

Troubleshooting BGP

Philip Smith <pfs@cisco.com>

NANOG 29, Chicago, October 2003

Presentation Slides

Cisco.com

- **Available on**

<ftp://ftp-eng.cisco.com/pfs/seminars/NANOG29-BGP-Troubleshooting.pdf>

<http://www.nanog.org/mtg-0310/pdf/smith.pdf>

Assumptions

- **Presentation assumes working knowledge of BGP**
- **Please feel free to ask questions at any time!**

Agenda

Cisco.com

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

Fundamentals: Problem Recognition

- **First step is to recognise what causes the problem**

BUT

- **Newcomers to BGP usually enter minor panic at this stage:**
 - BGP determines network connectivity**
 - Break BGP, and connectivity breaks**
 - Break connectivity, and customers complain**
- **The result is that many problems languish in the network, or have (often bizarre) “sticking plaster” workarounds**

Fundamentals: Problem Recognition

- **The best troubleshooter is the one who learns from:**
 - Experience**

fixing one problem leads to greater confidence at tackling the next
 - Mistakes**

We all learn from our mistakes – and troubleshooting does involve making lots of mistakes. But you'll get better at it!
 - Others**

Listen to what other operators say – plenty of BGP problem analysis on various lists
- **And the best troubleshooter creates some basic troubleshooting principles, based on what they've learned**

Fundamentals: Problem Areas

- **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 or executed maintenance activities, plus the above

Technical

Problems with hardware, software, inter-router link loads affecting protocol stability

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

- **Familiarise yourself with the routers tools:**

Is logging of the BGP process enabled?

Are the logs being stored somewhere useful

And do you know what the logs mean?

Are you familiar with the BGP debug process and commands (if available)

Check vendor documentation and operational recommendations before switching on full BGP debugging – you might get fewer surprises

Agenda

Cisco.com

- **Fundamentals**
- **Local Configuration Problems**
- **Internet Reachability Problems**

Local Configuration Problems

Cisco.com

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

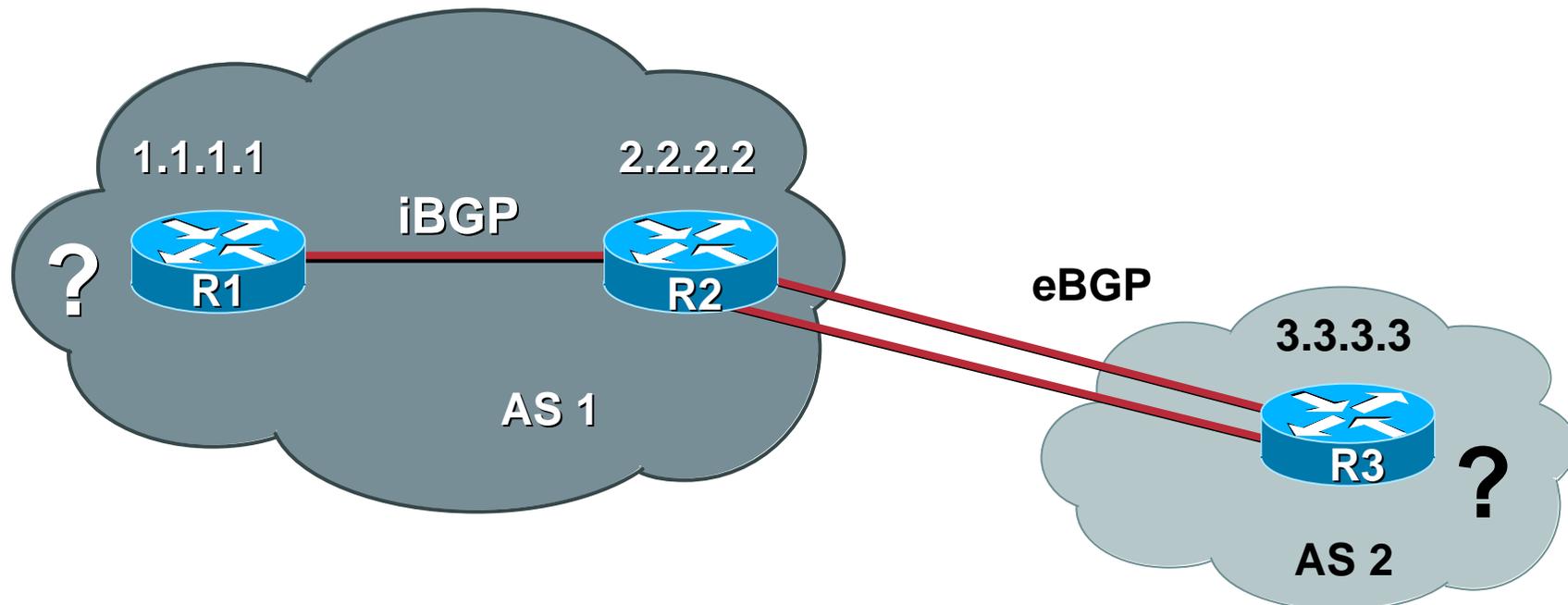
Peer Establishment: ACLs and Connectivity

- **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**
 - Link saturation problems**
 - CPU utilisation 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 Problems

- **Assume that IP connectivity has been checked**
- **Check TCP to find out what connections we are accepting**
 - Check the ports (TCP/179)**
 - Check 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 P**
 - 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?**

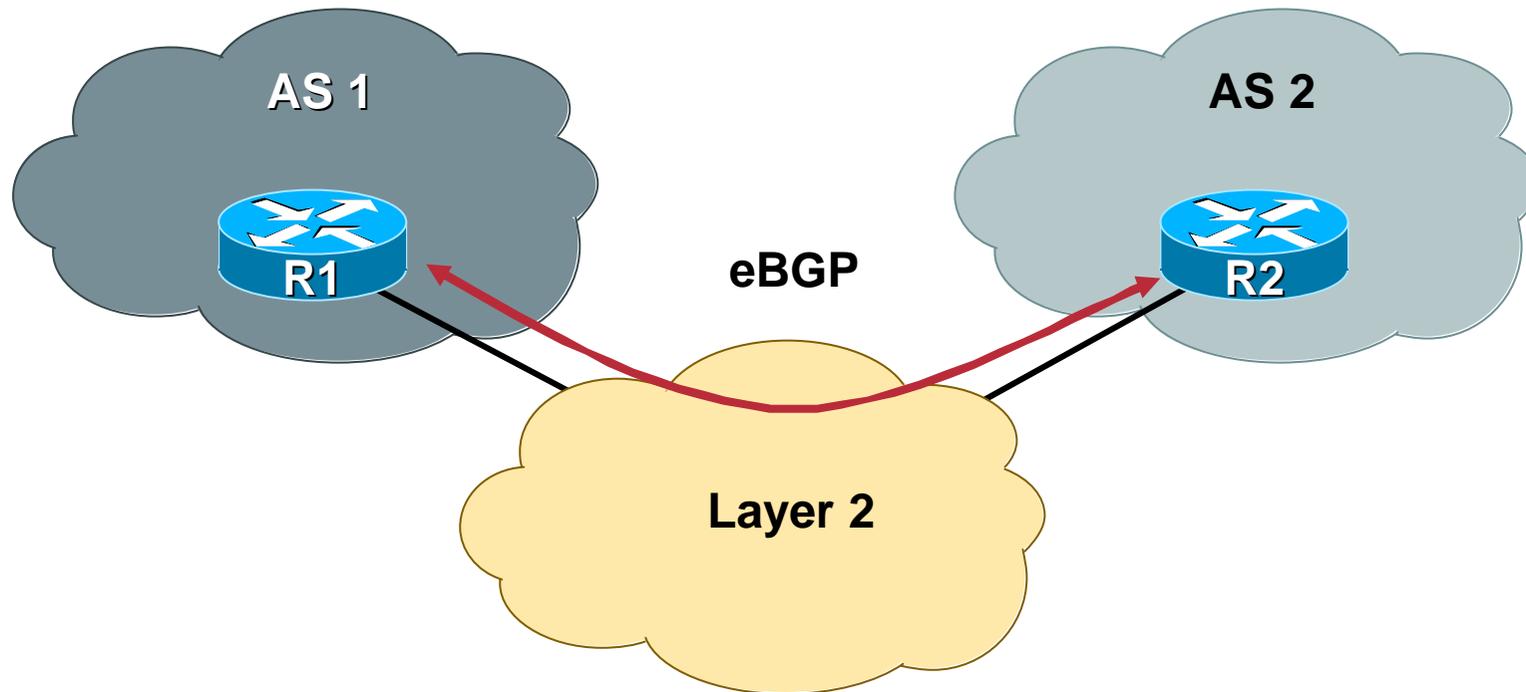
If this crosses multiple providers, this needs their cooperation

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

Cisco.com



- **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 clues**
- **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 can 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 \Rightarrow 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

Local Configuration Problems

Cisco.com

- 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
(assuming we are not using route-reflectors or BGP confederations)

Missing Routes—Agenda

Cisco.com

- **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

Cisco.com

- **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?!

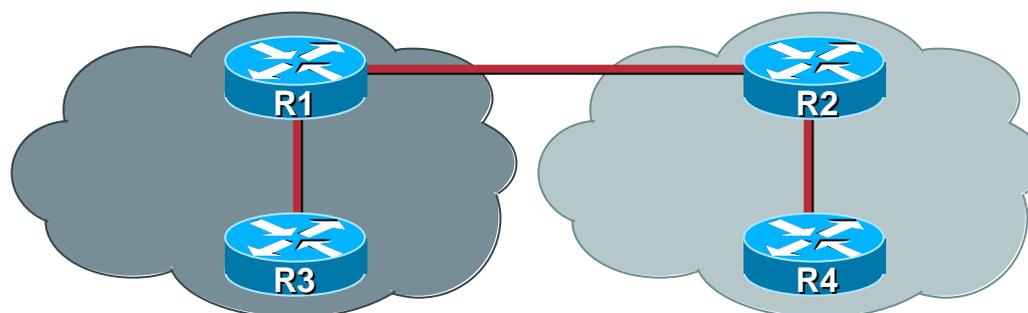
- **Common issues**

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

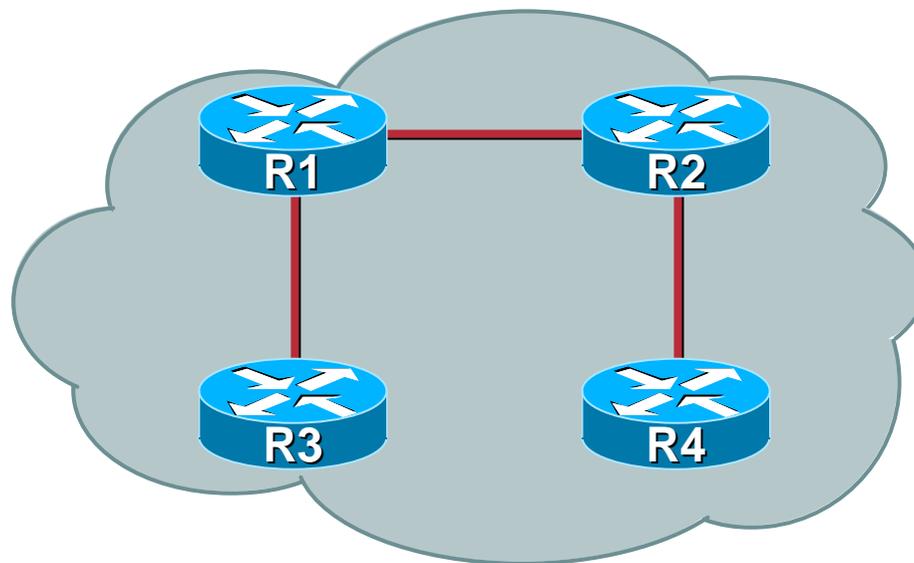
Router-ID is usually calculated automatically by router

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**

Some early documentation claimed that RR redundancy could only 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 RR clusters for redundancy, and stay with 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

Update Filtering: Common Problems

- **Communities (more)**

Each ISP has different policies

Never assume that because communities exist that your peers will use them

Often peers will advertise that they support RFC1998-style communities – worthwhile confirming this before you use them!

Never assume that your peers will pay attention to the communities you send

The “no-export” problem – just because you send a prefix with “no-export” set does not mean that your neighbour will obey it. Cooperation, not assumption

Missing Routes: General Problems

- **Make and then Stay with 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

Cisco.com

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

Missing Routes—iBGP

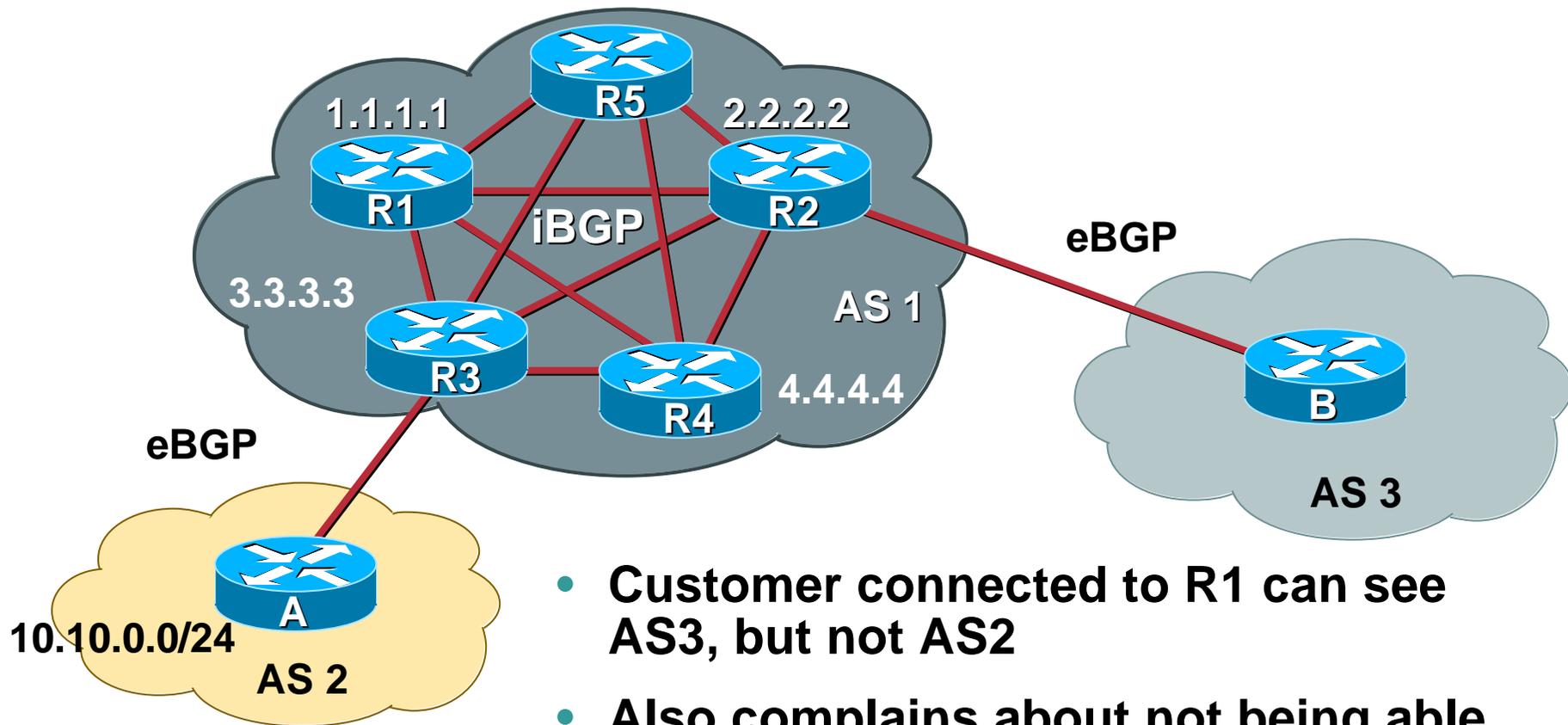
Cisco.com

- **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 \Rightarrow R3 is somehow not passing routing information about AS2 to R1

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

But can see rest of the Internet \Rightarrow 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

Cisco.com

- **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 defines 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**

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

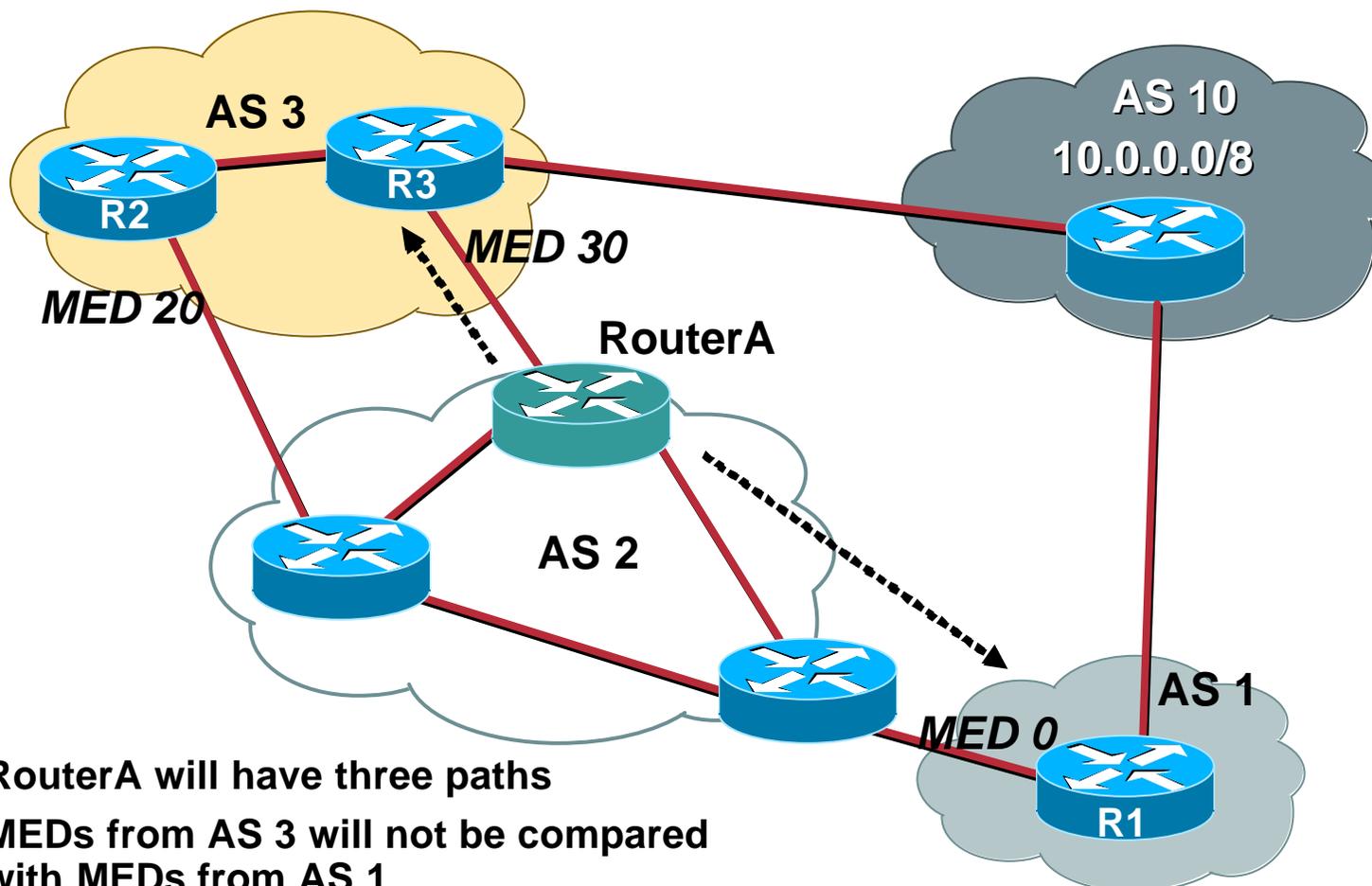
- **Difficult to catch and troubleshoot**

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

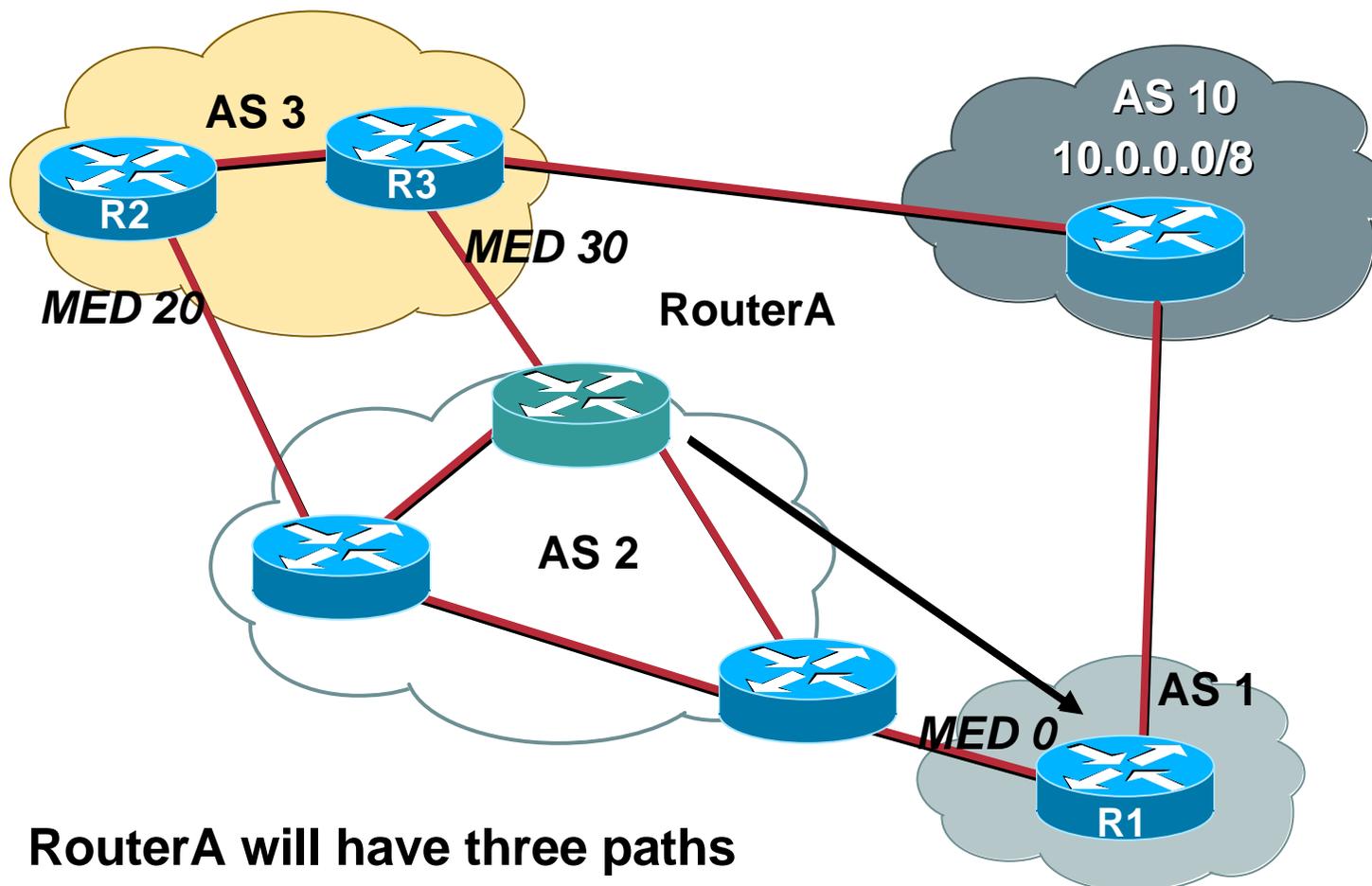


- RouterA will have three paths
- MEDs from AS 3 will not be compared with MEDs from AS 1
- RouterA will sometimes select the path from R1 as best and but may also select the path from R3 as best

Deterministic MED—Operation

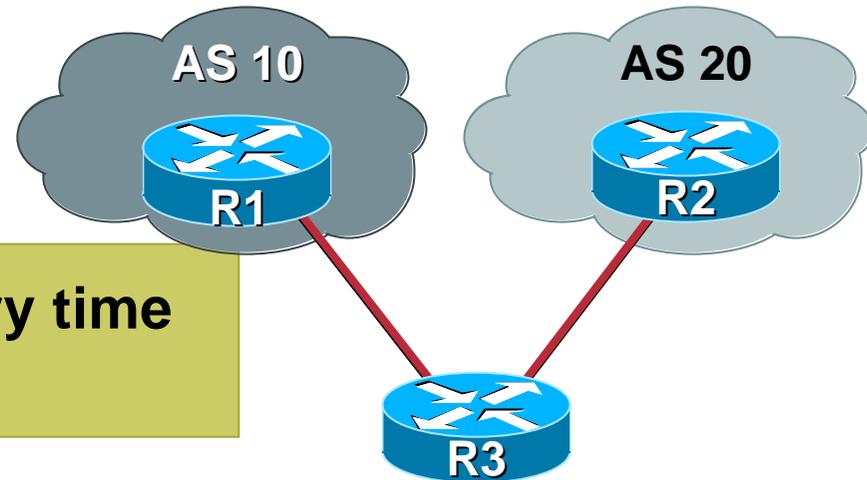
- **The paths are ordered by Neighbour AS**
- **The bestpath for each Neighbour AS group is selected**
- **The overall bestpath results from comparing the winners from each group**
- **The bestpath will be consistent because paths will be placed in a deterministic order**

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 isn't perfect when it comes to path selection – years of operational experience have shown this

Vendors and ISPs have worked to put in stability enhancements

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

Cisco.com

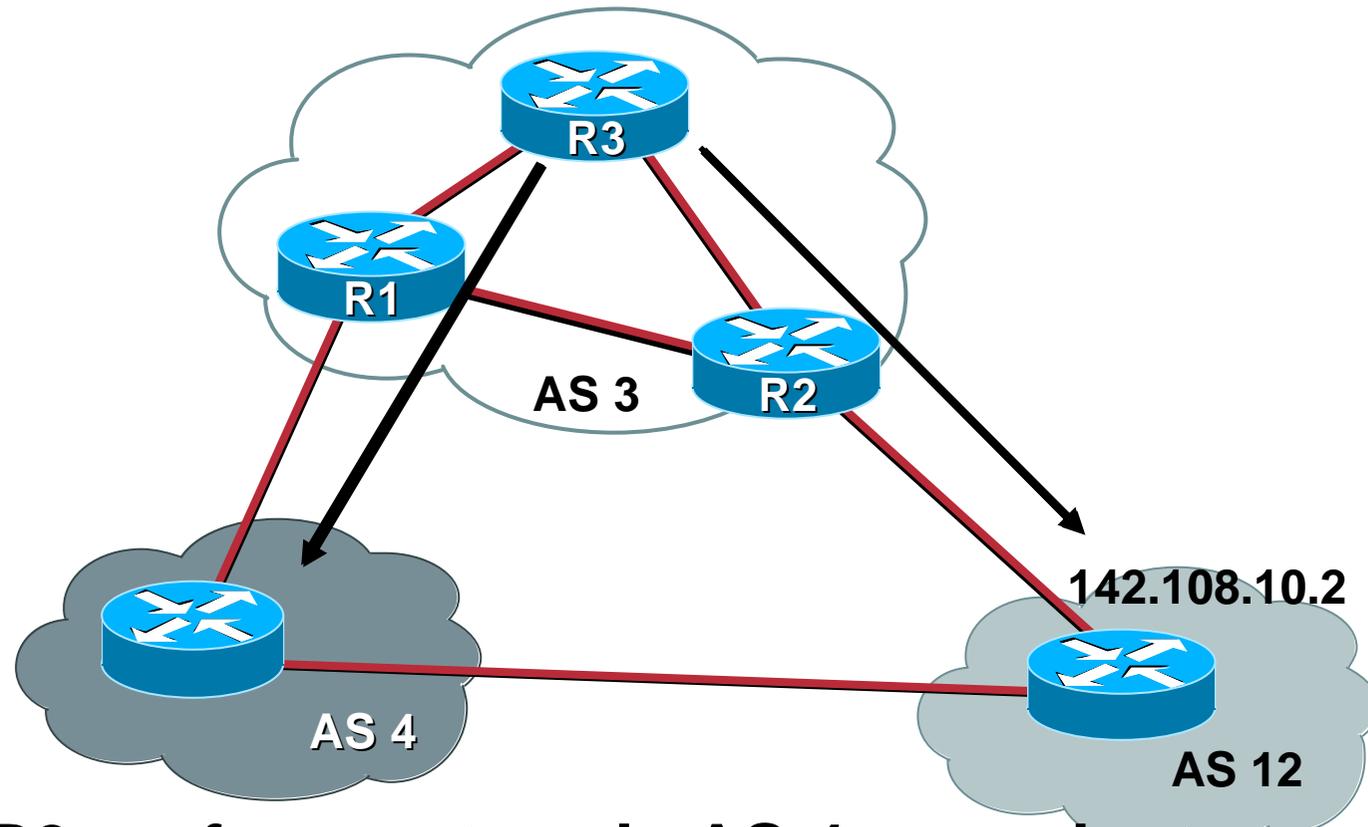
- **Peer Establishment**
- **Missing Routes**
- **Inconsistent Route Selection**
- **Loops and Convergence Issues**

Route Oscillation

- **One of the most common problems!**
- **Every minute routes flap in the routing table from one nexthop to another**
- **With full routes the most obvious symptom is high CPU in “BGP Router” process**

Route Oscillation—Diagram

Cisco.com



- 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—Symptom

- **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—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**

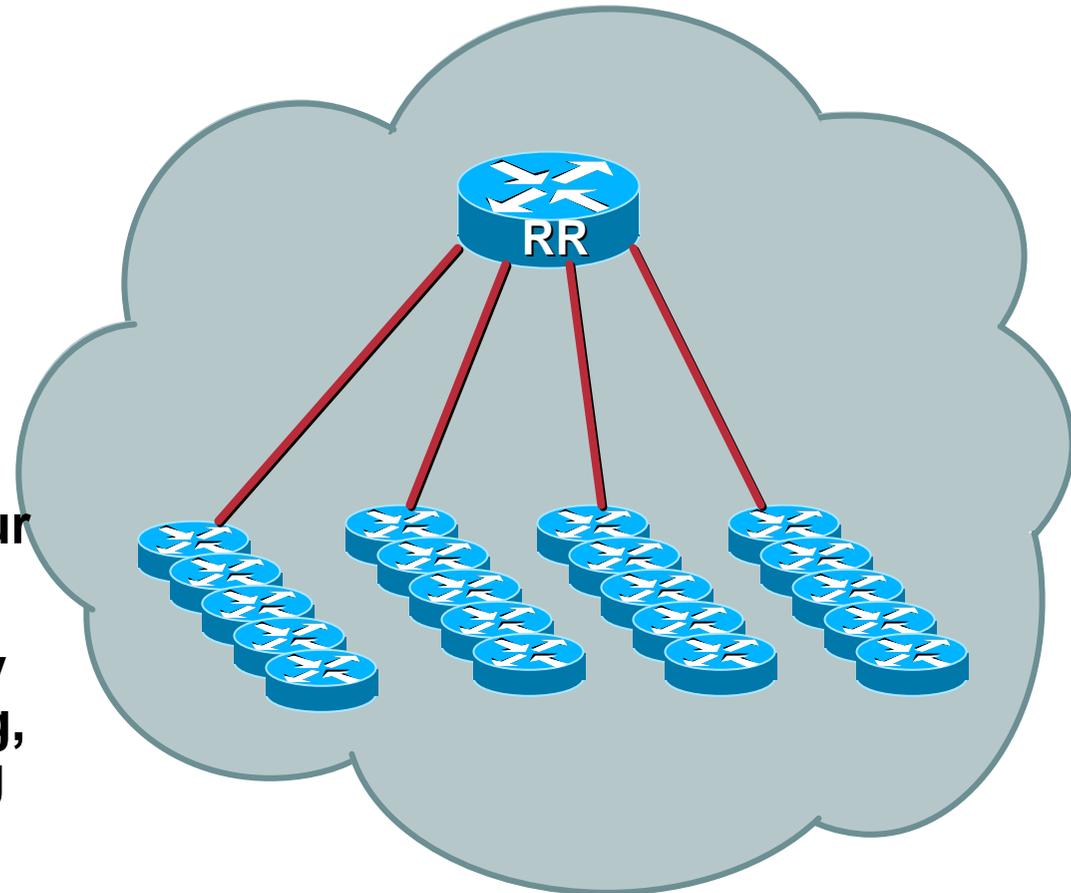
Troubleshooting Tips

- **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

Cisco.com

- **Fundamentals**
- **Local Configuration Problems**
- **Internet Reachability Problems**

Internet Reachability Problems

- **BGP Attribute Confusion**

- To Control Traffic in (R) Send MEDs and AS-PATH prepends on outbound announcements**

- To Control Traffic out (R) 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

Cisco.com

- **BGP Path Selection Process**

Each vendor has “tweaked” the path selection process

Know it, learn it, for your router equipment – saves time later

- **MED confusion**

Default MED on Cisco IOS is ZERO – it may not be this on your router, or your peer’s router

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 your BGP path selection algorithm

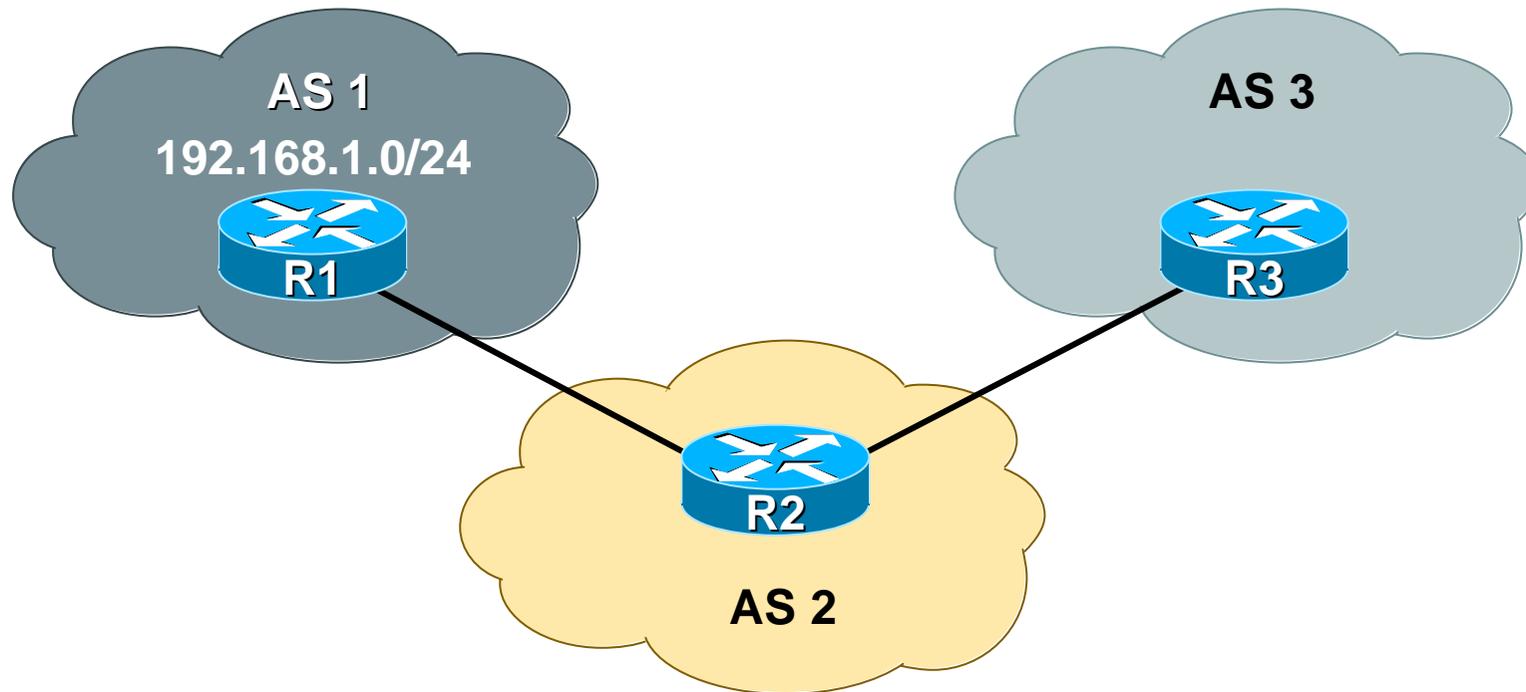
Some ISPs use `bgp maxas-path 15` to drop prefixes with ridiculously long AS-paths

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

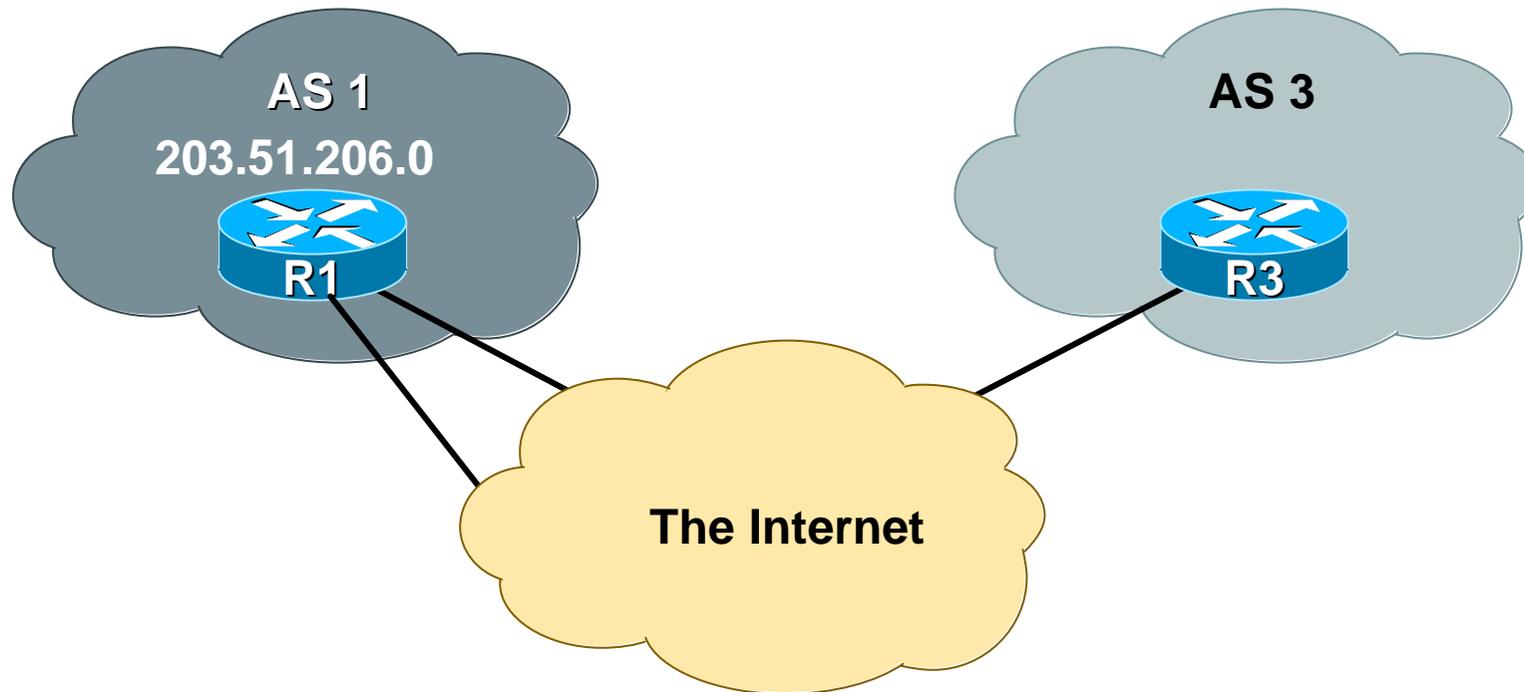
Troubleshooting Connectivity – Example I

- **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 203.51.206.0/24 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

www.traceroute6.org for IPv6

- **Many still use the original: nitrous.digex.net**
- **Next slides have some examples of a typical looking glass in action**



[homepage](#) | [what's new](#) | [whois db](#) | [search](#) | [site map](#) | [f.a.q.](#)

RIS - Looking Glass

RRC Box:

- RRC00, Amsterdam
- RRC01, LINX
- RRC02, SFINX
- RRC03, AMS-IX
- RRC04, CIXP
- RRC05, VIX
- RRC06, NSPIXP2
- RRC07, Netnod
- RRC08, MAE-West

Query:

- bgp
- bgp sum
- bgp neig
- bgp rege
- bgp path
- version
- traceroute
- ping

Argument:

Multi-Router Looking Glass version 3.3.2 Beta

Written by: John Fraizer - [EnterZone, Inc](#)

[Contact Webmaster](#)

Copyright © RIPE NCC

[Mail RIPE NCC](#)



RIS Looking Glass - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://www.ris.ripe.net/cgi-bin/lg/index.cgi

RRC Box: RRC01, LINX

Query:

- bgp
- bgp summary
- bgp neighbors
- bgp regexp
- bgp paths
- version
- traceroute
- ping

Argument: 203.51.206.0

BGP routing table entry for 203.48.0.0/14
Paths: (3 available, best #2, table Default-IP-Routing-Table)
Not advertised to any peer

```
13237 1 4637 1221, (aggregated by 1221 203.62.252.26)
  195.66.224.99 from 195.66.224.99 (80.245.35.6)
    Origin IGP, localpref 100, valid, external, atomic-aggregate
    Community: 13237:44693
    Last update: Fri Oct 18 09:24:43 2002
```

```
286 209 4637 1221, (aggregated by 1221 203.62.252.26)
  195.66.224.54 from 195.66.224.54 (134.222.86.174)
    Origin IGP, localpref 100, valid, external, atomic-aggregate, best
    Last update: Wed Oct 16 18:16:29 2002
```

```
8406 8210 1239 4637 1221, (aggregated by 1221 203.62.252.26)
  195.66.226.71 from 195.66.226.71 (62.72.156.25)
    Origin IGP, metric 0, localpref 100, valid, external, atomic-aggregate
    Last update: Tue Oct 15 14:33:06 2002
```

rrc01

Multi-Router Looking Glass version 3.3.2 Beta
Written by John Fraizer, EnterZone, Inc.

Document: Done (2,033 secs)

Troubleshooting Connectivity – Example II

- **Hmmm....**
- **Looking Glass can see 203.48.0.0/14**
This includes 203.51.206.0/24
So the problem must be with AS3, or AS3's upstream
- **A traceroute confirms the connectivity**

RIS Looking Glass - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://www.ris.ripe.net/cgi-bin/lg/index.cgi

RIS - Looking Glass

RRC Box:

Query:

- bgp
- bgp summary
- bgp neighbors
- bgp regexp
- bgp paths
- version
- traceroute
- ping

Argument:

Traceroute from **RRC011** to **203.51.206.206**.

```
traceroute to 203.51.206.206 (203.51.206.206), 30 hops max, 40 byte packets
 1  collector.linx.net (195.66.225.254)  0.629 ms  0.580 ms  0.607 ms
 2  195.66.224.166 (195.66.224.166)  0.615 ms  0.464 ms  0.487 ms
 3  pos3-0.lnx01.London.net.reach.com (202.40.148.34)  0.685 ms  0.680 ms  0.585 ms
 4  202.84.143.138 (202.84.143.138)  72.142 ms  72.283 ms  72.363 ms
 5  202.84.143.85 (202.84.143.85)  104.273 ms  104.375 ms  104.274 ms
 6  202.84.143.57 (202.84.143.57)  170.109 ms  170.195 ms  169.894 ms
 7  202.84.143.22 (202.84.143.22)  357.587 ms  357.811 ms  357.514 ms
 8  GigabitEthernet1-2.pad-core4.Sydney.telstra.net (203.50.13.245)  357.674 ms  359.339 ms  357.256 ms
 9  Pos12-0.ken-core4.Sydney.telstra.net (203.50.6.21)  357.431 ms  357.326 ms  357.311 ms
10  Pos4-0.woo-core1.Brisbane.telstra.net (203.50.6.222)  368.377 ms  369.083 ms  371.041 ms
11  GigabitEthernet1-2.cha23.Brisbane.telstra.net (203.50.50.33)  369.113 ms  368.800 ms  370.978 ms
12  tcbroad1.lnk.telstra.net (139.130.193.118)  368.616 ms  369.616 ms  368.675 ms
13  gigabit-msfc1.qld-remote.bigpond.net.au (61.9.209.4)  368.455 ms  368.517 ms  370.156 ms
14  * * *
15  * * *
16  CPE-203-51-206-206.qld.bigpond.net.au (203.51.206.206)  387.364 ms  387.103 ms *
```

Document: Done (73.836 secs)

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

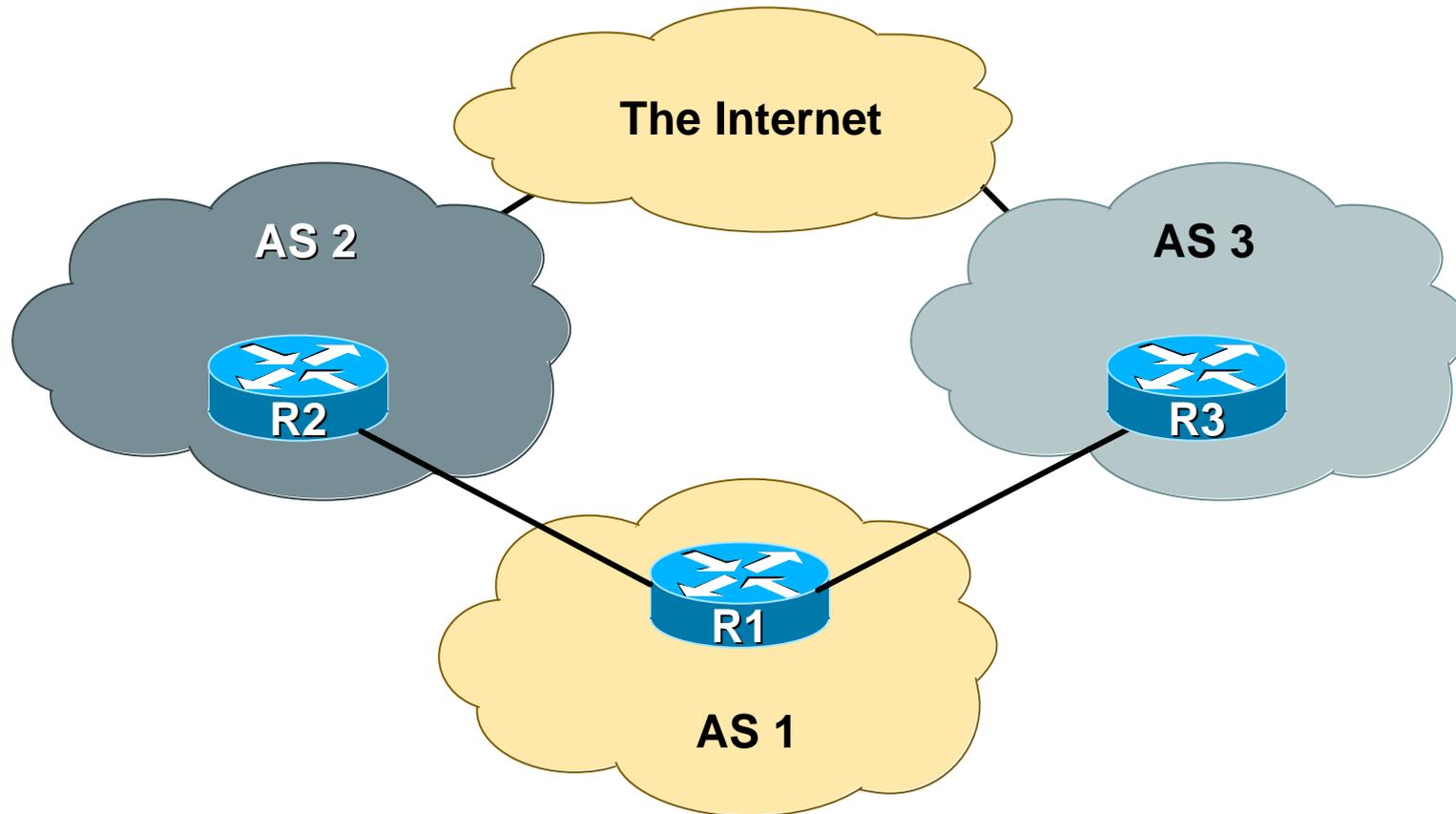
Troubleshooting Connectivity – Example II

- **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

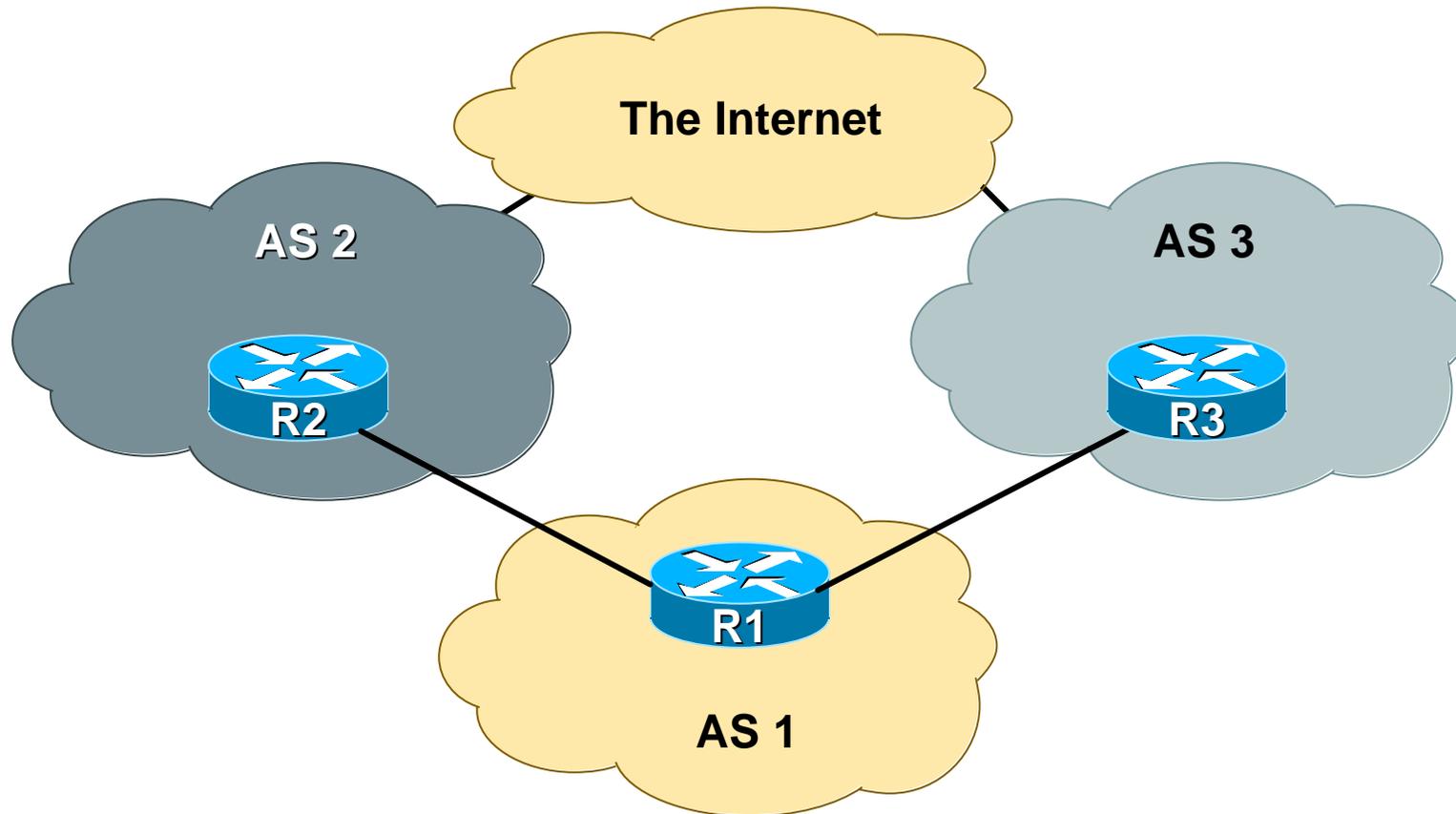
LG will let you see what AS2, or AS2's upstreams are announcing

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 even now some real concern that the “more lenient” RIPE-229 values are too severe

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

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

Mozilla

File Edit View Go Bookmarks Tools Window Help

http://nitrous.digex.net/cgi-bin/looking_glass.pl

h		198.32.176.38	1	00:04:36	6453	4755														
h	202.86.170.0	216.200.249.89	1	00:03:25	11466	6461	1239	6453	4755											
h		198.32.176.38	1	00:04:36	6453	4755														
h	202.86.174.0	216.200.249.89	1	00:03:25	11466	6461	1239	6453	4755											
h		198.32.176.38	1	00:04:36	6453	4755														
h	202.88.128.0/20	198.32.176.38	2	00:06:57	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.129.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.130.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.131.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.133.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.134.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.135.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.136.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.137.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.138.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.139.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.140.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.141.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.142.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.143.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.160.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.160.0/21	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488						
h	202.88.161.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.162.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.163.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.164.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.165.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.166.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
h	202.88.167.0	198.32.176.38	2	00:06:58	6453	4755	4755	4755	4755	4755	4755	4755	4755	17488	17488					
*>	202.133.39.0	198.32.176.50	1	00:07:53	7473	17557														
*>	202.133.44.0	198.32.176.50	1	00:07:53	7473	17557														
h	202.134.192.0/22	198.32.176.50	2	00:15:39	7473	10029	10029	10029	10029	10029	10029	10029	10029	10029	10029					
h	202.134.196.0/22	198.32.176.50	2	00:15:39	7473	10029	10029	10029	10029	10029	10029	10029	10029	10029	10029					
h	202.134.197.0	198.32.176.38	2	00:14:30	6453	4755	10029	10029	10029	10029	10029	10029	10029	10029	10029					
h		198.32.176.50	2	00:15:39	7473	10029	10029	10029	10029	10029	10029	10029	10029	10029	10029					
h	202.134.200.0/22	198.32.176.50	2	00:15:39	7473	10029	10029	10029	10029	10029	10029	10029	10029	10029	10029					
h	202.134.204.0/22	198.32.176.50	2	00:15:39	7473	10029	10029	10029	10029	10029	10029	10029	10029	10029	10029					
h	202.140.142.0	216.200.249.89	1	00:04:19	11466	6461	1239	6453	4755	9910										
h		198.32.176.38	1	00:05:32	6453	4755	9910													

Document: Done (0.731 secs)

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

Agenda

Cisco.com

- **Fundamentals**
- **Local Configuration Problems**
- **Internet Reachability Problems**

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**

Presentation Slides

Cisco.com

- **Available on**

<ftp://ftp-eng.cisco.com/pfs/seminars/NANOG29-BGP-Troubleshooting.pdf>

<http://www.nanog.org/mtg-0310/pdf/smith.pdf>

Troubleshooting BGP

The End! 😊