Cisco Systems

Troubleshooting BGP

Philip Smith <pfs@cisco.com>

APRICOT 2006

22 Feb - 3 Mar 2006

Perth, Australia

Presentation Slides

- Slides are at:

  ftp://ftp-eng.cisco.com/pfs/seminars/APRICOT2006-BGP-part4.pdf

  And on the APRICOT 2006 website

- Feel free to ask questions any time

Assumptions

- **Presentation assumes working knowledge of BGP**

  Beginner and Intermediate experience of protocol

- **If in any doubt, please ask!**

Agenda

- Fundamentals of Troubleshooting
- Local Configuration Problems
- Internet Reachability Problems

Fundamentals:
Problem Areas

- First step is to recognise what causes the problem
- Possible Problem Areas:

  Misconfiguration

  Configuration errors caused by bad documentation, misunderstanding of concepts, poor communication between colleagues or departments

  Human error

  Typos, using wrong commands, accidents, poorly planned maintenance activities

Fundamentals: Problem Areas

- **More Possible Problem Areas:**
  - "feature behaviour"
    - Or – "it used to do this with Release X.Y(a) but Release X.Y(b) does that"
  - Interoperability issues
    - Differences in interpretation of RFC1771 and its developments
  - Those beyond your control
    - Upstream ISP or peers make a change which has an unforeseen impact on your network

Fundamentals: Working on Solutions

- Next step is to try and fix the problem

  And this is not about diving into network and trying random commands on random routers, just to "see what difference this makes"

- Before we begin/Troubleshooting is about:

  Not panicking

  Creating a checklist

  Working to that checklist

  Starting at the bottom and working up

Fundamentals: Checklists

- This presentation will have references in the later stages to checklists
  - They are the best way to work to a solution
  - They are what many NOC staff follow when diagnosing and solving network problems
  - It may seem daft to start with simple tests when the problem looks complex
    - But quite often the apparently complex can be solved quite easily

Fundamentals:
Tools

- **Use system and network logs as an aid**
- **Record keeping:**
  - **Good and detailed system logs**
  - **Last known good configuration**
  - **History trail of working configurations and all intermediate changes**
  - **Record of commands entered on routers and other network devices**

Fundamentals:
Tools

- **Familiarise yourself with the routers tools:**
  - **Is logging of the BGP process enabled?**
    - **(And is it captured/recorded off the router?)**
  - **Are you familiar with the BGP debug process and commands (if available)**
    - **Check vendor documentation before switching on full BGP debugging – you might get fewer surprises**

Fundamentals:
Tools

- **Traffic and traffic flow measurement in the network**

  Unexplained change in traffic levels on an interface, a connection, a peering,…

  Correlation of customer feedback on network or connectivity issues…

Agenda

- Fundamentals
- Local Configuration Problems
- Internet Reachability Problems

Local Configuration Problems

- Peer Establishment
- Missing Routes
- Inconsistent Route Selection
- Loops and Convergence Issues

Peer Establishment:
ACLs and Connectivity

- Routers establish a TCP session

  Port 179—Permit in interface packet filters

  IP connectivity (route from IGP)

- OPEN messages are exchanged

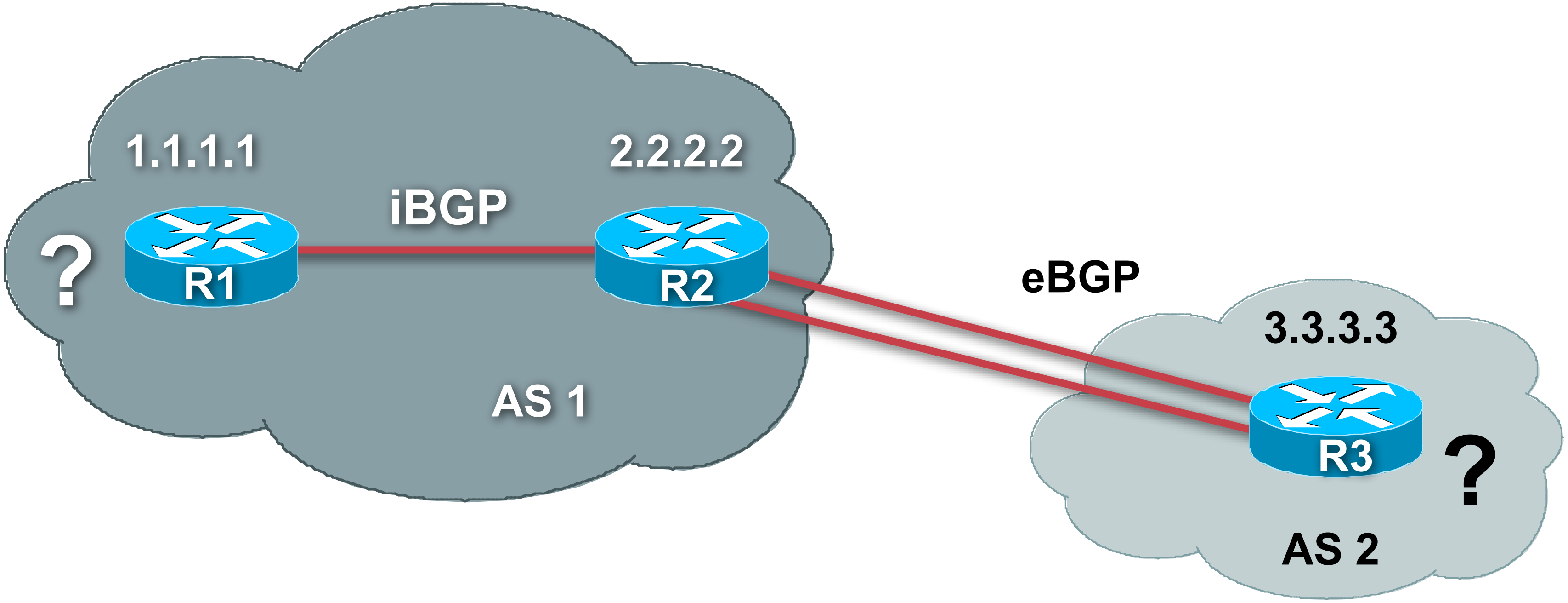
  Peering addresses must match the TCP session

  Local AS configuration parameters

Peer Establishment:
Common Problems

- Sessions are not established
  - No IP reachability
  - Incorrect configuration
- Peers are flapping
  - Layer 2 problems

Peer Establishment



Is the Local AS configured correctly?
Is the remote-as assigned correctly?
Verify with your diagram or other documentation!

Peer Establishment: iBGP – Summary

- Assume that IP connectivity has been checked
  - Including IGP reachability between peers
- Check TCP to find out what connections we are accepting
  - Check the ports and source/destination addresses
  - Do they match the configuration?

- Common problem:
  - iBGP is run between loopback interfaces on router (for stability), but the configuration is missing from the router ⇒ iBGP fails to establish
  - Remember that source address is the IP address of the outgoing interface unless otherwise specified

Peer Establishment: eBGP Problems

- eBGP by and large is problem free for single point to point links
  - Source address is that of the outbound interface
  - Destination address is that of the outbound interface on the remote router
    - And is directly connected (TTL is set to 1 for eBGP peers)
  - Filters permit TCP/179 in both directions

Peer Establishment: eBGP Problems

- Load balancing over multiple links and/or use of eBGP multihop gives potential for so many problems

  IP Connectivity to the remote address

  Filters somewhere in the path

  eBGP by default sets TTL to 1, so you need to change this to permit multiple hops

- Some ISPs won't even allow their customers to use eBGP multihop due to the potential for problems

Peer Establishment: eBGP Problems

- eBGP multihop problems

  IP Connectivity to the remote address

  is a route in the local *routing* table?

  is a route in the remote *routing* table?

  Check this using ping, including the extended options that it has in most implementations

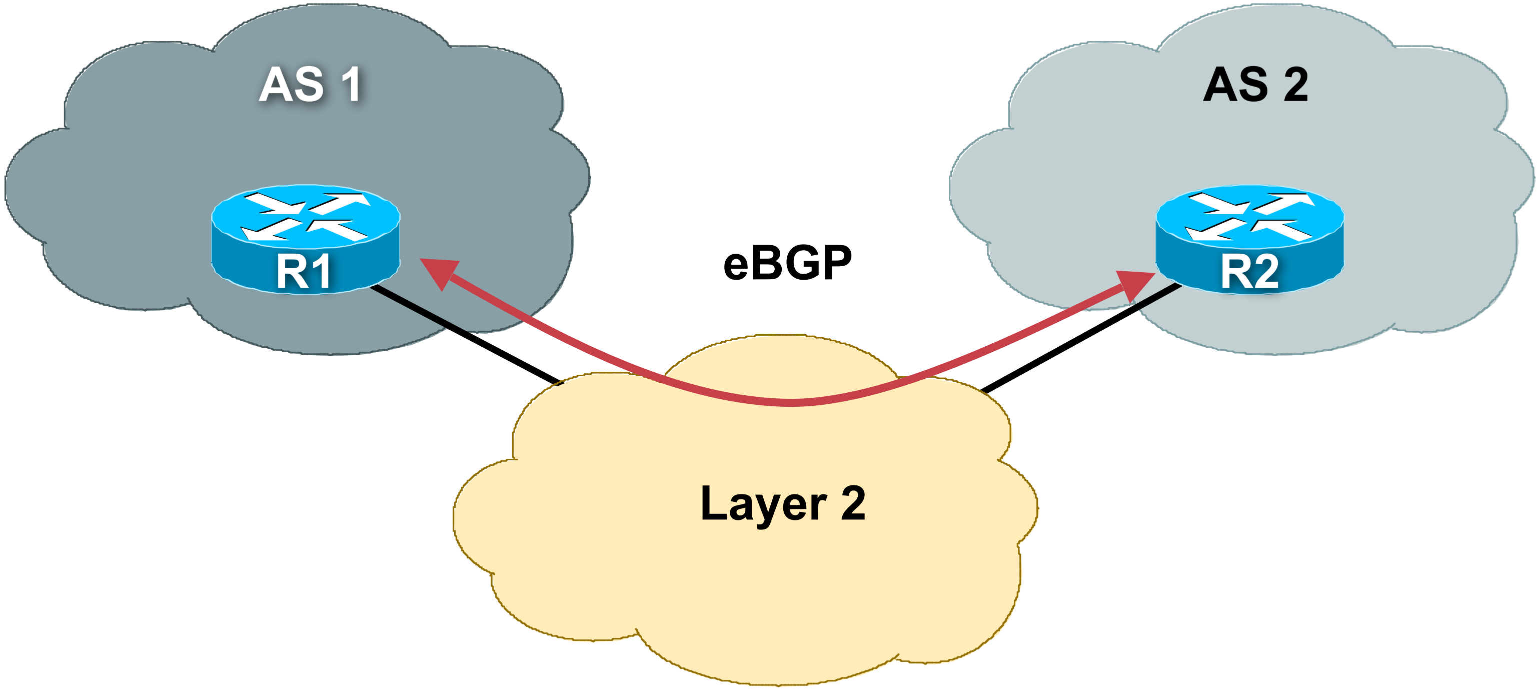
- Filters in the path?

  If this crosses multiple providers, this needs their cooperation

Peer Establishment:
Passwords

- Using passwords on iBGP and eBGP sessions
  - Link won’t come up
  - Been through all the previous troubleshooting steps
- Common problems:
  - Missing password – needs to be on both ends
  - Cut and paste errors – don’t!
  - Typographical errors
  - Capitalisation, extra characters, white space…
- Common solutions:
  - Check for symptoms/messages in the logs
  - Re-enter passwords from scratch – don’t cut&paste

Flapping Peer: Common Symptoms



- Symptoms – the eBGP session flaps
- eBGP peering establishes, then drops, re-establishes, then drops,…

Flapping Peer:
Common Symptoms

- **Ensure logging is enabled – no logs → no clue**
- **What do the logs say?**

  Problems are usually caused because BGP keepalives are lost

  No keepalive ⇒ local router assumes remote has gone down, so tears down the BGP session

  Then tries to re-establish the session – which succeeds

  Then tries to exchange UPDATEs – fails, keepalives get lost, session falls over again

  WHY??

Flapping Peer: Diagnosis and Solution

- Diagnosis

  Keepalives get lost because they get stuck in the router's queue behind BGP update packets.

  BGP update packets are packed to the size of the MTU – keepalives and BGP OPEN packets are not packed to the size of the MTU ⇒ Path MTU problems

  Use ping with different size packets to confirm the above – 100byte ping succeeds, 1500byte ping fails = MTU problem somewhere

- Solution

  Pass the problem to the L2 folks – but be helpful, try and pinpoint using ping where the problem might be in the network

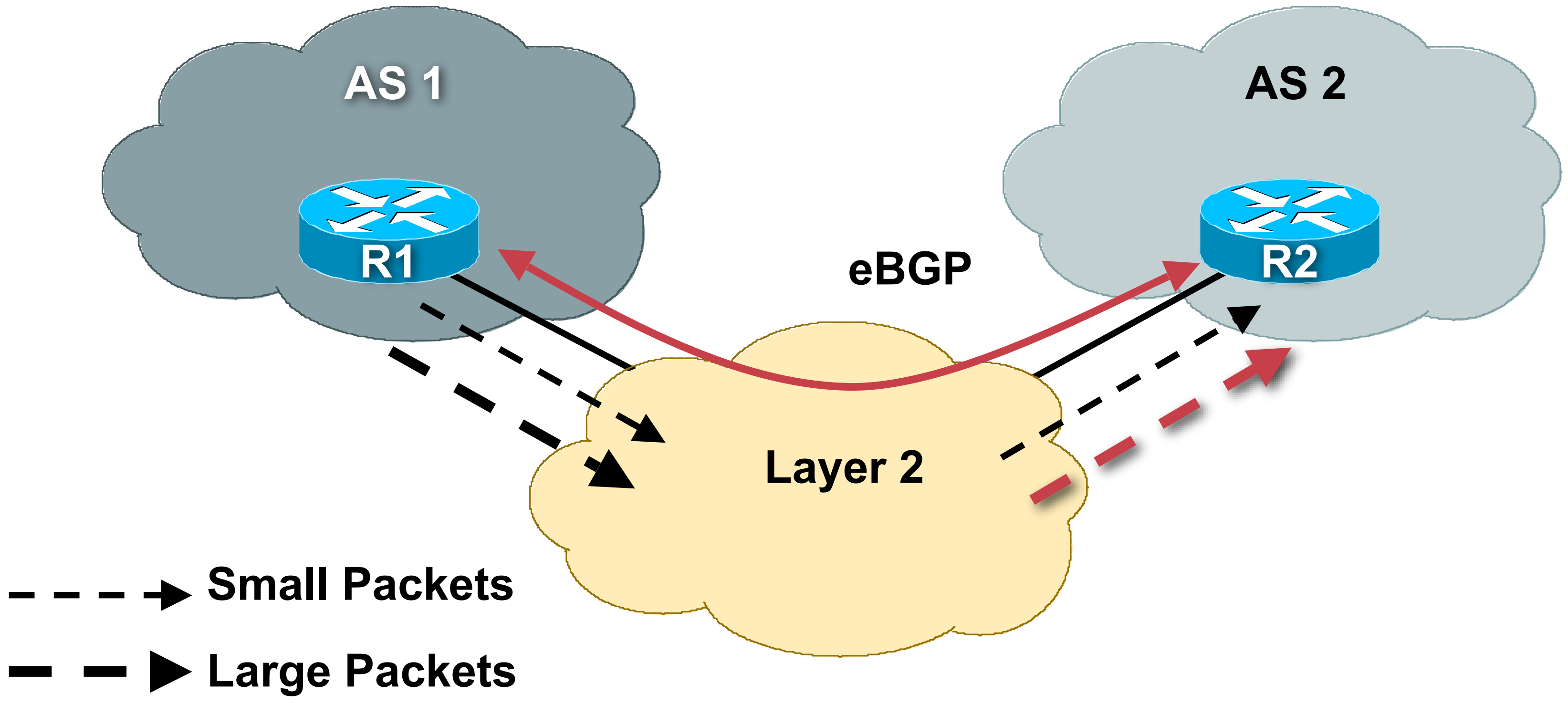
Flapping Peer: Other Common Problems

- Remote router rebooting continually (typical with a 3-5 minute BGP peering cycle time)
- Remote router BGP process unstable, restarting
- Traffic Shaping & Rate Limiting parameters
- MTU incorrectly set on links, PMTU discovery disabled on router
- For non-ATM/FR links, instability in the L2 point-to-point circuits

  Faulty MUXes, bad connectors, interoperability problems, PPP problems, satellite or radio problems, weather, etc. The list is endless – your L2 folks should know how to solve them

  For you, *ping* is the tool to use

Flapping Peer: Fixed!



- Large packets are ok now
- BGP session is stable!

Local Configuration Problems

- Peer Establishment
- Missing Routes
- Inconsistent Route Selection
- Loops and Convergence Issues

Quick Review

- Once the session has been established, UPDATEs are exchanged
  - All the locally known routes
  - Only the bestpath is advertised
- Incremental UPDATE messages are exchanged afterwards

Quick Review

- **Bestpath received from eBGP peer**
  - **Advertise to all peers**
- **Bestpath received from iBGP peer**
  - **Advertise only to eBGP peers**
  - **A full iBGP mesh must exist**

Missing Routes

- Route Origination
- UPDATE Exchange
- Filtering
- iBGP mesh problems

Missing Routes: Route Origination

- Common problem occurs when putting prefixes into the BGP table
- BGP table is NOT the RIB

  BGP table, as with OSPF table, ISIS table, static routes, etc, is used to feed the RIB, and hence the FIB

- To get a prefix into BGP, it must exist in another routing process too, typically:

  Static route pointing to customer (for customer routes into your iBGP)

  Static route pointing to Null (for aggregates you want to put into your eBGP)

Missing Routes

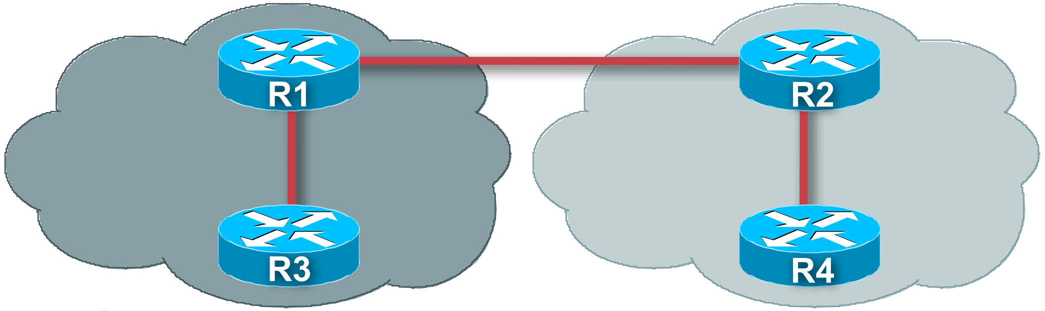
- Route Origination
- UPDATE Exchange
- Filtering
- iBGP mesh problems

Missing Routes:
Update Exchange

- **Ah, Route Reflectors…**
  - **Such a nice solution to help scale BGP**
  - **But why do people insist in breaking the rules all the time?!**
- **Common issues**
  - **Clashing router IDs**
  - **Clashing cluster IDs**

Missing Routes: Example I

- Two RR clusters
- R1 is a RR for R3
- R2 is a RR for R4
- R4 is advertising 7.0.0.0/8
- R2 has the route but R1 and R3 do not?



Missing Routes: Example I

- R1 is not accepting the route when R2 sends it on

  Clashing router ID!

  If R1 sees its own router ID in the originator attribute in any received prefix, it will reject that prefix

  - How a route reflector attempts to avoid routing loops

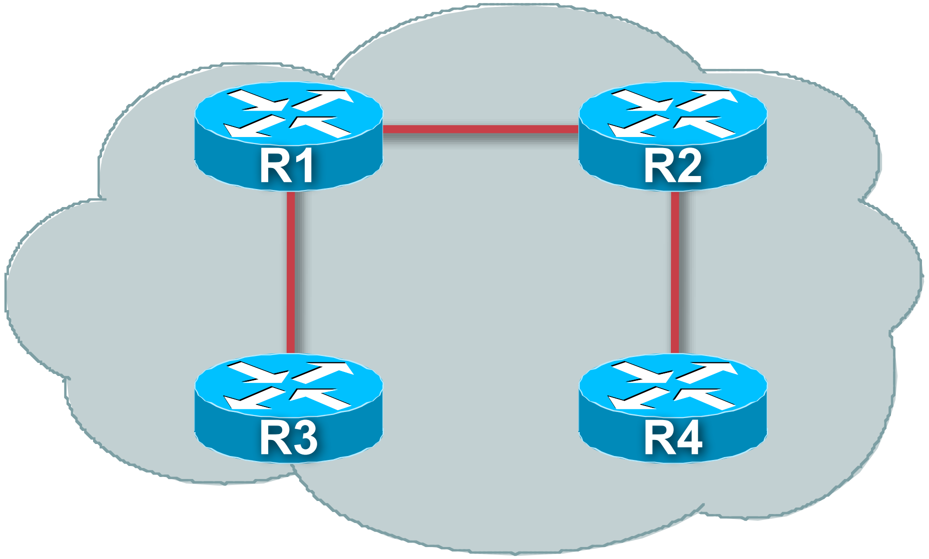
- Solution

  do NOT set the router ID by hand unless you have a very good reason to do so and have a very good plan for deployment

  Router-ID is usually calculated automatically by router

Missing Routes: Example II

- One RR cluster
- R1 and R2 are RRs
- R3 and R4 are RRCs
- R4 is advertising 7.0.0.0/8
  - R2 has it
  - R1 and R3 do not



Missing Routes: Example II

- **R1 is not accepting the route when R2 sends it on**

  If R1 sees its own router ID in the cluster-ID attribute in any received prefix, it will reject that prefix

  How a route reflector avoids redundant information

- **Reason**

  Early documentation claimed that RRC redundancy should be achieved by dual route reflectors in the same cluster

  This is fine and good, but then ALL clients must peer with both RRs, otherwise examples like this will occur

- **Solution**

  Use overlapping RRCs for redundancy, stick to defaults

Missing Routes

- Route Origination
- UPDATE Exchange
- Filtering
- iBGP mesh problems

Update Filtering

- Type of filters

    Prefix filters

    AS_PATH filters

    Community filters

    Policy/Attribute manipulation

- Applied incoming and/or outgoing

Update Filtering

- If you suspect a filtering problem, become familiar with the router tools to find out what BGP filters are applied
- Tip: don't cut and paste!

  Many filtering errors and diagnosis problems result from cut and paste buffer problems on the client, the connection, and even the router

Update Filtering: Common Problems

- Typos in regular expressions

  Extra characters, missing characters, white space, etc

  In regular expressions every character matters, so accuracy is highly important

- Typos in prefix filters

  Watch the router CLI, and the filter logic – it may not be as obvious as you think, or as simple as the manual makes out

  Watch netmask confusion, and 255 profusion – easy to muddle 255 with 0 and 225!

Update Filtering: Common Problems

- Communities
  - Each implementation has different defaults for when communities are sent
    - Some don’t send communities by default
    - Others do for iBGP and not for eBGP by default
    - Others do for all BGP peers by default
  - Watch how your implementation handles communities
    - There may be implicit filtering rules
  - Each ISP has different policies – never assume that because communities exist that people will use them, or pay attention to the ones you send

Missing Routes: General Problems

- **Make and then Stick to simple policy rules:**

  Most implementations have particular rules for filtering of prefixes, AS-paths, and for manipulating BGP attributes

  Try not to mix these rules

  Rules for manipulating attributes can also be used for filtering prefixes and ASNs – can be very powerful, but can also become very confusing

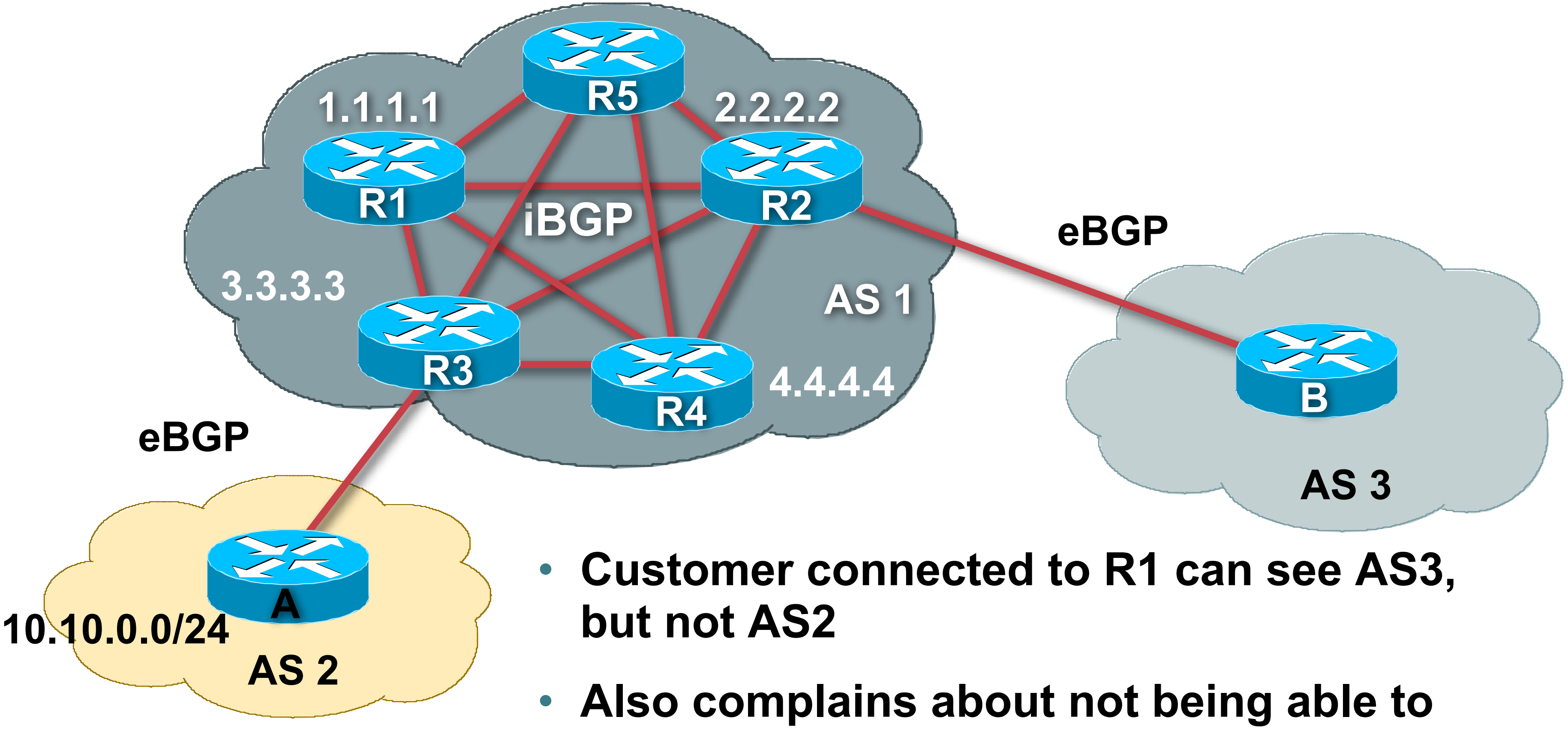
Missing Routes

- Route Origination
- UPDATE Exchange
- Filtering
- iBGP mesh problems

Missing Routes: iBGP

- **Symptom: customer complains about patchy Internet access**
  - **Can access some, but not all, sites connected to backbone**
  - **Can access some, but not all, of the Internet**

Missing Routes: iBGP



- Customer connected to R1 can see AS3, but not AS2
- Also complains about not being able to see sites connected to R5
- No complaints from other customers

Missing Routes:
iBGP

- Diagnosis: This is the classic iBGP mesh problem

  The full mesh isn't complete – how do we know this?

- Customer is connected to R1

  Can't see AS2 ⇒ R3 is somehow not passing routing information about AS2 to R1

  Can't see R5 ⇒ R5 is somehow not passing routing information about sites connected to R5

  But can see rest of the Internet ⇒ his prefix is being announced to some places, so not an iBGP origination problem

Missing Routes: iBGP

- When using full mesh iBGP, check on every iBGP speaker that it has a neighbour relationship with every other iBGP speaker

  In this example, R3 peering with R1 is down as R1 isn't seeing any of the routes connected through R3

- Try and use configuration shorthand if available in your implementation

  Peering between R1 and R5 was down as there was a typo in the shorthand, resulting in the incorrect configuration being used

Troubleshooting Tips

- Use configuration shorthand both for efficiency and to avoid making policy errors within the iBGP mesh
  - This is especially true for full iBGP mesh networks
  - But be careful of not introducing typos into names of these “subroutines” – common problem
- Use route reflectors to avoid accidentally missing iBGP peers, especially as the mesh grows in size
  - But stick to the route reflector rules and the defaults in the implementation – changing defaults and ignoring BCP techniques introduces complexity and causes problems

Local Configuration Problems

- Peer Establishment
- Missing Routes
- Inconsistent Route Selection
- Loops and Convergence Issues

Inconsistent Route Selection

- Two common problems with route selection
  - Inconsistency
  - Appearance of an incorrect decision
- RFC 1771 defined the decision algorithm
- Every vendor has tweaked the algorithm
  - http://www.cisco.com/warp/public/459/25.shtml
- Route selection problems can result from oversights by RFC 1771
- RFC1771 is now obsoleted by RFC4271
  - Hopefully compliance with RFC4271 will help avoid future issues

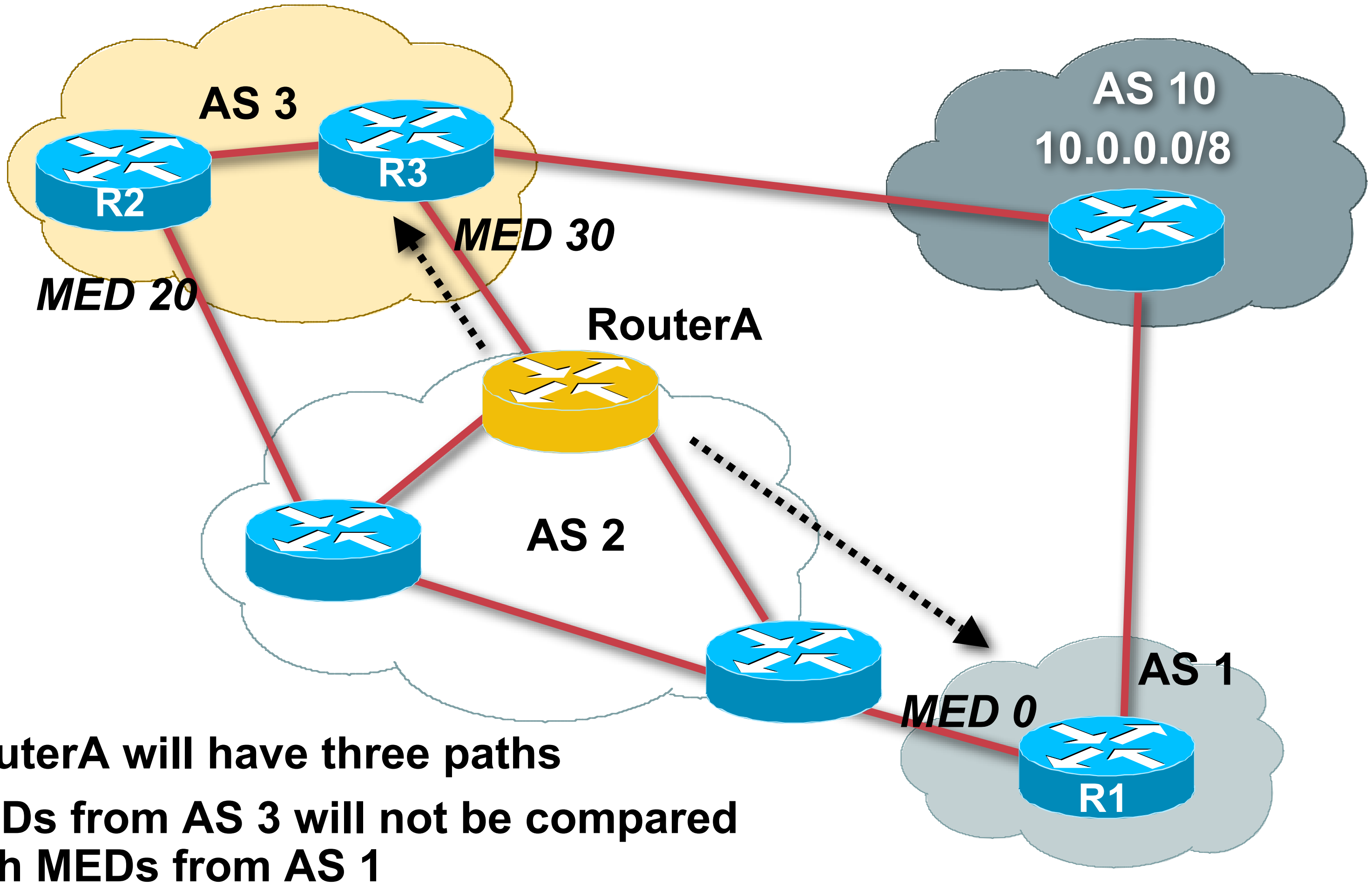
Inconsistent: Example I

- RFC says that MED is not always compared
- As a result, the ordering of the paths can effect the decision process
- For example, the default in Cisco IOS is to compare the prefixes in order of arrival (most recent to oldest)
  - This can result in inconsistent route selection
  - Symptom is that the best path chosen after each BGP reset is different

Inconsistent: Example I

- Inconsistent route selection may cause problems
  - Routing loops
  - Convergence loops—i.e. the protocol continuously sends updates in an attempt to converge
  - Changes in traffic patterns
- Difficult to catch and troubleshoot
  - In Cisco IOS, the deterministic-med configuration command is used to order paths consistently
    - Enable in all the routers in the AS
  - The bestpath is recalculated as soon as the command is entered

Symptom I: Diagram

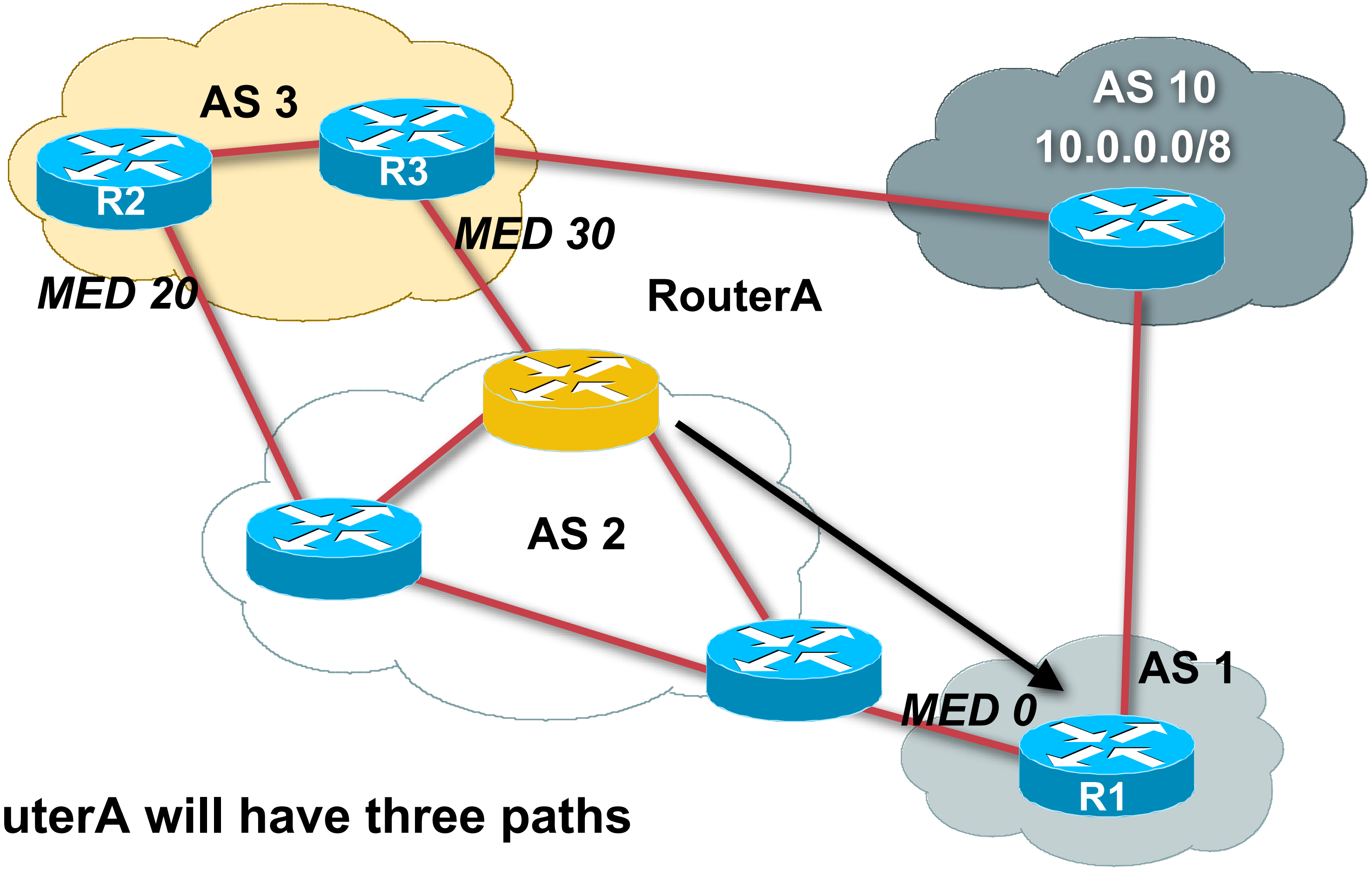


- RouterA will have three paths
- MEDs from AS 3 will not be compared with MEDs from AS 1
- RouterA will sometimes select the path from R1 as best and but may also select the path from R3 as best

Deterministic MED: Operation

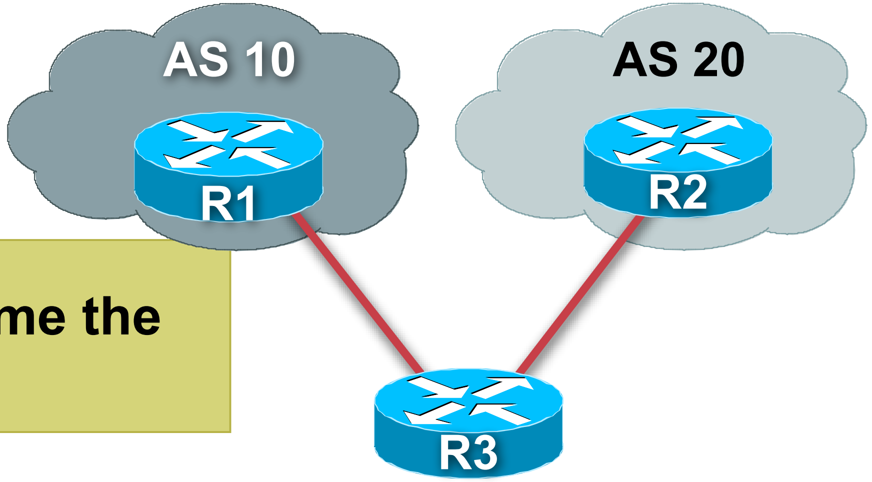
- The paths are ordered by Neighbour AS
- The bestpath for each Neighbour AS group is selected
- The overall bestpath results from comparing the winners from each group
- The bestpath will be consistent because paths will be placed in a deterministic order

Solution: Diagram



- RouterA will have three paths
- RouterA will consistently select the path from R1 as best!

Inconsistent: Example II



- The bestpath changes every time the peering is reset

- By default, the “oldest” external is the bestpath
  - All other attributes are the same
  - Stability Enhancement in Cisco IOS
- The BGP sub-command “bestpath compare-router-id” will disable this enhancement

Inconsistent: Example III

- Path 1 has higher localpref but path 2 is better???
- This appears to be incorrect…

- It's because Cisco IOS has "synchronization" on by default

  …and if a prefix is not synchronized (i.e. appearing in IGP as well as BGP), its path won't be included in the bestpath process

Inconsistent Path Selection

- **Summary:**

  RFC1771 wasn’t prefect when it came to path selection – years of operational experience have shown this

  Vendors and ISPs have worked to put in stability enhancements, now reflected in RFC4271

  But these can lead to interesting problems

  And of course some defaults linger much longer than they ought to – so never assume that an out of the box default configuration will be perfect for your network

Local Configuration Problems

- Peer Establishment
- Missing Routes
- Inconsistent Route Selection
- Loops and Convergence Issues

Route Oscillation: Symptom

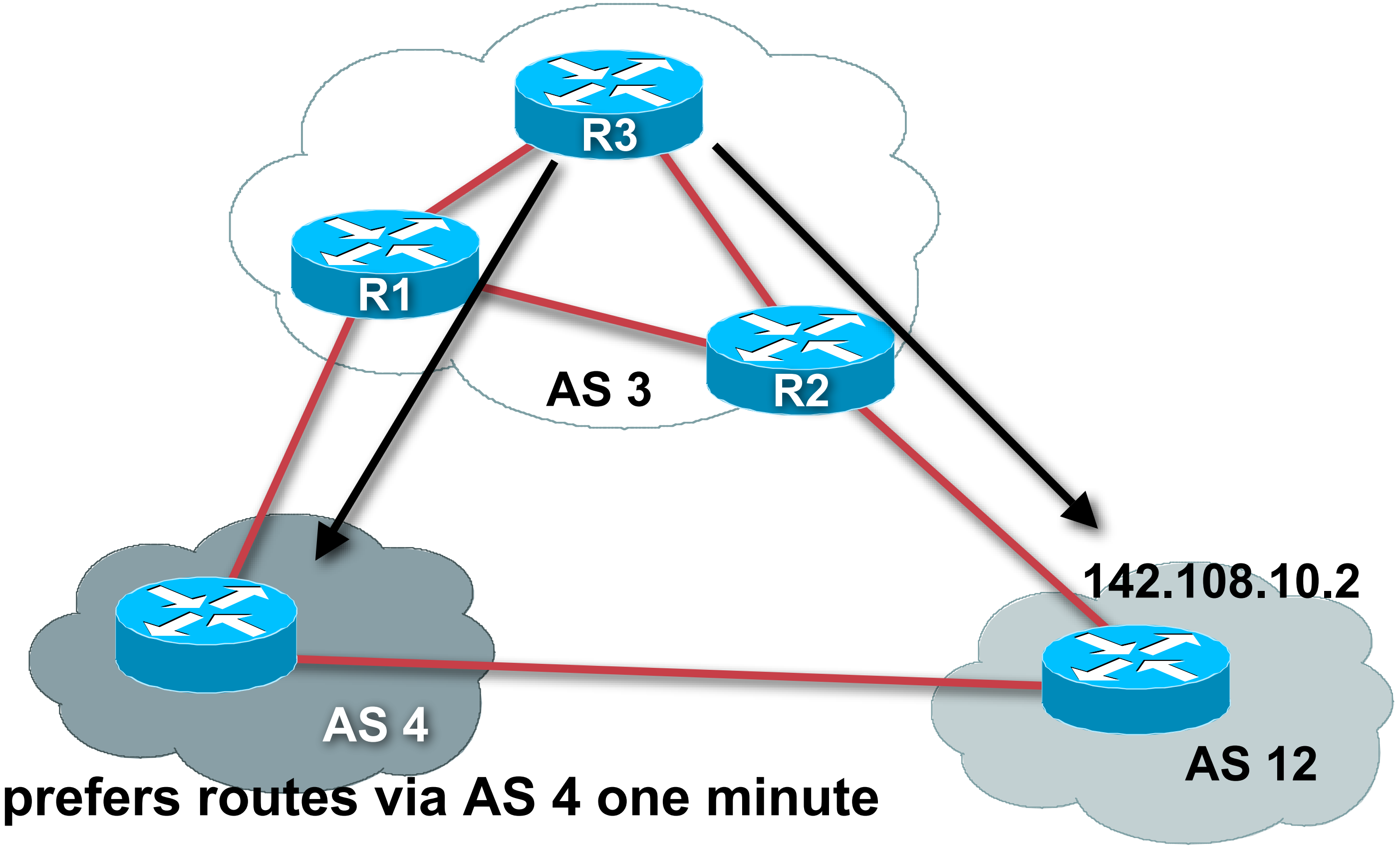
- **One of the most common problems**
- **Main symptom is that traffic exiting the network oscillates every minute between two exit points**

  This is almost *always* caused by the BGP NEXT_HOP being known only by BGP

  Common problem in ISP networks – but if you have never seen it before, it can be a nightmare to debug and fix

- **Other symptom is high CPU utilisation for the BGP router process**

Route Oscillation: Diagram



- R3 prefers routes via AS 4 one minute
- 1 minute later R3 prefers routes via AS 12
- And 1 minute after that R3 prefers AS 4 again

Route Oscillation: Cause

- BGP nexthop is known via BGP

  This is an illegal recursive lookup

- Scanner will notice, drop this path, and install the other path in the RIB
- Route to the nexthop is now valid
- Scanner will detect this and re-install the other path
- Routes will oscillate forever

  One minute cycle in Cisco IOS as scanner runs every minute

Route Oscillation: Solution

- Make sure that all the BGP NEXT_HOPs are known by the IGP

  (whether OSPF/ISIS, static or connected routes)

  If NEXT_HOP is also in iBGP, ensure the iBGP distance is longer than the IGP distance

—or—

- Don’t carry external NEXT_HOPs in your network

  Use “next-hop-self” concept on all the edge BGP routers

- Two simple solutions

Troubleshooting Tips

- **High CPU utilisation in the BGP process is normally a sign of a convergence problem**
- **Find a prefix that changes every minute**
- **Troubleshoot/debug that one prefix**

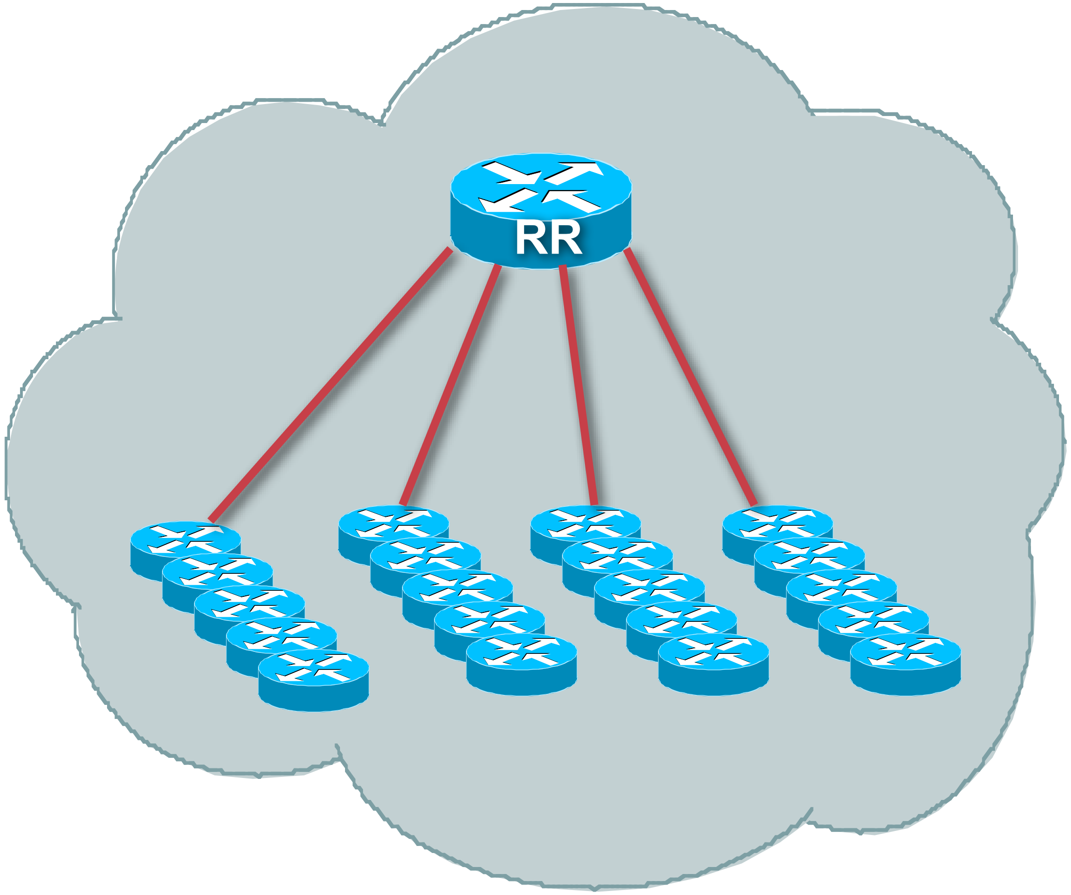
Troubleshooting Tips

- **BGP routing loop?**
  - First, check for IGP routing loops to the BGP NEXT_HOPs
- **BGP loops are normally caused by**
  - Not following physical topology in RR environment
  - Multipath with confederations
  - Lack of a full iBGP mesh
- **Get the following from each router in the loop path**
  - The routing table entry
  - The BGP table entry
  - The route to the NEXT_HOP

Convergence Problems: Example I

- Route reflector with 250 route reflector clients
- 100k routes
- BGP will not converge
- Logs show that neighbour hold times have expired
- The BGP router summary shows peers establishing, dropping, re-establishing

  And it’s not the MTU problem we saw earlier!



Convergence Problems: Example I

- We are either missing hellos or our peers are not sending them
- Check for interface input drops

  If the number is large, and the interface counters show recent history, then this is probably the cause of the peers going down
- Large drops is usually due to the input queue being too small

  Large numbers of peers can easily overflow the queue, resulting in lost hellos
- Solution is to increase the size of the input queues to be considerably larger than the number of peers

Convergence Problems:
Example II

- BGP converges in 25 minutes for 250 peers and 100k routes
    - Seems like a long time
    - What is TCP doing?

- Check the MSS size
    - And enable Path MTU discovery on the router if it is not on by default
    - MSS of 536 means that router needs to send almost three times the amount of packets compared with an MSS of 1460
- Result:
    - Should see BGP converging in about half the time – which is respectable for 250 peers and 100k routes

Agenda

- Fundamentals
- Local Configuration Problems
- Internet Reachability Problems

Internet Reachability Problems

- **BGP Attribute Confusion**

  To Control Traffic in → Send MEDs and AS-PATH prepends on outbound announcements

  To Control Traffic out → Attach local-preference to inbound announcements

- **Troubleshooting of multihoming and transit is often hampered because the relationship between routing information flow and traffic flow is forgotten**

Internet Reachability Problems
BGP Path Selection Process

- Each vendor has “tweaked” the path selection process
  - Know it for your router equipment – saves time later
  - Especially applies with networks with more than one BGP implementation present
  - Best policy is to use supplied “knobs” to ensure consistency – and avoid steps in the process which can lead to inconsistency

Internet Reachability Problems
MED Confusion

- **Default MED on Cisco IOS is ZERO**
  - **It may not be this on your router, or your peer's router**
- **Best not to rely on MEDs for multihoming on multiple links to upstream**
  - **Their default might be 2^{32}-1 resulting in your hoped for best path being their worst path**
  - **"Workaround", i.e. current good practice, is to use communities rather than MEDs**

Internet Reachability Problems

- Community confusion

  set community does just that – it overwrites any other community set on the prefix

  Use additive keyword to add community to existing list

  Use Internet format for community (AS:xx) not the 32-bit IETF format

  Cisco IOS never sends community by default

  Other implementations may send community by default for iBGP and/or eBGP

  Never assume that your neighbouring AS will honour your no-export community – ask first!

Internet Reachability Problems

- AS-PATH prepends

  20 prepends won't lessen the priority of your path any more than 10 prepends will – check it out at a Looking Glass

    The Internet is on average only 5 ASes deep, maximum AS prepend most ISPs have to use is around this too

    Know you BGP path selection algorithm

  Some ISPs use bgp maxas-limit 15 to drop prefixes with ridiculously long AS-paths

Internet Reachability Problems

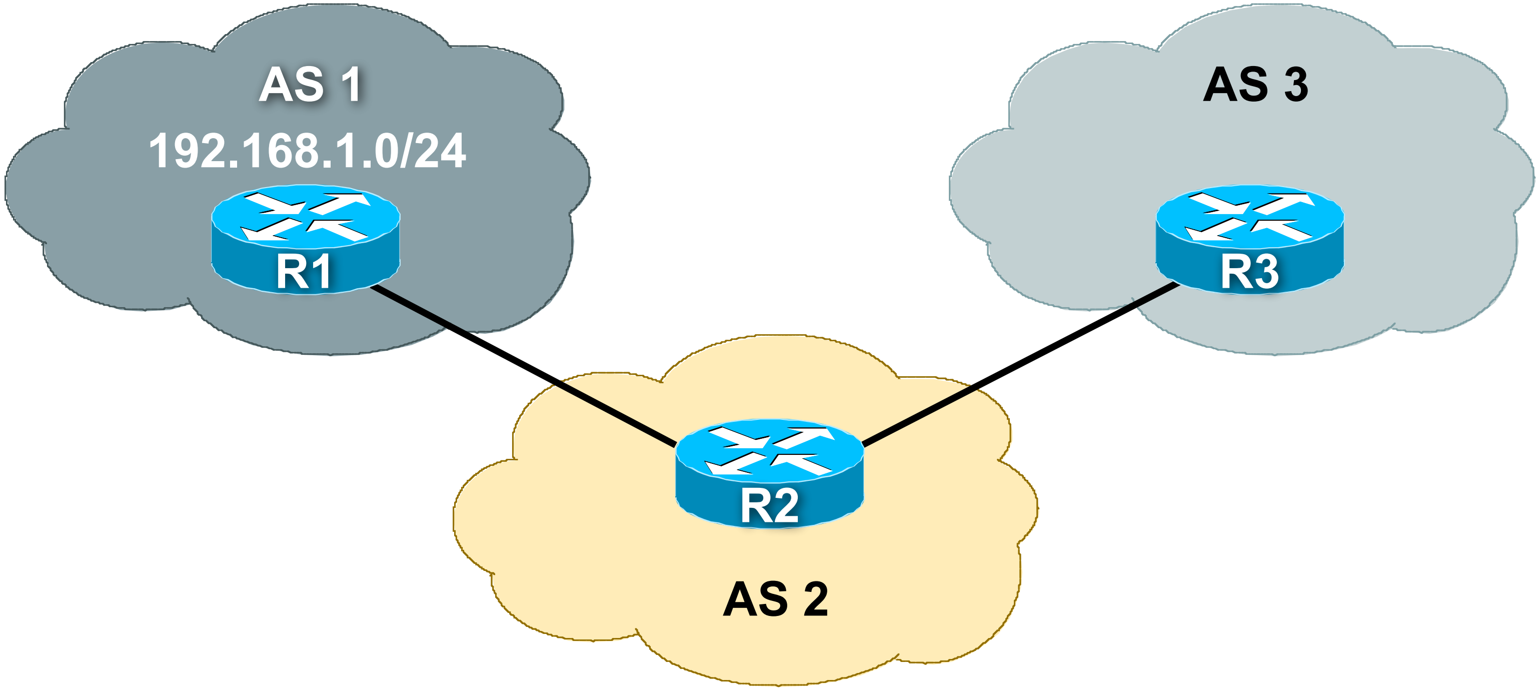
- Private ASes should not ever appear in the Internet
- Cisco IOS remove-private-AS command does not remove every instance of a private AS

  e.g. won’t remove private AS appearing in the middle of a path surrounded by public ASNs

  www.cisco.com/warp/public/459/32.html

- Apparent non-removal of private-ASNs may not be a bug, but a configuration error somewhere else

Troubleshooting Connectivity – Example I



- **Symptom: AS1 announces 192.168.1.0/24 to AS2 but AS3 cannot see the network**

Troubleshooting Connectivity – Example I

- **Checklist:**

  AS1 announces, but does AS2 see it?

  We are checking eBGP filters on R1 and R2. Remember that R2 access will require cooperation and assistance from your peer

  Does AS2 see it over entire network?

  We are checking iBGP across AS2's network (unneeded step in this case, but usually the next consideration). Quite often iBGP is misconfigured, lack of full mesh, problems with RRs, etc.

Troubleshooting Connectivity – Example I

- **Checklist:**

  Does AS2 send it to AS3?

  We are checking eBGP configuration on R2. There may be a configuration error with as-path filters, or prefix-lists, or communities such that only local prefixes get out

  Does AS3 see all of AS2's originated prefixes?

  We are checking eBGP configuration on R3. Maybe AS3 does not know to expect prefixes from AS1 in the peering with AS2, or maybe it has similar errors in as-path or prefix or community filters

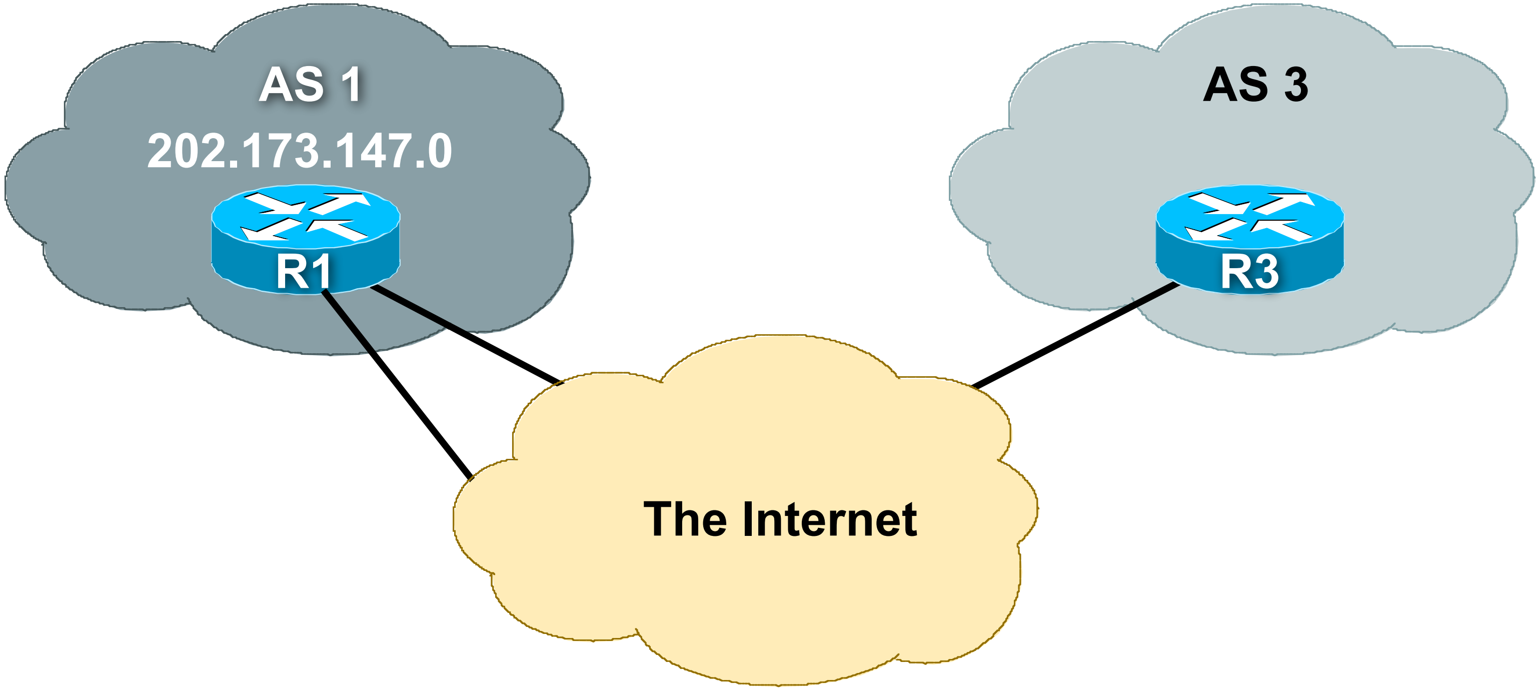
Troubleshooting Connectivity – Example I

- **Troubleshooting connectivity beyond immediate peers is much harder**

  Relies on your peer to assist you – they have the relationship with their BGP peers, not you

  Quite often connectivity problems are due to the private business relationship between the two neighbouring ASNs

Troubleshooting Connectivity – Example II



- **Symptom: AS1 announces 202.173.147.0/24 to its upstreams but AS3 cannot see the network**

Troubleshooting Connectivity – Example II

- **Checklist:**

  AS1 announces, but do its upstreams see it?

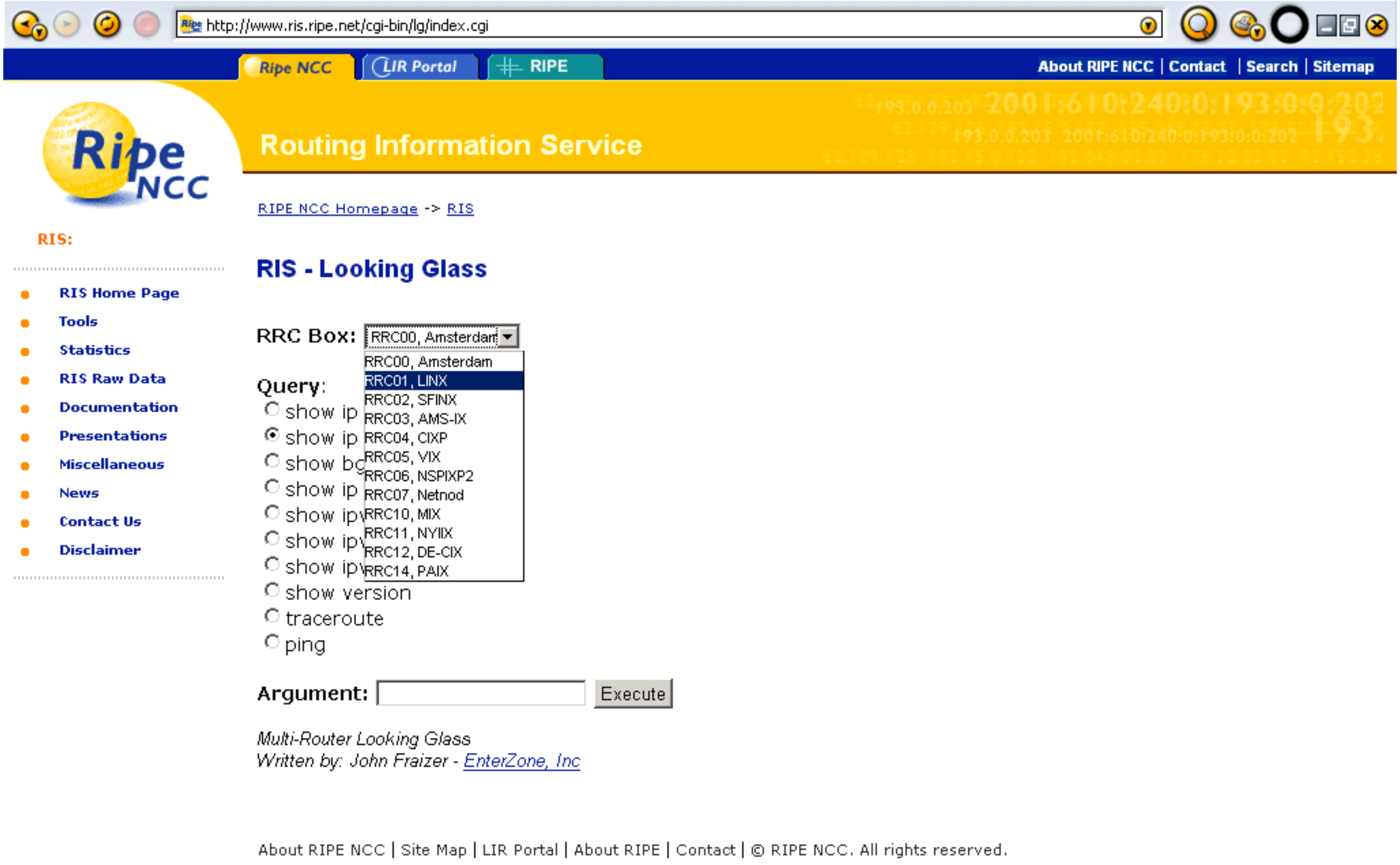
  We are checking eBGP filters on R1 and upstreams. Remember that upstreams will need to be able to help you with this

  Is the prefix visible anywhere on the Internet?

  We are checking if the upstreams are announcing the network to anywhere on the Internet. See next slides on how to do this.

Troubleshooting Connectivity – Example II

- Help is at hand – the Looking Glass
- Many networks around the globe run Looking Glasses
  - These let you see the BGP table and often run simple ping or traceroutes from their sites
  - www.traceroute.org for IPv4
  - Some IPv6 Looking Glasses listed at www.bgp4.as/looking-glasses
- Some ISPs, especially those with large and diverse networks, run their own internal Looking Glass to aid internal troubleshooting
- Next slides have some examples of a typical looking glass in action



RIS - Looking Glass

RRC Box: RRC00, Amsterdam

- RRC00, Amsterdam
- RRC01, LINX
- RRC02, SFINX
- RRC03, AMS-IX
- RRC04, CIXP
- RRC05, VIX
- RRC06, NSPIXP2
- RRC07, Netnod
- RRC10, MIX
- RRC11, NYIIX
- RRC12, DE-CIX
- RRC14, PAIX

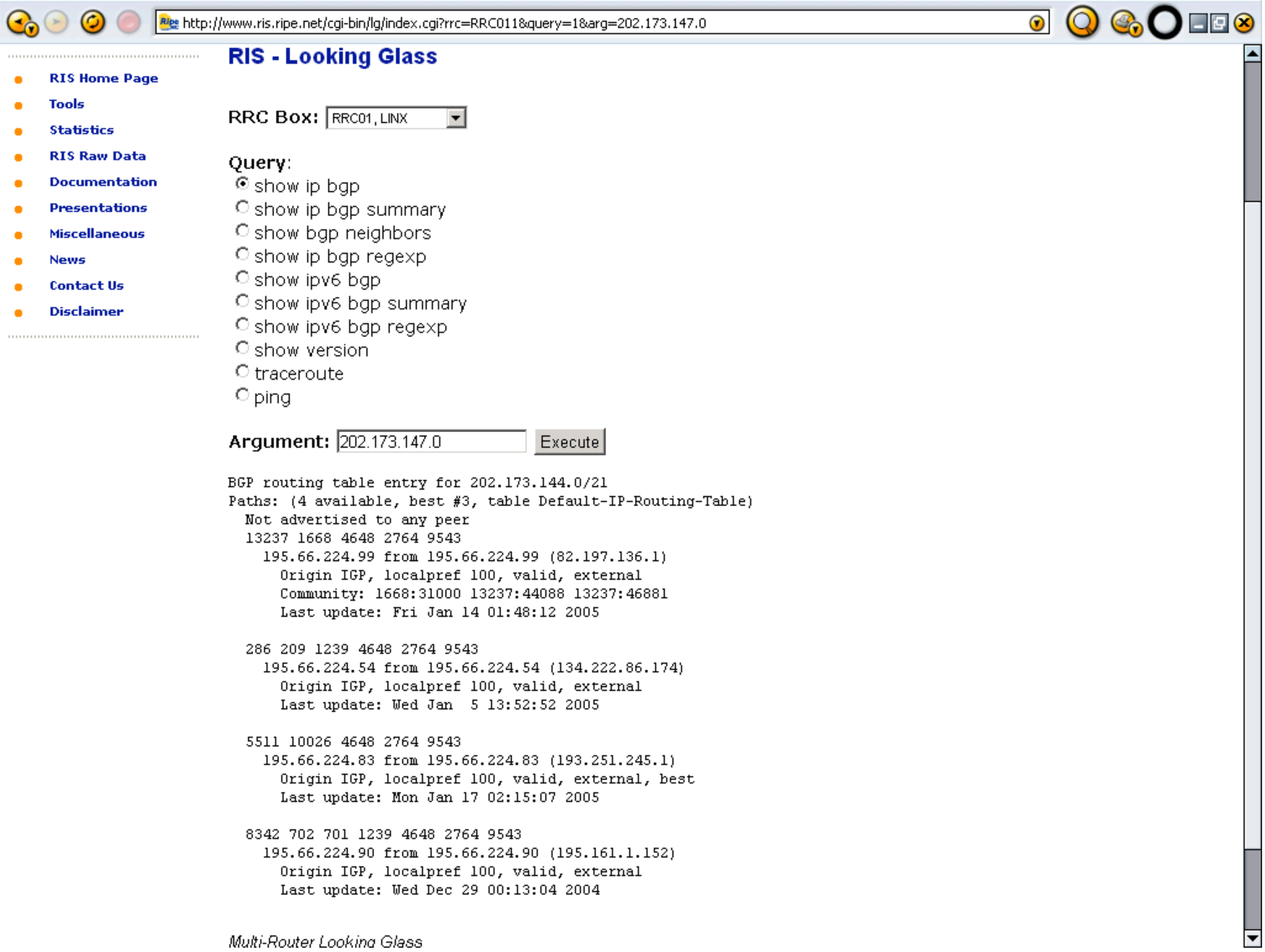
Query:

○ show ip
● show ip
○ show bg
○ show ip
○ show ipv
○ show ipv
○ show ipv
○ show version
○ traceroute
○ ping

Argument: [ ] Execute

Multi-Router Looking Glass
Written by: John Fraizer - EnterZone, Inc

About RIPE NCC | Site Map | LIR Portal | About RIPE | Contact | © RIPE NCC. All rights reserved.



http://www.ris.ripe.net/cgi-bin/lg/index.cgi?rrc=RRC01&query=1&arg=202.173.147.0

RIS - Looking Glass

RRC Box: RRC01, LINX

Query:

- ● show ip bgp
- ○ show ip bgp summary
- ○ show bgp neighbors
- ○ show ip bgp regexp
- ○ show ipv6 bgp
- ○ show ipv6 bgp summary
- ○ show ipv6 bgp regexp
- ○ show version
- ○ traceroute
- ○ ping

Argument: 202.173.147.0

```
BGP routing table entry for 202.173.144.0/21
Paths: (4 available, best #3, table Default-IP-Routing-Table)
  Not advertised to any peer
  13237 1668 4648 2764 9543
    195.66.224.99 from 195.66.224.99 (82.197.136.1)
      Origin IGP, localpref 100, valid, external
      Community: 1668:31000 13237:44088 13237:46881
      Last update: Fri Jan 14 01:48:12 2005

  286 209 1239 4648 2764 9543
    195.66.224.54 from 195.66.224.54 (134.222.86.174)
      Origin IGP, localpref 100, valid, external
      Last update: Wed Jan  5 13:52:52 2005

  5511 10026 4648 2764 9543
    195.66.224.83 from 195.66.224.83 (193.251.245.1)
      Origin IGP, localpref 100, valid, external, best
      Last update: Mon Jan 17 02:15:07 2005

  8342 702 701 1239 4648 2764 9543
    195.66.224.90 from 195.66.224.90 (195.161.1.152)
      Origin IGP, localpref 100, valid, external
      Last update: Wed Dec 29 00:13:04 2004
```

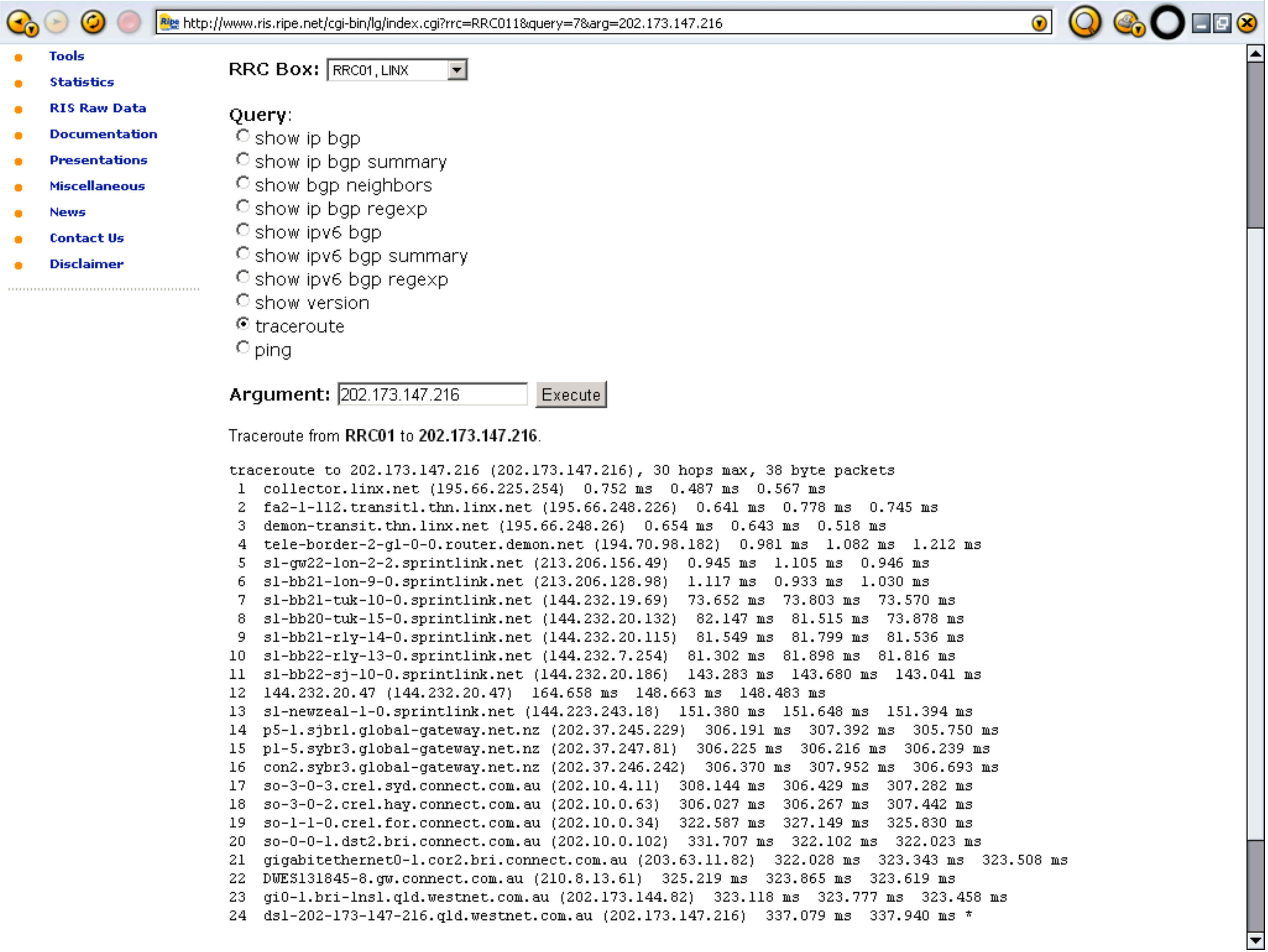
Multi-Router Looking Glass

Troubleshooting Connectivity – Example II

- Hmmm….
- Looking Glass can see 202.173.144.0/21

  This includes 202.173.147.0/24

  So the problem must be with AS3, or AS3's upstream
- A traceroute confirms the connectivity



RRC Box: RRC01, LINX

Query:

○ show ip bgp
○ show ip bgp summary
○ show bgp neighbors
○ show ip bgp regexp
○ show ipv6 bgp
○ show ipv6 bgp summary
○ show ipv6 bgp regexp
○ show version
● traceroute
○ ping

Argument: 202.173.147.216

Traceroute from **RRC01** to **202.173.147.216**.

```
traceroute to 202.173.147.216 (202.173.147.216), 30 hops max, 38 byte packets
 1  collector.linx.net (195.66.225.254)  0.752 ms  0.487 ms  0.567 ms
 2  fa2-1-112.transit1.thn.linx.net (195.66.248.226)  0.641 ms  0.778 ms  0.745 ms
 3  demon-transit.thn.linx.net (195.66.248.26)  0.654 ms  0.643 ms  0.518 ms
 4  tele-border-2-g1-0-0.router.demon.net (194.70.98.182)  0.981 ms  1.082 ms  1.212 ms
 5  sl-gw22-lon-2-2.sprintlink.net (213.206.156.49)  0.945 ms  1.105 ms  0.946 ms
 6  sl-bb21-lon-9-0.sprintlink.net (213.206.128.98)  1.117 ms  0.933 ms  1.030 ms
 7  sl-bb21-tuk-10-0.sprintlink.net (144.232.19.69)  73.652 ms  73.803 ms  73.570 ms
 8  sl-bb20-tuk-15-0.sprintlink.net (144.232.20.132)  82.147 ms  81.515 ms  73.878 ms
 9  sl-bb21-rly-14-0.sprintlink.net (144.232.20.115)  81.549 ms  81.799 ms  81.536 ms
10  sl-bb22-rly-13-0.sprintlink.net (144.232.7.254)  81.302 ms  81.898 ms  81.816 ms
11  sl-bb22-sj-10-0.sprintlink.net (144.232.20.186)  143.283 ms  143.680 ms  143.041 ms
12  144.232.20.47 (144.232.20.47)  164.658 ms  148.663 ms  148.483 ms
13  sl-newzeal-1-0.sprintlink.net (144.223.243.18)  151.380 ms  151.648 ms  151.394 ms
14  p5-1.sjbr1.global-gateway.net.nz (202.37.245.229)  306.191 ms  307.392 ms  305.750 ms
15  p1-5.sybr3.global-gateway.net.nz (202.37.247.81)  306.225 ms  306.216 ms  306.239 ms
16  con2.sybr3.global-gateway.net.nz (202.37.246.242)  306.370 ms  307.952 ms  306.693 ms
17  so-3-0-3.cre1.syd.connect.com.au (202.10.4.11)  308.144 ms  306.429 ms  307.282 ms
18  so-3-0-2.cre1.hay.connect.com.au (202.10.0.63)  306.027 ms  306.267 ms  307.442 ms
19  so-1-1-0.cre1.for.connect.com.au (202.10.0.34)  322.587 ms  327.149 ms  325.830 ms
20  so-0-0-1.dst2.bri.connect.com.au (202.10.0.102)  331.707 ms  322.102 ms  322.023 ms
21  gigabitethernet0-1.cor2.bri.connect.com.au (203.63.11.82)  322.028 ms  323.343 ms  323.508 ms
22  DWES131845-8.gw.connect.com.au (210.8.13.61)  325.219 ms  323.865 ms  323.619 ms
23  gi0-1.bri-lns1.qld.westnet.com.au (202.173.144.82)  323.118 ms  323.777 ms  323.458 ms
24  dsl-202-173-147-216.qld.westnet.com.au (202.173.147.216)  337.079 ms  337.940 ms *
```

Troubleshooting Connectivity – Example II

- Help is at hand – RouteViews
- The RouteViews router has BGP feeds from around 60 peers
  - www.routeviews.org explains the project
  - Gives access to a real router, and allows any provider to find out how their prefixes are seen in various parts of the Internet
  - Complements the Looking Glass facilities
- Anyway, back to our problem…

Troubleshooting Connectivity – Example II

- **Checklist:**

  Does AS3's upstream send it to AS3?

  We are checking eBGP configuration on AS3's upstream. There may be a configuration error with as-path filters, or prefix-lists, or communities such that only local prefixes get out. This needs AS3's assistance.

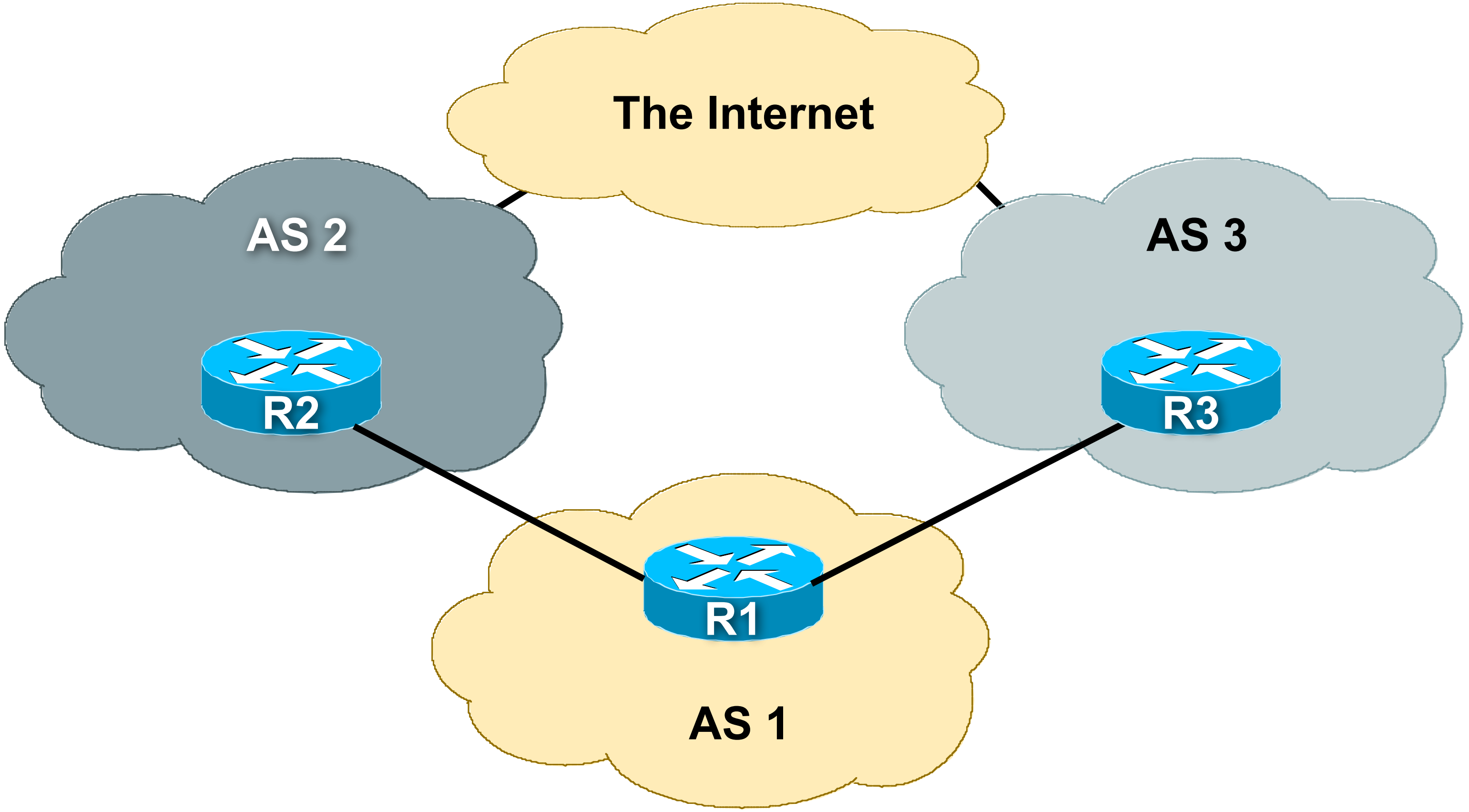
  Does AS3 see any of AS1's originated prefixes?

  We are checking eBGP configuration on R3. Maybe AS3 does not know to expect the prefix from AS1 in the peering with its upstream, or maybe it has some errors in as-path or prefix or community filters

Troubleshooting Connectivity – Example II

- Troubleshooting across the Internet is harder
  - But tools are available
- Looking Glasses, offering traceroute, ping and BGP status are available all over the globe
  - Most connectivity problems seem to be found at the edge of the network, rarely in the transit core
  - Problems with the transit core are usually intermittent and short term in nature

Troubleshooting Connectivity – Example III



- **Symptom: AS1 is trying to loadshare between its upstreams, but has trouble getting traffic through the AS2 link**

Troubleshooting Connectivity – Example III

- **Checklist:**

  What does “trouble” mean?

- **Is outbound traffic loadsharing okay?**

  Can usually fix this with selectively rejecting prefixes, and using local preference

  Generally easy to fix, local problem, simple application of policy

- **Is inbound traffic loadsharing okay?**

  Errummm, bigger problem if not

  Need to do some troubleshooting if configuration with communities, AS-PATH prepends, MEDs and selective leaking of subprefixes don’t seem to help

Troubleshooting Connectivity – Example III

- **Checklist:**

  AS1 announces, but does AS2 see it?

  We are checking eBGP filters on R1 and R2. Remember that R2 access will require cooperation and assistance from your peer

  Does AS2 see it over entire network?

  We are checking iBGP across AS2's network. Quite often iBGP is misconfigured, lack of full mesh, problems with RRs, etc.

Troubleshooting Connectivity – Example III

- **Checklist:**

  Does AS2 send it to its upstream?

  We are checking eBGP configuration on R2. There may be a configuration error with as-path filters, or prefix-lists, or communities such that only local prefixes get out

  Does the Internet see all of AS2's originated prefixes?

  We are checking eBGP configuration on other Internet routers. This means using looking glasses. And trying to find one as close to AS2 as possible.

Troubleshooting Connectivity – Example III

- **Checklist:**

  Repeat all of the above for AS3

- **Stopping here and resorting to a huge prepend towards AS3 won't solve the problem**

- **There are many common problems – listed on next slide**

  And tools to help decipher the problem

Troubleshooting Connectivity – Example III

- No inbound traffic from AS2

  AS2 is not seeing AS1’s prefix, or is blocking it in inbound filters

- A trickle of inbound traffic

  Switch on NetFlow (if the router has it) and check the origin of the traffic

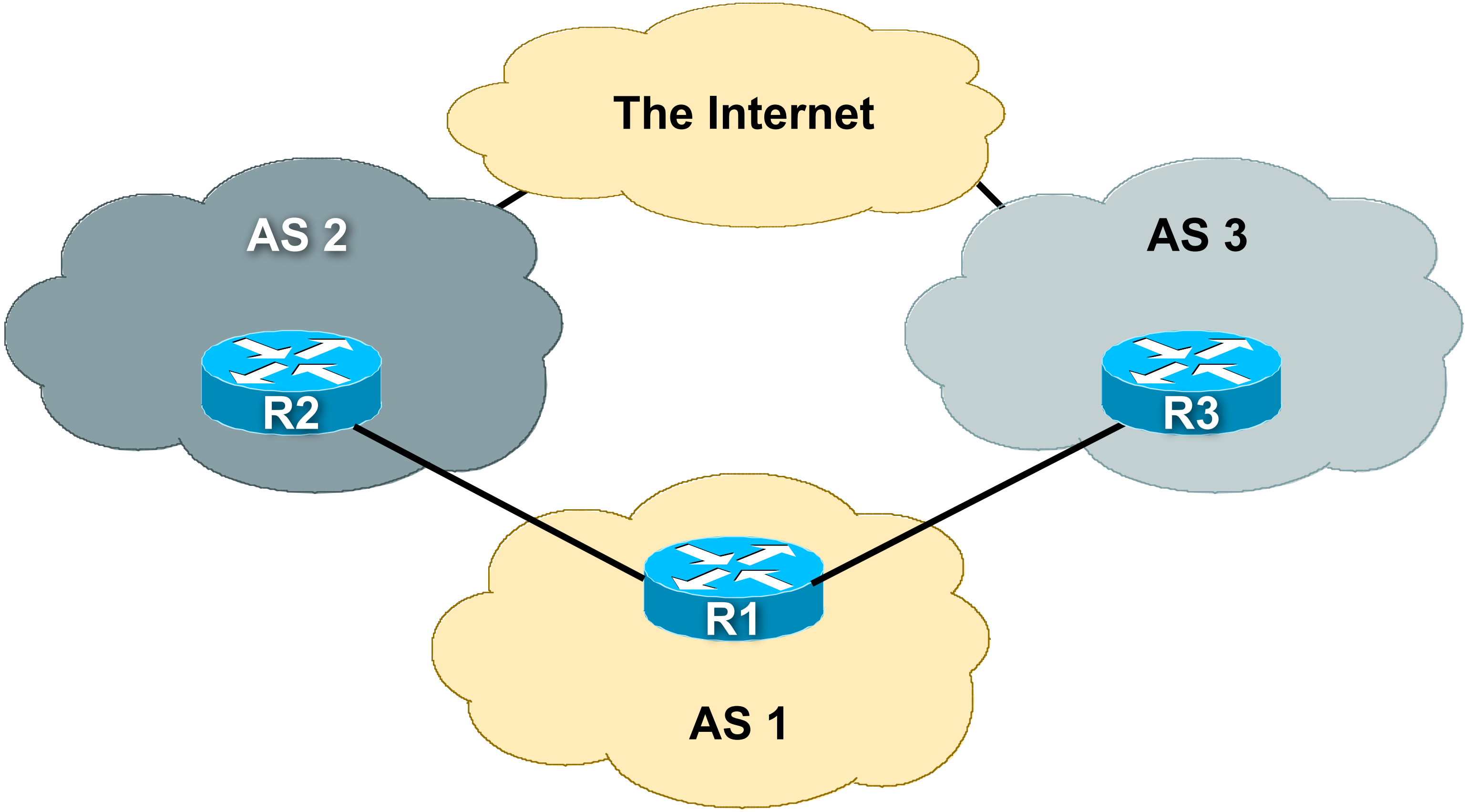
  If it is just from AS2’s network blocks, then is AS2 announcing the prefix to its upstreams?

  If they claim they are, ask them to ask their upstream for their BGP table – or use a Looking Glass to check

Troubleshooting Connectivity – Example III

- A light flow of traffic from AS2, but 50% less than from AS3
  - Looking Glass comes to the rescue
    - LG will let you see what AS2, or AS2's upstreams are announcing
    - AS1 may choose this as primary path, but AS2 relationship with their upstream may decide otherwise
  - NetFlow comes to the rescue
    - Allows AS1 to see what the origins are, and with the LG, helps AS1 to find where the prefix filtering culprit might be

Troubleshooting Connectivity – Example IV



- **Symptom: AS1 is loadsharing between its upstreams, but the traffic load swings randomly between AS2 and AS3**

Troubleshooting Connectivity – Example IV

- **Checklist:**

  Assume AS1 has done everything in this tutorial so far

  All the configurations look fine, the Looking Glass outputs look fine, life is wonderful… Apart from those annoying traffic swings every hour or so

  L2 problem? Route Flap Damping?

  Since BGP is configured fine, and the net has been stable for so long, can only be an L2 problem, or Route Flap Damping side-effect

Troubleshooting Connectivity – Example IV

- **L2 – upstream somewhere has poor connectivity between themselves and the rest of the Internet**
  - **Only real solution is to impress upon upstream that this isn't good enough, and get them to fix it**
  - **Or change upstreams**

Troubleshooting Connectivity – Example IV

- **Route Flap Damping**

  Many ISPs implement route flap damping

  Many ISPs simply use the vendor defaults

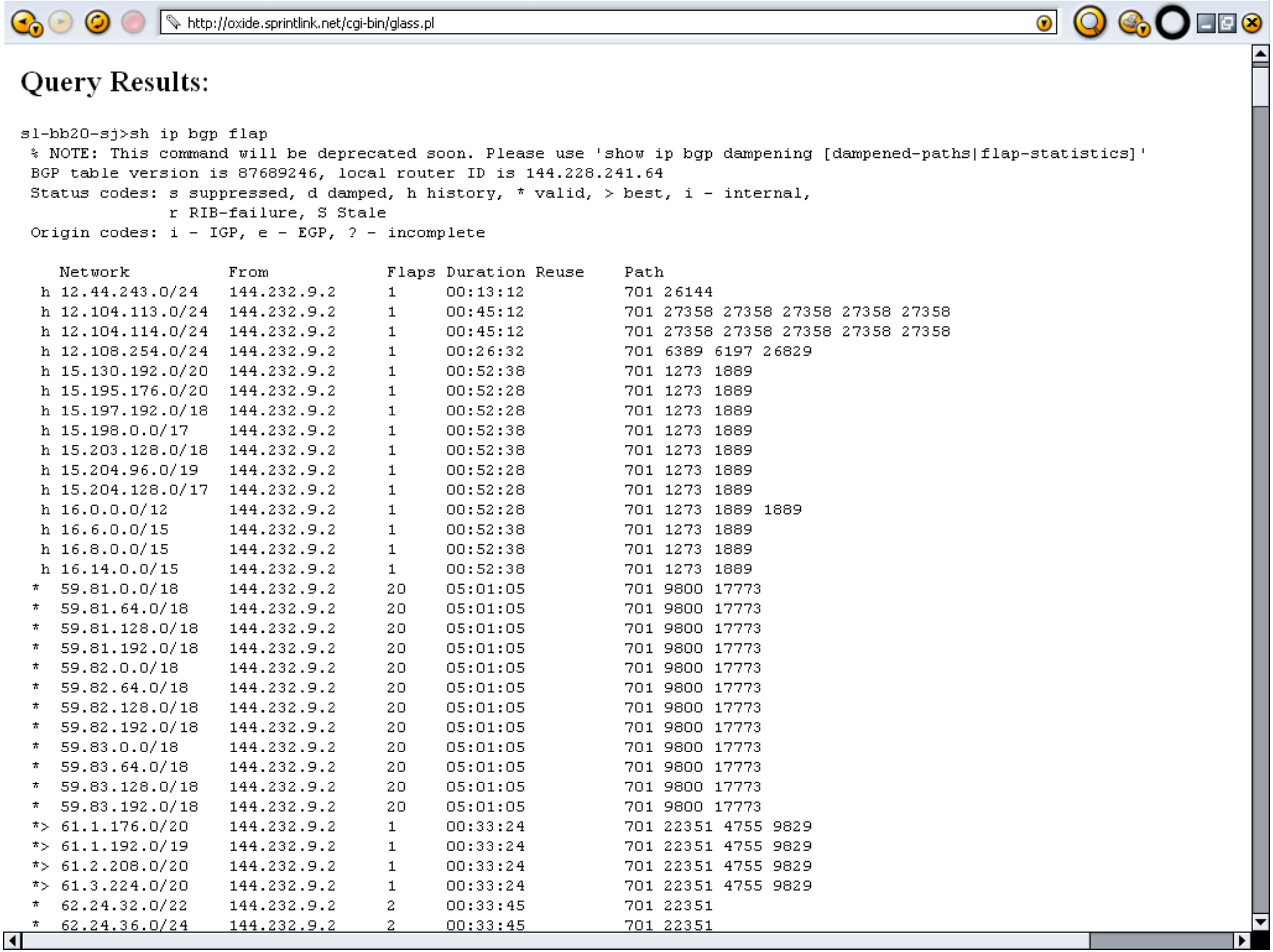
  Vendor defaults are generally far too severe

  There is real concern that the "more lenient" RIPE-229 values are too severe

  Opinion is growing that flap damping does more harm than good

  e.g. www.cs.berkeley.edu/~zmao/Papers/sig02.pdf

- **Again Looking Glasses come to the operator's assistance**



Query Results:

```
sl-bb20-sj>sh ip bgp flap
 % NOTE: This command will be deprecated soon. Please use 'show ip bgp dampening [dampened-paths|flap-statistics]'
 BGP table version is 87689246, local router ID is 144.228.241.64
 Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
 Origin codes: i - IGP, e - EGP, ? - incomplete

    Network          From            Flaps Duration Reuse    Path
  h 12.44.243.0/24   144.232.9.2      1     00:13:12          701 26144
  h 12.104.113.0/24  144.232.9.2      1     00:45:12          701 27358 27358 27358 27358 27358
  h 12.104.114.0/24  144.232.9.2      1     00:45:12          701 27358 27358 27358 27358 27358
  h 12.108.254.0/24  144.232.9.2      1     00:26:32          701 6389 6197 26829
  h 15.130.192.0/20  144.232.9.2      1     00:52:38          701 1273 1889
  h 15.195.176.0/20  144.232.9.2      1     00:52:28          701 1273 1889
  h 15.197.192.0/18  144.232.9.2      1     00:52:28          701 1273 1889
  h 15.198.0.0/17    144.232.9.2      1     00:52:38          701 1273 1889
  h 15.203.128.0/18  144.232.9.2      1     00:52:38          701 1273 1889
  h 15.204.96.0/19   144.232.9.2      1     00:52:28          701 1273 1889
  h 15.204.128.0/17  144.232.9.2      1     00:52:28          701 1273 1889
  h 16.0.0.0/12      144.232.9.2      1     00:52:28          701 1273 1889 1889
  h 16.6.0.0/15      144.232.9.2      1     00:52:38          701 1273 1889
  h 16.8.0.0/15      144.232.9.2      1     00:52:38          701 1273 1889
  h 16.14.0.0/15     144.232.9.2      1     00:52:38          701 1273 1889
 *  59.81.0.0/18     144.232.9.2      20    05:01:05          701 9800 17773
 *  59.81.64.0/18    144.232.9.2      20    05:01:05          701 9800 17773
 *  59.81.128.0/18   144.232.9.2      20    05:01:05          701 9800 17773
 *  59.81.192.0/18   144.232.9.2      20    05:01:05          701 9800 17773
 *  59.82.0.0/18     144.232.9.2      20    05:01:05          701 9800 17773
 *  59.82.64.0/18    144.232.9.2      20    05:01:05          701 9800 17773
 *  59.82.128.0/18   144.232.9.2      20    05:01:05          701 9800 17773
 *  59.82.192.0/18   144.232.9.2      20    05:01:05          701 9800 17773
 *  59.83.0.0/18     144.232.9.2      20    05:01:05          701 9800 17773
 *  59.83.64.0/18    144.232.9.2      20    05:01:05          701 9800 17773
 *  59.83.128.0/18   144.232.9.2      20    05:01:05          701 9800 17773
 *  59.83.192.0/18   144.232.9.2      20    05:01:05          701 9800 17773
 *> 61.1.176.0/20    144.232.9.2      1     00:33:24          701 22351 4755 9829
 *> 61.1.192.0/19    144.232.9.2      1     00:33:24          701 22351 4755 9829
 *> 61.2.208.0/20    144.232.9.2      1     00:33:24          701 22351 4755 9829
 *> 61.3.224.0/20    144.232.9.2      1     00:33:24          701 22351 4755 9829
 *  62.24.32.0/22    144.232.9.2      2     00:33:45          701 22351
 *  62.24.36.0/24    144.232.9.2      2     00:33:45          701 22351
```

Troubleshooting Connectivity – Example IV

- **Most Looking Glasses allow the operators to check the flap or damped status of their announcements**
    - **Many oscillating connectivity issues are usually caused by L2 problems**
    - **Route flap damping will cause connectivity to persist via alternative paths even though primary paths have been restored**
    - **Quite often, the exponential back off of the flap damping timer will give rise to bizarre routing**
        - **Common symptom is that bizarre routing will often clear away by itself**

Troubleshooting Summary

- Most troubleshooting is about:
- Experience
    - Recognising the common problems
- Not panicking
- Logical approach
    - Check configuration first
    - Check locally first before blaming the peer
    - Troubleshoot layer 1, then layer 2, then layer 3, etc

Troubleshooting Summary

- **Most troubleshooting is about:**
- **Using the available tools**
  - **The debugging tools on the router hardware**
  - **Internet Looking Glasses**
  - **Colleagues and their knowledge**
  - **Public mailing lists where appropriate**

Closing Comments

- **Presentation has covered the most common troubleshooting techniques used by ISPs today**
- **Once these have been mastered, more complex or arcane problems are easier to solve**
- **Feedback and input for future improvements is encouraged and very welcome**

Troubleshooting BGP

Philip Smith <pfs@cisco.com>

APRICOT 2006

22 Feb - 3 Mar 2006

Perth, Australia