Multihoming: Introduction

ISP Workshops

Last updated 30th January 2021

Acknowledgements

- This material originated from the Cisco ISP/IXP Workshop Programme developed by Philip Smith & Barry Greene

- Use of these materials is encouraged as long as the source is fully acknowledged and this notice remains in place

- Bug fixes and improvements are welcomed
  - Please email *workshop (at) bgp4all.com*

Philip Smith

Agenda

- Why Multihome?
- The Multihoming Toolset
- How to Multihome – Options
- Basic Principles of Multihoming
- IP Addressing & Multihoming

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?

- Changing upstream provider
- With one upstream, migration means:
  - Disconnecting existing connection
  - Moving the link to the new upstream
  - Reconnecting the link
  - Reannouncing address space
  - Break in service for end users (hours, days,...?)
- With two upstreams, migration means:
  - Bring up link with new provider (including BGP and address announcements)
  - Disconnect link with original upstream
  - No break in service for end users

Why Multihome?

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

Why Multihome?

- Summary:
  - Multihoming is easy to demand as requirement of any operation
  - But what does it really mean:
    - In real life?
    - For the network?
    - For the Internet?
  - And how do we do it?

Multihoming Definition

- More than one link external to the local network
  - Two or more links to the same AS
  - Two or more links to different ASes
- Usually **two** external facing routers
  - One router gives link and provider redundancy only

Multihoming

- The scenarios described here apply equally well to:
  - End-sites being customers of network operators *and*
  - Network operators being customers of other network operators

- Implementation details may be different, for example:
  - End site → ISP Configuration on End-Site
  - ISP1 → ISP2 Network Operators share config

Multihoming: Number Resources

- BGP handles the relationship between Autonomous Systems
  - Each autonomous system is represented by an Autonomous System Number (ASN)
  - Each multihoming organisation requires their own unique ASN
- Address space (IPv4/IPv6) for each autonomous system comes from either:
  - Their upstream *or*
  - A Regional Internet Registry

Autonomous System Number (ASN)

Range:	
0-4294967295	(32-bit range – RFC6793)
	(0-65535 was original 16-bit range

Usage:	
0 and 65535	(reserved)
1-64495	(public Internet)
64496-64511	(documentation – RFC5398)
64512-65534	(private use only)
23456	(represent 32-bit range in 16-bit world)
65536-65551	(documentation – RFC5398)
65552-4199999999	(public Internet)
4200000000-4294967295	(private use only)

- 32-bit range representation specified in RFC5396
  - Defines “asplain” (traditional format) as standard notation

Autonomous System Number

- ASNs are distributed by the Regional Internet Registries
  - They are also available from upstream ISPs who are members of one of the RIRs
- The entire 16-bit ASN pool has been assigned to the RIRs
  - Around 41500 16-bit ASNs are visible on the Internet
    - (this number is dropping slightly as 32-bit ASN numbers increase)
- Each RIR has also received a block of 32-bit ASNs
  - Out of 35000 assignments, around 29000 are visible on the Internet (January 2021)
- See www.iana.org/assignments/as-numbers

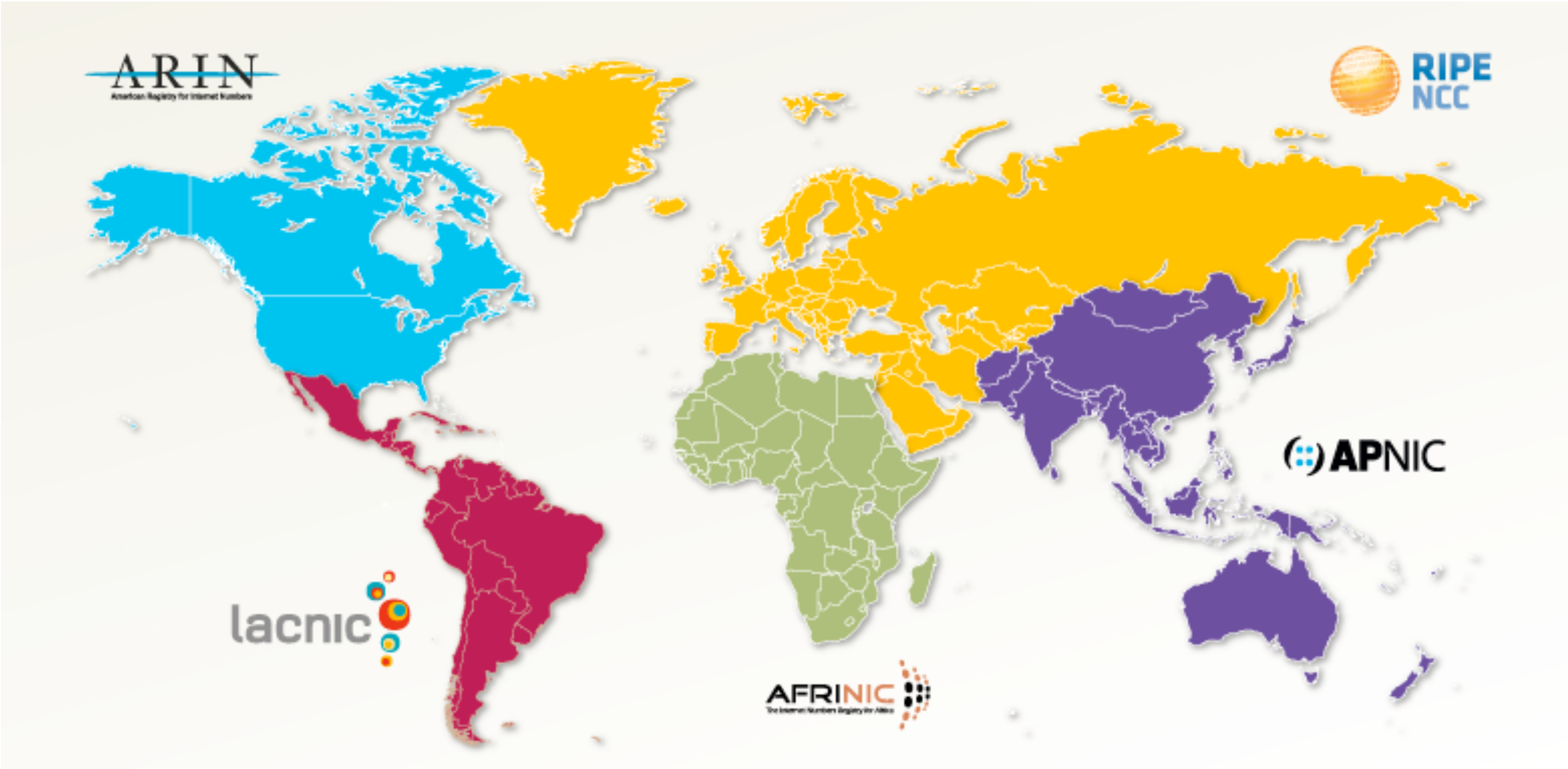
IP Addressing

- IP addresses are also distributed by the Regional Internet Registries
  - They are also available from upstream providers who are members of one of the RIRs
- The entire IPv4 address pool has been almost exhausted
  - The RIRs are operating in “IPv4 runout” mode now
- IPv6 address space is plentiful
  - Network operators receive at least a /32
  - End sites/users receive at least a /48

Where to get Internet Numbering Resources

- Your upstream provider
- Africa
  - AfriNIC – http://www.afrinic.net
- Asia and the Pacific
  - APNIC – http://www.apnic.net
- North America
  - ARIN – http://www.arin.net
- Latin America and the Caribbean
  - LACNIC – http://www.lacnic.net
- Europe and Middle East
  - RIPE NCC – http://www.ripe.net/info/ncc

Internet Registry Regions



Private AS – Application

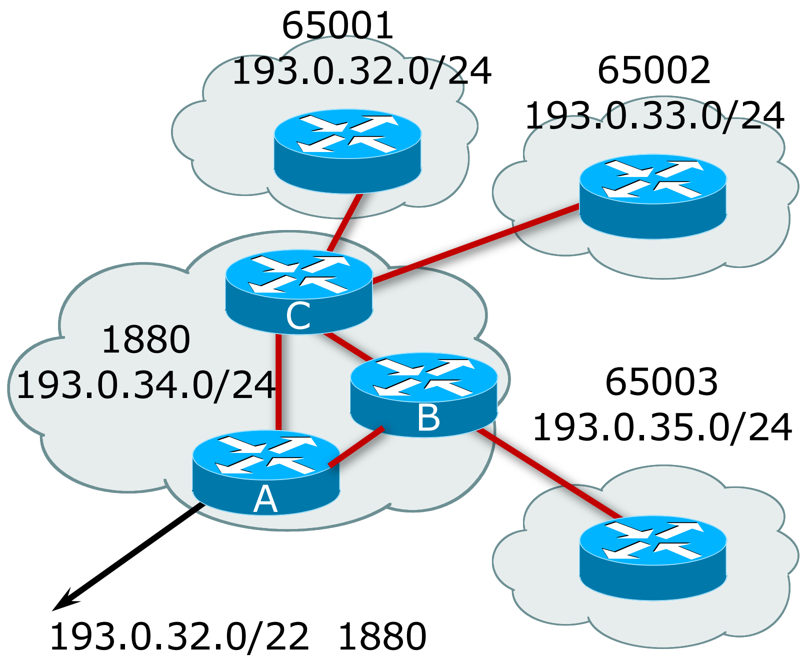
- A network operator with end-sites multihomed on their backbone (RFC2270)

  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 must not be leaked to or used on the public Internet
- Cisco IOS

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

More Definitions

- Transit
  - Carrying traffic across a network
  - Usually **for a fee**
- Peering
  - Exchanging routing information and traffic
  - Usually **for no fee**
  - Sometimes called **settlement free peering**
- Default
  - Where to send traffic when there is no explicit match in the routing table

Configuring Policy – Cisco IOS

- Assumptions:
  - Prefix-lists are used throughout
  - Easier/better/faster than access-lists
- Three BASIC Principles
  - Prefix-lists to filter prefixes
  - Filter-lists to filter ASNs
  - Route-maps to apply policy
- Route-maps can be used for filtering, but this is more "advanced" configuration

Policy Tools

- Local preference
  - Outbound traffic flows
- Metric (MED)
  - Inbound traffic flows (local scope)
- AS-PATH prepend
  - Inbound traffic flows (Internet scope)
- Subdividing Aggregates
  - Inbound traffic flows (local & 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
- Minimum allocations:
  - IPv4 is /24
  - IPv6 is /48 (endsite) and /32 (operator)
  - Several Network Operators filter RIR blocks on published minimum allocation boundaries
  - Several Network Operators filter the rest of address space according to the IANA assignments
  - This activity is called “Net Police” by some

Originating Prefixes

- The RIRs publish their minimum allocation sizes per /8 address block
  - AfriNIC: www.afrinic.net/library/policies/126-afpub-2005-v4-001
  - 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" Network Operators 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 Team Cymru or other reputable bogon BGP feed:
  - https://www.team-cymru.com/bogon-reference-bgp.html

How to Multihome

Some choices…

Transits

- Transit provider is another autonomous system which is used to provide the local network with access to other networks
  - Might be local or regional only
  - But more usually the whole Internet
- Transit providers need to be chosen wisely:
  - Only one
    - No redundancy
  - Too many
    - More difficult to load balance
    - No economy of scale (costs more per Mbps)
    - Hard to provide service quality
- **Recommendation: at least two, no more than three**

Common Mistakes

- Network Operators sign up with too many transit providers
  - Lots of small circuits (cost more per Mbps than larger ones)
  - Transit rates per Mbps reduce with increasing transit bandwidth purchased
  - Hard to implement reliable traffic engineering that doesn't need daily fine tuning depending on customer activities
- No diversity
  - Chosen transit providers all reached over same satellite or same submarine cable
  - Chosen transit providers have poor onward transit and peering

Peers

- A peer is another autonomous system with which the local network has agreed to exchange locally sourced routes and traffic
- Private peer
  - Private link between two providers for the purpose of interconnecting
- Public peer
  - Internet Exchange Point, where providers meet and freely decide who they will interconnect with
- **Recommendation: peer as much as possible!**

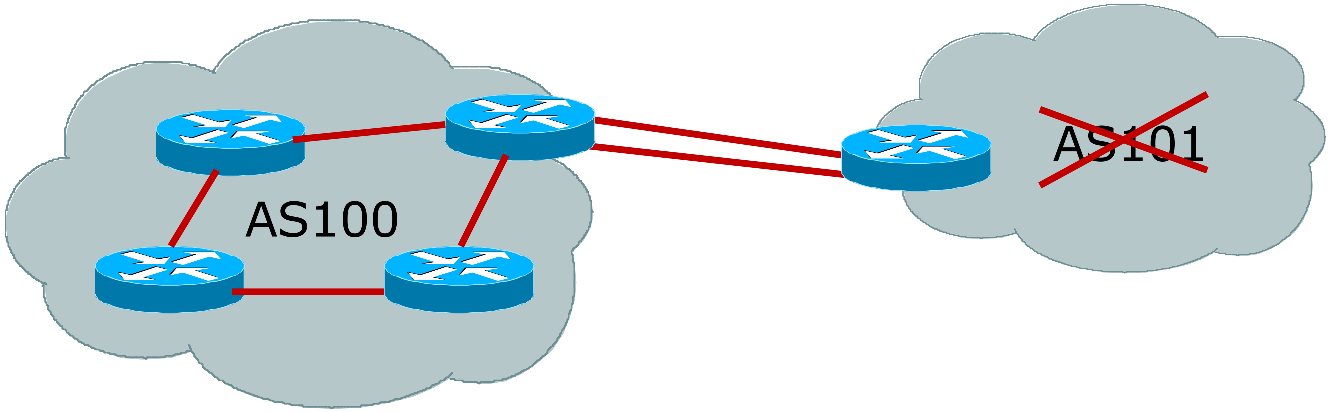
Common Mistakes

- Mistaking a transit provider's “Exchange” business for a no-cost public peering point
- Not working hard to get as much peering as possible
  - Physically near a peering point (IXP) but not present at it
  - (Transit sometimes is cheaper than peering!!)
- Ignoring/avoiding competitors because they are competition
  - Even though potentially valuable peering partner to give customers a better experience

Multihoming Scenarios

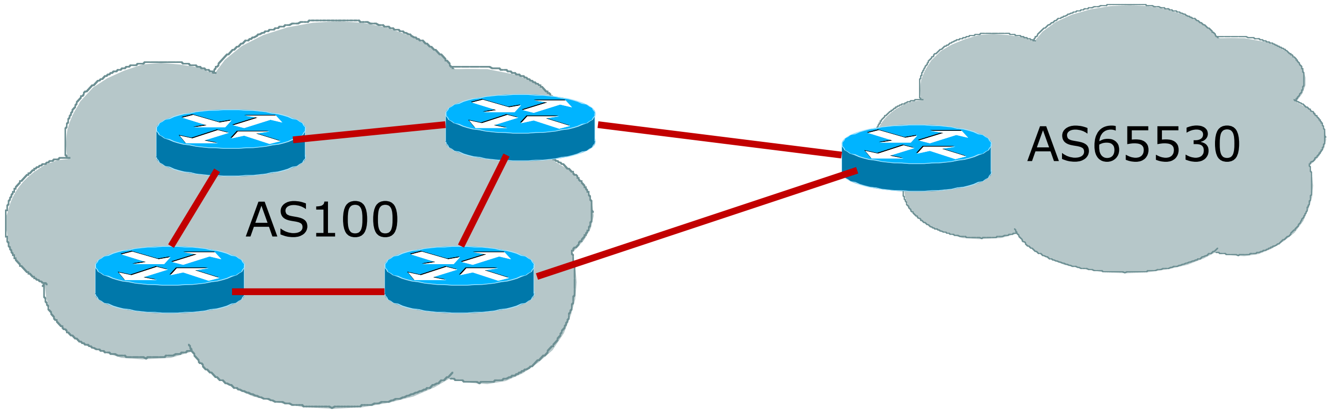
- Stub network
- Multi-homed stub network
- Multi-homed network
- Multiple Sessions between two ASes

Stub Network



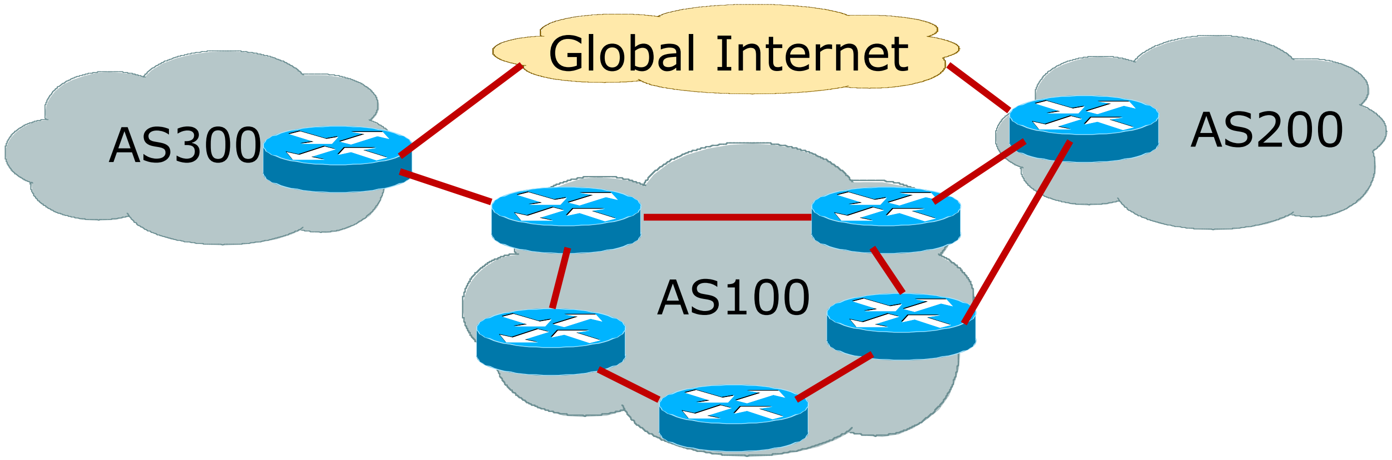
- No need for BGP
- Point static default to upstream AS
- Upstream AS advertises stub network
- Policy confined within upstream AS's policy

Multi-homed Stub Network



- Use BGP (not IGP or static) to loadshare
- Use private AS number (see earlier for ranges)
- Upstream AS advertises stub network
- Policy confined within upstream AS's policy

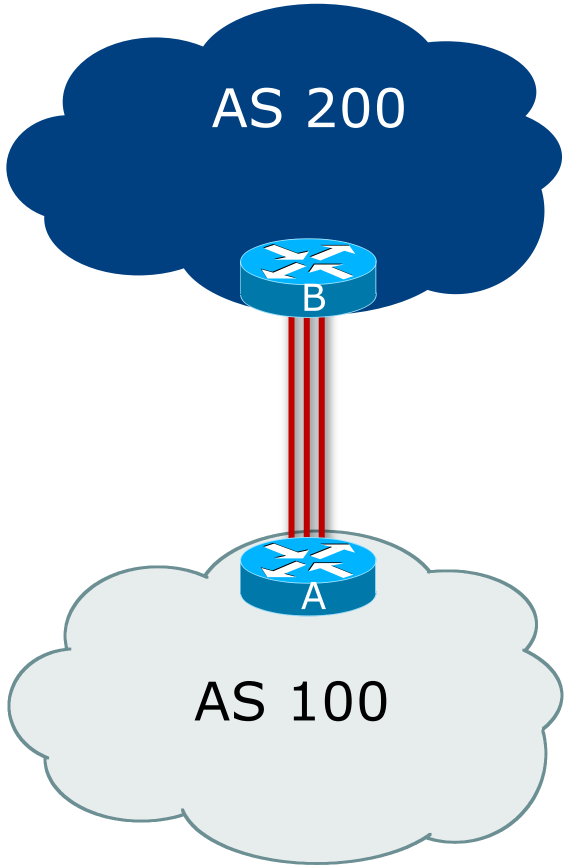
Multi-homed Network



- Several situations possible, including:
  1. Multiple sessions to same AS
  2. Secondary for backup only
  3. Load-share between primary and secondary
  4. Selectively use different ASes

Multiple Sessions between two ASes

- Several options
  - EBGP multihop
  - BGP multipath
  - CEF loadsharing
  - BGP attribute manipulation

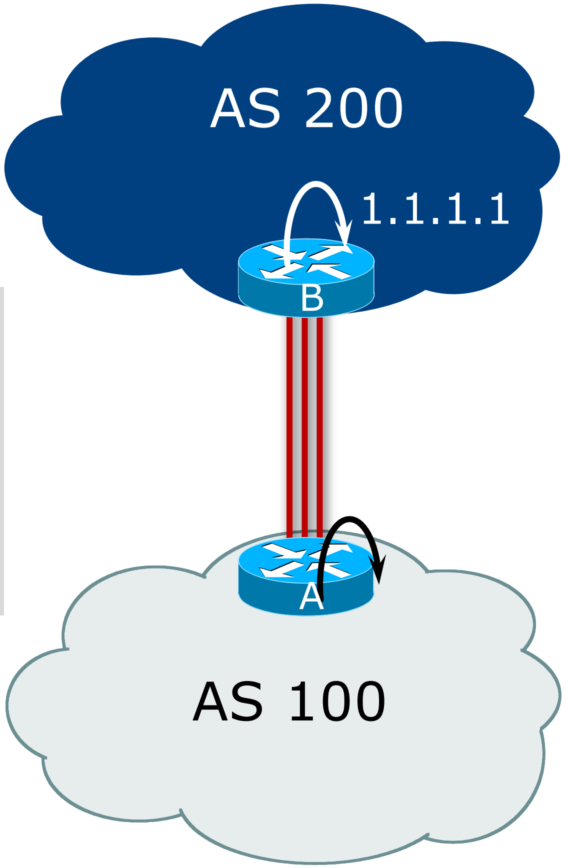


Multiple Sessions between two ASes – EBGP multihop

- Use ebgp-multihop
  - Run EBGP between loopback addresses
  - EBGP prefixes learned with loopback address as next hop
- Cisco IOS

```
router bgp 100
 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
```

- Common error made is to point remote loopback route at IP address rather than specific link

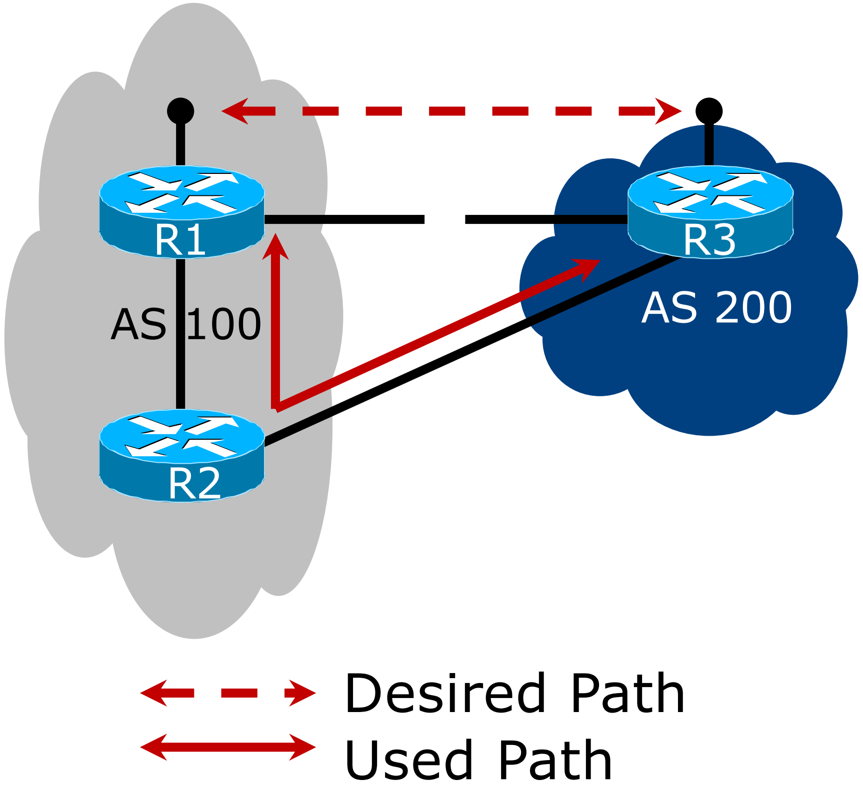


Multiple Sessions between two ASes – EBGP multihop

- One serious ebgp-multihop caveat:
  - R1 and R3 are EBGP peers that are loopback peering
  - Configured with:

```
neighbor x.x.x.x ebgp-multihop 2
```

  - If the R1 to R3 link goes down the session could establish via R2
- Usually happens when routing to remote loopback is dynamic, rather than static pointing at a link



Multiple Sessions between two ASes – EBGP multihop

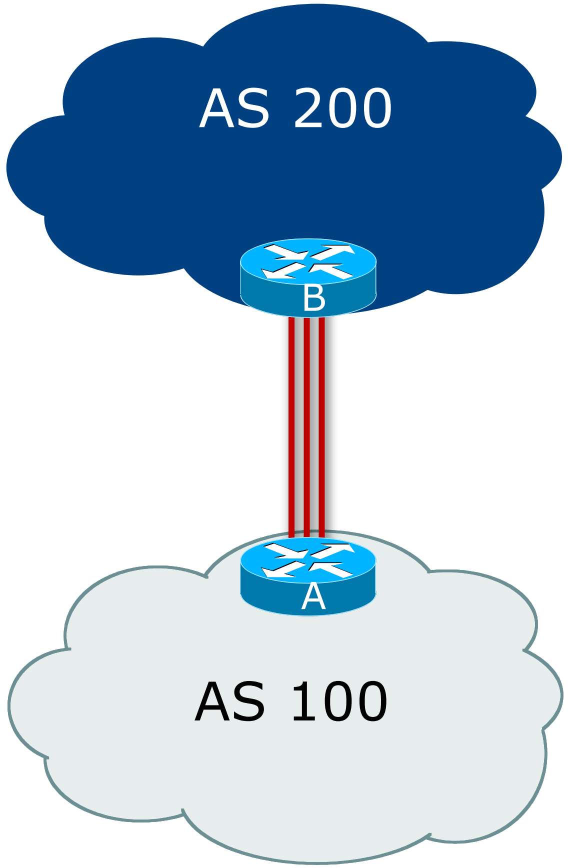
- Try and avoid use of ebgp-multihop unless:
  - It's absolutely necessary –or–
  - Loadsharing across multiple links
- Many Network Operators discourage its use, for example:

> 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 between two ASes – bgp multi path

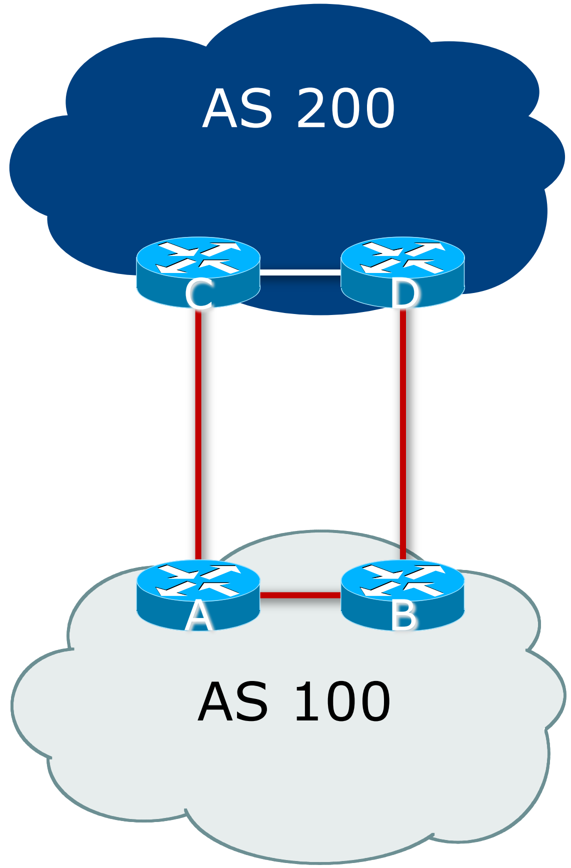
- Three BGP sessions required
- Platform limit on number of paths (could be as little as 6)
- Full BGP feed makes this unwieldy
  - 3 copies of Internet Routing Table goes into the FIB

```
router bgp 100
 neighbor 100.64.2.1 remote-as 200
 neighbor 100.64.2.5 remote-as 200
 neighbor 100.64.2.9 remote-as 200
 maximum-paths 3
```



Multiple Sessions between two ASes – BGP attributes & filters

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



Basic Principles of Multihoming

Let’s learn to walk before we try running…

The Basic Principles

- Announcing address space attracts traffic
  - (Unless policy in upstream providers interferes)
- Announcing the AS aggregate out a link will result in traffic for that aggregate coming in that link
- Announcing a subprefix of an aggregate out a link means that all traffic for that subprefix will come in that link, even if the aggregate is announced somewhere else
  - The most specific announcement wins!

The Basic Principles

- To split traffic between two links:
  - Announce the aggregate on both links – ensures redundancy
  - Announce one half of the address space on each link
  - (This is the first step, all things being equal)
- Results in:
  - Traffic for first half of address space comes in first link
  - Traffic for second half of address space comes in second link
  - If either link fails, the fact that the aggregate is announced ensures there is a backup path

The Basic Principles

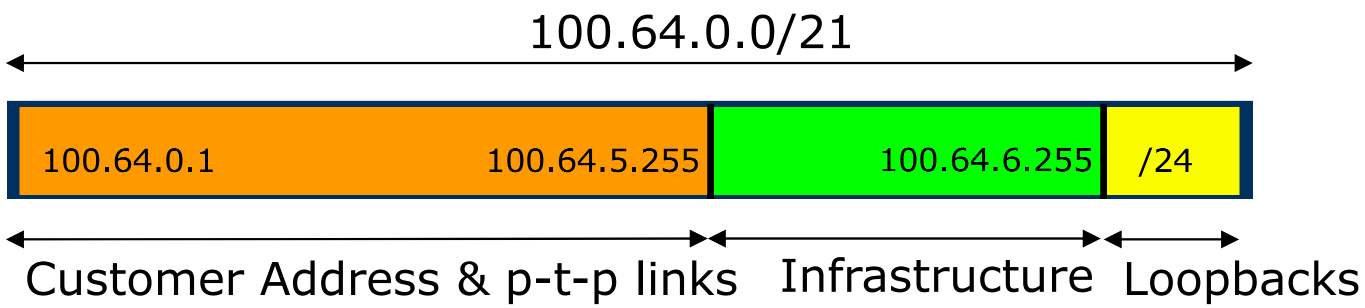
- The keys to successful multihoming configuration:
  - Keeping traffic engineering prefix announcements independent of customer IBGP
  - Understanding how to announce aggregates
  - Understanding the purpose of announcing subprefixes of aggregates
  - Understanding how to manipulate BGP attributes
  - Too many upstreams/external paths makes multihoming harder (2 or 3 is enough!)

IP Addressing & Multihoming

How Good IP Address Plans assist with Multihoming

IP Addressing & Multihoming

- IP Address planning is an important part of Multihoming
- Previously have discussed separating:
  - Customer address space
  - Customer p-t-p link address space
  - Infrastructure p-t-p link address space
  - Loopback address space

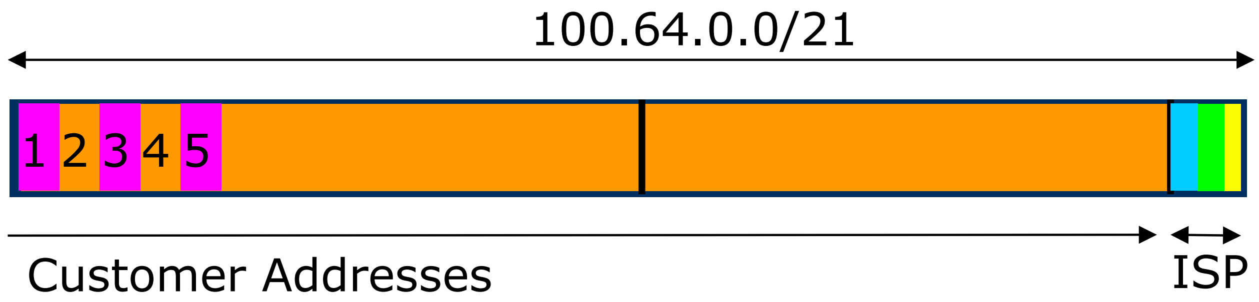


IP Addressing & Multihoming

- Router loopbacks and backbone point-to-point links make up a small part of total address space
  - And they don't attract traffic, unlike customer address space
- Links from the Network Operator's Aggregation edge to customer router needs one /30
  - Small requirements compared with total address space
  - Some operators use IP unnumbered
- Planning customer assignments is a very important part of multihoming
  - Traffic engineering involves subdividing aggregate into pieces until load balancing works

Unplanned IP addressing

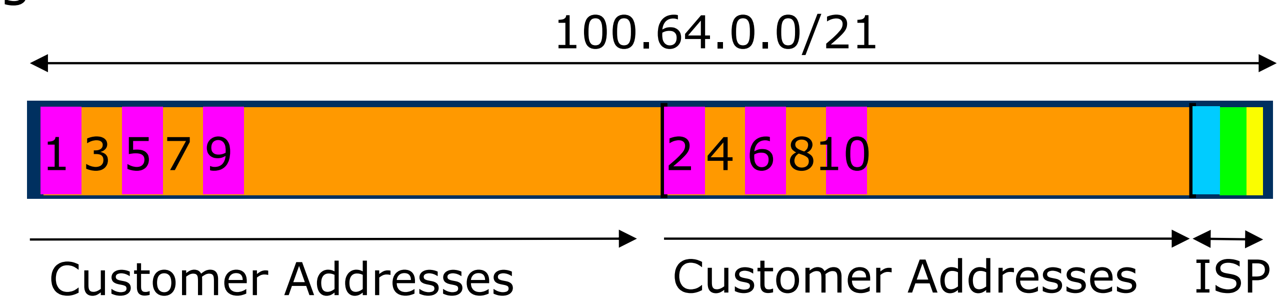
- Network Operator fills up customer IP addressing from one end of the range:



- Customers generate traffic
  - Dividing the range into two pieces will result in one /22 with all the customers, and one /22 with just the Network Operator infrastructure the addresses
  - No loadbalancing as all traffic will come in the first /22
  - Means further subdivision of the first /22 = harder work

Planned IP addressing

- If Network Operator fills up customer addressing from both ends of the range:



- Scheme then is:
  - First customer from first /22, second customer from second /22, third from first /22, etc
- This works also for residential versus commercial customers:
  - Residential from first /22
  - Commercial from second /22

Planned IP Addressing

- This works fine for multihoming between two upstream links (same or different providers)
- Can also subdivide address space to suit more than two upstreams
  - Follow a similar scheme for populating each portion of the address space
- Don’t forget to always announce an aggregate out of each link

Summary

- Presentation has covered:
  - Why Multihome?
  - The Multihoming Toolset
  - How to Multihome – Options
  - Basic Principles of Multihoming
  - IP Addressing & Multihoming

Multihoming: Introduction

ISP Workshops