BGP Multihoming Techniques

Philip Smith <pfs@cisco.com>

APNIC 22

4th-8th September 2006

Kaohsiung, Taiwan

Presentation Slides

- **Available on**

  ftp://ftp-eng.cisco.com

  /pfs/seminars/APNIC22-BGP-part3.pdf

  And on the APNIC 22 website

- **Feel free to ask questions any time**
- **Aimed at Service Providers**

  Techniques can be used by many enterprises too

BGP Multihoming Techniques

- **Why Multihome?**
- **Definition & Options**
- **Basic Multihoming**
- **Service Provider Multihoming**
- **Using Communities**

Why Multihome?

It's all about redundancy, diversity & reliability

Why Multihome?

- **Redundancy**

  One connection to internet means the network is dependent on:

  Local router (configuration, software, hardware)

  WAN media (physical failure, carrier failure)

  Upstream Service Provider (configuration, software, hardware)

Why Multihome?

- **Reliability**

  Business critical applications demand continuous availability

  Lack of redundancy implies lack of reliability implies loss of revenue

Why Multihome?

- **Supplier Diversity**
  - **Many businesses demand supplier diversity as a matter of course**
  - **Internet connection from two or more suppliers**
    - **With two or more diverse WAN paths**
    - **With two or more exit points**
    - **With two or more international connections**
    - **Two of everything**

Why Multihome?

- **Not really a reason, but oft quoted…**
- **Leverage:**
  - **Playing one ISP off against the other for:**
    - **Service Quality**
    - **Service Offerings**
    - **Availability**

Why Multihome?

- **Summary:**

  Multihoming is easy to demand as requirement for any service provider or end-site network

  But what does it really mean:

    In real life?

    For the network?

    For the Internet?

  And how do we do it?

BGP Multihoming Techniques

- Why Multihome?
- Definition & Options
- Basic Multihoming
- Service Provider Multihoming
- Using Communities

Multihoming: Definitions & Options

What does it mean, what do we need, and how do we do it?

Multihoming Definition

- More than one link external to the local network
  - two or more links to the same ISP
  - two or more links to different ISPs
- Usually two external facing routers
  - one router gives link and provider redundancy only

AS Numbers

- An Autonomous System Number is required by BGP
- Obtained from upstream ISP or Regional Registry (RIR)
  - AfriNIC, APNIC, ARIN, LACNIC, RIPE NCC
- Necessary when you have links to more than one ISP or to an exchange point
- 16 bit integer, ranging from 1 to 65534
  - Zero and 65535 are reserved
  - 64512 through 65534 are called Private ASNs

Private-AS – Application

- **Applications**

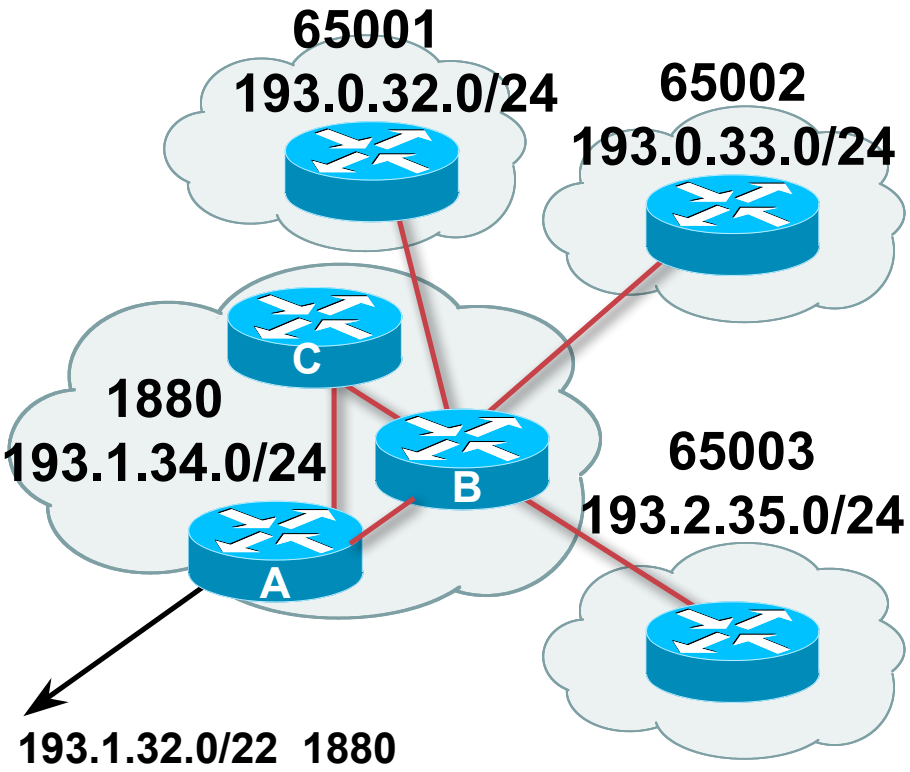
  An ISP with customers multihomed on their backbone (RFC2270)

  -or-

  A corporate network with several regions but connections to the Internet only in the core

  -or-

  Within a BGP Confederation



Private-AS – Removal

- **Private ASNs MUST be removed from all prefixes announced to the public Internet**

  Include configuration to remove private ASNs in the eBGP template

- **As with RFC1918 address space, private ASNs are intended for internal use**

  They should not be leaked to the public Internet

- **Cisco IOS**

  neighbor x.x.x.x remove-private-AS

Policy Tools

- **Local preference**
  - **outbound traffic flows**
- **Metric (MED)**
  - **inbound traffic flows (local scope)**
- **AS-PATH prepend**
  - **inbound traffic flows (Internet scope)**
- **Communities**
  - **specific inter-provider peering**

Originating Prefixes: Assumptions

- **MUST announce assigned address block to Internet**
- **MAY also announce subprefixes – reachability is not guaranteed**
- **Current RIR minimum allocation is /21**
  - **Several ISPs filter RIR blocks on this boundary**
  - **Several ISPs filter the rest of address space according to the IANA assignments**
  - **This activity is called “Net Police” by some**

Originating Prefixes

- Some ISPs publish their minimum allocation sizes per /8 address block

  AfriNIC: www.afrinic.net/docs/policies/afpol-v4200407-000.htm

  APNIC: www.apnic.net/db/min-alloc.html

  ARIN: www.arin.net/reference/ip_blocks.html

  LACNIC: lacnic.net/en/registro/index.html

  RIPE NCC: www.ripe.net/ripe/docs/smallest-alloc-sizes.html

  Note that AfriNIC only publishes its current minimum allocation size, not the allocation size for its address blocks

- IANA publishes the address space it has assigned to end-sites and allocated to the RIRs:

  www.iana.org/assignments/ipv4-address-space

- Several ISPs use this published information to filter prefixes on:

  What should be routed (from IANA)

  The minimum allocation size from the RIRs

“Net Police” prefix list issues

- meant to “punish” ISPs who pollute the routing table with specifics rather than announcing aggregates
- impacts legitimate multihoming especially at the Internet’s edge
- impacts regions where domestic backbone is unavailable or costs $$$ compared with international bandwidth
- hard to maintain – requires updating when RIRs start allocating from new address blocks
- don’t do it unless consequences understood and you are prepared to keep the list current

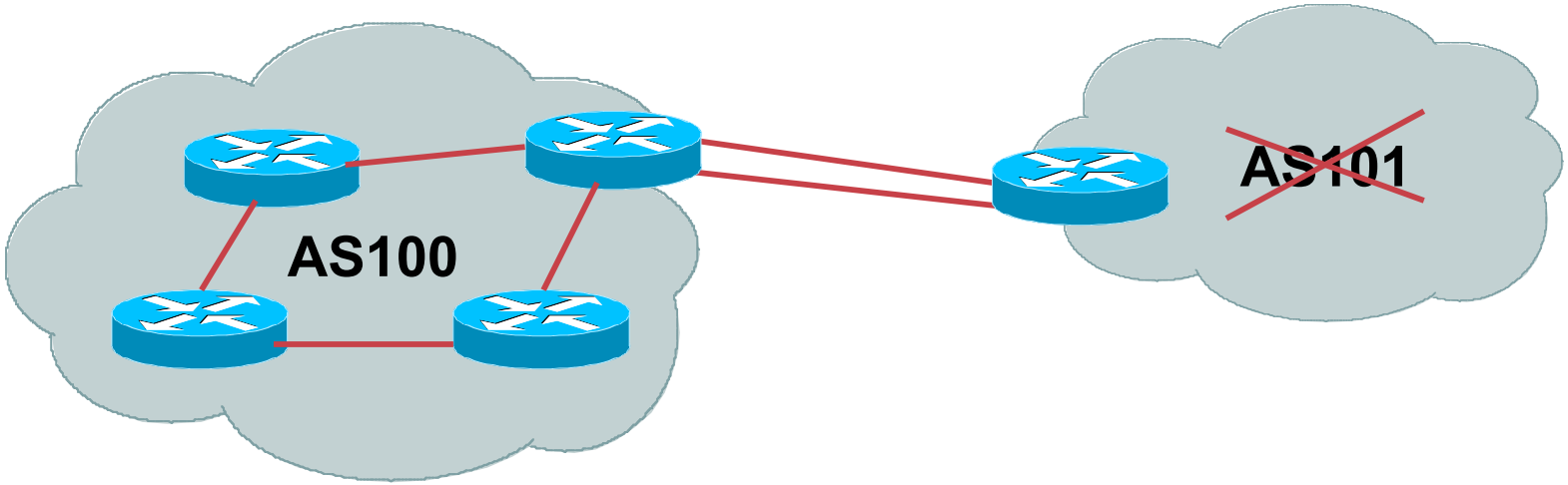
  Consider using the Project Cymru bogon BGP feed

  http://www.cymru.com/BGP/bogon-rs.html

Multihoming Scenarios

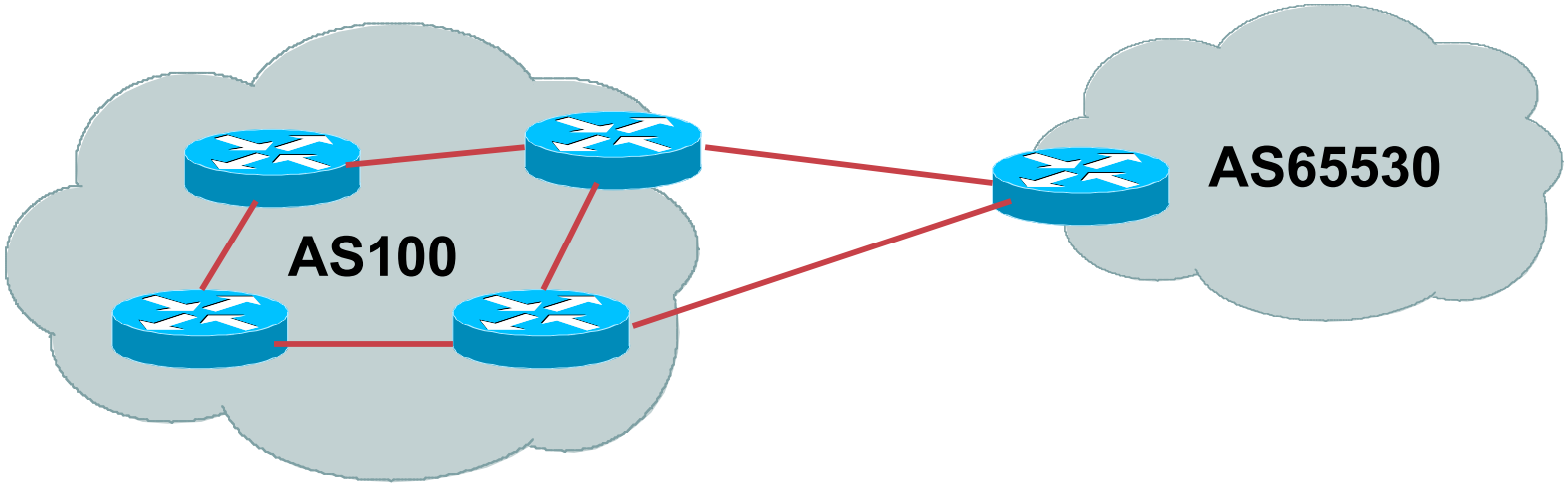
- **Stub network**
- **Multi-homed stub network**
- **Multi-homed network**
- **Load-balancing**

Stub Network



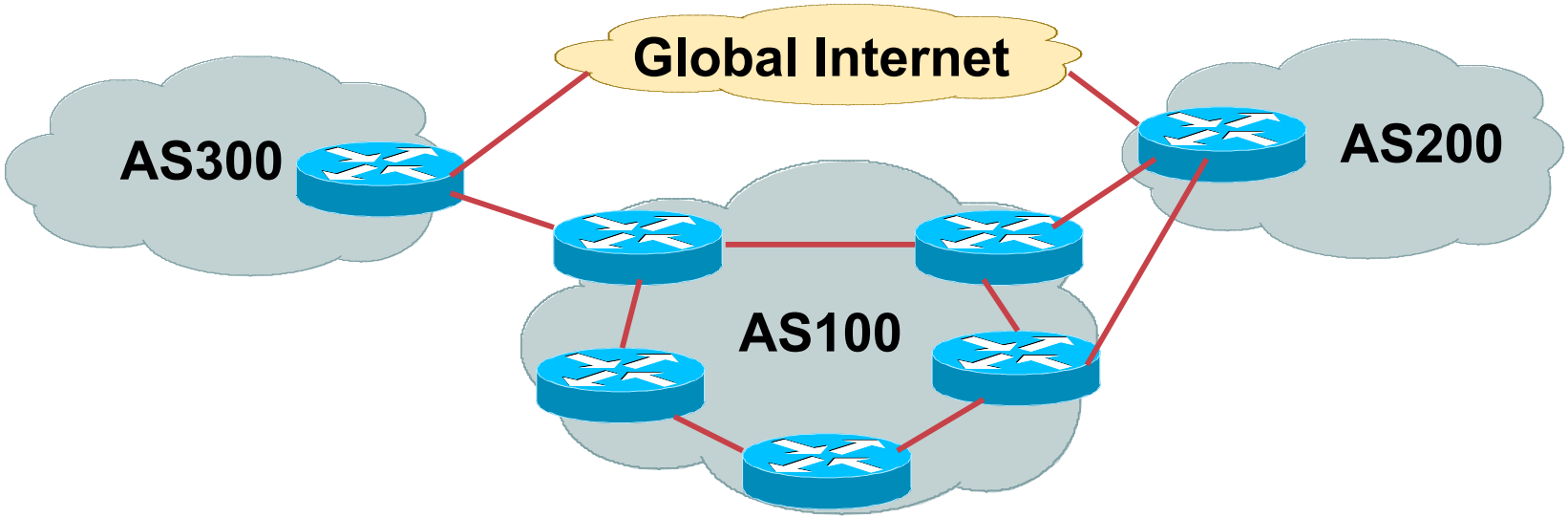
- No need for BGP
- Point static default to upstream ISP
- Router will load share on the two parallel circuits
- Upstream ISP advertises stub network
- Policy confined within upstream ISP's policy

Multi-homed Stub Network



- Use BGP (not IGP or static) to loadshare
- Use private AS (ASN > 64511)
- Upstream ISP advertises stub network
- Policy confined within upstream ISP's policy

Multi-Homed Network

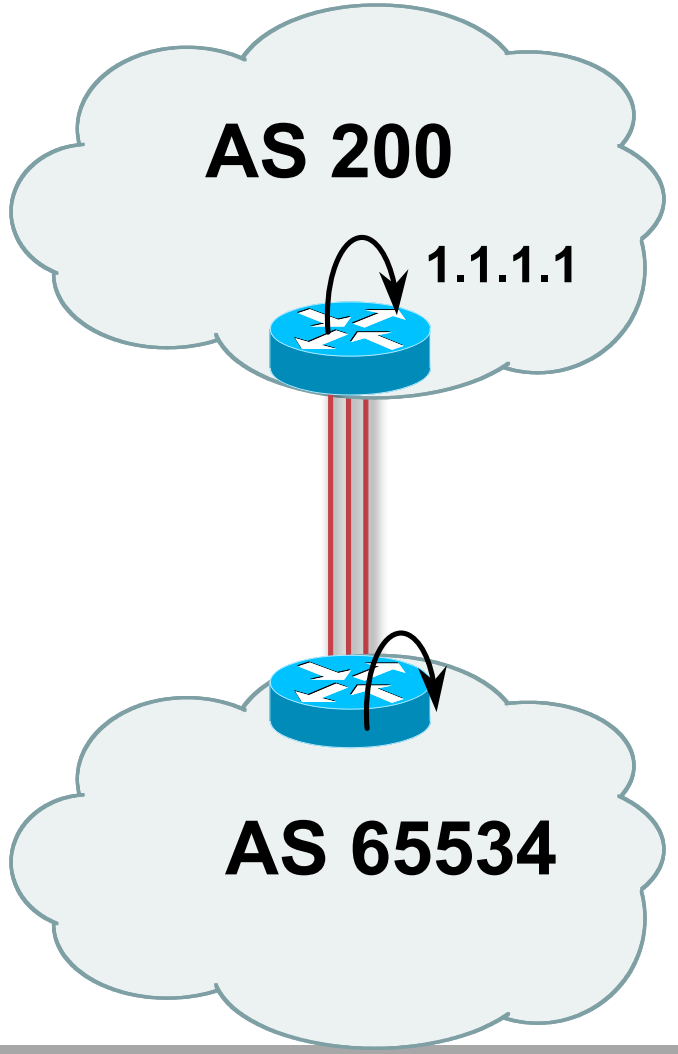


- **Many situations possible**
  - **multiple sessions to same ISP**
  - **secondary for backup only**
  - **load-share between primary and secondary**
  - **selectively use different ISPs**

Multiple Sessions to an ISP

- Use eBGP multihop
  - eBGP to loopback addresses
  - eBGP prefixes learned with loopback address as next hop
- Cisco IOS

```
router bgp 65534
 neighbor 1.1.1.1 remote-as 200
 neighbor 1.1.1.1 ebgp-multihop 2
!
ip route 1.1.1.1 255.255.255.255 serial 1/0
ip route 1.1.1.1 255.255.255.255 serial 1/1
ip route 1.1.1.1 255.255.255.255 serial 1/2
```



Multiple Sessions to an ISP

- **Try and avoid use of ebgp-multihop unless:**

  It's absolutely necessary –or–

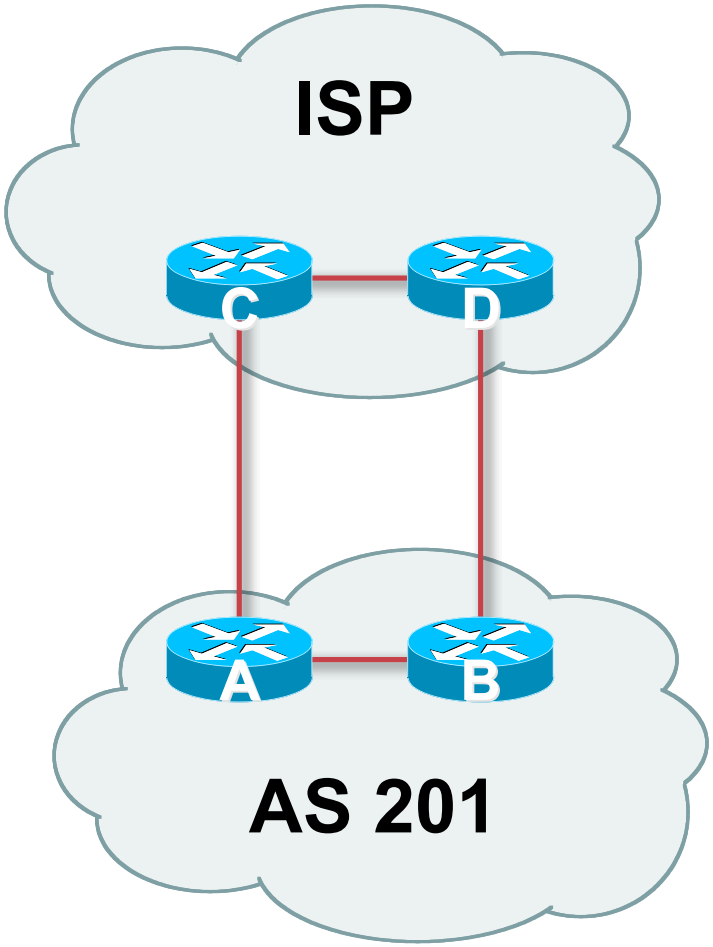
  Loadsharing across multiple links

- **Many ISPs discourage its use, for example:**

We will run eBGP multihop, but do not support it as a standard offering because customers generally have a hard time managing it due to:
• routing loops
• failure to realise that BGP session stability problems are usually due connectivity problems between their CPE and their BGP speaker

Multiple Sessions to an ISP

- Simplest scheme is to use defaults
- Learn/advertise prefixes for better control
- Planning and some work required to achieve loadsharing
  - Point default towards one ISP
  - Learn selected prefixes from second ISP
  - Modify the number of prefixes learnt to achieve acceptable load sharing
- No magic solution



BGP Multihoming Techniques

- Why Multihome?
- Definition & Options
- Basic Multihoming
- “BGP Traffic Engineering”
- Using Communities

Basic Multihoming

Learning to walk before we try running

Basic Multihoming

- **No frills multihoming**
- **Will look at two cases:**
  - **Multihoming with the same ISP**
  - **Multihoming to different ISPs**
- **Will keep the examples easy**
  - **Understanding easy concepts will make the more complex scenarios easier to comprehend**
  - **All assume that the site multihoming has a /19 address block**

Basic Multihoming

- **This type is most commonplace at the edge of the Internet**
  - **Networks here are usually concerned with inbound traffic flows**
  - **Outbound traffic flows being “nearest exit” is usually sufficient**
- **Can apply to the leaf ISP as well as Enterprise networks**

Basic Multihoming

Multihoming to the Same ISP

Basic Multihoming: Multihoming to the same ISP

- **Use BGP for this type of multihoming**
  - **use a private AS (ASN > 64511)**
  - **There is no need or justification for a public ASN**
    - **Making the nets of the end-site visible gives no useful information to the Internet**
- **Upstream ISP proxy aggregates**
  - **in other words, announces only your address block to the Internet from their AS (as would be done if you had one statically routed connection)**

Two links to the same ISP

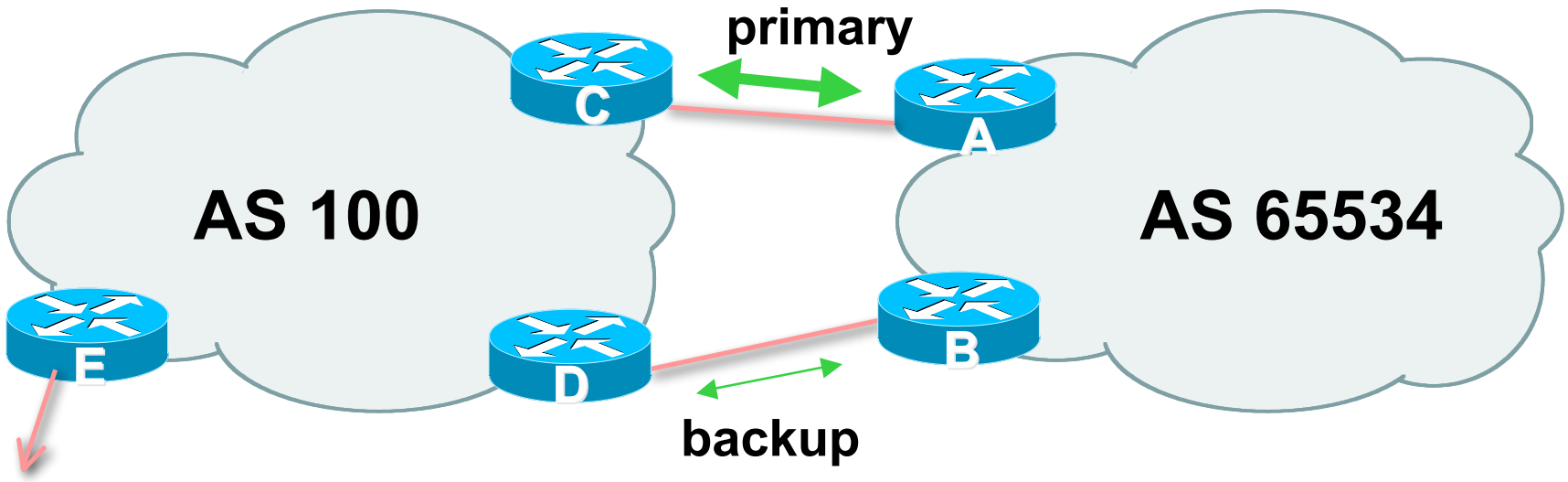
One link primary, the other link backup only

Two links to the same ISP (one as backup only)

- **Applies when end-site has bought a large primary WAN link to their upstream a small secondary WAN link as the backup**

  For example, primary path might be an E1, backup might be 64kbps

Two links to the same ISP (one as backup only)



- **Border router E in AS100 removes private AS and any customer subprefixes from Internet announcement**

Two links to the same ISP (one as backup only)

- **Announce /19 aggregate on each link**

  primary link:

  Outbound – announce /19 unaltered

  Inbound – receive default route

  backup link:

  Outbound – announce /19 with increased metric

  Inbound – received default, and reduce local preference

- **When one link fails, the announcement of the /19 aggregate via the other link ensures continued connectivity**

Two links to the same ISP (one as backup only)

- Router E removes the private AS and customer's subprefixes from external announcements
- Private AS still visible inside AS100

Two links to the same ISP

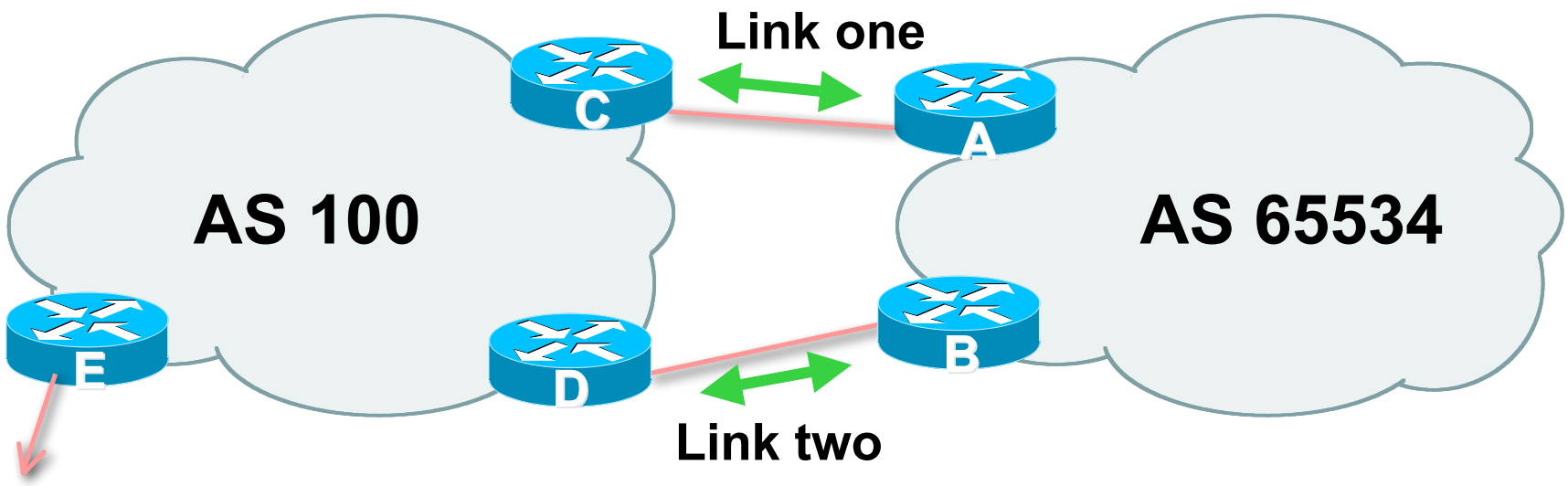
With Loadsharing

Loadsharing to the same ISP

- More common case
- End sites tend not to buy circuits and leave them idle, only used for backup as in previous example
- This example assumes equal capacity circuits

    Unequal capacity circuits requires more refinement – see later

Loadsharing to the same ISP



- Border router E in AS100 removes private AS and any customer subprefixes from Internet announcement

Loadsharing to the same ISP

- Announce /19 aggregate on each link
- Split /19 and announce as two /20s, one on each link
  - basic inbound loadsharing
  - assumes equal circuit capacity and even spread of traffic across address block
- Vary the split until “perfect” loadsharing achieved
- Accept the default from upstream
  - basic outbound loadsharing by nearest exit
  - okay in first approx as most ISP and end-site traffic is inbound

Loadsharing to the same ISP

- **Loadsharing configuration is only on customer router**
- **Upstream ISP has to**
  - **remove customer subprefixes from external announcements**
  - **remove private AS from external announcements**
- **Could also use BGP communities**

Basic Multihoming

Multihoming to different ISPs

Two links to different ISPs

- **Use a Public AS**
  - **Or use private AS if agreed with the other ISP**
  - **But some people don't like the "inconsistent-AS" which results from use of a private-AS**
- **Address space comes from**
  - **both upstreams or**
  - **Regional Internet Registry**
- **Configuration concepts very similar**

Inconsistent-AS?

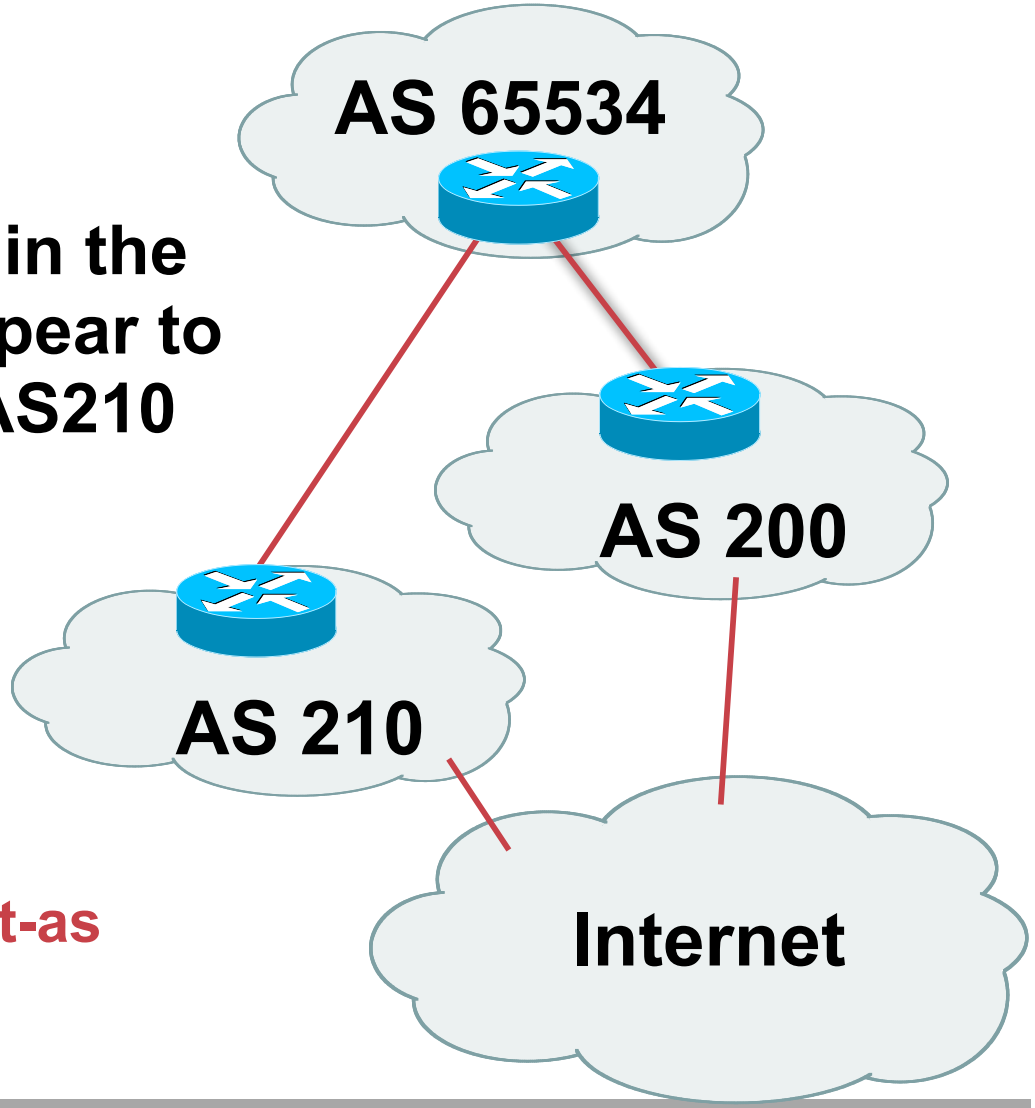
- Viewing the prefixes originated by AS65534 in the Internet shows they appear to be originated by both AS210 and AS200

  This is NOT bad

  Nor is it illegal

- Cisco IOS command is

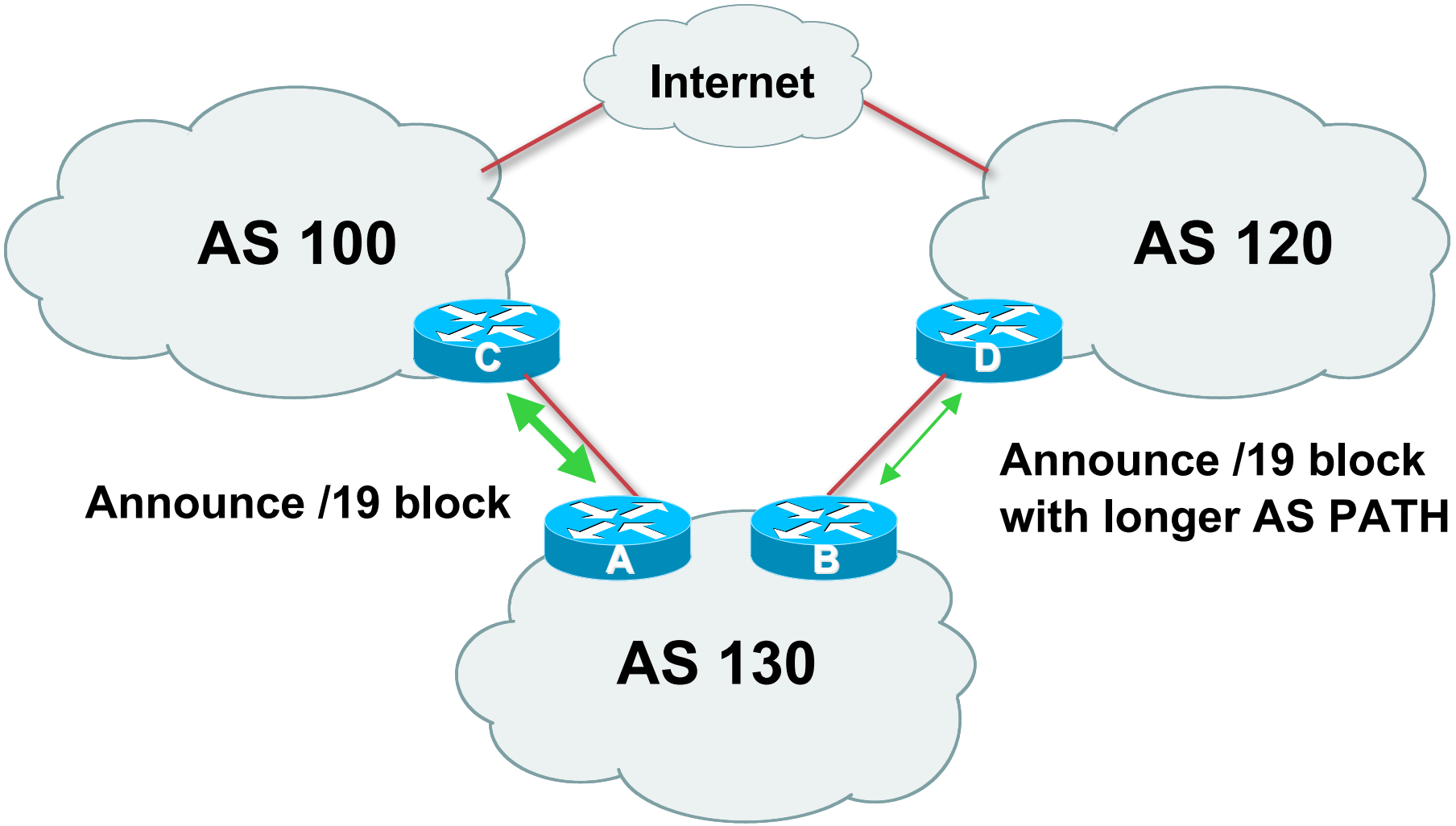
  show ip bgp inconsistent-as



Two links to different ISPs

One link primary, the other link backup only

Two links to different ISPs (one as backup only)



Two links to different ISPs (one as backup only)

- Announce /19 aggregate on each link
  - primary link makes standard announcement
  - backup link lengthens the AS PATH by using AS PATH prepend
- When one link fails, the announcement of the /19 aggregate via the other link ensures continued connectivity

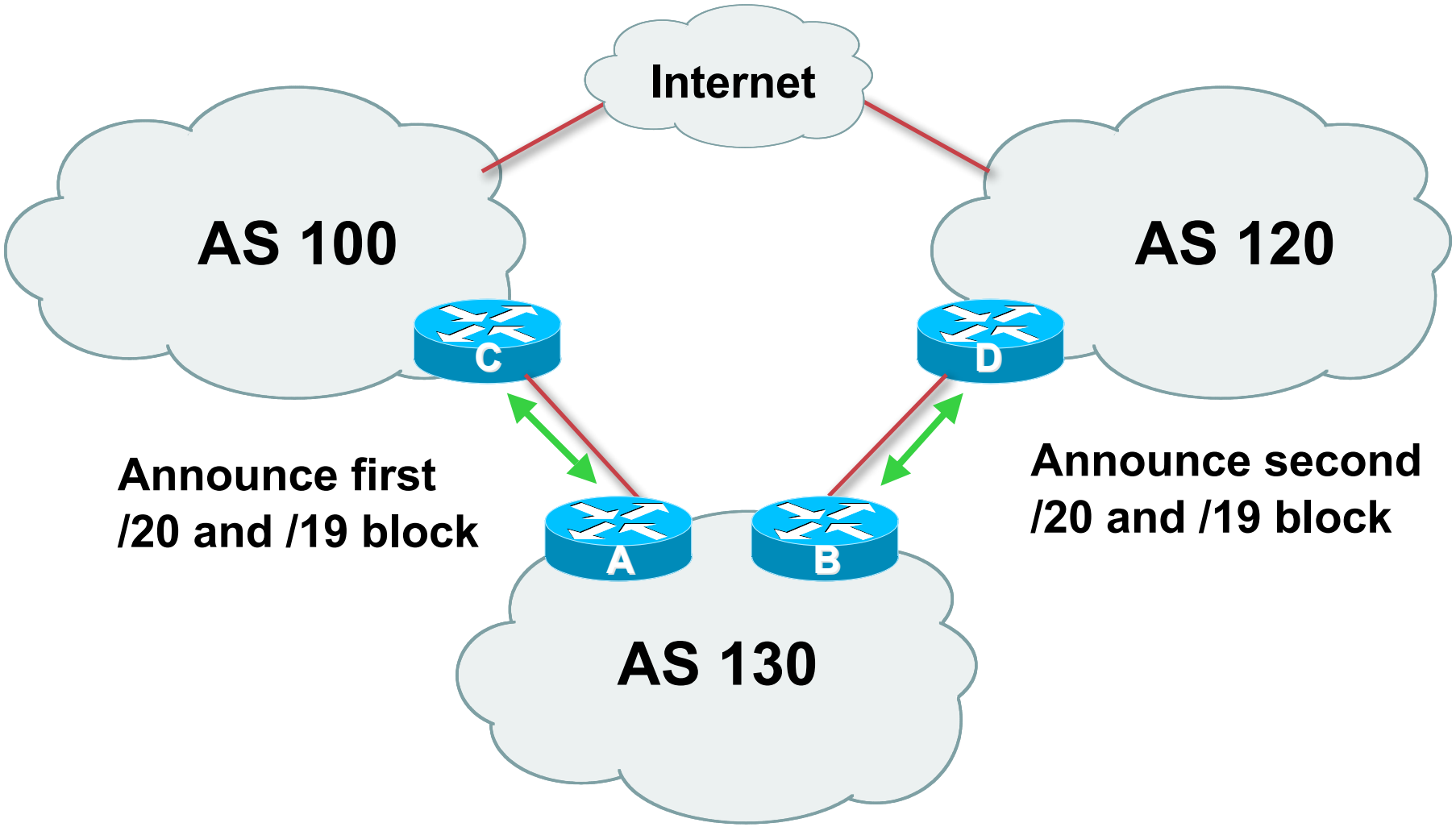
Two links to different ISPs (one as backup only)

- Not a common situation as most sites tend to prefer using whatever capacity they have
- But it shows the basic concepts of using local-prefs and AS-path prepends for engineering traffic in the chosen direction

Two links to different ISPs

With Loadsharing

Two links to different ISPs (with loadsharing)



Two links to different ISPs (with loadsharing)

- Announce /19 aggregate on each link
- Split /19 and announce as two /20s, one on each link

  basic inbound loadsharing

- When one link fails, the announcement of the /19 aggregate via the other ISP ensures continued connectivity

Two links to different ISPs (with loadsharing)

- Loadsharing in this case is very basic
- But shows the first steps in designing a load sharing solution

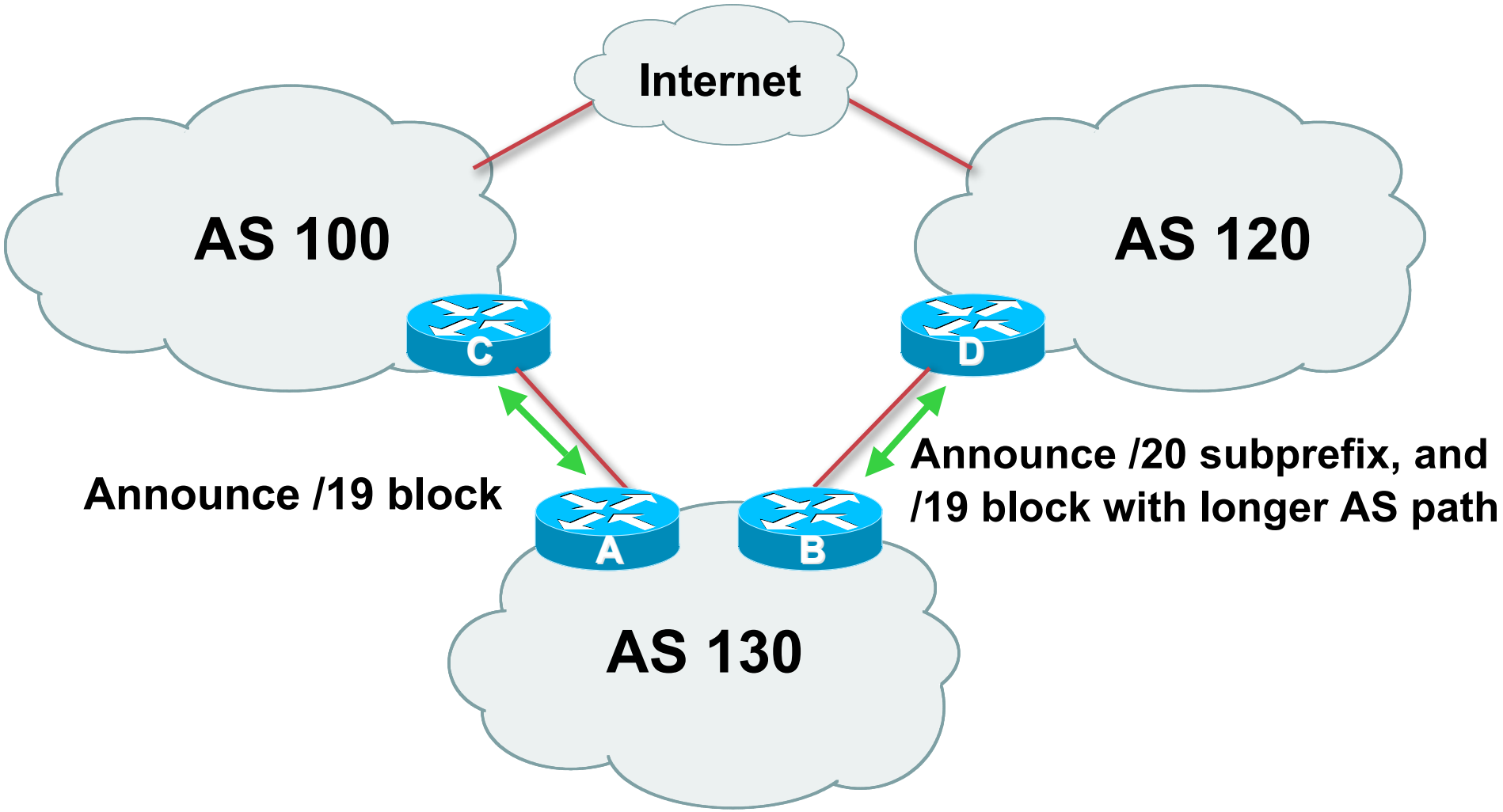
  Start with a simple concept

  And build on it…!

Two links to different ISPs

More Controlled Loadsharing

Loadsharing with different ISPs



Loadsharing with different ISPs

- **Announce /19 aggregate on each link**
  - **On first link, announce /19 as normal**
  - **On second link, announce /19 with longer AS PATH, and announce one /20 subprefix**
    - **controls loadsharing between upstreams and the Internet**
- **Vary the subprefix size and AS PATH length until "perfect" loadsharing achieved**
- **Still require redundancy!**

Loadsharing with different ISPs

- This example is more commonplace
- Shows how ISPs and end-sites subdivide address space frugally, as well as use the AS-PATH prepend concept to optimise the load sharing between different ISPs
- Notice that the /19 aggregate block is ALWAYS announced

BGP Multihoming Techniques

- **Why Multihome?**
- **Definition & Options**
- **Basic Multihoming**
- **“BGP Traffic Engineering”**
- **Using Communities**

Service Provider Multihoming

BGP Traffic Engineering

Service Provider Multihoming

- Previous examples dealt with loadsharing inbound traffic
  - Of primary concern at Internet edge
  - What about outbound traffic?
- Transit ISPs strive to balance traffic flows in both directions
  - Balance link utilisation
  - Try and keep most traffic flows symmetric
  - Some edge ISPs try and do this too
- The original “Traffic Engineering”

Service Provider Multihoming

- **Balancing outbound traffic requires inbound routing information**
  - **Common solution is "full routing table"**
  - **Rarely necessary**
    - **Why use the "routing mallet" to try solve loadsharing problems?**
  - **"Keep It Simple" is often easier (and $$$ cheaper) than carrying N-copies of the full routing table**

Service Provider Multihoming MYTHS!!

- Common MYTHS
- 1: You need the full routing table to multihome
  - People who sell router memory would like you to believe this
  - Only true if you are a transit provider
  - Full routing table can be a significant hindrance to multihoming
- 2: You need a BIG router to multihome
  - Router size is related to data rates, not running BGP
  - In reality, to multihome, your router needs to:
    - Have two interfaces,
    - Be able to talk BGP to at least two peers,
    - Be able to handle BGP attributes,
    - Handle at least one prefix
- 3: BGP is complex
  - In the wrong hands, yes it can be! Keep it Simple!

Service Provider Multihoming: Some Strategies

- Take the prefixes you need to aid traffic engineering
  - Look at NetFlow data for popular sites
- Prefixes originated by your immediate neighbours and their neighbours will do more to aid load balancing than prefixes from ASNs many hops away
  - Concentrate on local destinations
- Use default routing as much as possible
  - Or use the full routing table with care

Service Provider Multihoming

- **Examples**
  - **One upstream, one local peer**
  - **One upstream, local exchange point**
  - **Two upstreams, one local peer**
- **Require BGP and a public ASN**
- **Examples assume that the local network has their own /19 address block**

Service Provider Multihoming

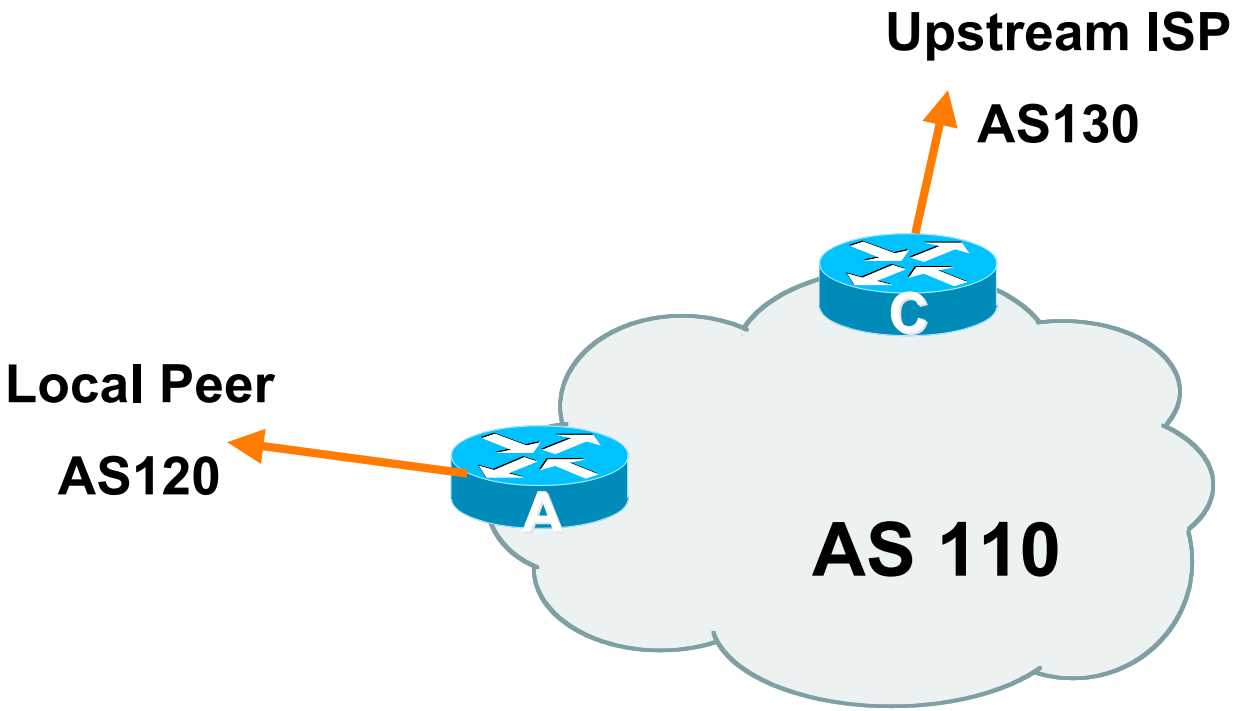
One upstream, one local peer

One Upstream, One Local Peer

- **Very common situation in many regions of the Internet**
- **Connect to upstream transit provider to see the "Internet"**
- **Connect to the local competition so that local traffic stays local**

  Saves spending valuable $ on upstream transit costs for local traffic

One Upstream, One Local Peer



One Upstream, One Local Peer

- Announce /19 aggregate on each link
- Accept default route only from upstream
  - Either 0.0.0.0/0 or a network which can be used as default
- Accept all routes from local peer

One Upstream, One Local Peer

- **Two configurations possible for Router A**
  - **Use of AS Path Filters assumes peer knows what they are doing**
  - **Prefix Filters are higher maintenance, but safer**
  - **Some ISPs use both**
- **Local traffic goes to and from local peer, everything else goes to upstream**

Aside:
Configuration Recommendation

- **Private Peers**
  - **The peering ISPs exchange prefixes they originate**
  - **Sometimes they exchange prefixes from neighbouring ASNs too**
- **Be aware that the private peer eBGP router should carry only the prefixes you want the private peer to receive**
  - **Otherwise they could point a default route to you and unintentionally transit your backbone**

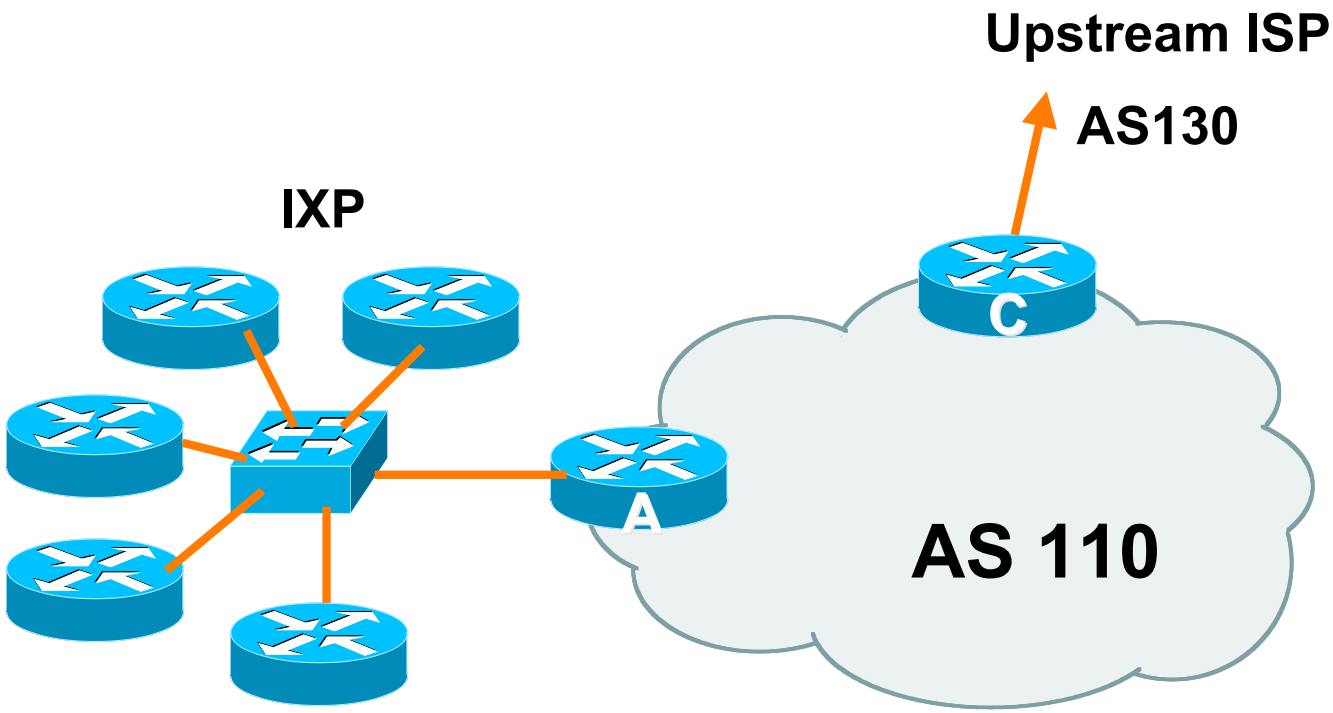
Service Provider Multihoming

One Upstream, Local Exchange Point

One Upstream, Local Exchange Point

- **Very common situation in many regions of the Internet**
- **Connect to upstream transit provider to see the "Internet"**
- **Connect to the local Internet Exchange Point so that local traffic stays local**
  - **Saves spending valuable $ on upstream transit costs for local traffic**

One Upstream, Local Exchange Point



One Upstream, Local Exchange Point

- Announce /19 aggregate to every neighbouring AS
- Accept default route only from upstream
  - Either 0.0.0.0/0 or a network which can be used as default
- Accept all routes originated by IXP peers

One Upstream, Local Exchange

- Router A does not generate the aggregate for AS110
  - If Router A becomes disconnected from backbone, then the aggregate is no longer announced to the IX
  - BGP failover works as expected
- Note that the local preference for for inbound announcements from the IX is set higher than the default
  - This ensures that local traffic crosses the IXP
  - (And avoids potential problems with any uRPF check)

Aside:
IXP Configuration Recommendation

- **IXP peers**
  - **The peering ISPs at the IXP exchange prefixes they originate**
  - **Sometimes they exchange prefixes from neighbouring ASNs too**
- **Be aware that the IXP border router should carry only the prefixes you want the IXP peers to receive and the destinations you want them to be able to reach**
  - **Otherwise they could point a default route to you and unintentionally transit your backbone**
- **If IXP router is at IX, and distant from your backbone**
  - **Don't originate your address block at your IXP router**

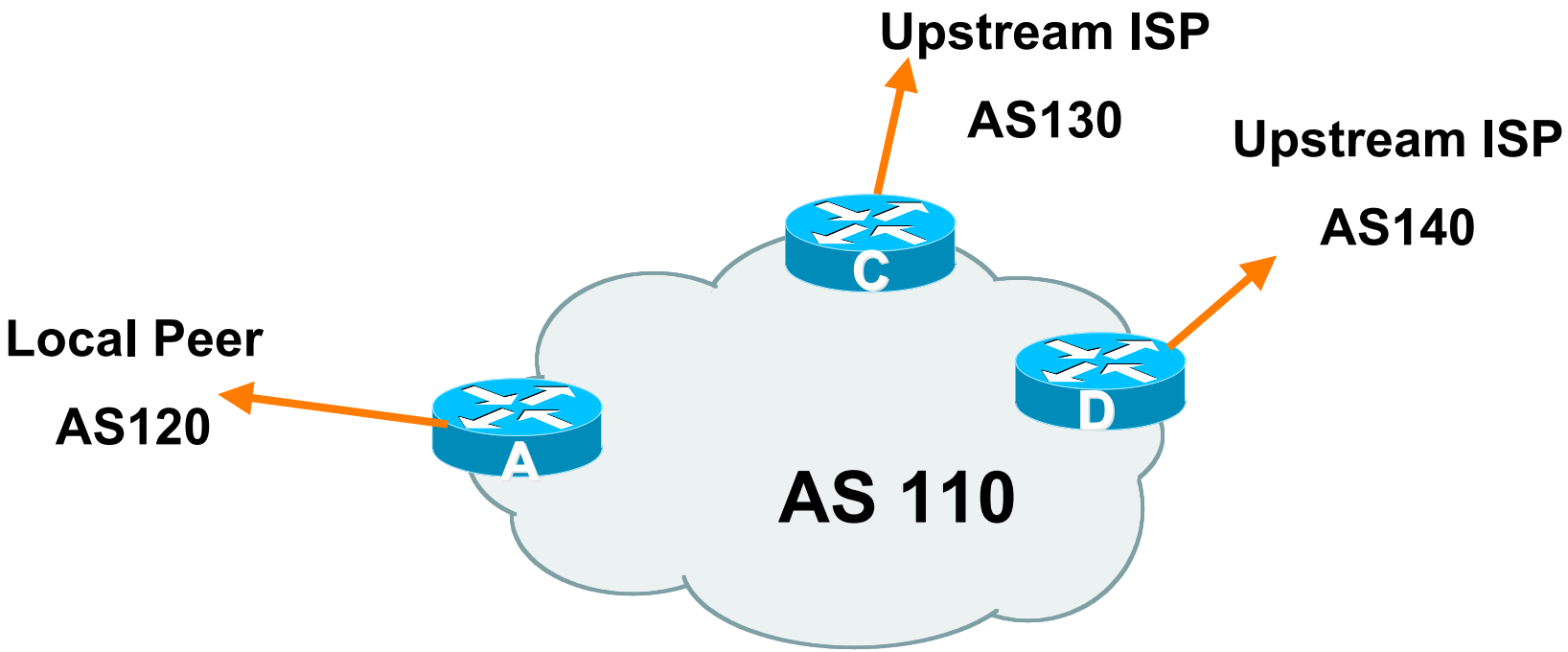
Service Provider Multihoming

Two Upstreams, One local peer

Two Upstreams, One Local Peer

- **Connect to both upstream transit providers to see the “Internet”**
  - **Provides external redundancy and diversity – the reason to multihome**
- **Connect to the local peer so that local traffic stays local**
  - **Saves spending valuable $ on upstream transit costs for local traffic**

Two Upstreams, One Local Peer



Two Upstreams, One Local Peer

- Announce /19 aggregate on each link
- Accept default route only from upstreams
  - Either 0.0.0.0/0 or a network which can be used as default
- Accept all routes from local peer

Two Upstreams, One Local Peer

- Router A has same routing configuration as in example with one upstream and one local peer
- Two configuration options for Routers C and D:

  Accept full routing from both upstreams

  Expensive & unnecessary!

  Accept default from one upstream and some routes from the other upstream

  The way to go!

Two Upstreams, One Local Peer
Full Routes

- **Router C configuration:**
  - **Accept full routes from AS130**
  - **Tag prefixes originated by AS130 and AS130's neighbouring ASes with local preference 120**
    - **Traffic to those ASes will go over AS130 link**
  - **Remaining prefixes tagged with local preference of 80**
    - **Traffic to other all other ASes will go over the link to AS140**
- **Router D configuration same as Router C without setting any preferences**

Two Upstreams, One Local Peer
Full Routes

- **Full routes from upstreams**
  - **Expensive – needs lots of memory and CPU**
  - **Need to play preference games**
  - **Previous example is only an example – real life will need improved fine-tuning!**
  - **Previous example doesn't consider inbound traffic – see earlier in presentation for examples**

Two Upstreams, One Local Peer
Partial Routes

- **Strategy:**
  - **Ask one upstream for a default route**
    - **Easy to originate default towards a BGP neighbour**
  - **Ask other upstream for a full routing table**
    - **Then filter this routing table based on neighbouring ASN**
    - **E.g. want traffic to their neighbours to go over the link to that ASN**
    - **Most of what upstream sends is thrown away**
    - **Easier than asking the upstream to set up custom BGP filters for you**

Two Upstreams, One Local Peer Partial Routes

- **Router C configuration:**
  - **Accept full routes from AS130**
    - **(or get them to send less)**
  - **Filter ASNs so only AS130 and AS130's neighbouring ASes are accepted**
  - **Allow default, and set it to local preference 80**
  - **Traffic to those ASes will go over AS130 link**
  - **Traffic to other all other ASes will go over the link to AS140**
  - **If AS140 link fails, backup via AS130 – and vice-versa**
- **Router D configuration:**
  - **Accept only the default route**

Two Upstreams, One Local Peer
Partial Routes

- **Partial routes from upstreams**
  - **Not expensive – only carry the routes necessary for loadsharing**
  - **Need to filter on AS paths**
  - **Previous example is only an example – real life will need improved fine-tuning!**
  - **Previous example doesn't consider inbound traffic – see earlier in presentation for examples**

Two Upstreams, One Local Peer

- **When upstreams cannot or will not announce default route**

  Because of operational policy against using “default-originate” on BGP peering

  Solution is to use IGP to propagate default from the edge/peering routers

Aside:
Configuration Recommendation

- **When distributing internal default by iBGP or OSPF**

  Make sure that routers connecting to private peers or to IXPs do NOT carry the default route

  Otherwise they could point a default route to you and unintentionally transit your backbone

  Simple fix for Private Peer/IXP routers:

```
ip route 0.0.0.0 0.0.0.0 null0
```

BGP Multihoming Techniques

- Why Multihome?
- Definition & Options
- Basic Multihoming
- “BGP Traffic Engineering”
- Using Communities

Communities

How they are used in practice

Using Communities: RFC1998

- Informational RFC
- Describes how to implement loadsharing and backup on multiple inter-AS links
  - BGP communities used to determine local preference in upstream's network
- Gives control to the customer
- Simplifies upstream's configuration
  - simplifies network operation!

RFC1998

- **Community values defined to have particular meanings:**

ASx:100	**set local pref 100**	**preferred route**
ASx:90	**set local pref 90**	**backup route if dualhomed on ASx**
ASx:80	**set local pref 80**	**main link is to another ISP with same AS path length**
ASx:70	**set local pref 70**	**main link is to another ISP**

RFC1998

- **Supporting RFC1998**
  - **Many ISPs do, more should**
  - **Check AS object in the Internet Routing Registry**
  - **If you do, insert comment in AS object in the IRR**
  - **Or make a note on your website**

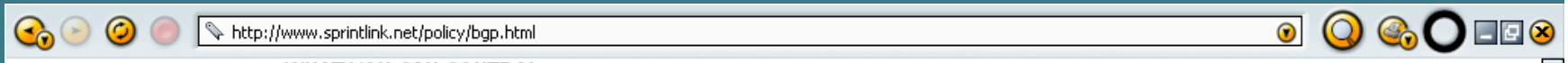
Beyond RFC1998

- **RFC1998 is okay for “simple” multihomed customers**
  - **assumes that upstreams are interconnected**
- **ISPs have created many other communities to handle more complex situations**
  - **Simplify ISP BGP configuration**
  - **Give customer more policy control**

ISP BGP Communities

- There are no recommended ISP BGP communities apart from
  - RFC1998
  - The four standard communities
    - www.iana.org/assignments/bgp-well-known-communities
- Efforts have been made to document from time to time
  - totem.info.ucl.ac.be/publications/papers-elec-versions/draft-quoitin-bgp-comm-survey-00.pdf
  - But so far… nothing more… ☹
  - Collection of ISP communities at www.onesc.net/communities
- ISP policy is usually published
  - On the ISP's website
  - Referenced in the AS Object in the IRR

Some ISP Examples: Sprintlink



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 ...

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

Some ISP Examples
AAPT

- Australian ISP
- Run their own Routing Registry

  Whois.connect.com.au

- Offer 6 different communities to customers to aid with their traffic engineering

Some ISP Examples
AAPT

```
aut-num:        AS2764
as-name:        ASN-CONNECT-NET
descr:          AAPT Limited
admin-c:        CNO2-AP
tech-c:         CNO2-AP
remarks:        Community support definitions
remarks:
remarks:        Community  Definition
remarks:        ----------------------------------------
remarks:        2764:2 Don't announce outside local POP
remarks:        2764:4 Lower local preference by 15
remarks:        2764:5 Lower local preference by 5
remarks:        2764:6 Announce to customers and all peers
                       (incl int'l peers), but not transit
remarks:        2764:7 Announce to customers only
remarks:        2764:14 Announce to AANX
notify:         routing@connect.com.au
mnt-by:         CONNECT-AU
changed:        nobody@connect.com.au 20050225
source:         CCAIR
```

More at http://info.connect.com.au/docs/routing/general/multi-faq.shtml#q13

Some ISP Examples
MCI Europe

- MCI's European operation
- 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 Ignite

- One of the most comprehensive community lists around

  Seems to be based on definitions originally used in Tiscali's network

  whois –h whois.ripe.net AS5400 reveals all

- Extensive community definitions allow sophisticated traffic engineering by customers

Some ISP 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:        5400:1004 C&W EU (1273)                         5400:2004
<snip>
notify:         notify@eu.bt.net
mnt-by:         CIP-MNT
source:         RIPE
```

And many many more!

Some ISP Examples
Level 3

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

  whois –h whois.ripe.net AS3356 reveals all

Some ISP Examples
Level 3

```
aut-num:        AS3356
descr:          Level 3 Communications
<snip>
remarks:        ----------------------------------------------------
remarks:        customer traffic engineering communities - Suppression
remarks:        ----------------------------------------------------
remarks:        64960:XXX - announce to AS XXX if 65000:0
remarks:        65000:0   - announce to customers but not to peers
remarks:        65000:XXX - do not announce at peerings to AS XXX
remarks:        ----------------------------------------------------
remarks:        customer traffic engineering communities - Prepending
remarks:        ----------------------------------------------------
remarks:        65001:0   - prepend once  to all peers
remarks:        65001:XXX - prepend once  at peerings to AS XXX
<snip>
remarks:        3356:70   - set local preference to 70
remarks:        3356:80   - set local preference to 80
remarks:        3356:90   - set local preference to 90
remarks:        3356:9999 - blackhole (discard) traffic
<snip>
mnt-by:         LEVEL3-MNT
source:         RIPE
```

And many many more!

Creating your own community policy

- **Consider creating communities to give policy control to customers**
  - **Reduces technical support burden**
  - **Reduces the amount of router reconfiguration, and the chance of mistakes**
  - **Use the previous examples as a guideline**

Summary

- **Multihoming:**
  - **Inbound traffic engineering**
  - **Outbound traffic engineering**
- **Think of:**
  - **Aggregation**
  - **Frugal announcements**

BGP Multihoming Techniques
Next: BGP Troubleshooting

Philip Smith <pfs@cisco.com>

APNIC 22

4th - 8th September 2006

Kaohsiung, Taiwan