BGP Multihoming Techniques

Philip Smith <pfs@cisco.com>

MENOG 1

3-5 April 2007

Bahrain

Presentation Slides

- **Available on**
  - **ftp://ftp-eng.cisco.com**
  - **/pfs/seminars/MENOG1-Multihoming.pdf**
  - **And on the MENOG 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
- Preparing the Network
- Basic Multihoming
- Service Provider Multihoming

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
- Preparing the Network
- Basic Multihoming
- Service Provider Multihoming

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**

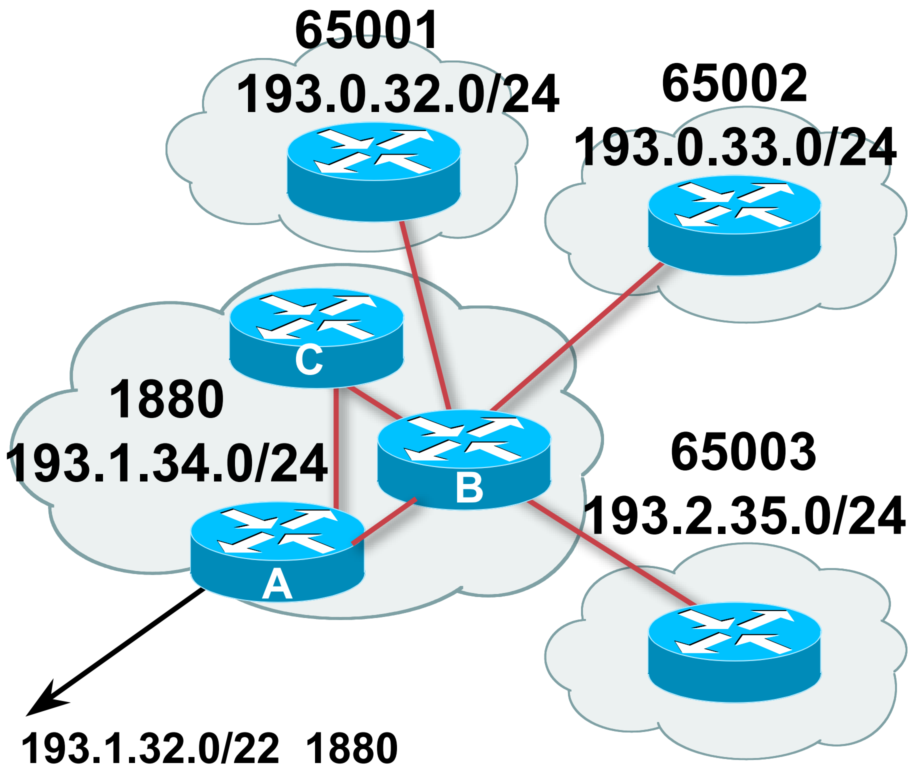
  An ISP with customers multihomed on their backbone (RFC2270)

  -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 RIRs 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

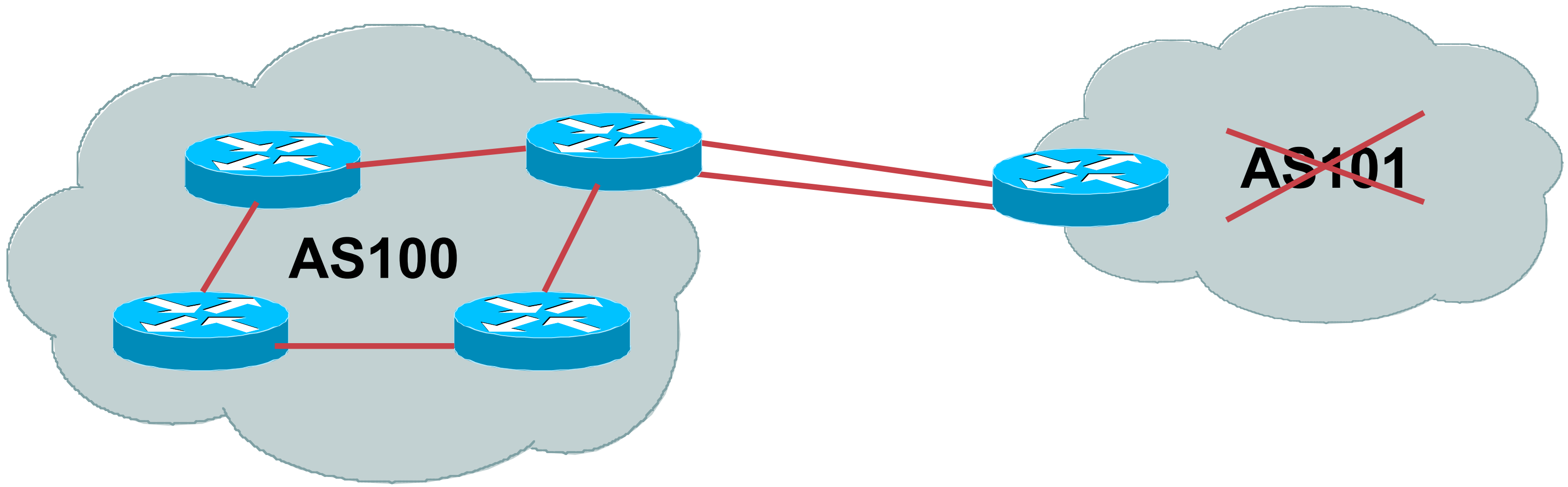
  Consider using the Project Cymru bogon BGP feed

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

Multihoming Scenarios

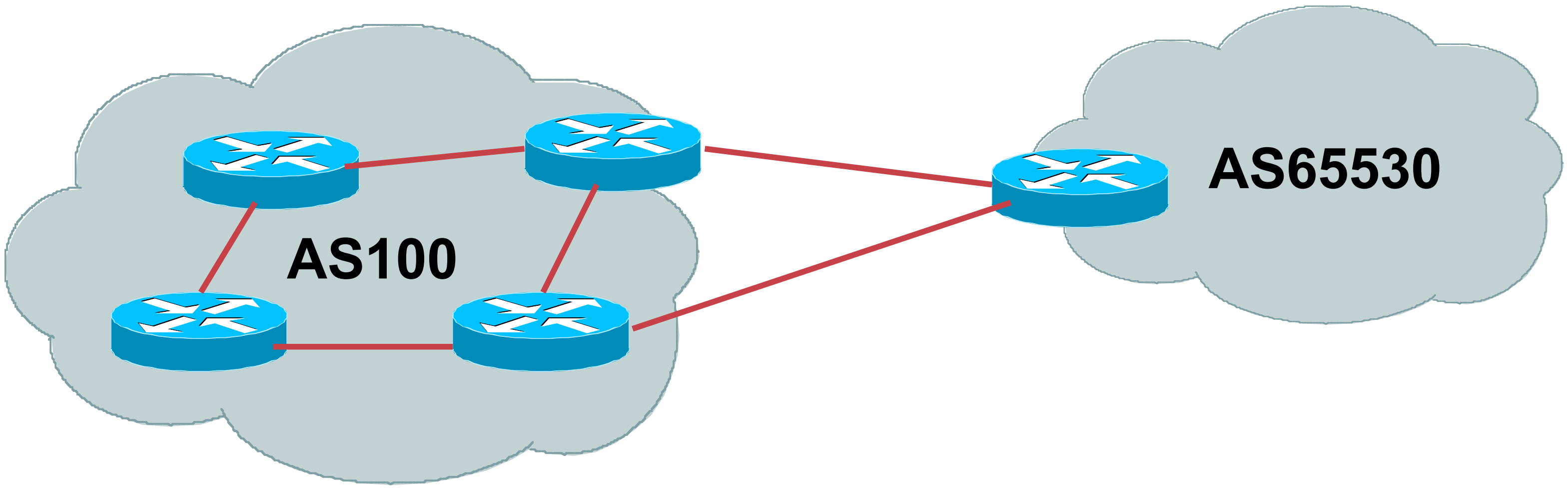
- **Stub network**
- **Multi-homed stub network**
- **Multi-homed network**
- **Load-balancing**

Stub Network



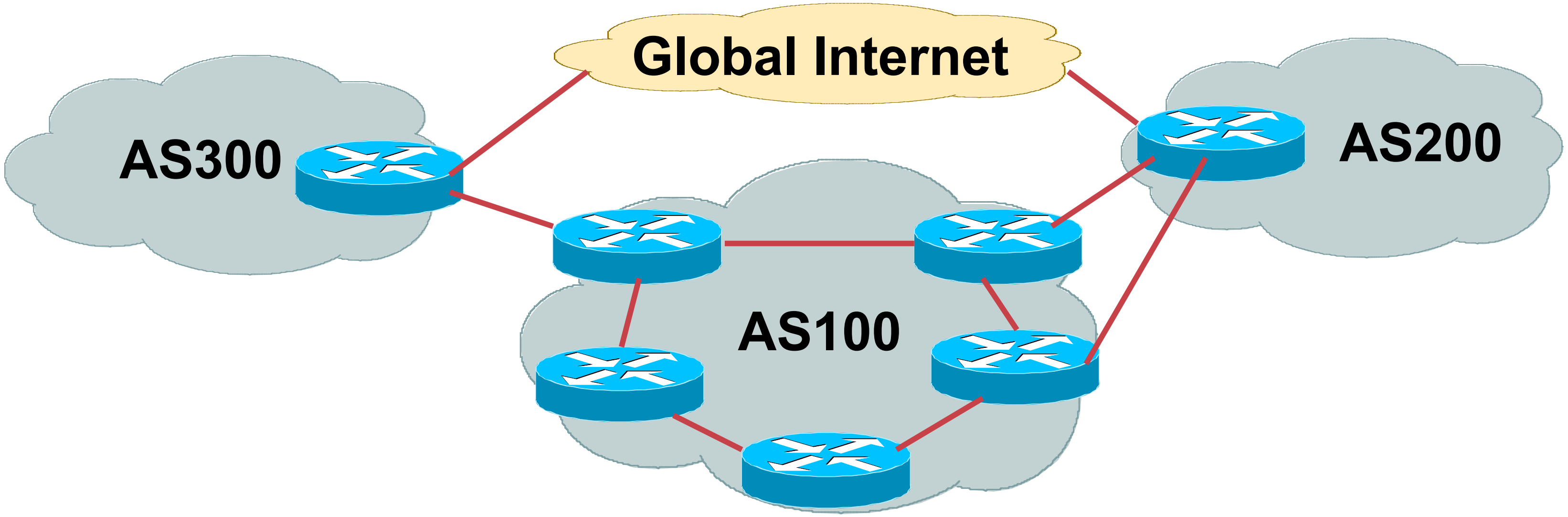
- No need for BGP
- Point static default to upstream ISP
- Router will load share on the two parallel circuits
- Upstream ISP advertises stub network
- Policy confined within upstream ISP’s policy

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

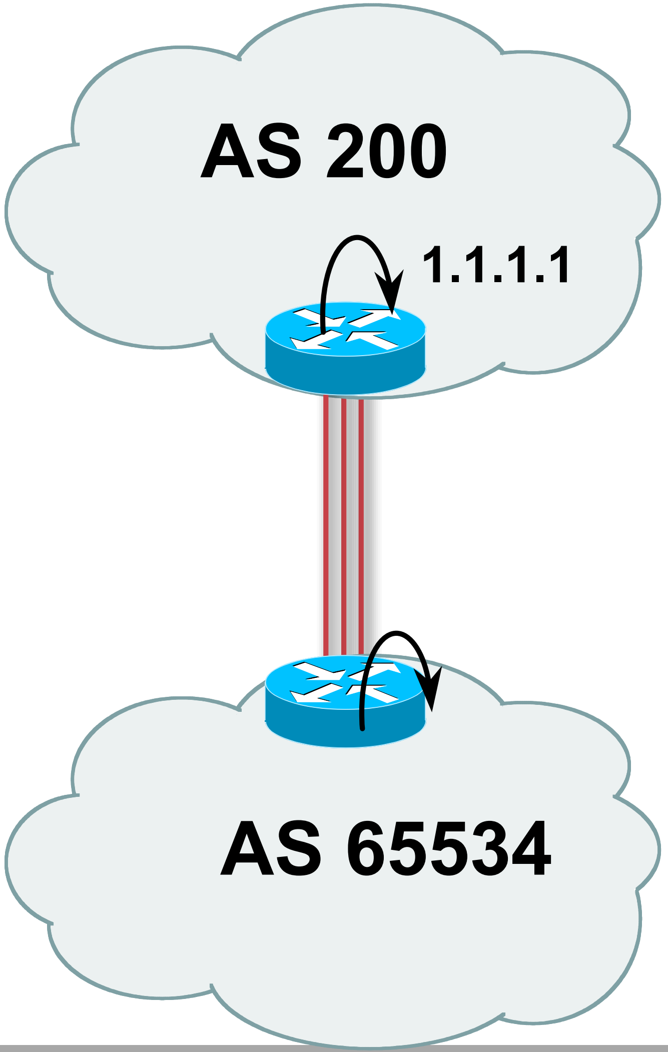


- **Many situations possible**
  - multiple sessions to same ISP
  - secondary for backup only
  - load-share between primary and secondary
  - selectively use different ISPs

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
```



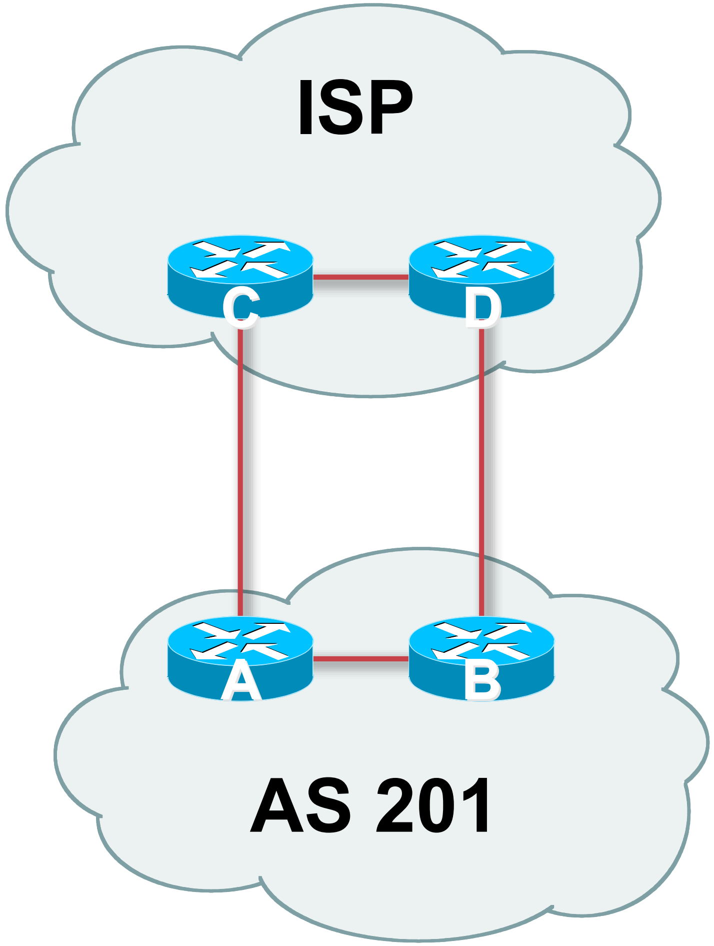
Multiple Sessions to an ISP

- 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
- Preparing the Network
- Basic Multihoming
- Service Provider Multihoming

Preparing the Network

Putting our own house in order first…

Preparing the Network

- We will deploy BGP across the network before we try and multihome
- BGP will be used therefore an ASN is required
- If multihoming to different ISPs, public ASN needed:
  - Either go to upstream ISP who is a registry member, or
  - Apply to the RIR yourself for a one off assignment, or
  - Ask an ISP who is a registry member, or
  - Join the RIR and get your own IP address allocation too (this option strongly recommended)!

Preparing the Network

- The network is not running any BGP at the moment
  - single statically routed connection to upstream ISP
- The network is not running any IGP at all
  - Static default and routes through the network to do "routing"

Preparing the Network
IGP

- Decide on IGP: OSPF or ISIS ☺
- Assign loopback interfaces and /32 addresses to each router which will run the IGP
  - Loopback is used for OSPF and BGP router id anchor
  - Used for iBGP and route origination
- Deploy IGP (e.g. OSPF)
  - IGP can be deployed with NO IMPACT on the existing static routing
    - For Cisco IOS, OSPF distance is 110 & static distance is 1
    - Smallest distance wins

Preparing the Network
IGP (cont)

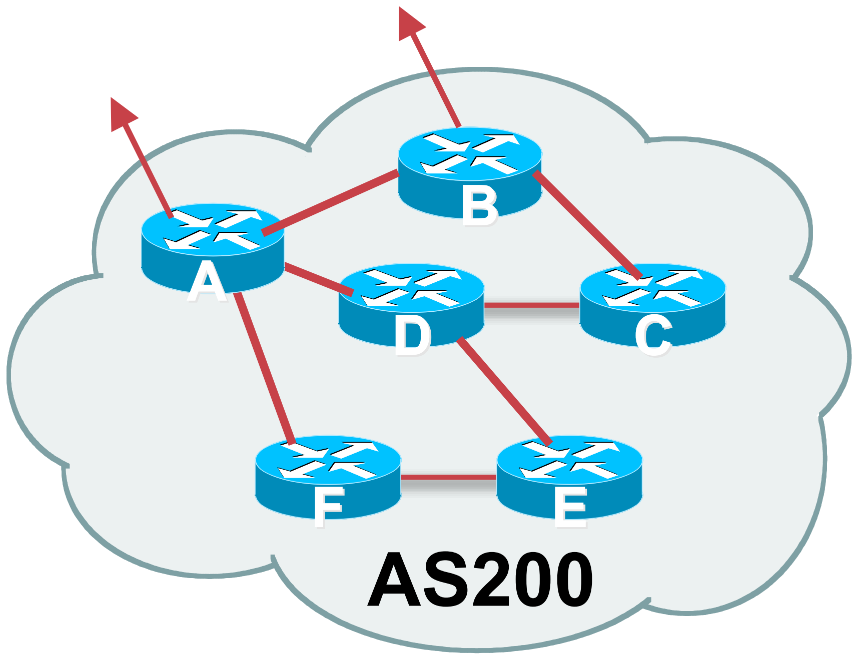
- **Be prudent deploying IGP – keep the Link State Database Lean!**
  - **Router loopbacks go in IGP**
  - **WAN point to point links go in IGP**
  - **(In fact, any link where IGP dynamic routing will be run should go into IGP)**
  - **Summarise on area/level boundaries (if possible) – i.e. think about your IGP address plan**

Preparing the Network
IGP (cont)

- **Routes which don't go into the IGP include:**
  - **Dynamic assignment pools (DSL/Cable/Dial)**
  - **Customer point to point link addressing**
    - **(using next-hop-self in iBGP ensures that these do NOT need to be in IGP)**
  - **Static/Hosting LANs**
  - **Customer assigned address space**
  - **Anything else not listed in the previous slide**

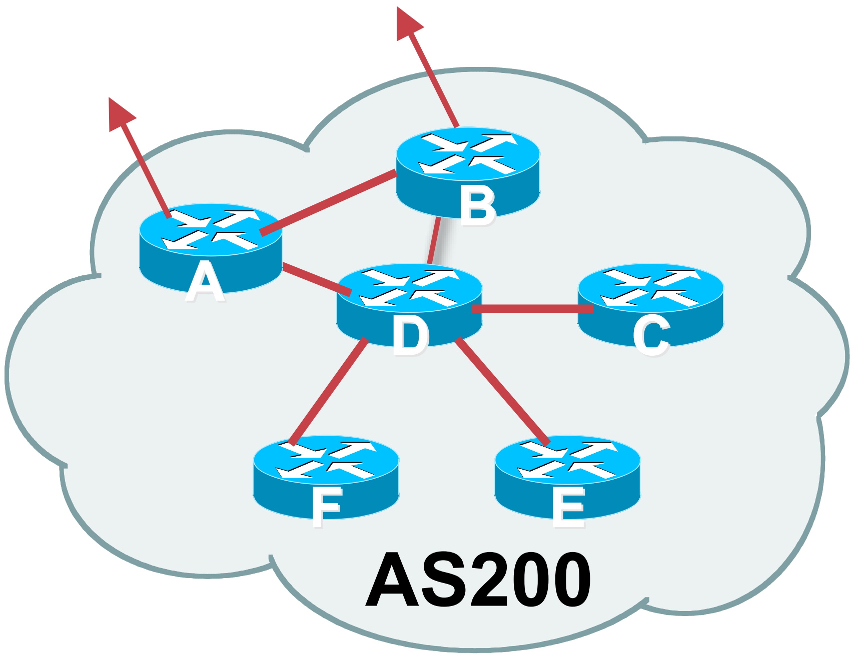
Preparing the Network
iBGP

- **Second step is to configure the local network to use iBGP**
- **iBGP can run on**
  - **all routers, or**
  - **a subset of routers, or**
  - **just on the upstream edge**
- ***iBGP must run on all routers which are in the transit path between external connections***



Preparing the Network
iBGP (Transit Path)

- ***iBGP must run on all routers which are in the transit path between external connections***
- **Routers C, E and F are not in the transit path**
  - **Static routes or IGP will suffice**
- **Router D is in the transit path**
  - **Will need to be in iBGP mesh, otherwise routing loops will result**



Preparing the Network
Layers

- **Typical SP networks have three layers:**
  - **Core – the backbone, usually the transit path**
  - **Distribution – the middle, PoP aggregation layer**
  - **Aggregation – the edge, the devices connecting customers**

Preparing the Network Aggregation Layer

- iBGP is optional
  - Many ISPs run iBGP here, either partial routing (more common) or full routing (less common)
  - Full routing is not needed unless customers want full table
  - Partial routing is cheaper/easier, might usually consist of internal prefixes and, optionally, external prefixes to aid external load balancing
    - Communities and peer-groups make this administratively easy
- Many aggregation devices can't run iBGP
  - Static routes from distribution devices for address pools
  - IGP for best exit

Preparing the Network
Distribution Layer

- **Usually runs iBGP**
  - **Partial or full routing (as with aggregation layer)**
- **But does not have to run iBGP**
  - **IGP is then used to carry customer prefixes (does not scale)**
  - **IGP is used to determine nearest exit**
- **Networks which plan to grow large should deploy iBGP from day one**
  - **Migration at a later date is extra work**
  - **No extra overhead in deploying iBGP, indeed IGP benefits**

Preparing the Network
Core Layer

- Core of network is usually the transit path
- iBGP necessary between core devices
  - Full routes or partial routes:
    - Transit ISPs carry full routes in core
    - Edge ISPs carry partial routes only
- Core layer includes AS border routers

Preparing the Network
iBGP Implementation

Decide on:

- **Best iBGP policy**

  Will it be full routes everywhere, or partial, or some mix?

- **iBGP scaling technique**

  Community policy?

  Route-reflectors?

  Techniques such as peer groups and peer templates?

Preparing the Network
iBGP Implementation

- **Then deploy iBGP:**

    Step 1: Introduce iBGP mesh on chosen routers

        make sure that iBGP distance is greater than IGP distance (it usually is)

    Step 2: Install "customer" prefixes into iBGP

    Check! Does the network still work?

    Step 3: Carefully remove the static routing for the prefixes now in IGP and iBGP

    Check! Does the network still work?

    Step 4: Deployment of eBGP follows

Preparing the Network
iBGP Implementation

Install "customer" prefixes into iBGP?

- **Customer assigned address space**
  - **Network statement/static route combination**
  - **Use unique community to identify customer assignments**
- **Customer facing point-to-point links**
  - **Redistribute connected through filters which only permit point-to-point link addresses to enter iBGP**
  - **Use a unique community to identify point-to-point link addresses (these are only required for your monitoring system)**
- **Dynamic assignment pools & local LANs**
  - **Simple network statement will do this**
  - **Use unique community to identify these networks**

Preparing the Network
iBGP Implementation

Carefully remove static routes?

- **Work on one router at a time:**

  Check that static route for a particular destination is also learned either by IGP or by iBGP

  If so, remove it

  If not, establish why and fix the problem

  (Remember to look in the RIB, not the FIB!)

- **Then the next router, until the whole PoP is done**
- **Then the next PoP, and so on until the network is now dependent on the IGP and iBGP you have deployed**

Preparing the Network Completion

- **Previous steps are NOT flag day steps**
  - **Each can be carried out during different maintenance periods, for example:**
  - **Step One on Week One**
  - **Step Two on Week Two**
  - **Step Three on Week Three**
  - **And so on**
  - **And with proper planning will have NO customer visible impact at all**

Preparing the Network
Configuration Summary

- IGP essential networks are in IGP
- Customer networks are now in iBGP
  - iBGP deployed over the backbone
  - Full or Partial or Upstream Edge only
- BGP distance is greater than any IGP
- Now ready to deploy eBGP

BGP Multihoming Techniques

- Why Multihome?
- Definition & Options
- Preparing the Network
- Basic Multihoming
- “BGP Traffic Engineering”

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

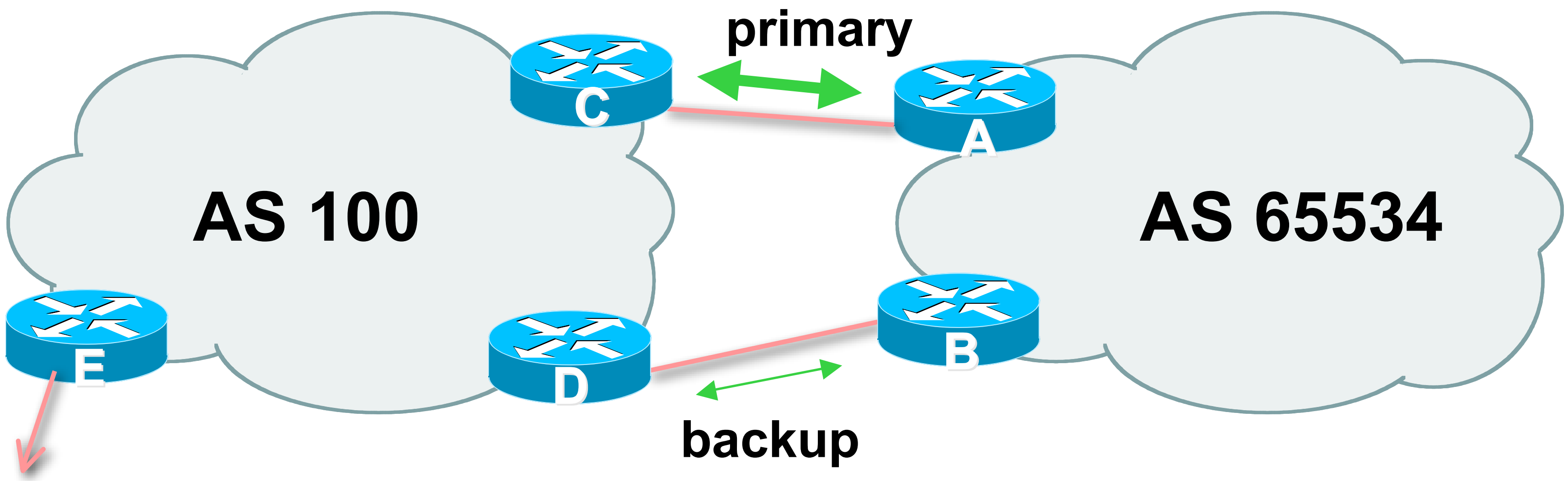
One link primary, the other link backup only

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**

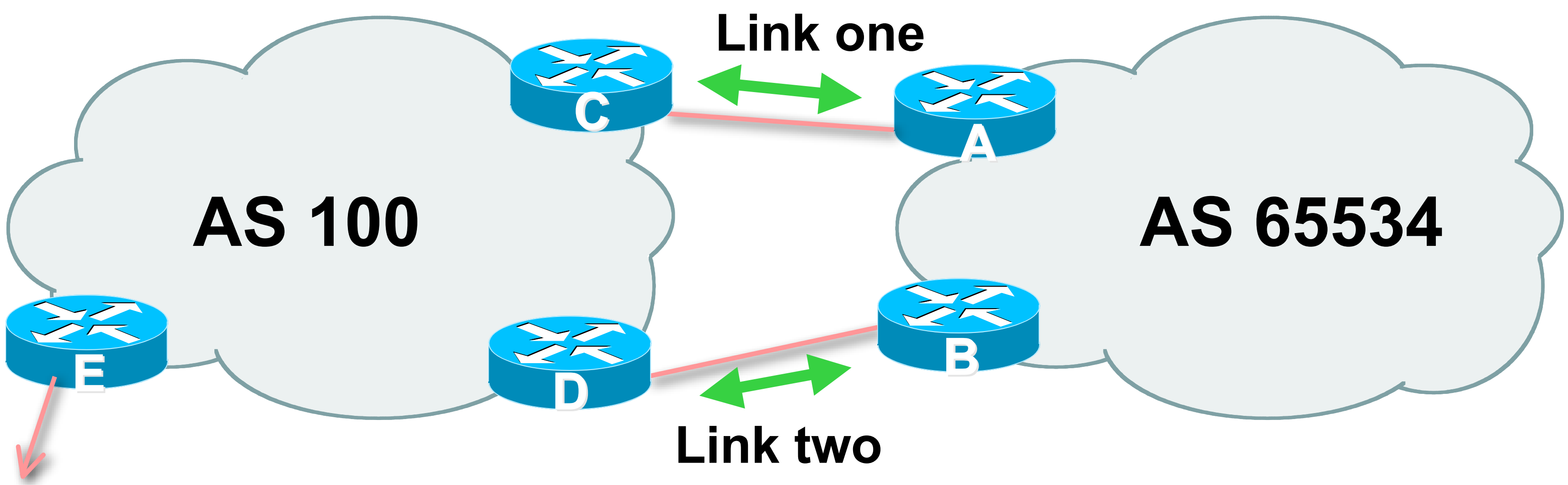
Two links to the same ISP

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

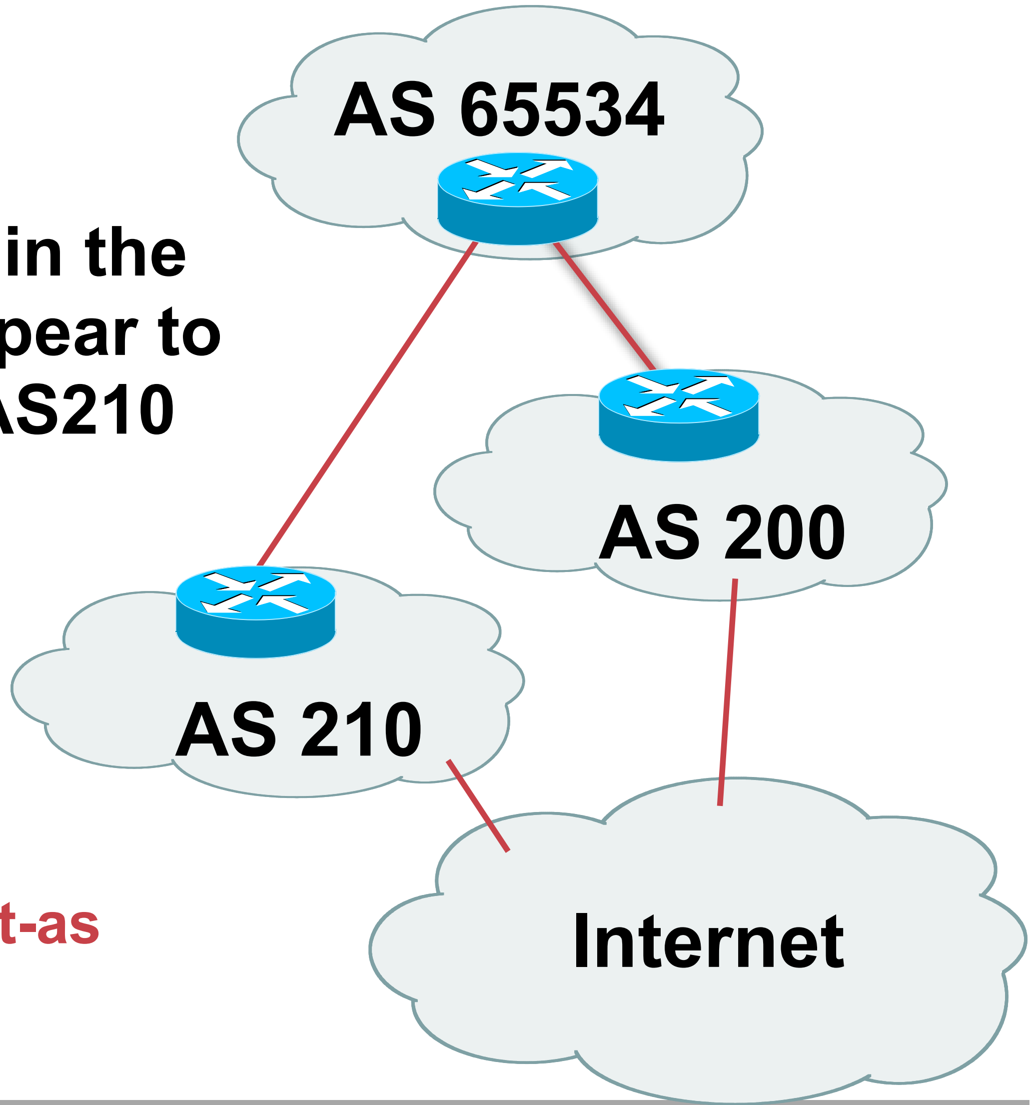
Multihoming to different ISPs

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?

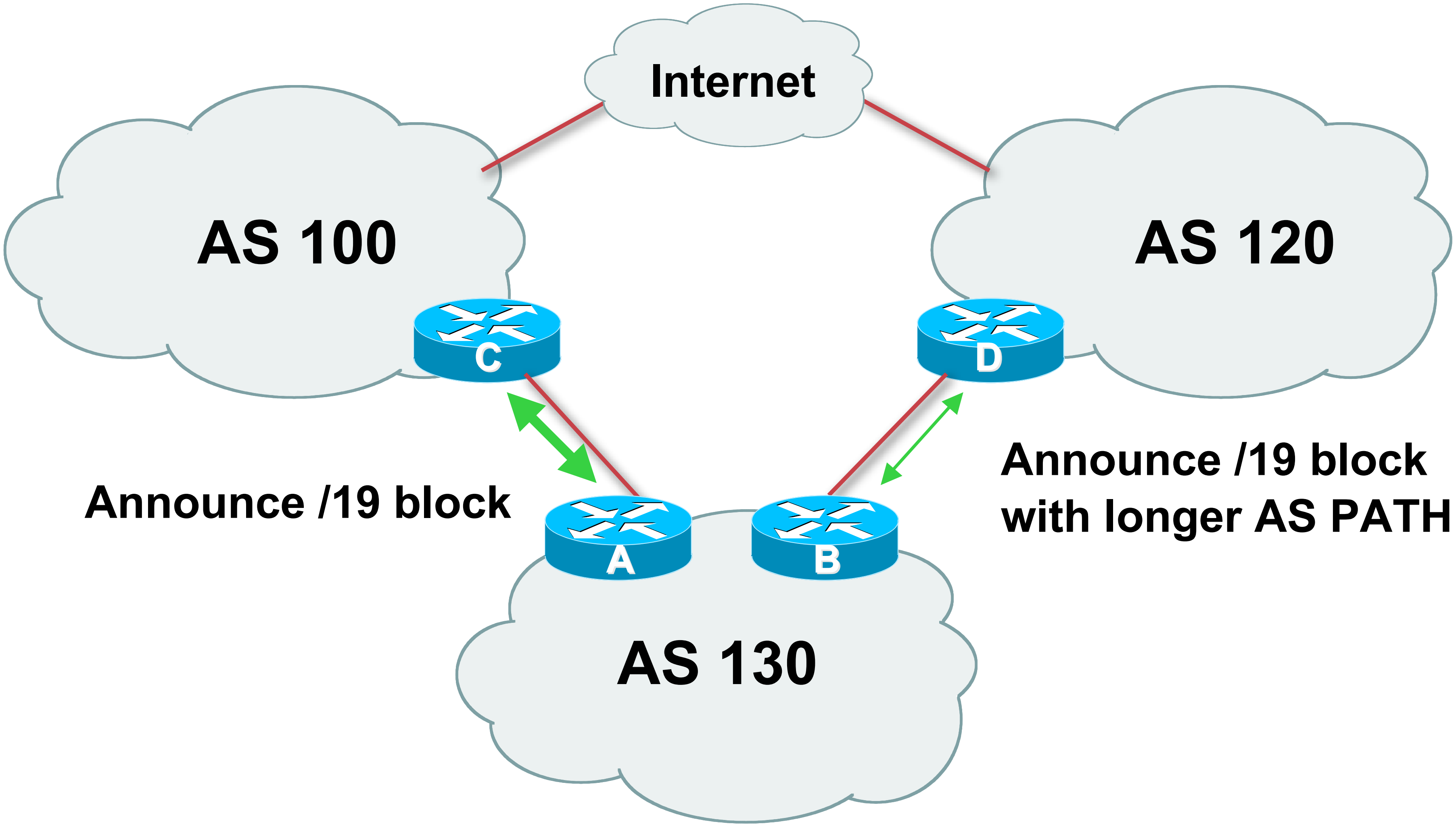
- **Viewing the prefixes originated by AS65534 in the Internet shows they appear to be originated by both AS210 and AS200**
  - **This is NOT bad**
  - **Nor is it illegal**
- **Cisco IOS command is**
  - **show ip bgp inconsistent-as**



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**

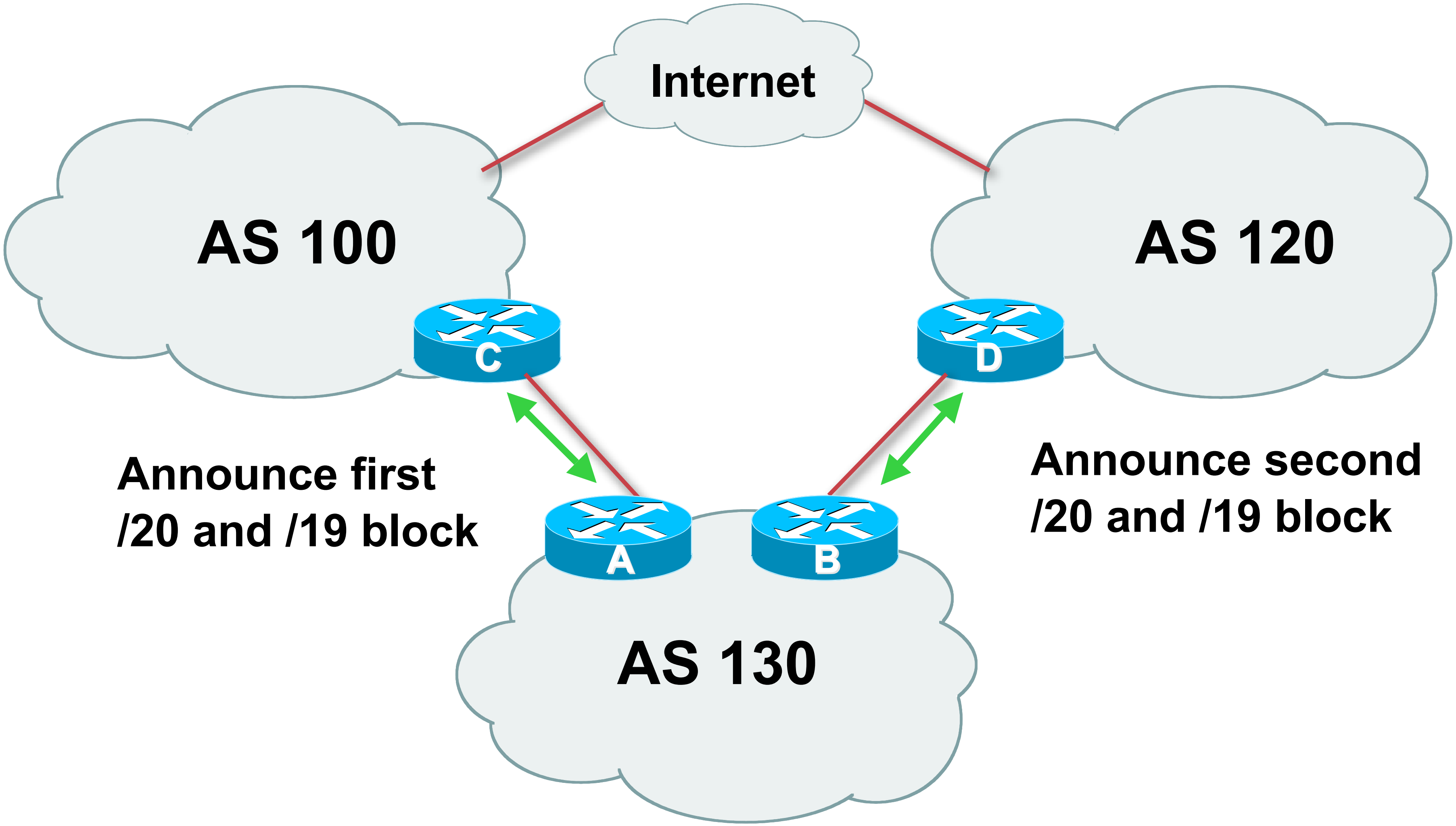
Two links to different ISPs (one as backup only)

- **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

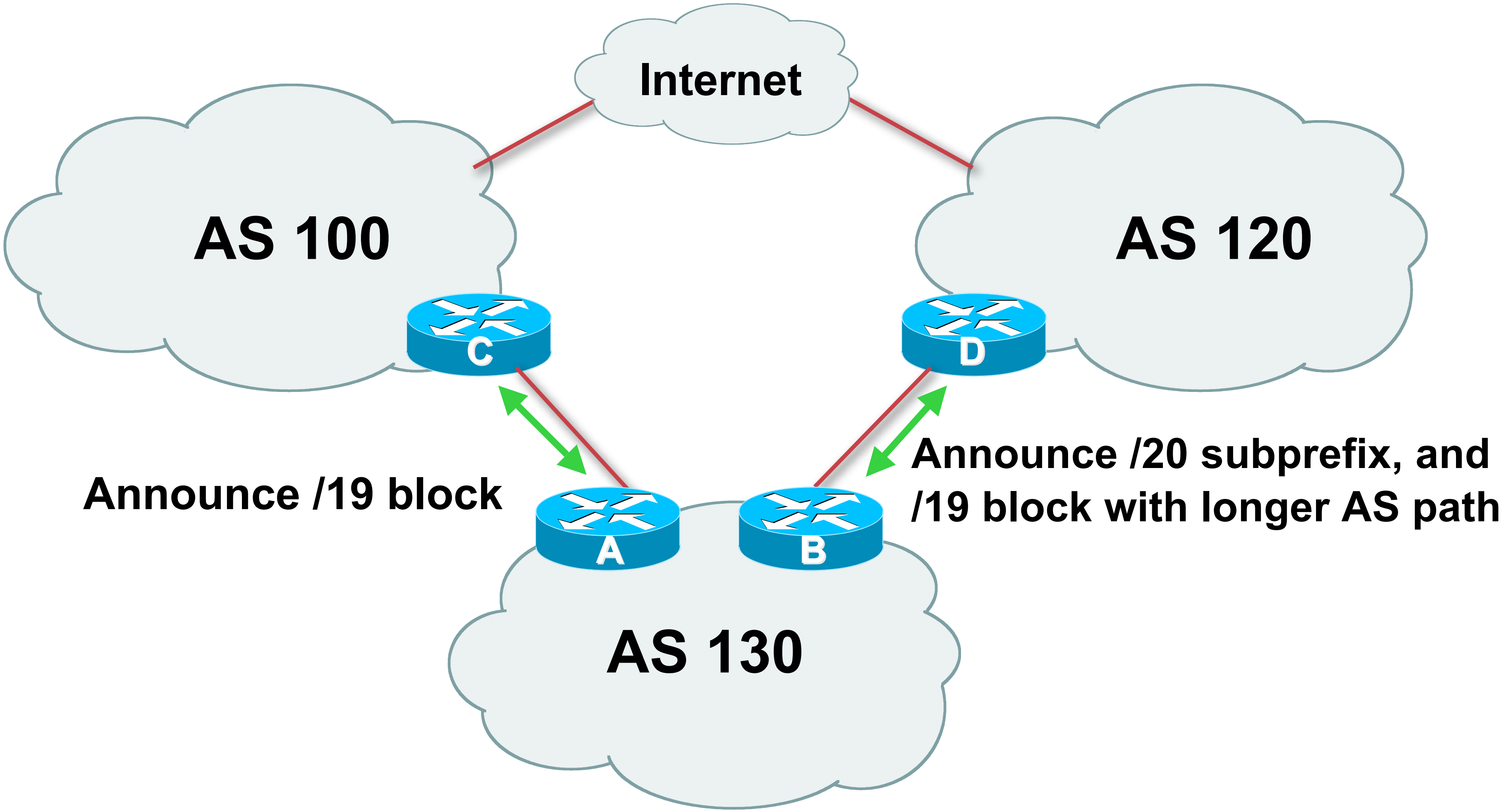
  Start with a simple concept

  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
- Preparing the Network
- Basic Multihoming
- "BGP Traffic Engineering"

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

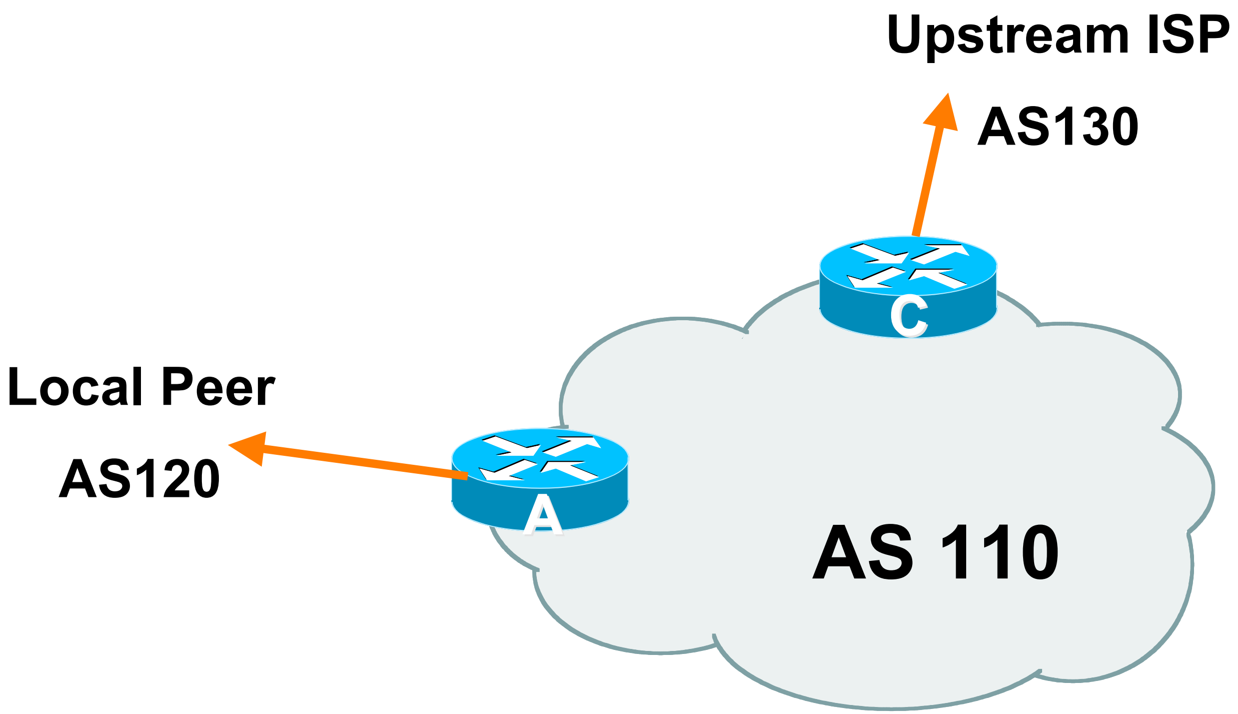
One upstream, one local peer

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

Cisco Systems

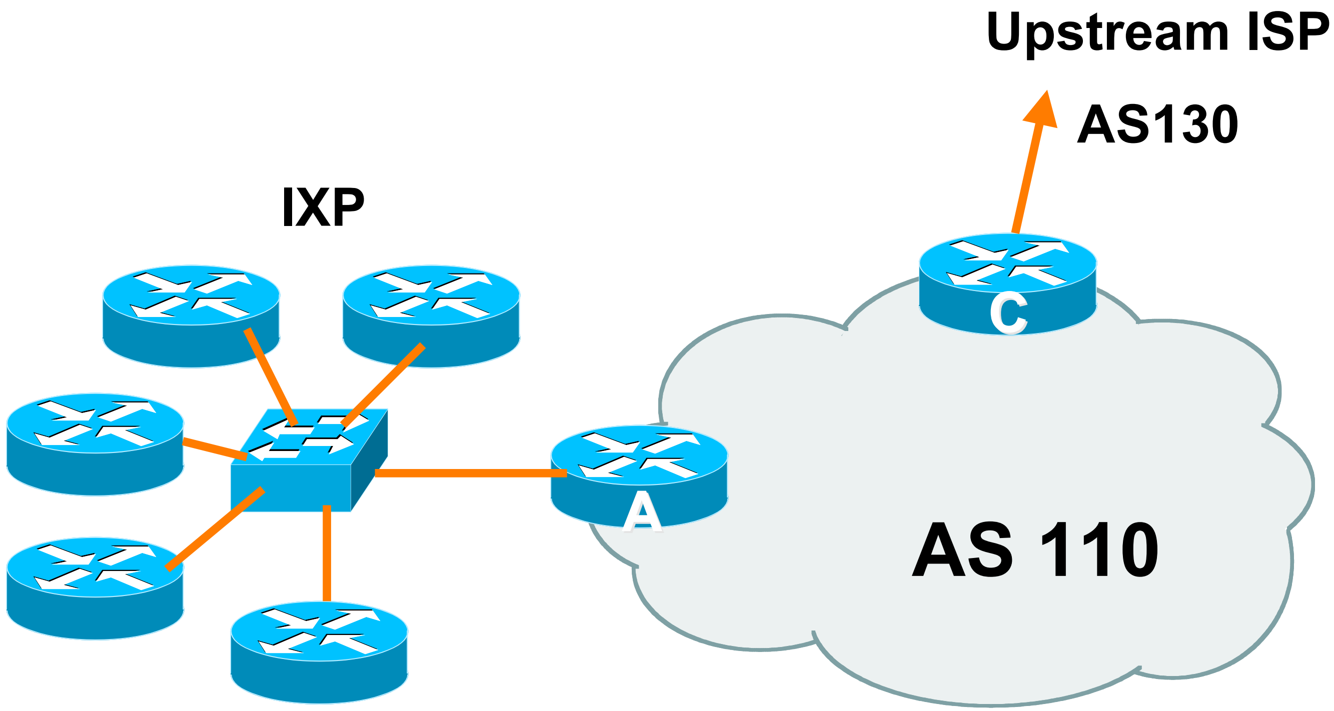
Service Provider Multihoming

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

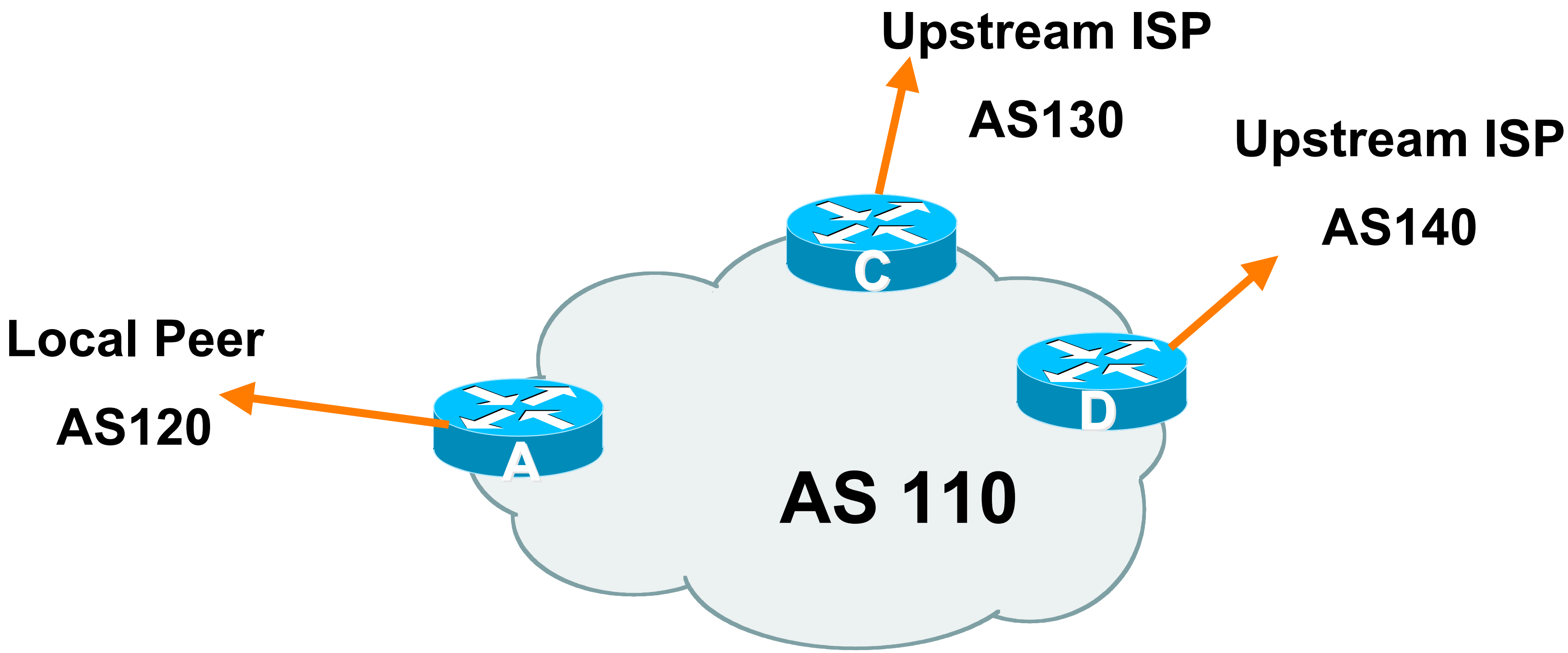
- Connect to both upstream transit providers to see the "Internet"

  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
- Preparing the Network
- Basic Multihoming
- “BGP Traffic Engineering”

BGP Multihoming Techniques

Philip Smith <pfs@cisco.com>

MENOG 1

3-5 April 2007

Bahrain