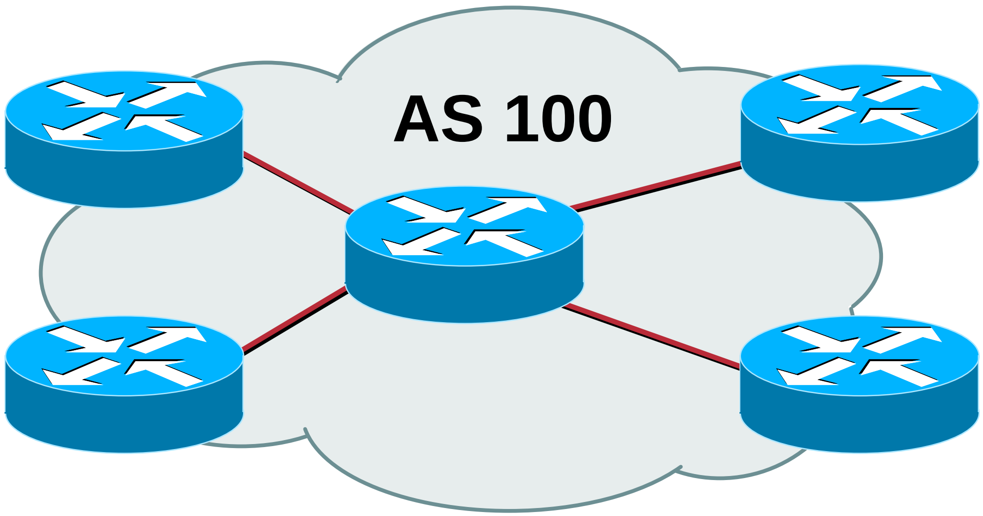
BGP Basics

Reminder…

Border Gateway Protocol

- **Routing Protocol used to exchange routing information between networks**
  - **exterior gateway protocol**
- **Described in RFC1771**
  - **work in progress to update**
  - **www.ietf.org/internet-drafts/draft-ietf-idr-bgp4-26.txt**
- **The Autonomous System is BGP's fundamental operating unit**
  - **It is used to uniquely identify networks with common routing policy**

Autonomous System (AS)



- Collection of networks with same routing policy
- Single routing protocol
- Usually under single ownership, trust and administrative control
- Identified by a unique number

Autonomous System Number (ASN)

- An ASN is a 16 bit integer

  1-64511 are for public network use

  64512-65534 are for private use and should never appear on the Internet

  0 and 65535 are reserved

- 32 bit ASNs are coming soon

  www.ietf.org/internet-drafts/draft-ietf-idr-as4bytes-09.txt

  With ASN 23456 reserved for the transition

Autonomous System Number (ASN)

- ASNs are distributed by the Regional Internet Registries
  - Also available from upstream ISPs who are members of one of the RIRs
- Current ASN allocations up to 37887 have been made to the RIRs
  - Of these, around 19500 are visible on the Internet
- Current estimates are that 4-byte ASNs will be required by July 2010

Applying Policy with BGP

Control!

Applying Policy in BGP: Why?

- **Policies are applied to:**
  - **Influence BGP Path Selection by setting BGP attributes**
  - **Determine which prefixes are announced or blocked**
  - **Determine which AS-paths are preferred, permitted, or denied**
  - **Determine route groupings and their effects**
- **Decisions are generally based on prefix, AS-path and community**

Applying Policy with BGP: Tools

- **Most implementations have tools to apply policies to BGP:**
  - **Prefix manipulation/filtering**
  - **AS-PATH manipulation/filtering**
  - **Community Attribute setting and matching**
- **Implementations also have policy language which can do various match/set constructs on the attributes of chosen BGP routes**

BGP Capabilities

Extending BGP

BGP Capabilities

- **Documented in RFC2842**
- **Capabilities parameters passed in BGP open message**
- **Unknown or unsupported capabilities will result in NOTIFICATION message**
- **Codes:**

  0 to 63 are assigned by IANA by IETF consensus

  64 to 127 are assigned by IANA “first come first served”

  128 to 255 are vendor specific

BGP Capabilities

Current capabilities are:

0	Reserved	[RFC3392]
1	Multiprotocol Extensions for BGP-4	[RFC2858]
2	Route Refresh Capability for BGP-4	[RFC2918]
3	Cooperative Route Filtering Capability	[ID]
4	Multiple routes to a destination capability	[RFC3107]
64	Graceful Restart Capability	[ID]
65	Support for 4 octet ASNs	[ID]
66	Deprecated 2003-03-06	
67	Support for Dynamic Capability	[ID]

See www.iana.org/assignments/capability-codes

BGP Capabilities

- **Multiprotocol extensions**

  This is a whole different world, allowing BGP to support more than IPv4 unicast routes

  Examples include: v4 multicast, IPv6, v6 multicast, VPNs

  Another tutorial (or many!)

- **Route refresh is a well known scaling technique – covered shortly**
- **The other capabilities are still in development or not widely implemented or deployed yet**

BGP for Internet Service Providers

- BGP Basics
- Scaling BGP
- Using Communities
- Deploying BGP in an ISP network

BGP Scaling Techniques

BGP Scaling Techniques

- How does a service provider:
  - Scale the iBGP mesh beyond a few peers?
  - Implement new policy without causing flaps and route churning?
  - Keep the network stable, scalable, as well as simple?

Route Refresh

Route Refresh

- **BGP peer reset required after every policy change**
  - **Because the router does not store prefixes which are rejected by policy**
- **Hard BGP peer reset:**
  - **Terminates BGP peering & Consumes CPU**
  - **Severely disrupts connectivity for all networks**
- **Soft BGP peer reset (or Route Refresh):**
  - **BGP peering remains active**
  - **Impacts only those prefixes affected by policy change**

Route Refresh Capability

- **Facilitates non-disruptive policy changes**
- **For most implementations, no configuration is needed**
  - **Automatically negotiated at peer establishment**
- **No additional memory is used**
- **Requires peering routers to support “route refresh capability” – RFC2918**

Route Refresh

- **Use Route Refresh capability if supported**
  - **find out from the BGP neighbour status display**
  - **Non-disruptive, "Good For the Internet"**
- **If not supported, see if implementation has a workaround**
- **Only hard-reset a BGP peering as a last resort**

Consider the impact to be equivalent to a router reboot

Route Flap Damping

Stabilising the Network

Route Flap Damping

- **Route flap**
  - **Going up and down of path or change in attribute**
    - **BGP WITHDRAW followed by UPDATE = 1 flap**
    - **eBGP neighbour peering reset is NOT a flap**
  - **Ripples through the entire Internet**
  - **Causes instability, wastes CPU**
- **Damping aims to reduce scope of route flap propagation**

Route Flap Damping (continued)

- Requirements
  - Fast convergence for normal route changes
  - History predicts future behaviour
  - Suppress oscillating routes
  - Advertise stable routes
- Documented in RFC2439

Operation

- Add penalty for each flap

  NB: Change in attribute is also penalized

- Exponentially decay penalty

  half life determines decay rate

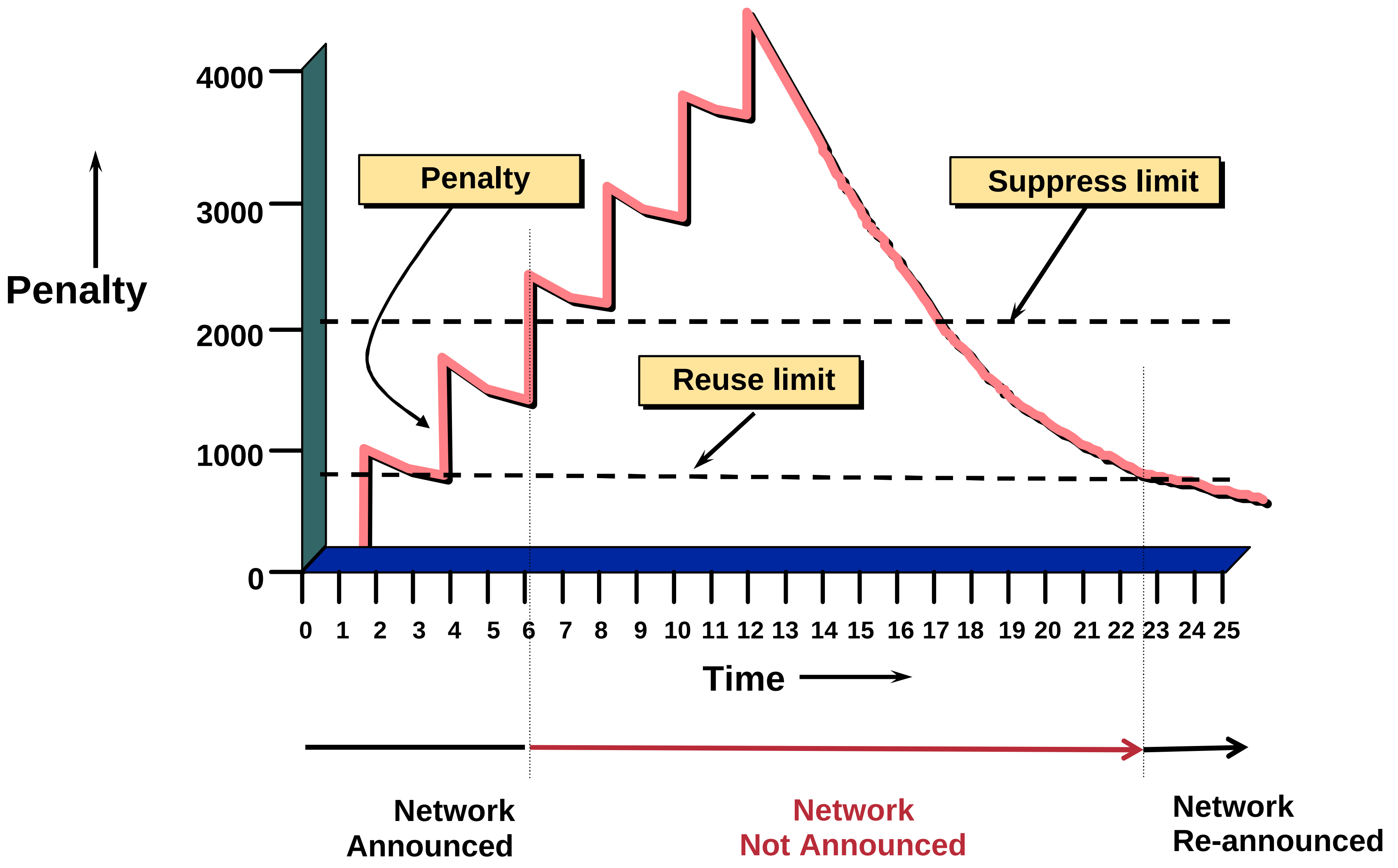
- Penalty above suppress-limit

  do not advertise route to BGP peers

- Penalty decayed below reuse-limit

  re-advertise route to BGP peers

Operation



Penalty
4000
3000
2000
1000
0
Penalty
Suppress limit
Reuse limit
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Time
Network Announced
Network Not Announced
Network Re-announced

Operation

- Only applied to inbound announcements from eBGP peers
- Alternate paths still usable
- Controllable by at least:
  - Half-life
  - reuse-limit
  - suppress-limit
  - maximum suppress time

Configuration

- **Implementations allow various policy control with flap damping**
  - **Fixed damping, same rate applied to all prefixes**
  - **Variable damping, different rates applied to different ranges of prefixes and prefix lengths**

Implementing Flap Damping

- Flap Damping should only be implemented to address a specific network stability problem
- Flap Damping can and does make stability worse

  "Flap Amplification" from AS path attribute changes caused by BGP exploring alternate paths being unnecessarily penalised

  "Route Flap Damping Exacerbates Internet Routing Convergence"

    Zhuoqing Morley Mao, Ramesh Govindan, George Varghese & Randy H. Katz, August 2002

Implementing Flap Damping

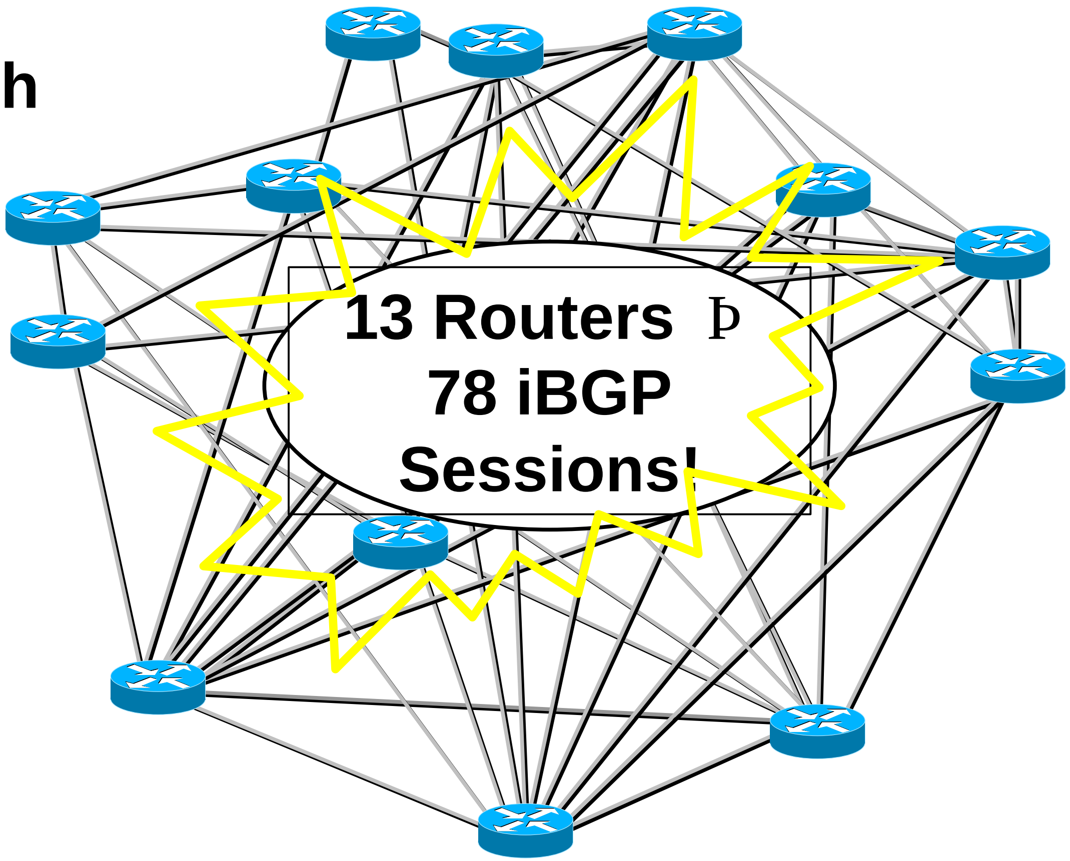
- **If you have to implement flap damping, understand the impact on the network**
  - **Vendor defaults are very severe**
  - **Variable flap damping can bring benefits**
  - **Transit provider flap damping impacts peer ASes more harshly due to flap amplification**
- **Recommendations for ISPs**
  - **http://www.ripe.net/docs/ripe-229.html**
  - **(work by European and US ISPs a few years ago as vendor defaults were considered to be too aggressive)**

Route Reflectors

Scaling iBGP mesh

Avoid ½n(n-1) iBGP mesh

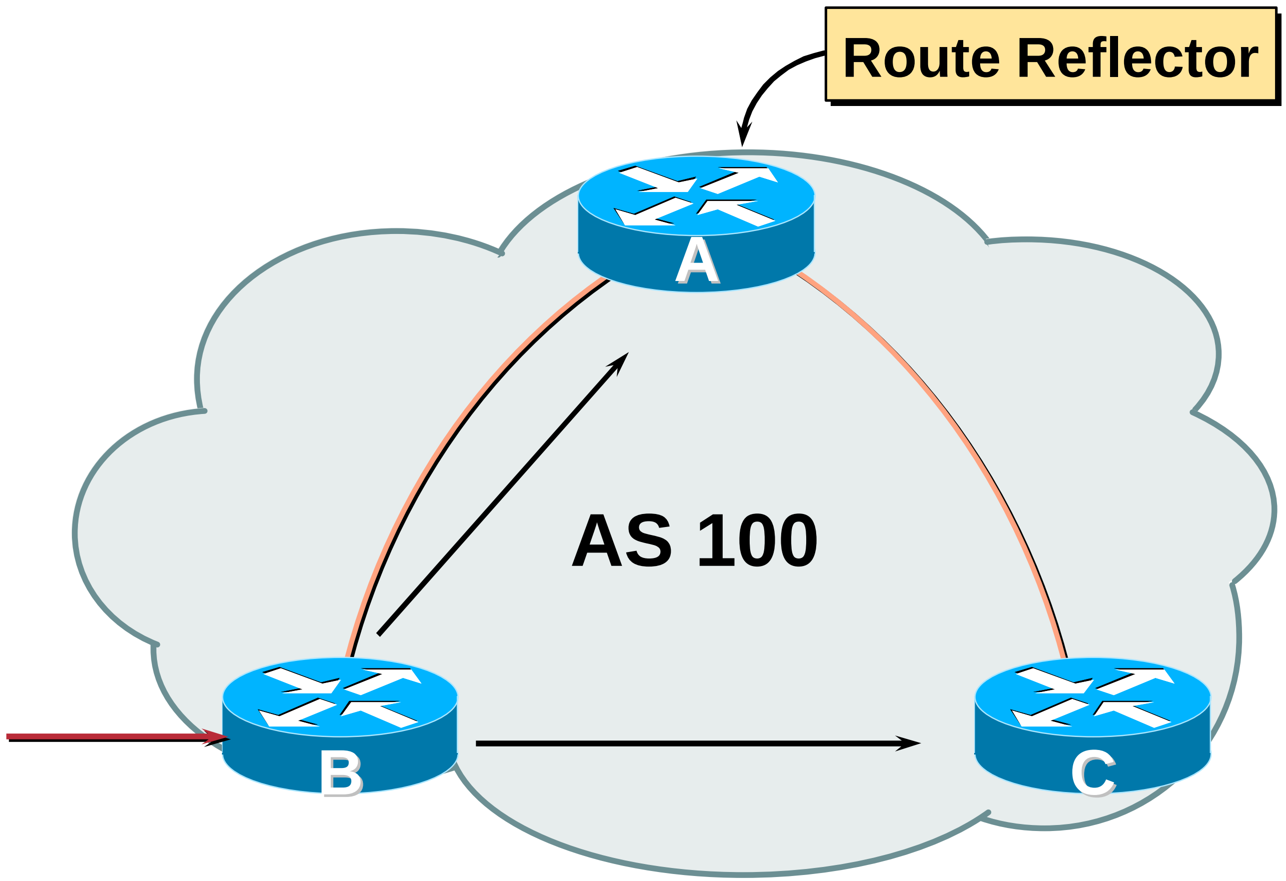
n=1000 Þ nearly half a million ibgp sessions!



Two solutions

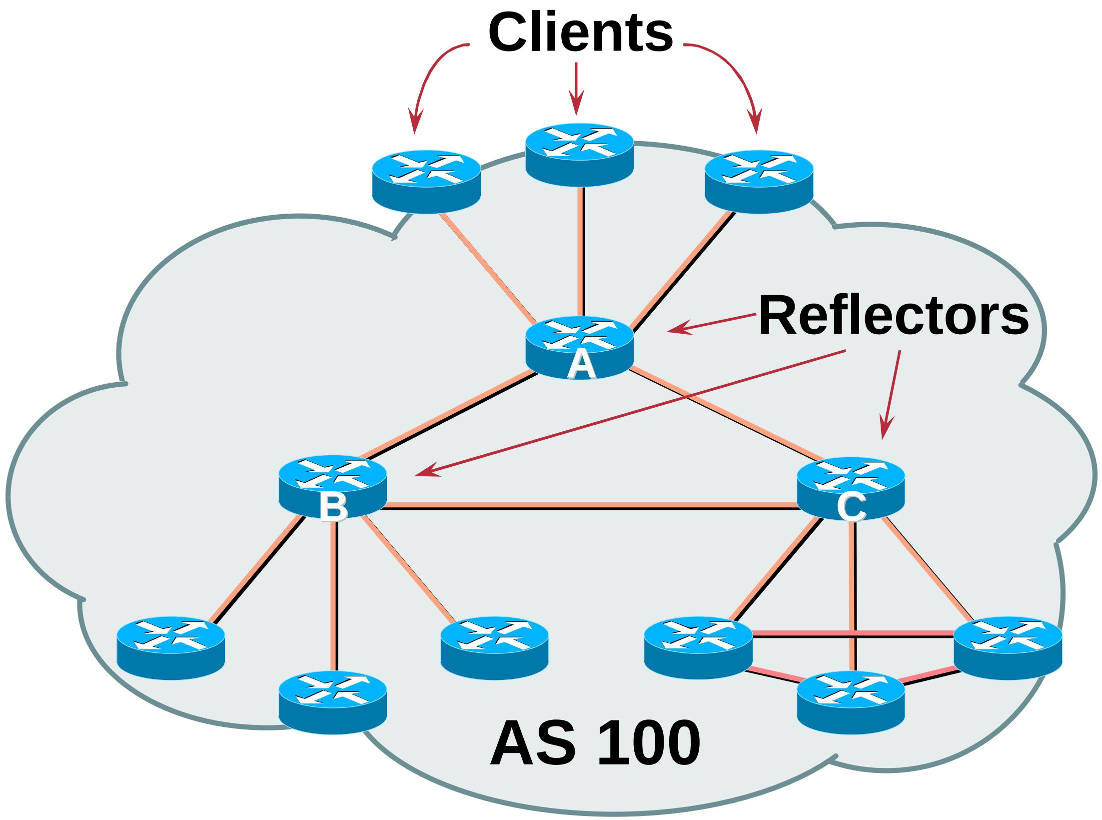
- Route reflector – simpler to deploy and run
- Confederation – more complex, corner case benefits

Route Reflector: Principle



Route Reflector

- Reflector receives path from clients and non-clients
- Selects best path
- If best path is from client, reflect to other clients and non-clients
- If best path is from non-client, reflect to clients only
- Non-meshed clients
- Described in RFC2796



Route Reflector Topology

- Divide the backbone into multiple clusters
- At least one route reflector and few clients per cluster
- Route reflectors are fully meshed
- Clients in a cluster could be fully meshed
- Single IGP to carry next hop and local routes

Route Reflectors: Loop Avoidance

- **Originator_ID attribute**

  Carries the RID of the originator of the route in the local AS (created by the RR)

- **Cluster_list attribute**

  The local cluster-id is added when the update is sent by the RR

  Best to set cluster-id is from router-id (address of loopback)

  (Some ISPs use their own cluster-id assignment strategy – but needs to be well documented!)

Route Reflectors: Redundancy

- **Multiple RRs can be configured in the same cluster – not advised!**

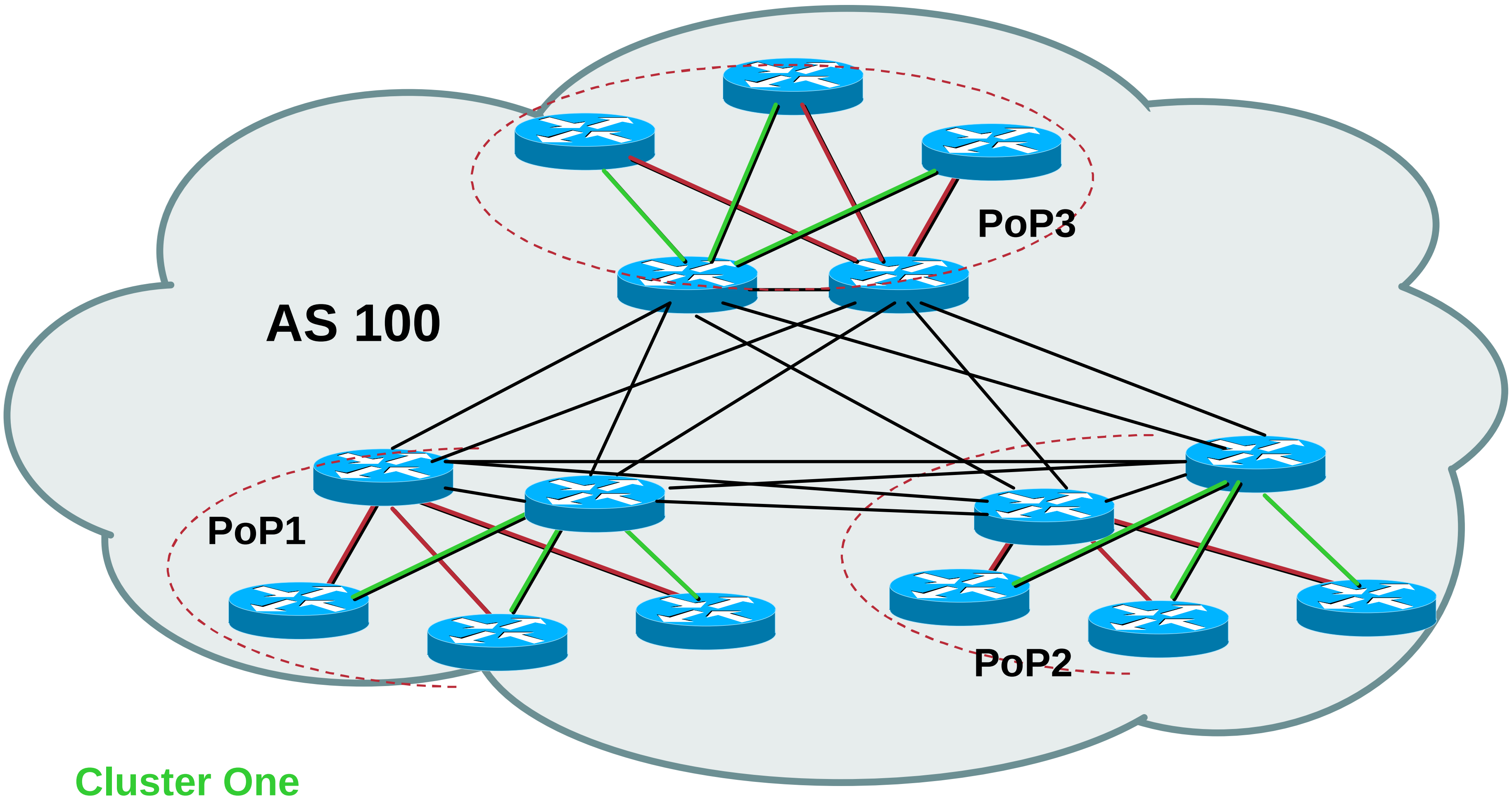
  All RRs in the cluster must have the same cluster-id (otherwise it is a different cluster)

- **A router may be a client of RRs in different clusters**

  Common today in ISP networks to overlay two clusters – redundancy achieved that way

  ® **Each client has two RRs = redundancy**

Route Reflectors: Redundancy



Cluster One

Cluster Two

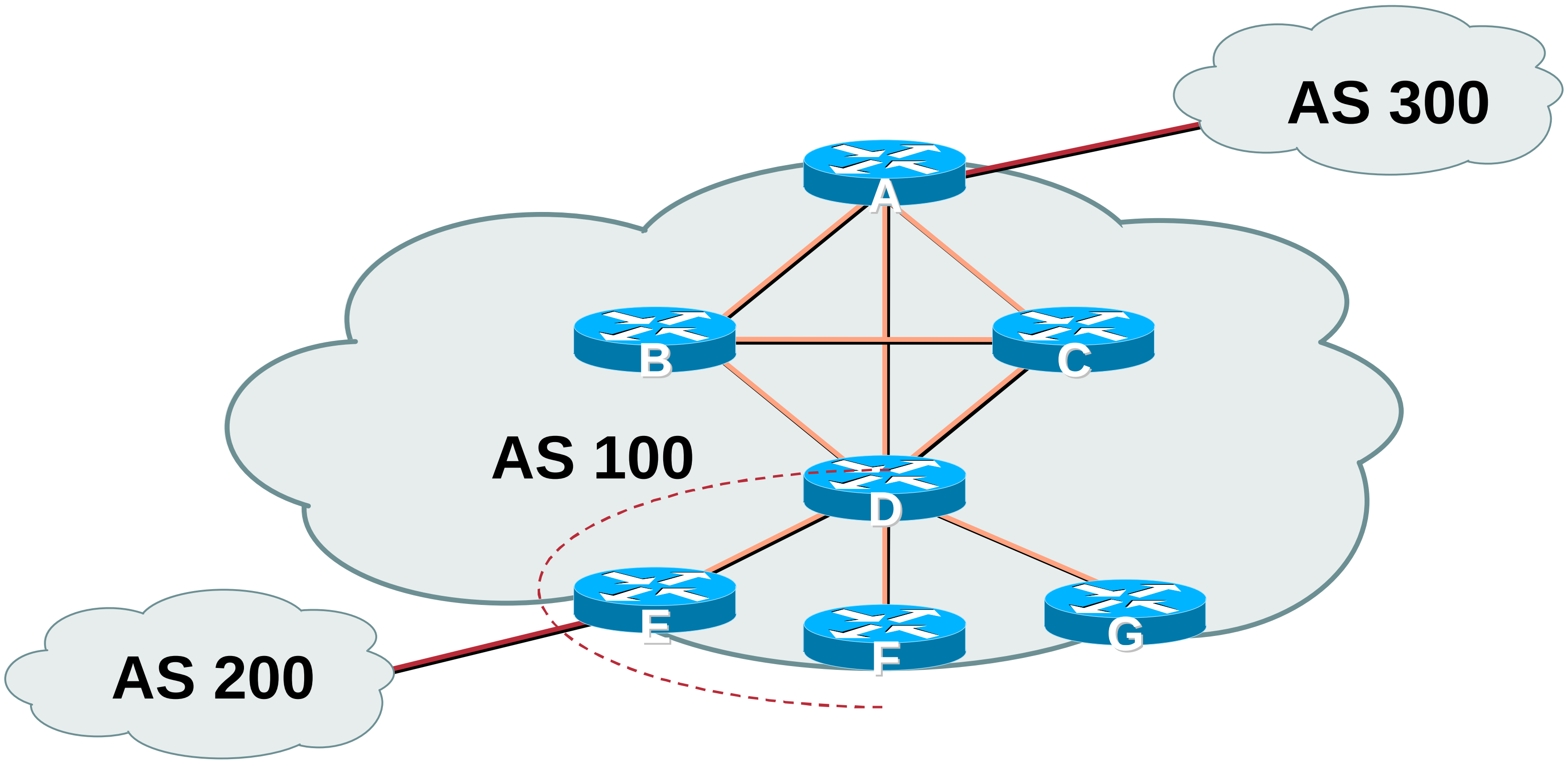
Route Reflectors: Migration

- **Where to place the route reflectors?**
  - **Always follow the physical topology!**
  - **This will guarantee that the packet forwarding won't be affected**
- **Typical ISP network:**
  - **PoP has two core routers**
  - **Core routers are RR for the PoP**
  - **Two overlaid clusters**

Route Reflectors: Migration

- **Typical ISP network:**
  - **Core routers have fully meshed iBGP**
  - **Create further hierarchy if core mesh too big**
    - **Split backbone into regions**
- **Configure one cluster pair at a time**
  - **Eliminate redundant iBGP sessions**
  - **Place maximum of one RR per cluster**
  - **Easy migration, multiple levels**

Route Reflector: Migration



- Migrate small parts of the network, one part at a time.

BGP Scaling Techniques

- Route Refresh

  Use should be mandatory

- Route flap damping

  Only use if you understand why

- Route Reflectors

  The way to scale the iBGP mesh

BGP for Internet Service Providers

- **BGP Basics**
- **Scaling BGP**
- **Using Communities**
- **Deploying BGP in an ISP network**

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

- **One simple example follows 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:**

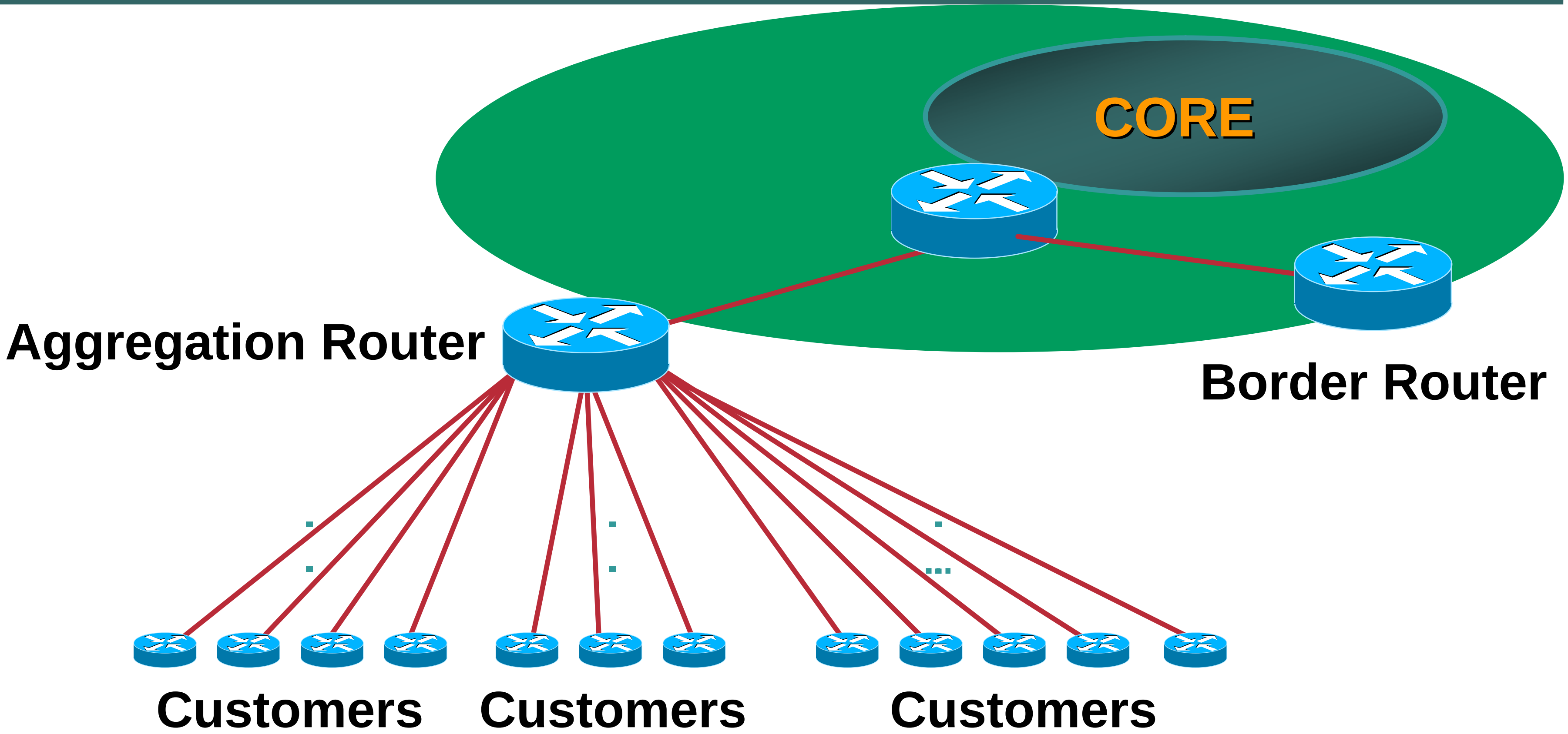
  IXP connection: community 100:2100

  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

- **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!
- **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: Sprintlink

http://www.sprintlink.net/policy/bgp.html

WHAT YOU CAN CONTROL

AS-PATH PREPENDS

Sprint allows customers to use AS-path prepending to adjust route preference on the network. Such prepending will be received and passed on properly without notifiying Sprint of your change in announcments.

Additionally, Sprint will prepend AS1239 to eBGP sessions with certain autonomous systems depending on a received community. Currently, the following ASes are supported: 1668, 209, 2914, 3300, 3356, 3549, 3561, 4635, 701, 7018, 702 and 8220.

String	Resulting AS Path to ASXXX
65000:XXX	Do not advertise to ASXXX
65001:XXX	1239 (default) ...
65002:XXX	1239 1239 ...
65003:XXX	1239 1239 1239 ...
65004:XXX	1239 1239 1239 1239 ...

String	Resulting AS Path to ASXXX in Asia
65070:XXX	Do not advertise to ASXXX
65071:XXX	1239 (default) ...
65072:XXX	1239 1239 ...
65073:XXX	1239 1239 1239 ...
65074:XXX	1239 1239 1239 1239 ...

String	Resulting AS Path to ASXXX in Europe
65050:XXX	Do not advertise to ASXXX
65051:XXX	1239 (default) ...
65052:XXX	1239 1239 ...
65053:XXX	1239 1239 1239 ...
65054:XXX	1239 1239 1239 1239 ...

String	Resulting AS Path to ASXXX in North America
65010:XXX	Do not advertise to ASXXX
65011:XXX	1239 (default) ...
65012:XXX	1239 1239 ...
65013:XXX	1239 1239 1239 ...
65014:XXX	1239 1239 1239 1239 ...

String	Resulting AS Path to all supported ASes
65000:0	Do not advertise
65001:0	1239 (default) ...
65002:0	1239 1239 ...
65003:0	1239 1239 1239

More info at
www.sprintlink.net/policy/bgp.html

Some ISP Examples
MCI Europe

- **Permits customers to send communities which determine**
  - **local preferences within MCI’s network**
  - **Reachability of the prefix**
  - **How the prefix is announced outside of MCI’s network**

Some ISP Examples
MCI Europe

```
aut-num: AS702
descr:   MCI EMEA - Commercial IP service provider in Europe
remarks: MCI uses the following communities with its customers:
         702:80    Set Local Pref 80 within AS702
         702:120   Set Local Pref 120 within AS702
         702:20    Announce only to MCI AS'es and MCI customers
         702:30    Keep within Europe, don't announce to other MCI AS's
         702:1     Prepend AS702 once at edges of MCI to Peers
         702:2     Prepend AS702 twice at edges of MCI to Peers
         702:3     Prepend AS702 thrice at edges of MCI to Peers
         Advanced communities for customers
         702:7020  Do not announce to AS702 peers with a scope of
                   National but advertise to Global Peers, European
                   Peers and MCI customers.
         702:7001  Prepend AS702 once at edges of MCI to AS702
                   peers with a scope of National.
         702:7002  Prepend AS702 twice at edges of MCI to AS702
                   peers with a scope of  National.
(more)
```

Some ISP Examples
MCI Europe

```
(more)
          702:7003 Prepend AS702 thrice at edges of MCI to AS702
                   peers with a scope  of National.
          702:8020 Do not announce to AS702 peers with a scope of
                   European but advertise to Global Peers, National
                   Peers and MCI  customers.
          702:8001 Prepend AS702 once at edges of MCI to AS702
                   peers with a scope of European.
          702:8002 Prepend AS702 twice at edges of MCI to AS702
                   peers with a scope of  European.
          702:8003 Prepend AS702 thrice at edges of MCI to AS702
                   peers with a scope  of European.
          --------------------------------------------------------------
          Additional details of the MCI communities are located at:
          http://global.mci.com/uk/customer/bgp/
          --------------------------------------------------------------
mnt-by:   WCOM-EMEA-RICE-MNT
changed:  rice@lists.mci.com 20040523
source:   RIPE
```

Some ISP Examples
BT

- **One of the most comprehensive community lists around**

  whois –h whois.ripe.net AS5400 reveals all

- **Extensive community definitions allow sophisticated traffic engineering by customers**

Some ISP Examples
BT Ignite

```
aut-num:      AS5400
descr:        BT Ignite European Backbone
remarks:
remarks:      Community to                         Community to
remarks:      Not announce       To peer:          AS prepend 5400
remarks:
remarks:      5400:1000 All peers & Transits       5400:2000
remarks:
remarks:      5400:1500 All Transits               5400:2500
remarks:      5400:1501 Sprint Transit (AS1239)    5400:2501
remarks:      5400:1502 SAVVIS Transit (AS3561)    5400:2502
remarks:      5400:1503 Level 3 Transit (AS3356)   5400:2503
remarks:      5400:1504 AT&T Transit (AS7018)      5400:2504
remarks:      5400:1505 UUnet Transit (AS701)      5400:2505
remarks:
remarks:      5400:1001 Nexica (AS24592)           5400:2001
remarks:      5400:1002 Fujitsu (AS3324)           5400:2002
remarks:      5400:1003 Unisource (AS3300)         5400:2003
<snip>
notify:       notify@eu.bt.net
mnt-by:       CIP-MNT
source:       RIPE
```

And many many more!

BGP for Internet Service Providers

- **BGP Basics**
- **Scaling BGP**
- **Using Communities**
- **Deploying BGP in an ISP network**

Deploying BGP in an ISP Network

Okay, so we've learned all about BGP now; how do we use it on our network??

Deploying BGP

- The role of IGPs and iBGP
- Aggregation
- Receiving Prefixes
- Configuration Tips

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

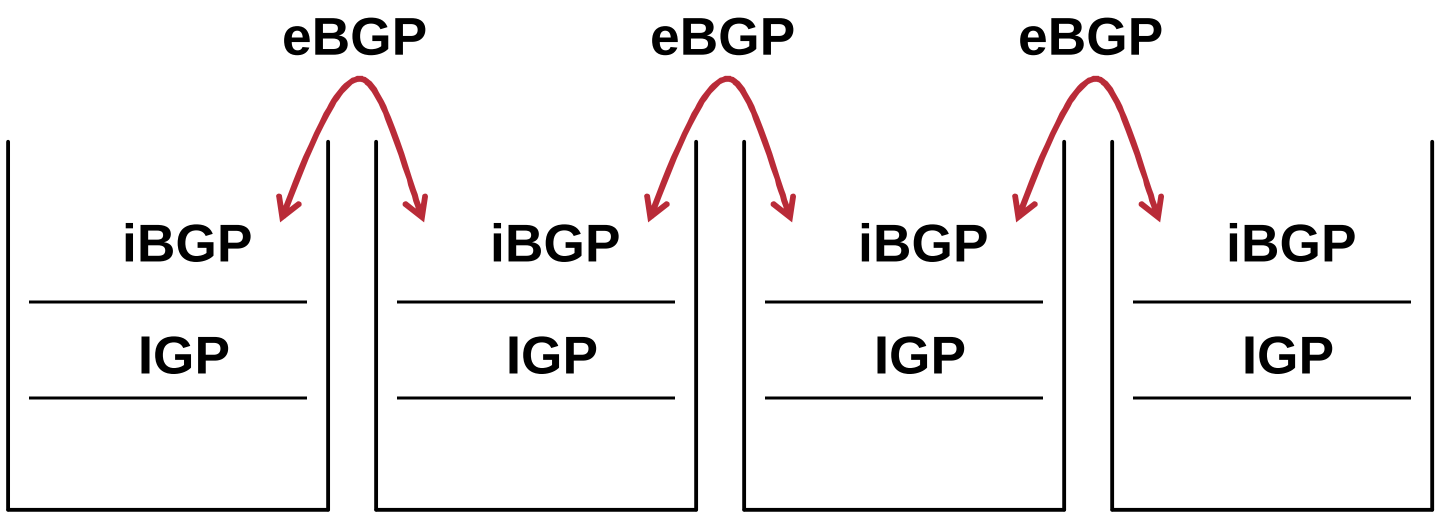
  design goal is to minimise number of prefixes in IGP to aid scalability and rapid convergence

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

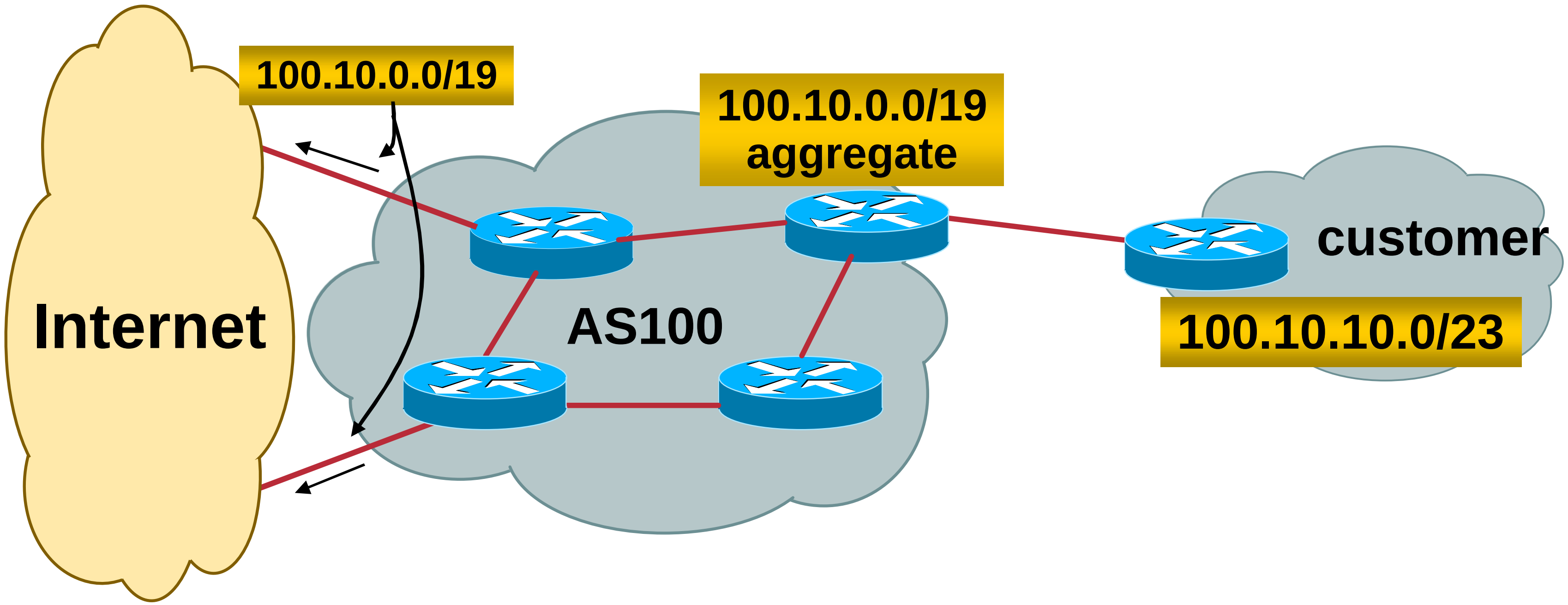
- **Address block should be announced to the Internet as an aggregate**
- **Subprefixes of address block should NOT be announced to Internet unless special circumstances (more later)**
- **Aggregate should be generated internally**

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 /22 depending on RIR
  - Different sizes for different address blocks
- No real reason to see anything longer than a /22 prefix in the Internet
  - BUT there are currently >87000 /24s!

Aggregation – Example

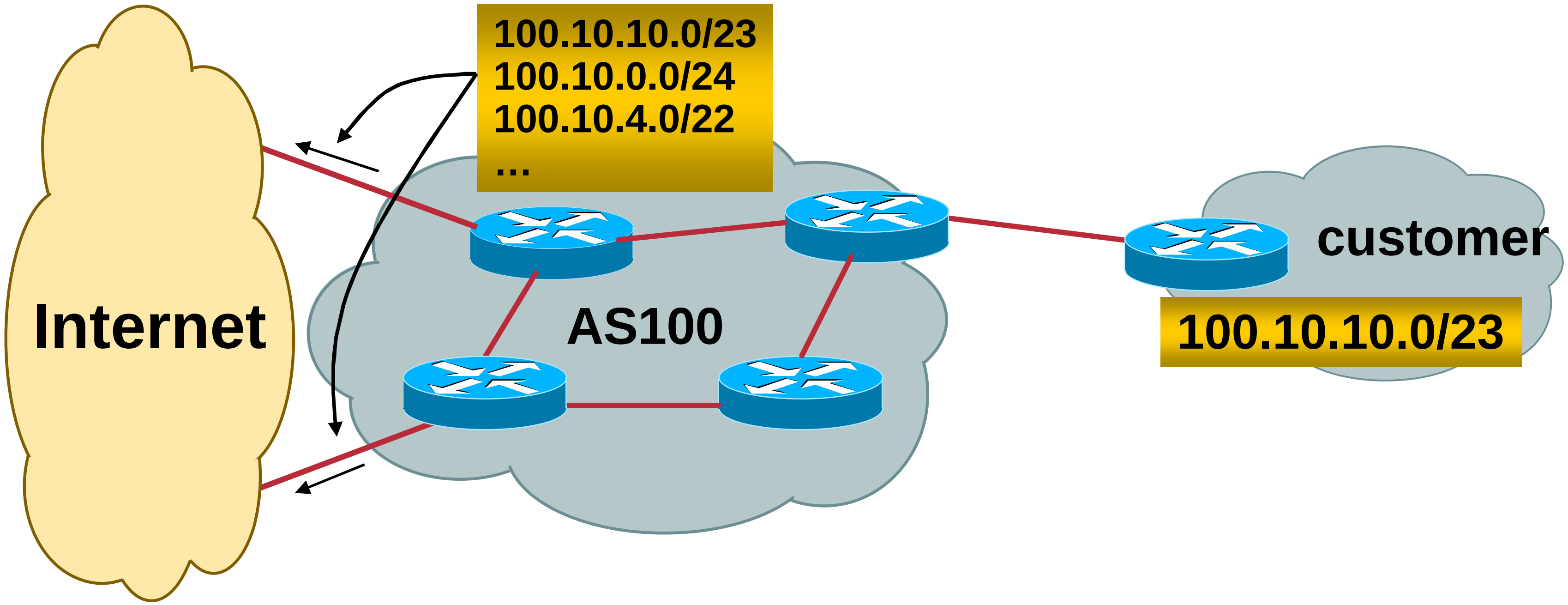


- **Customer has /23 network assigned from AS100's /19 address block**
- **AS100 announced /19 aggregate to the Internet**

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 – 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 – 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?**

The Internet Today (May 2005)

- **Current Internet Routing Table Statistics**

BGP Routing Table Entries	162009
Prefixes after maximum aggregation	94157
Unique prefixes in Internet	78129
Prefixes smaller than registry alloc	75990
/24s announced	88342
only 5702 /24s are from 192.0.0.0/8	
ASes in use	19627

"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 – May 1999

RIR blocks contribute 50891 prefixes or 86% of the Internet Routing Table

Block	Networks	Block	Networks	Block	Networks	Block	Networks
24/8	24	80/8	0	192/8	6287	208/8	2995
		81/8	0	193/8	2439	209/8	3083
60/8	0	82/8	0	194/8	2921	210/8	700
61/8	2	83/8	0	195/8	1429	211/8	0
62/8	100	84/8	0	196/8	550	212/8	840
63/8	78	85/8	0	197/8	0	213/8	1
64/8	0	86/8	0	198/8	4015	214/8	2
65/8	0	87/8	0	199/8	3503	215/8	4
66/8	0	88/8	0	200/8	1459	216/8	1218
67/8	0			201/8	0	217/8	0
68/8	0	124/8	0	202/8	2398	218/8	0
69/8	0	125/8	0	203/8	3782	219/8	0
70/8	0	126/8	0	204/8	3936	220/8	0
71/8	1			205/8	2694	221/8	1
72/8	0			206/8	3421	222/8	0
73/8	0			207/8	3014	223/8	0

“The New Swamp” RIR Space – May 2005

RIR blocks contribute 142575 prefixes or 88% of the Internet Routing Table

Block	Networks	Block	Networks	Block	Networks	Block	Networks
24/8	2989	80/8	1522	192/8	6867	208/8	3297
		81/8	1258	193/8	4818	209/8	5124
60/8	249	82/8	1158	194/8	3776	210/8	3622
61/8	2419	83/8	1653	195/8	3168	211/8	2355
62/8	1628	84/8	642	196/8	965	212/8	2731
63/8	2782	85/8	806	197/8	0	213/8	2776
64/8	4768	86/8	35	198/8	4977	214/8	316
65/8	3586	87/8	2	199/8	4184	215/8	356
66/8	5953	88/8	2	200/8	6445	216/8	6217
67/8	1731			201/8	654	217/8	2495
68/8	2474	124/8	0	202/8	8659	218/8	1114
69/8	2632	125/8	0	203/8	8629	219/8	948
70/8	869	126/8	7	204/8	5252	220/8	1273
71/8	250			205/8	2834	221/8	618
72/8	479			206/8	3990	222/8	731
73/8	0			207/8	4162	223/8	0

"The New Swamp" Summary

- **RIR space shows creeping deaggregation**
  - **It seems that an RIR /8 block averages around 4000 prefixes once fully allocated**
  - **So their existing 58 /8s will eventually result in 232000 prefix announcements**
- **Food for thought:**
  - **Remaining 74 unallocated /8s and the 58 RIR /8s combined will cause:**
  - **528000 prefixes with density of 4000 prefixes per /8**
  - **Plus ~10% due to "non RIR space deaggregation"**

"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

BGP Report (bgp.potaroo.net)

- **157000 total announcements**
- **108000 prefixes**

  After aggregating including full AS PATH info

  i.e. including each ASN's traffic engineering

  33% saving possible

- **93000 prefixes**

  After aggregating by Origin AS

  i.e. ignoring each ASN's traffic engineering

  10% saving possible

The excuses

- **Traffic engineering causes 10% of the Internet Routing table**
- **Deliberate deaggregation causes 33% of the Internet Routing table**

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

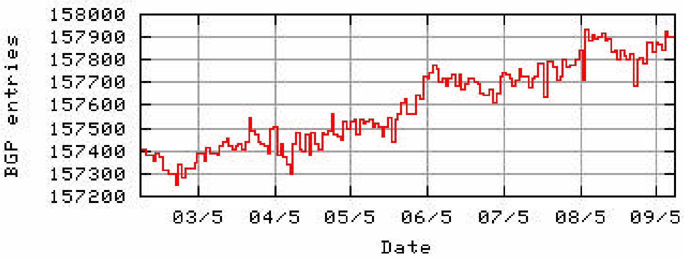
    Very effectively challenges the traffic engineering excuse

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

Status Summary

Table History

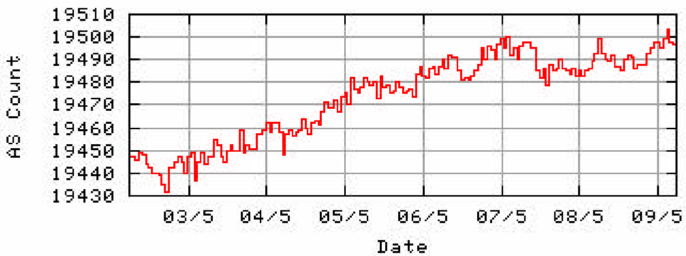
Date	Prefixes	CIDR Aggregated
02-05-05	157356	108023
03-05-05	157392	108044
04-05-05	157505	108133
05-05-05	157530	108201
06-05-05	157716	108341
07-05-05	157747	108272
08-05-05	157845	108355
09-05-05	157874	108388



Plot: BGP Table Size

AS Summary

19498	Number of ASes in routing system
7996	Number of ASes announcing only one prefix
1467	Largest number of prefixes announced by an AS AS7018: ATT-INTERNET4 - AT&T WorldNet Services
90497280	Largest address span announced by an AS (/32s) AS721: DLA-ASNBLOCK-AS - DoD Network Information Center



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

Aggregation Summary

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

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').

--- 09May05 ---

ASnum	NetsNow	NetsAggr	NetGain	% Gain	Description
Table	157925	108381	49544	31.4%	All ASes
AS4323	1098	223	875	79.7%	TWTC - Time Warner Telecom
AS18566	805	8	797	99.0%	COVAD - Covad Communications
AS4134	893	220	673	75.4%	CHINANET-BACKBONE No.31,Jin-rong Street
AS721	1117	564	553	49.5%	DLA-ASNBLOCK-AS - DoD Network Information Center
AS7018	1467	939	528	36.0%	ATT-INTERNET4 - AT&T WorldNet Services
AS27364	539	22	517	95.9%	ACS-INTERNET - Armstrong Cable Services
AS22773	483	23	460	95.2%	CCINET-2 - Cox Communications Inc.
AS6197	900	506	394	43.8%	BATI-ATL - BellSouth Network Solutions, Inc
AS3602	509	146	363	71.3%	SPRINT-CA-AS - Sprint Canada Inc.
AS17676	431	78	353	81.9%	JPNIC-JP-ASN-BLOCK Japan Network Information Center
AS9929	350	46	304	86.9%	CNCNET-CN China Netcom Corp.
AS4766	574	279	295	51.4%	KIXS-AS-KR Korea Telecom
AS6478	416	123	293	70.4%	ATT-INTERNET3 - AT&T WorldNet Services
AS6140	399	135	264	66.2%	IMPSAT-USA - ImpSat
AS14654	264	6	258	97.7%	WAYPORT - Wayport
AS9583	735	483	252	34.3%	SIFY-AS-IN Sify Limited
AS9443	374	123	251	67.1%	INTERNETPRIMUS-AS-AP Primus Telecommunications
AS7545	493	247	246	49.9%	TPG-INTERNET-AP TPG Internet Pty Ltd
AS1239	886	644	242	27.3%	SPRINTLINK - Sprint
AS15270	272	37	235	86.4%	AS-PAETEC-NET - PaeTec.net -a division of PaeTecCommunications, Inc.
AS23126	254	23	231	90.9%	KMCTELCOM-DIA - KMC Telecom, Inc.
AS4755	516	287	229	44.4%	VSNL-AS Videsh Sanchar Nigam Ltd. Autonomous System
AS7725	415	186	229	55.2%	CCH-AS7 - Comcast Cable Communications Holdings, Inc
AS6198	464	236	228	49.1%	BATI-MIA - BellSouth Network Solutions, Inc
AS5668	488	264	224	45.9%	AS-5668 - CenturyTel Internet Holdings, Inc.
AS2386	853	634	219	25.7%	INS-AS - AT&T Data Communications Services
AS9498	296	79	217	73.3%	BBIL-AP BHARTI BT INTERNET LTD.
AS11456	319	110	209	65.5%	NUVOX - NuVox Communications, Inc.
AS6167	264	67	197	74.6%	CELLCO-PART - Cellco Partnership
AS6517	319	128	191	59.9%	YIPESCOM - Yipes Communications, Inc.
Total	17193	6866	10327	60.1%	Top 30 total

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

Top 20 Added Routes this week per Originating AS

Prefixes	ASnum	AS Description
154	AS7725	CCH-AS7 - Comcast Cable Communications Holdings, Inc
108	AS4755	VSNL-AS Videsh Sanchar Nigam Ltd. Autonomous System
52	AS35911	BNQ-1 - Telebec
36	AS13645	BROADBANDONE - BroadbandONE, Inc.
19	AS17488	HATHWAY-NET-AP Hathway IP Over Cable Internet
16	AS9576	SOOKMYUNG-AS SOOKMYUNG WOMEN'S UNIVERSITY
16	AS174	COGENT Cogent/PSI
16	AS18633	GIANTWEB - Giant Technologies Inc.
16	AS18042	KBT Koos Broadband Telecom
16	AS32613	IWEB-AS - Groupe iWeb Technologies inc.
15	AS19632	Metropolis Intercom
15	AS30340	AS-LLIX - Liberty Lake Internet Portal
13	AS19916	ASTRUM-0001 - OLM LLC
13	AS22047	VTR BANDA ANCHA S.A.
13	AS21882	PRIORITYNETWORKS - Priority Networks Inc.
12	AS9940	WOLCST-AS-AP World online AS, Cybersoft Technologies.
12	AS12715	JAZZNET Jazz Telecom S.A.
12	AS22927	Telefonica de Argentina
11	AS30533	CONNEXION-BY-BOEING-LTN - Connexion by Boeing
11	AS25454	TELEMEDIAAS Telemedia SA Autonomous System

Top 20 Withdrawn Routes this week per Originating AS

Prefixes	ASnum	AS Description
-45	AS10970	LH - Lighthouse Communications, Inc.
-33	AS7496	WEBCENTRAL-AS WebCentral
-31	AS8921	I-CONNEXION ICX Autonomous System
-23	AS4513	Globix Corporation
-20	AS1239	SPRINTLINK - Sprint
-18	AS14103	ACDNET-ASN1 - ACD.net
-17	AS29257	CBB-IE-AS Connexion by Boeing Ireland, Ltd.
-16	AS20115	CHARTER-NET-HKY-NC - Charter Communications
-16	AS6167	CELLCO-PART - Cellco Partnership
-15	AS17557	PKTELECOM-AS-AP Pakistan Telecom
-14	AS9152	MEGADAT Autonomous System
-14	AS16154	TELECOMS-AS Telecoms-Net Ltd.
-14	AS24219	NFI-AS-AP No Fuss Internet
-13	AS174	COGENT Cogent/PSI
-13	AS10125	DACCESS-AP DATA ACCESS INDIA LIMITED
-13	AS30857	TAURUS-AS Taurus Telecom PJSC
-12	AS17854	CABLELINE-AS-KR BANDOCABLELINE
-12	AS7049	S&M International S.A.
-12	AS4323	TWTC - Time Warner Telecom
-12	AS3561	SAVVIS - Savvis

Adds and Wdls per Prefix Length

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

Report: Announced Route count per Originating AS
Report: Withdrawn Route count per Originating AS

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
1103	1467	AS7018	ATT-INTERNET4 - AT&T WorldNet Services
1012	1180	AS174	COGENT Cogent/PSI
974	1098	AS4323	TWTC - Time Warner Telecom
880	900	AS6197	BATI-ATL - BellSouth Network Solutions, Inc
801	1117	AS721	DLA-ASNBLOCK-AS - DoD Network Information Center
798	805	AS18566	COVAD - Covad Communications
780	853	AS2386	INS-AS - AT&T Data Communications Services
742	893	AS4134	CHINANET-BACKBONE No.31,Jin-rong Street
730	735	AS9583	SIFY-AS-IN Sify Limited
621	886	AS1239	SPRINTLINK - Sprint
594	994	AS701	ALTERNET-AS - UUNET Technologies, Inc.
583	595	AS20115	CHARTER-NET-HKY-NC - Charter Communications
540	574	AS4766	KIXS-AS-KR Korea Telecom
533	539	AS27364	ACS-INTERNET - Armstrong Cable Services
500	516	AS4755	VSNL-AS Videsh Sanchar Nigam Ltd. Autonomous System
475	488	AS5668	AS-5668 - CenturyTel Internet Holdings, Inc.
470	483	AS22773	CCINET-2 - Cox Communications Inc.
456	493	AS7545	TPG-INTERNET-AP TPG Internet Pty Ltd
453	509	AS3602	SPRINT-CA-AS - Sprint Canada Inc.
452	464	AS6198	BATI-MIA - BellSouth Network Solutions, Inc

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

http://www.cidr-report.org/cgi-bin/as-report?as=AS1239&view=4637

Rank	AS	Type	Originate Addr Space (pfx)		Transit Addr space (pfx)		Description
24	AS1239	ORG+TRN	Originate:	11982080 /8.49	Transit:	145498112 /4.88	SPRINTLINK - Sprint

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	%
20	AS1239	SPRINTLINK - Sprint	886	307	65	644	242	27.31%

AS 1239: SPRINTLINK - Sprint

Prefix	(AS Path)	Aggregation Action
12.9.182.0/23	4637 1239	
12.22.206.0/24	4637 1239	
24.56.144.0/21	4637 1239	
24.137.128.0/21	4637 1239	
24.221.0.0/17	4637 1239	+ Announce - aggregate of 24.221.0.0/18 (4637 1239) and 24.221.64.0/18 (4637 1239)
24.221.0.0/18	4637 1239	- Withdrawn - aggregated with 24.221.64.0/18 (4637 1239)
24.221.64.0/19	4637 1239	- Withdrawn - aggregated with 24.221.96.0/19 (4637 1239)
24.221.96.0/19	4637 1239	- Withdrawn - aggregated with 24.221.64.0/19 (4637 1239)
24.221.128.0/18	4637 1239	+ Announce - aggregate of 24.221.128.0/19 (4637 1239) and 24.221.160.0/19 (4637 1239)
24.221.128.0/19	4637 1239	- Withdrawn - aggregated with 24.221.160.0/19 (4637 1239)
24.221.160.0/19	4637 1239	- Withdrawn - aggregated with 24.221.128.0/19 (4637 1239)
24.221.192.0/20	4637 1239	
24.221.220.0/22	4637 1239	
24.221.224.0/20	4637 1239	+ Announce - aggregate of 24.221.224.0/21 (4637 1239) and 24.221.232.0/21 (4637 1239)
24.221.224.0/21	4637 1239	- Withdrawn - aggregated with 24.221.232.0/21 (4637 1239)
24.221.232.0/22	4637 1239	- Withdrawn - aggregated with 24.221.236.0/22 (4637 1239)
24.221.236.0/22	4637 1239	- Withdrawn - aggregated with 24.221.232.0/22 (4637 1239)
24.221.242.0/23	4637 1239	
24.221.244.0/22	4637 1239	
24.221.248.0/21	4637 1239	
38.113.4.0/24	4637 1239	
63.90.4.0/24	4637 1239	
63.113.210.0/24	4637 1239	
63.122.77.0/24	4637 1239	
63.122.78.0/23	4637 1239	
63.134.0.0/17	4637 1239	
63.160.0.0/12	4637 1239	
63.178.251.0/24	4637 1239	
63.237.89.0/24	4637 1239	
64.6.224.0/19	4637 1239	
64.9.45.0/24	4637 1239	
64.9.86.0/24	4637 1239	
64.17.64.0/22	4637 1239	

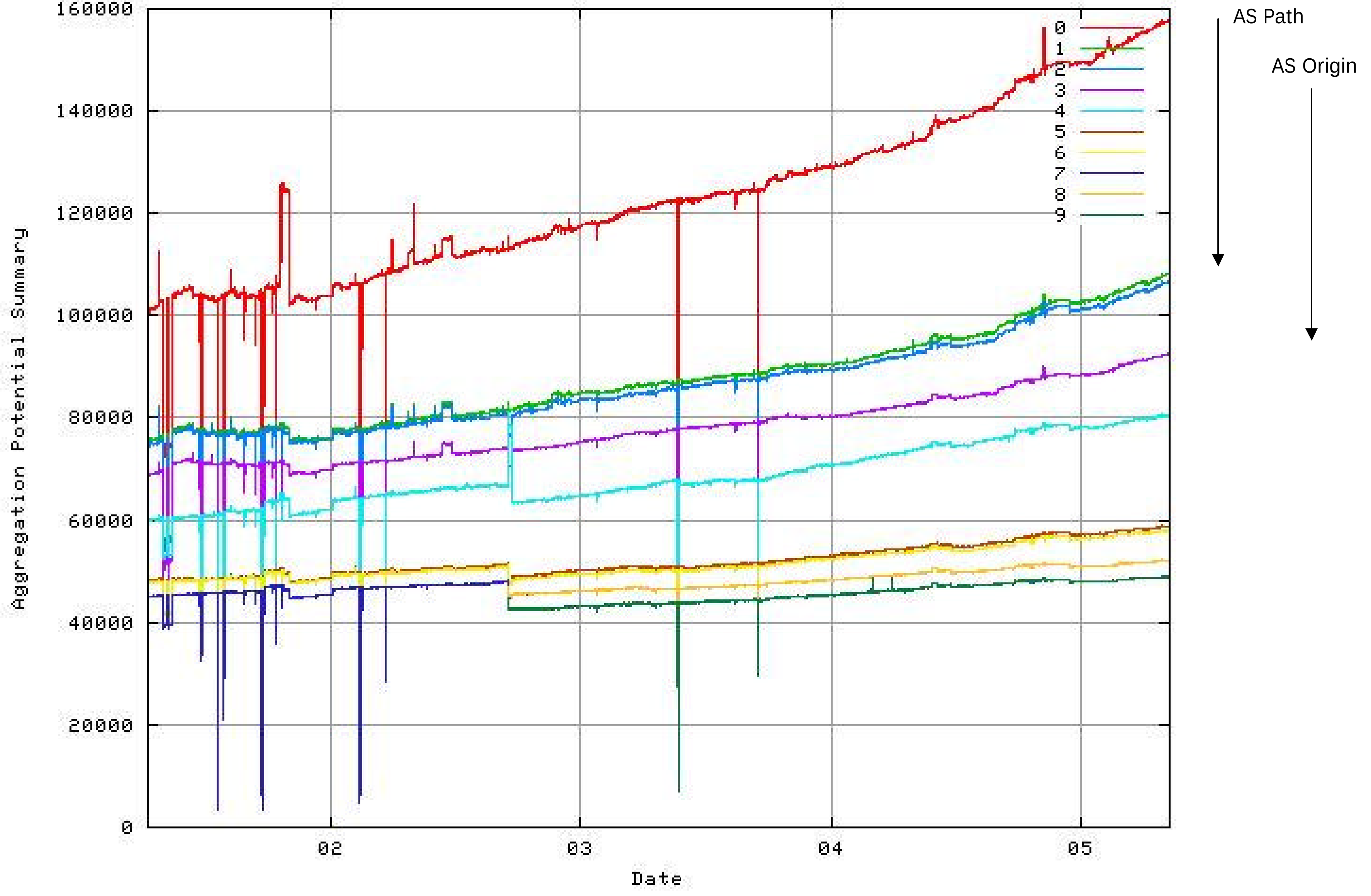
http://www.cidr-report.org/cgi-bin/as-report?as=AS701&view=4637

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
49	AS701	ALTERNET-AS - UUNET Technologies, Inc.	994	208	68	854	140	14.08%

AS 701: ALTERNET-AS - UUNET Technologies, Inc.

Prefix	(AS Path)	Aggregation Action
17.255.232.0/24	4637 701	
24.32.66.0/24	4637 701	
24.32.68.0/22	4637 701	+ Announce - aggregate of 24.32.68.0/23 (4637 701) and 24.32.70.0/23 (4637 701)
24.32.68.0/24	4637 701	- Withdrawn - aggregated with 24.32.69.0/24 (4637 701)
24.32.69.0/24	4637 701	- Withdrawn - aggregated with 24.32.68.0/24 (4637 701)
24.32.70.0/24	4637 701	- Withdrawn - aggregated with 24.32.71.0/24 (4637 701)
24.32.71.0/24	4637 701	- Withdrawn - aggregated with 24.32.70.0/24 (4637 701)
24.32.130.0/24	4637 701	
24.32.144.0/22	4637 701	+ Announce - aggregate of 24.32.144.0/23 (4637 701) and 24.32.146.0/23 (4637 701)
24.32.144.0/23	4637 701	- Withdrawn - aggregated with 24.32.146.0/23 (4637 701)
24.32.146.0/23	4637 701	- Withdrawn - aggregated with 24.32.144.0/23 (4637 701)
24.32.163.0/24	4637 701	
24.32.164.0/24	4637 701	
24.206.172.0/24	4637 701	
24.216.0.0/16	4637 701	
24.216.82.0/24	4637 701	- Withdrawn - matching aggregate 24.216.0.0/16 4637 701
24.216.94.0/23	4637 701	- Withdrawn - matching aggregate 24.216.0.0/16 4637 701
24.216.174.0/24	4637 701	
24.240.0.0/15	4637 701	
55.191.7.0/24	4637 701	
62.70.23.0/24	4637 701	
63.0.0.0/9	4637 701	+ Announce - aggregate of 63.0.0.0/10 (4637 701) and 63.64.0.0/10 (4637 701)
63.0.0.0/12	4637 701	- Withdrawn - aggregated with 63.16.0.0/12 (4637 701)
63.16.0.0/12	4637 701	- Withdrawn - aggregated with 63.0.0.0/12 (4637 701)
63.32.0.0/12	4637 701	- Withdrawn - aggregated with 63.48.0.0/12 (4637 701)
63.48.0.0/12	4637 701	- Withdrawn - aggregated with 63.32.0.0/12 (4637 701)
63.64.0.0/12	4637 701	- Withdrawn - aggregated with 63.80.0.0/12 (4637 701)
63.80.0.0/12	4637 701	- Withdrawn - aggregated with 63.64.0.0/12 (4637 701)
63.96.0.0/12	4637 701	- Withdrawn - aggregated with 63.112.0.0/12 (4637 701)
63.112.0.0/12	4637 701	- Withdrawn - aggregated with 63.96.0.0/12 (4637 701)
63.134.153.0/24	4637 701	
63.134.154.0/24	4637 701	
63.134.161.0/24	4637 701	
63.134.162.0/23	4637 701	+ Announce - aggregate of 63.134.162.0/24 (4637 701) and 63.134.163.0/24 (4637 701)
63.134.162.0/24	4637 701	- Withdrawn - aggregated with 63.134.163.0/24 (4637 701)
63.134.163.0/24	4637 701	- Withdrawn - aggregated with 63.134.162.0/24 (4637 701)
63.134.164.0/24	4637 701	
63.134.168.0/23	4637 701	
63.134.176.0/24	4637 701	
63.134.179.0/24	4637 701	
63.141.42.0/24	4637 701	

Aggregation Potential (source: bgp.potaroo.net/as4637/)



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 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 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**
  - **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

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

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 templates, passwords, tricks, and more 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!**

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 unnumbered interfaces**
  - **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**
- **eBGP and iBGP examples follow**
  - **Also see Project Cymru's BGP templates**
    - **www.cymru.com/Documents**

iBGP Template
Example

- **iBGP between loopbacks!**
  - **So IGP can do intelligent re-route**
- **Next-hop-self**
  - **Keep DMZ and external point-to-point out of IGP**
- **Always send community attribute for 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**

eBGP Template
Example

- **BGP damping**
  - Do NOT use it unless you understand why
  - Use RIPE-229 parameters, or something even weaker
  - Do NOT use the vendor defaults without thinking
- **Remove private ASes from announcements**
  - Private ASNs should not appear on the public Internet
- **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

- **Consider using maximum-prefix tracking**

  Router will warn you if there are sudden increases in BGP table size, bringing down eBGP if desired

- **Log changes of neighbour state**

  …and monitor those logs

  ...both on and off the router!

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

  Otherwise prefixes heard from outside your network could override your IGP!!

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

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

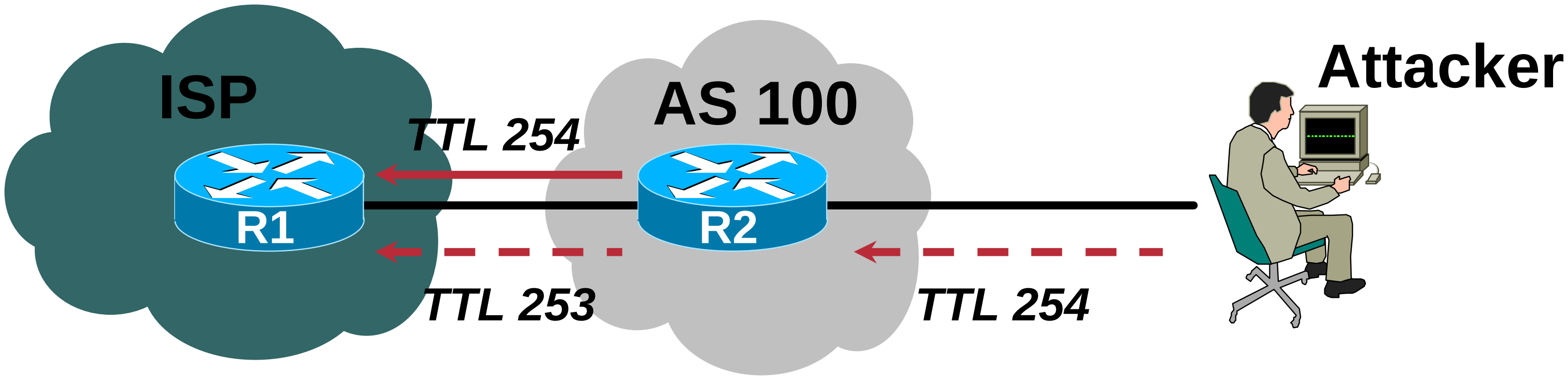
BGP TTL “hack”

- **Implement RFC3682 on BGP peerings**

  Neighbour sets TTL to 255

  Local router configured to expect 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**

Passwords on BGP sessions

- *Yes, I am mentioning passwords again*
- **Put password on the BGP session**
  - **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"**

Using Communities

- **Use communities to:**
  - **Scale iBGP management**
  - **Ease iBGP management**
- **Come up with a strategy for different classes of customers**
  - **Which prefixes stay inside network**
  - **Which prefixes are announced by eBGP**
  - **…etc…**

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

BGP Techniques for Internet Service Providers

Philip Smith <pfs@cisco.com>

NANOG 34

Seattle, 15-17 May 2005