BGP Multihoming Techniques

Philip Smith   <pfs@cisco.com>

SANOG I, Kathmandu, Nepal – January 2003

Presentation Slides

- **Available on**

    ftp://ftp-eng.cisco.com/pfs/seminars/
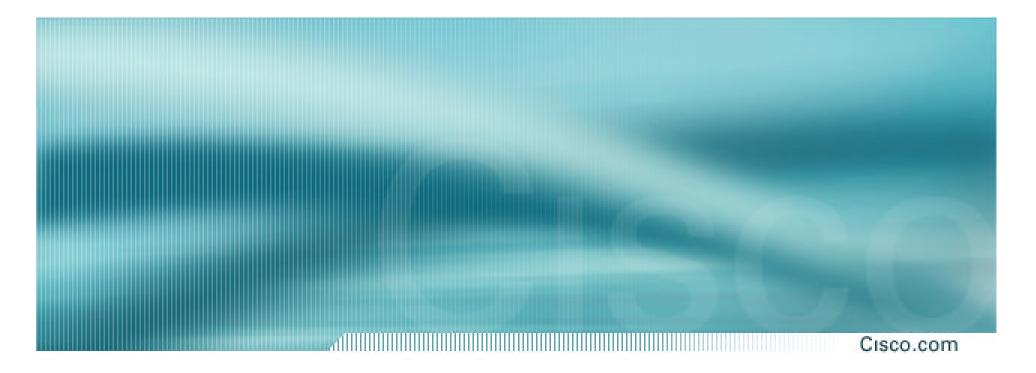
Introduction

- **Presentation has many configuration examples**

- **Uses Cisco IOS CLI**

- **Aimed at Service Providers**

    Techniques can be used by many enterprises too

- **Feel free to ask questions**

BGP Multihoming Techniques

- **Definition & Options**

- **Connecting to the same ISP**

- **Connecting to different ISPs**

- **Service Provider Multihoming**

- **Using Communities**

- **Case Study**

Multihoming Definition & Options

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**

- **Necessary when you have links to more than one ISP or exchange point**

Configuring Policy

- **Three BASIC Principles**

  prefix-lists to filter **prefixes**

  filter-lists to filter **ASNs**

  route-maps to apply **policy**

- **Avoids confusion!**

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

- **Basic Assumptions**

  MUST announce assigned address block to Internet

  MAY also announce subprefixes – reachability is not guaranteed

  RIR minimum allocation is /20

    several ISPs filter RIR blocks on this boundary

    called "Net Police" by some

Par

```
!! APNIC
ip prefix-list FILTER permit 61.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 202.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 210.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 218.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 220.0.0.0/7 ge 9 le 20
!! ARIN
ip prefix-list FILTER permit 24.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 63.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 64.0.0.0/6 ge 9 le 20
ip prefix-list FILTER permit 68.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 199.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 204.0.0.0/6 ge 9 le 20
ip prefix-list FILTER permit 208.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 216.0.0.0/8 ge 9 le 20
!! RIPE NCC
ip prefix-list FILTER permit 62.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 80.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 82.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 193.0.0.0/8 ge 9 le 20
ip prefix-list FILTER permit 194.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 212.0.0.0/7 ge 9 le 20
ip prefix-list FILTER permit 217.0.0.0/8 ge 9 le 20
!! LACNIC
ip prefix-list FILTER permit 200.0.0.0/8 ge 9 le 20
```
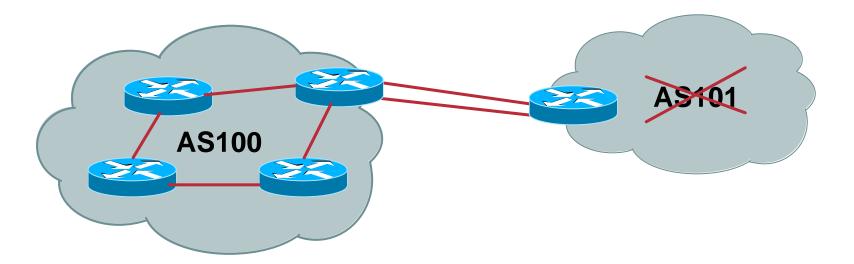
"Net Police" prefix list issues

- 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**
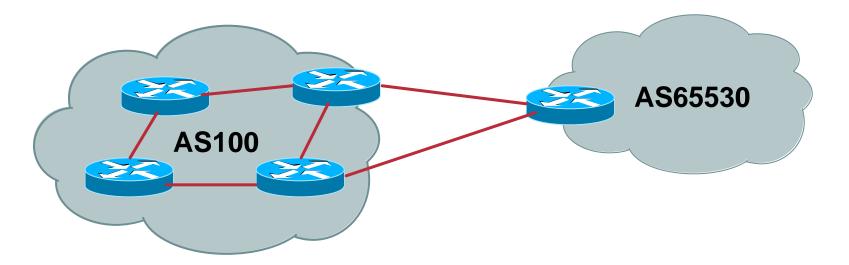
Multihoming Scenarios

- **Stub network**

- **Multi-homed stub network**

- **Multi-homed network**

- **Configuration Options**

Stub Network

AS100

AS101

- **No need for BGP**

- **Point static default to upstream ISP**

- **Upstream ISP advertises stub network**

- **Policy confined within upstream ISP's policy**
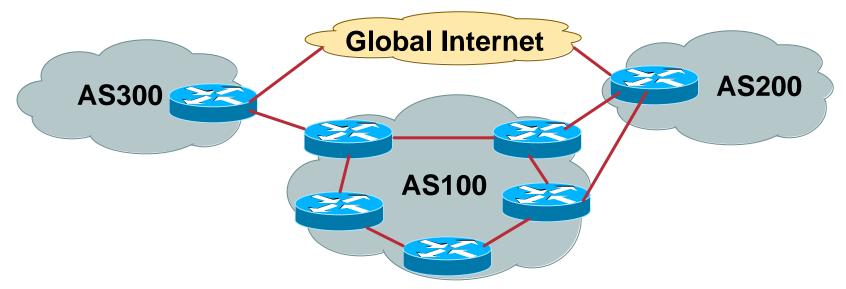
Multi-homed Stub Network

AS65530

AS100

- **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

Global Internet

AS300

AS200

AS100

- **Many situations possible**

    multiple sessions to same ISP

    secondary for backup only

    load-share between primary and secondary
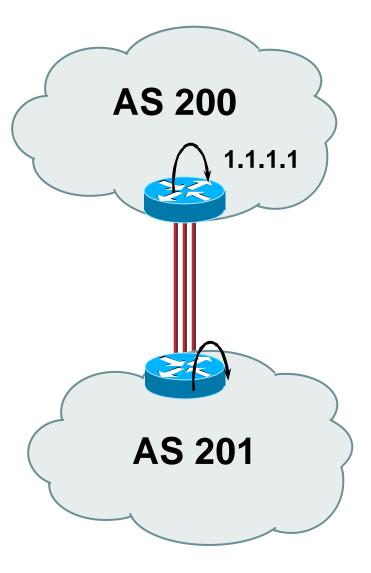
    selectively use different ISPs

Multiple Sessions to an ISP – Example One

- eBGP multihop

- eBGP to loopback addresses

- eBGP prefixes learned with loopback address as next hop

AS 200

1.1.1.1

```
router bgp 201
 neighbor 1.1.1.1 remote-as 200
 neighbor 1.1.1.1 ebgp-multihop 5
!
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
```

AS 201

Multiple Sessions to an ISP – Example Two

- **BGP multi-path**

- **Three BGP sessions required**

- **limit of 6 parallel paths**

```
router bgp 201
 neighbor 1.1.2.1 remote-as 200
 neighbor 1.1.2.5 remote-as 200
 neighbor 1.1.2.9 remote-as 200
 maximum-paths 3
```
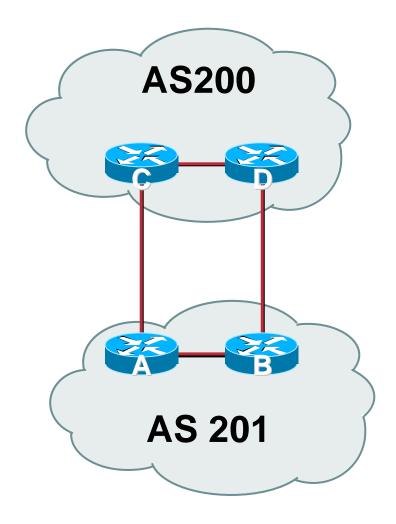
AS 200

AS 201

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**

- **No magic solution**
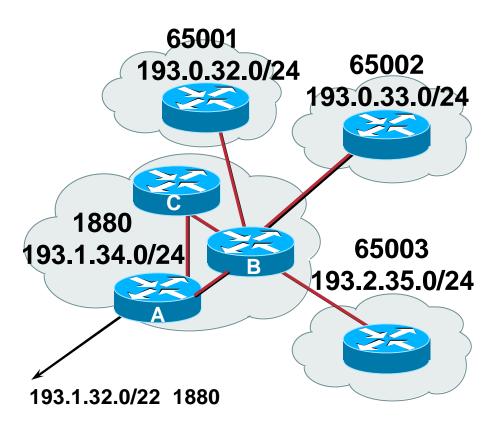
AS200

C   D

A   B

AS 201

Private-AS – Application

- ## Applications

    ISP with single-homed customers (RFC2270)

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

65001
193.0.32.0/24

65002
193.0.33.0/24

C

1880
193.1.34.0/24

B

65003
193.2.35.0/24

A

193.1.32.0/22 1880

Private-AS Removal

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

- **Please include in all eBGP configurations**

- **Rules:**

    Available for eBGP neighbours only

    if the update has AS_PATH made up of private-AS numbers, the private-AS will be dropped

    if the AS_PATH includes private and public AS numbers, private AS number will not be removed…it is a configuration error!

    if AS_PATH contains the AS number of the eBGP neighbor, the private-AS numbers will not be removed

    if used with confederations, it will work as long as the private AS numbers are after the confederation portion of the AS_PATH

BGP Multihoming Techniques

- **Definition & Options**

- **Connecting to the same ISP**

- **Connecting to different ISPs**

- **Service Provider Multihoming**

- **Using Communities**

- **Case Study**

Multihoming to the same ISP

Multihoming to the same ISP

- ## Use BGP for this type of multihoming

  use a private AS (ASN > 64511)

- ## upstream ISP proxy aggregates

  in other words, announces only your address block to the Internet (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

primary

AS 100

AS 65534

backup

- **AS109 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 makes standard announcement

  backup link increases metric on outbound, and reduces local-pref on inbound

- ## 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 A Configuration**

```
router bgp 65534
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.2 remote-as 100
 neighbor 222.222.10.2 description RouterC
 neighbor 222.222.10.2 prefix-list aggregate out
 neighbor 222.222.10.2 prefix-list default in
!
ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
```

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

- **Router B Configuration**

```
router bgp 65534
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.6 remote-as 100
  neighbor 222.222.10.6 description RouterD
  neighbor 222.222.10.6 prefix-list aggregate out
  neighbor 222.222.10.6 route-map routerD-out out
  neighbor 222.222.10.6 prefix-list default in
  neighbor 222.222.10.6 route-map routerD-in in
!
..next slide
```

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

```
ip prefix-list aggregate permit 221.10.0.0/19

ip prefix-list default permit 0.0.0.0/0

!

route-map routerD-out permit 10
 match ip address prefix-list aggregate
 set metric 10
route-map routerD-out permit 20

!

route-map routerD-in permit 10
 set local-preference 90

!
```

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

- **Router C Configuration (main link)**

```
router bgp 100
 neighbor 222.222.10.1 remote-as 65534
 neighbor 222.222.10.1 default-originate
 neighbor 222.222.10.1 prefix-list Customer in
 neighbor 222.222.10.1 prefix-list default out
!
ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```

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

- **Router D Configuration (backup link)**

```
router bgp 100
  neighbor 222.222.10.5 remote-as 65534
  neighbor 222.222.10.5 default-originate
  neighbor 222.222.10.5 prefix-list Customer in
  neighbor 222.222.10.5 prefix-list default out
!
ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```

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

- **Router E Configuration**

```
router bgp 100
  neighbor 222.222.10.17 remote-as 110
  neighbor 222.222.10.17 remove-private-AS
  neighbor 222.222.10.17 prefix-list Customer out
!
ip prefix-list Customer permit 221.10.0.0/19
```

- **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

Link one

C

A

AS 100

AS 65534

E

D

B

Link two

- **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

- ## Router A Configuration

```
router bgp 65534
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.0.0 mask 255.255.240.0
 neighbor 222.222.10.2 remote-as 100
 neighbor 222.222.10.2 prefix-list routerC out
 neighbor 222.222.10.2 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
ip prefix-list routerC permit 221.10.0.0/20
ip prefix-list routerC permit 221.10.0.0/19
!
ip route 221.10.0.0 255.255.240.0 null0
ip route 221.10.0.0 255.255.224.0 null0
```

Router B configuration is similar but with the other /20

Loadsharing to the same ISP

- **Router C Configuration**

```
router bgp 100

  neighbor 222.222.10.1 remote-as 65534

  neighbor 222.222.10.1 default-originate

  neighbor 222.222.10.1 prefix-list Customer in

  neighbor 222.222.10.1 prefix-list default out

!

ip prefix-list Customer permit 221.10.0.0/19 le 20

ip prefix-list default permit 0.0.0.0/0
```

- **Router C only allows in /19 and /20 prefixes from customer block**

- **Router D configuration is identical**

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**

Two links to the same ISP

Multiple Dualhomed Customers

(RFC2270)

Multiple Dualhomed Customers (RFC2270)

AS 100

C

A1

AS 65534

B1

E

D

A2

AS 65534

B2

A3

AS 65534

B3

- **AS100 removes private AS and any customer subprefixes from Internet announcement**

Multiple Dualhomed Customers

- **Customer announcements as per previous example**

- **Use the *same* private AS for each customer**

    documented in RFC2270

    address space is not overlapping

    each customer hears default only

- **Router A*n* and B*n* configuration same as Router A and B previously**

Two links to the same ISP

- ## Router A1 Configuration

```
router bgp 65534
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.0.0 mask 255.255.240.0
 neighbor 222.222.10.2 remote-as 100
 neighbor 222.222.10.2 prefix-list routerC out
 neighbor 222.222.10.2 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
ip prefix-list routerC permit 221.10.0.0/20
ip prefix-list routerC permit 221.10.0.0/19
!
ip route 221.10.0.0 255.255.240.0 null0
ip route 221.10.0.0 255.255.224.0 null0
```

Router B1 configuration is similar but for the other /20

Multiple Dualhomed Customers

- ## Router C Configuration

```
router bgp 100

  neighbor bgp-customers peer-group

  neighbor bgp-customers remote-as 65534

  neighbor bgp-customers default-originate

  neighbor bgp-customers prefix-list default out

  neighbor 222.222.10.1 peer-group bgp-customers

  neighbor 222.222.10.1 description Customer One

  neighbor 222.222.10.1 prefix-list Customer1 in

  neighbor 222.222.10.9 peer-group bgp-customers

  neighbor 222.222.10.9 description Customer Two

  neighbor 222.222.10.9 prefix-list Customer2 in
```

Multiple Dualhomed Customers

```
neighbor 222.222.10.17 peer-group bgp-customers

neighbor 222.222.10.17 description Customer Three

neighbor 222.222.10.17 prefix-list Customer3 in
!
ip prefix-list Customer1 permit 221.10.0.0/19 le 20

ip prefix-list Customer2 permit 221.16.64.0/19 le 20

ip prefix-list Customer3 permit 221.14.192.0/19 le 20

ip prefix-list default permit 0.0.0.0/0
```

- **Router C only allows in /19 and /20 prefixes from customer block**

- **Router D configuration is almost identical**

Multiple Dualhomed Customers

- **Router E Configuration**

  assumes customer address space is not part of upstream's address block

```
router bgp 100
 neighbor 222.222.10.17 remote-as 110
 neighbor 222.222.10.17 remove-private-AS
 neighbor 222.222.10.17 prefix-list Customers out
!
ip prefix-list Customers permit 221.10.0.0/19
ip prefix-list Customers permit 221.16.64.0/19
ip prefix-list Customers permit 221.14.192.0/19
```

- **Private AS still visible inside AS100**

Multiple Dualhomed Customers

- ## If customers' prefixes come from ISP's address block

    do **NOThttp** announce them to the Internet

    announce ISP aggregate only

- ## Router E configuration:

```
router bgp 100
 neighbor 222.222.10.17 remote-as 110
 neighbor 222.222.10.17 prefix-list my-aggregate out
!
ip prefix-list my-aggregate permit 221.8.0.0/13
```

BGP Multihoming Techniques

- **Definition & Options**

- **Connecting to the same ISP**

- **Connecting to different ISPs**

- **Service Provider Multihoming**

- **Using Communities**

- **Case Study**

Multihoming to different ISPs

Two links to different ISPs

- **Use a Public AS**

  or use private AS if agreed with the other ISP

- **Address space comes from**

  both upstreams or

  Regional Internet Registry

- **Configuration concepts very similar**

Two links to different ISPs

One link primary, the other link 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)

Internet

AS 100

AS 120

AS 130

Announce /19 block

Announce /19 block with longer AS PATH

C

D

A

B

Two links to different ISPs (one as backup only)

- **Router A Configuration**

```
router bgp 130
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 100
 neighbor 222.222.10.1 prefix-list aggregate out
 neighbor 222.222.10.1 prefix-list default in
!
ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```

Two links to different ISPs (one as backup only)

- **Router B Configuration**

```
router bgp 130
 network 221.10.0.0 mask 255.255.224.0
 neighbor 220.1.5.1 remote-as 120
 neighbor 220.1.5.1 prefix-list aggregate out
 neighbor 220.1.5.1 route-map routerD-out out
 neighbor 220.1.5.1 prefix-list default in
 neighbor 220.1.5.1 route-map routerD-in in
!
ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
route-map routerD-out permit 10
 set as-path prepend 130 130 130
!
route-map routerD-in permit 10
 set local-preference 80
```

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**

    <u>basic</u> 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)

Internet

AS 100

AS 120

C

D

**Announce first
/20 and /19 block**

**Announce second
/20 and /19 block**

A

B

AS 130

Two links to different ISPs (with loadsharing)

- ## Router A Configuration

```
router bgp 130
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.0.0 mask 255.255.240.0
 neighbor 222.222.10.1 remote-as 100
 neighbor 222.222.10.1 prefix-list firstblock out
 neighbor 222.222.10.1 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
!
ip prefix-list firstblock permit 221.10.0.0/20
ip prefix-list firstblock permit 221.10.0.0/19
```

Two links to different ISPs (with loadsharing)

- **Router B Configuration**

```
router bgp 130
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.16.0 mask 255.255.240.0
 neighbor 220.1.5.1 remote-as 120
 neighbor 220.1.5.1 prefix-list secondblock out
 neighbor 220.1.5.1 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
!
ip prefix-list secondblock permit 221.10.16.0/20
ip prefix-list secondblock permit 221.10.0.0/19
```

Cisco.com

Two links to different ISPs

More Controlled Loadsharing

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

Internet

AS 100

AS 120

C

D

Announce /19 block

Announce /20 subprefix, and
/19 block with longer AS path

A

B

AS 130

Loadsharing with different ISPs

- ## Router A Configuration

```
router bgp 130
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 100
 neighbor 222.222.10.1 prefix-list default in
 neighbor 222.222.10.1 prefix-list aggregate out
!
ip prefix-list aggregate permit 221.10.0.0/19
```

Loadsharing with different ISPs

- **Router B Configuration**

```
router bgp 130
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.16.0 mask 255.255.240.0
 neighbor 220.1.5.1 remote-as 120
 neighbor 220.1.5.1 prefix-list default in
 neighbor 220.1.5.1 prefix-list subblocks out
 neighbor 220.1.5.1 route-map routerD out
!
route-map routerD permit 10
 match ip address prefix-list aggregate
 set as-path prepend 130 130
route-map routerD permit 20
!
ip prefix-list subblocks permit 221.10.0.0/19 le 20
ip prefix-list aggregate permit 221.10.0.0/19
```
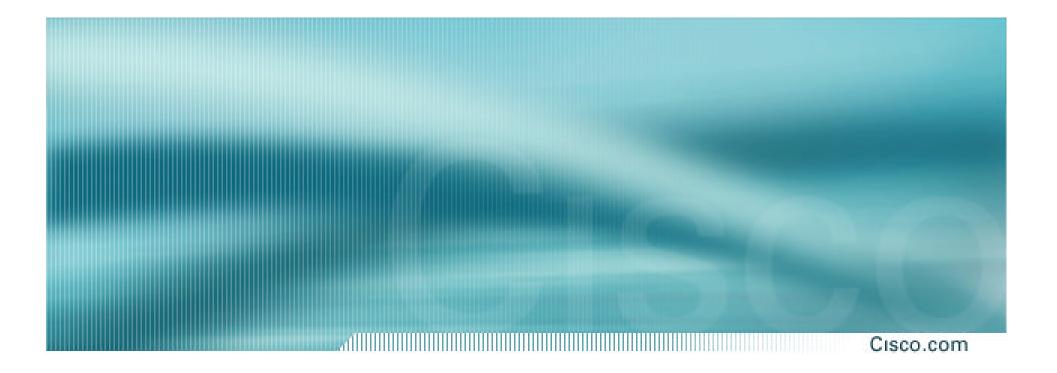
BGP Multihoming Techniques

- **Definition & Options**

- **Connecting to the same ISP**

- **Connecting to different ISPs**

- **Service Provider Multihoming**

- **Using Communities**

- **Case Study**

Service Provider Multihoming

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

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

- **Examples**

    One upstream, one local peer

    One upstream, local exchange point

    Two upstreams, one local peer

    Tier-1 and regional upstreams, with local peers

    Disconnected Backbone

    IDC Multihoming

- **All examples require BGP and a public ASN**

Service Provider Multihoming

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

Upstream ISP

AS130

Local Peer

AS120

C

A

AS 110

One Upstream, One Local Peer

- ## Router A Configuration

```
router bgp 110

 network 221.10.0.0 mask 255.255.224.0

 neighbor 222.222.10.2 remote-as 120

 neighbor 222.222.10.2 prefix-list my-block out

 neighbor 222.222.10.2 prefix-list AS120-peer in

!

ip prefix-list AS120peer permit 222.5.16.0/19

ip prefix-list AS120-peer permit 221.240.0.0/20

ip prefix-list my-block permit 221.10.0.0/19

!

ip route 221.10.0.0 255.255.224.0 null0
```

One Upstream, One Local Peer

- **Router A – Alternative Configuration**

```
router bgp 110

 network 221.10.0.0 mask 255.255.224.0

 neighbor 222.222.10.2 remote-as 120

 neighbor 222.222.10.2 prefix-list my-block out

 neighbor 222.222.10.2 filter-list 10 in

!

ip as-path access-list 10 permit ^(120_)+$

!

ip prefix-list my-block permit 221.10.0.0/19

!

ip route 221.10.0.0 255.255.224.0 null0
```

One Upstream, One Local Peer

- **Router C Configuration**

```
router bgp 110

  network 221.10.0.0 mask 255.255.224.0

  neighbor 222.222.10.1 remote-as 130

  neighbor 222.222.10.1 prefix-list default in

  neighbor 222.222.10.1 prefix-list my-block out
!
ip prefix-list my-block permit 221.10.0.0/19

ip prefix-list default permit 0.0.0.0/0
!
ip route 221.10.0.0 255.255.224.0 null0
```

One Upstream, One Local Peer

- **Two configurations possible for Router A**

  Filter-lists assume peer knows what they are doing

  Prefix-list higher maintenance, but safer

  Some ISPs use both

- **Local traffic goes to and from local peer, everything else goes to upstream**

Service Provider Multihoming

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 from IXP peers**

One Upstream, Local Exchange Point

Upstream ISP

AS130

IXP

C

A

AS 110

One Upstream, Local Exchange Point

- ## Router A Configuration

```
interface fastethernet 0/0
 description Exchange Point LAN
 ip address 220.5.10.1 mask 255.255.255.224
 ip verify unicast reverse-path
 no ip directed-broadcast
 no ip proxy-arp
 no ip redirects
!
router bgp 110
 network 221.10.0.0 mask 255.255.224.0
 neighbor ixp-peers peer-group
 neighbor ixp-peers soft-reconfiguration in
 neighbor ixp-peers prefix-list my-block out
..next slide
```

One Upstream, Local Exchange Point

```
neighbor 220.5.10.2 remote-as 100

neighbor 222.5.10.2 peer-group ixp-peers

neighbor 222.5.10.2 prefix-list peer100 in

neighbor 220.5.10.3 remote-as 101

neighbor 222.5.10.3 peer-group ixp-peers

neighbor 222.5.10.3 prefix-list peer101 in

neighbor 220.5.10.4 remote-as 102

neighbor 222.5.10.4 peer-group ixp-peers

neighbor 222.5.10.4 prefix-list peer102 in

neighbor 220.5.10.5 remote-as 103

neighbor 222.5.10.5 peer-group ixp-peers

neighbor 222.5.10.5 prefix-list peer103 in
..next slide
```

One Upstream, Local Exchange Point

```
ip route 221.10.0.0 255.255.224.0 null0
!
ip prefix-list my-block permit 221.10.0.0/19
ip prefix-list peer100 permit 222.0.0.0/19
ip prefix-list peer101 permit 222.30.0.0/19
ip prefix-list peer102 permit 222.12.0.0/19
ip prefix-list peer103 permit 222.18.128.0/19
!
```

One Upstream, Local Exchange Point

- **Router C Configuration**

```
router bgp 110
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.1 remote-as 130
  neighbor 222.222.10.1 prefix-list default in
  neighbor 222.222.10.1 prefix-list my-block out
!
ip prefix-list my-block permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
ip route 221.10.0.0 255.255.224.0 null0
```

One Upstream, Local Exchange Point

- **Note Router A configuration**

    Prefix-list higher maintenance, but safer

    uRPF on the FastEthernet interface

- **IXP traffic goes to and from local IXP, everything else goes to upstream**

Service Provider Multihoming

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

Upstream ISP

AS130

Upstream ISP

AS140

Local Peer

AS120

AS 110

Two Upstreams, One Local Peer

- **Router A**

    Same routing configuration as in example with one upstream and one local peer

    Same hardware configuration

Two Upstreams, One Local Peer

- **Router C Configuration**

```
router bgp 110
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.1 remote-as 130
  neighbor 222.222.10.1 prefix-list default in
  neighbor 222.222.10.1 prefix-list my-block out
!
ip prefix-list my-block permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
ip route 221.10.0.0 255.255.224.0 null0
```

Two Upstreams, One Local Peer

- **Router D Configuration**

```
router bgp 110
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.5 remote-as 140
  neighbor 222.222.10.5 prefix-list default in
  neighbor 222.222.10.5 prefix-list my-block out
!
ip prefix-list my-block permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
ip route 221.10.0.0 255.255.224.0 null0
```
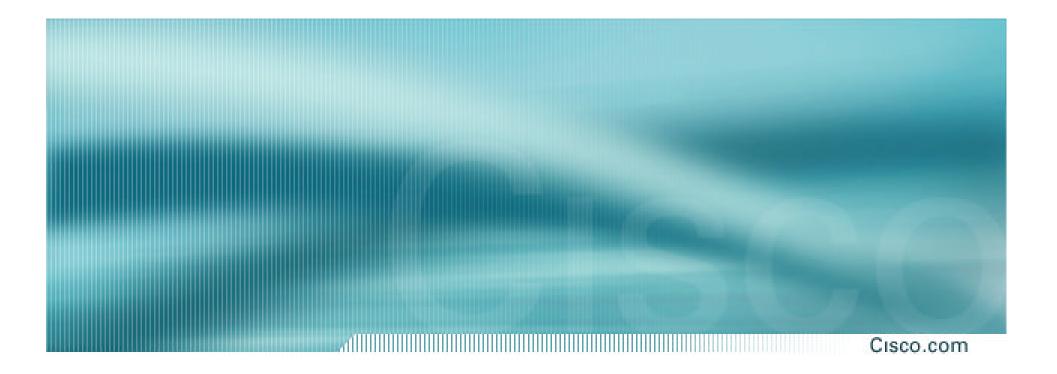
Two Upstreams, One Local Peer

- ## This is the simple configuration for Router C and D

- ## Traffic out to the two upstreams will take nearest exit

    ### Inexpensive routers required

    ### This is not useful in practice especially for international links

    ### Loadsharing needs to be better

Two Upstreams, One Local Peer

- **Better configuration options:**

   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**

```
router bgp 110
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.1 remote-as 130
  neighbor 222.222.10.1 prefix-list rfc1918-deny in
  neighbor 222.222.10.1 prefix-list my-block out
  neighbor 222.222.10.1 route-map AS130-loadshare in
!
ip prefix-list my-block permit 221.10.0.0/19
! See earlier presentation for RFC1918 list
..next slide
```

Two Upstreams, One Local Peer – Full Routes

```
ip route 221.10.0.0 255.255.224.0 null0

!

ip as-path access-list 10 permit ^(130_)+$

ip as-path access-list 10 permit ^(130_)+_[0-9]+$

!

route-map AS130-loadshare permit 10

 match ip as-path 10

 set local-preference 120

route-map AS130-loadshare permit 20

 set local-preference 80

!
```

Two Upstreams, One Local Peer – Full Routes

- **Router D Configuration**

```
router bgp 110
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.5 remote-as 140
 neighbor 222.222.10.5 prefix-list rfc1918-deny in
 neighbor 222.222.10.5 prefix-list my-block out
!
ip prefix-list my-block permit 221.10.0.0/19
! See earlier in presentation for RFC1918 list
```

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 the route-map**

Two Upstreams, One Local Peer – Full Routes

- **Full routes from upstreams**

    Expensive – needs 128Mbytes RAM today

    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

- ## Router C Configuration

```
router bgp 110

  network 221.10.0.0 mask 255.255.224.0

  neighbor 222.222.10.1 remote-as 130

  neighbor 222.222.10.1 prefix-list rfc1918-nodef-deny in

  neighbor 222.222.10.1 prefix-list my-block out

  neighbor 222.222.10.1 filter-list 10 in

  neighbor 222.222.10.1 route-map tag-default-low in

!

..next slide
```

Two Upstreams, One Local Peer – Partial Routes

```
ip prefix-list my-block permit 221.10.0.0/19

ip prefix-list default permit 0.0.0.0/0

!

ip route 221.10.0.0 255.255.224.0 null0

!

ip as-path access-list 10 permit ^(130_)+$

ip as-path access-list 10 permit ^(130_)+_[0-9]+$

!

route-map tag-default-low permit 10

 match ip address prefix-list default

 set local-preference 80

route-map tag-default-low permit 20

!
```

Two Upstreams, One Local Peer – Partial Routes

- ## Router D Configuration

```
router bgp 110

  network 221.10.0.0 mask 255.255.224.0

  neighbor 222.222.10.5 remote-as 140

  neighbor 222.222.10.5 prefix-list default in

  neighbor 222.222.10.5 prefix-list my-block out

!

ip prefix-list my-block permit 221.10.0.0/19

ip prefix-list default permit 0.0.0.0/0

!

ip route 221.10.0.0 255.255.224.0 null0
```

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 AS106 link fails, backup via AS130 – and vice-versa

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

Two Upstreams, One Local Peer – Partial Routes

- **Router C Configuration**

```
router ospf 110
 default-information originate metric 30
 passive-interface Serial 0/0
!
router bgp 110
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 130
 neighbor 222.222.10.1 prefix-list rfc1918-deny in
 neighbor 222.222.10.1 prefix-list my-block out
 neighbor 222.222.10.1 filter-list 10 in
!
..next slide
```

Two Upstreams, One Local Peer – Partial Routes

```
ip prefix-list my-block permit 221.10.0.0/19

! See earlier for RFC1918 list

!

ip route 221.10.0.0 255.255.224.0 null0

ip route 0.0.0.0 0.0.0.0 serial 0/0 254

!

ip as-path access-list 10 permit ^(130_)+$

ip as-path access-list 10 permit ^(130_)+_[0-9]+$

!
```

Two Upstreams, One Local Peer – Partial Routes

- ## Router D Configuration

```
router ospf 110
  default-information originate metric 10
  passive-interface Serial 0/0
!
router bgp 110
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.5 remote-as 140
  neighbor 222.222.10.5 prefix-list deny-all in
  neighbor 222.222.10.5 prefix-list my-block out
!
..next slide
```

Two Upstreams, One Local Peer – Partial Routes

```
ip prefix-list deny-all deny 0.0.0.0/0 le 32

ip prefix-list my-block permit 221.10.0.0/19

! See earlier in presentation for RFC1918 list

!

ip route 221.10.0.0 255.255.224.0 null0

ip route 0.0.0.0 0.0.0.0 serial 0/0 254

!
```

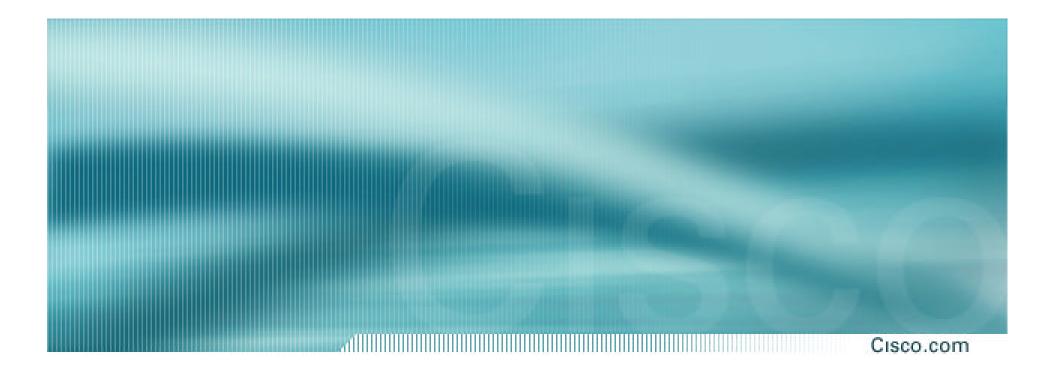# Two Upstreams, One Local Peer – Partial Routes

- **Partial routes from upstreams**

  Use OSPF to determine outbound path

  Router D default has metric 10 – primary outbound path

  Router C default has metric 30 – backup outbound path

  Serial interface goes down, static default is removed from routing table, OSPF default withdrawn

Service Provider Multihoming

Two Tier-1 upstreams, two regional upstreams, and local peers

Tier-1 & Regional Upstreams, Local Peers

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

- **Accept partial/default routes from upstreams**

  For default, use 0.0.0.0/0 or a network which can be used as default
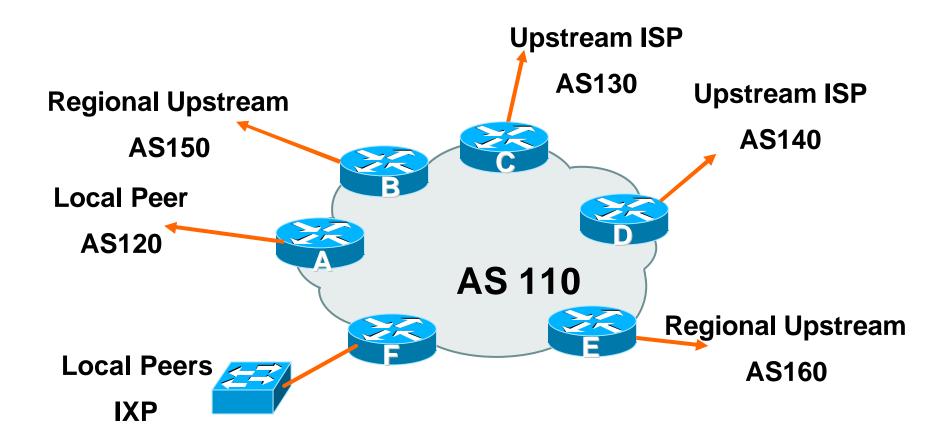
- **Accept all routes from local peer**

- **Accept all partial routes from regional upstreams**

- **This is more complex, but a very typical scenario**

Tier-1 & Regional Upstreams, Local Peers

Upstream ISP

AS130

Upstream ISP

AS140

Regional Upstream

AS150

Local Peer

AS120

AS 110

Regional Upstream

AS160

Local Peers

IXP

Tier-1 & Regional Upstreams, Local Peers – Detail

- **Router A – local private peer**

  Accept all (local) routes

  Local traffic stays local

  Use prefix and/or AS-path filters

  Use local preference (if needed)

- **Router F – local IXP peering**

  Accept all (local) routes

  Local traffic stays local

  Use prefix and/or AS-path filters

Tier-1 & Regional Upstreams, Local Peers – Detail

- **Router B – regional upstream**

  They provide transit to Internet, but longer AS path than Tier-1s

  Accept all regional routes from them

  e.g. ^150_[0-9]+$

  Ask them to send default, or send a network you can use as default

  Set local pref on "default" to 60

  Will provide backup to Internet only when direct Tier-1 links go down

Tier-1 & Regional Upstreams, Local Peers – Detail

- **Router E – regional upstream**

  They provide transit to Internet, but longer AS path than Tier-1s

  Accept all regional routes from them

  e.g. ^160_[0-9]+$

  Ask them to send default, or send a network you can use as default

  Set local pref on "default" to 70

  Will provide backup to Internet only when direct Tier-1 links go down

Tier-1 & Regional Upstreams, Local Peers – Detail

- **Router C – first Tier-1**

  Accept all their customer and AS neighbour routes from them

  e.g. ^130_[0-9]+$

  Ask them to send default, or send a network you can use as default

  Set local pref on "default" to 80

  Will provide backup to Internet only when link to second Tier-1 goes down

Tier-1 & Regional Upstreams, Local Peers – Detail

- **Router D – second Tier-1**

  Ask them to send default, or send a network you can use as default

  This has local preference 100 by default

  All traffic without any more specific path will go out this way

Tier-1 & Regional Upstreams, Local Peers – Summary

- **Local traffic goes to local peer and IXP**

- **Regional traffic goes to two regional upstreams**

- **Everything else is shared between the two Tier-1s**

- **To modify loadsharing tweak what is heard from the two regionals and the first Tier-1**

    Best way is through modifying the AS-path filter

Tier-1 & Regional Upstreams, Local Peers

- ## What about outbound announcement strategy?

  This is to determine incoming traffic flows

  /19 aggregate must be announced to everyone!

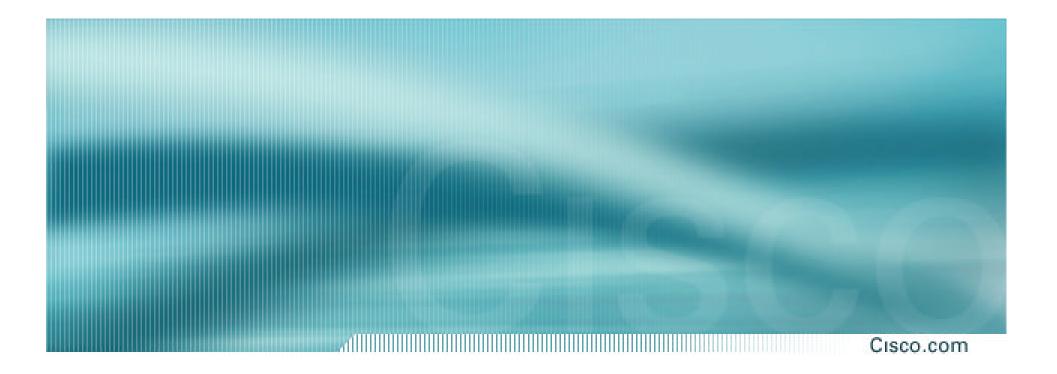 /20 or /21 more specifics can be used to improve or modify loadsharing

  See earlier for hints and ideas

Tier-1 & Regional Upstreams, Local Peers

- ## What about unequal circuit capacity?

    ### AS-path filters are very useful

- ## What if upstream will only give me full routing table or nothing

    ### AS-path and prefix filters are very useful

Service Provider Multihoming

Disconnected Backbone

Disconnected Backbone
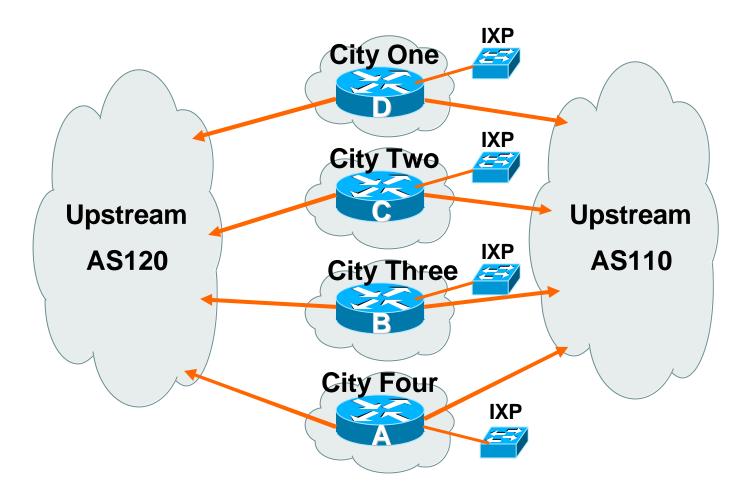
Cisco.com

- ## ISP runs large network

  Network has no backbone, only large PoPs in each location

  Each PoP multihomes to upstreams

  Common in some countries where backbone circuits are hard to obtain

Disconnected Backbone

City One

IXP

D

City Two

IXP

C

Upstream

AS120

Upstream

AS110

City Three

IXP

B

City Four

IXP

A

Disconnected Backbone

- ## Works with one AS number

    ### Not four – no BGP loop detection problem

- ## Each city operates as separate network

    ### Uses defaults and selected leaked prefixes for loadsharing

    ### Peers at local exchange point

Disconnected Backbone

- **Router A Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.248.0
 neighbor 222.200.0.1 remote-as 120
 neighbor 222.200.0.1 description AS120 – Serial 0/0
 neighbor 222.200.0.1 prefix-list default in
 neighbor 222.222.0.1 prefix-list my-block out
 neighbor 222.222.10.1 remote-as 110
 neighbor 222.222.10.1 description AS110 – Serial 1/0
 neighbor 222.222.10.1 prefix-list rfc1918-sua in
 neighbor 222.222.10.1 prefix-list my-block out
 neighbor 222.222.10.1 filter-list 10 in
!
```

…continued on next page…

Disconnected Backbone

```
ip prefix-list my-block permit 221.10.0.0/21

ip prefix-list default permit 0.0.0.0/0

!

ip as-path access-list 10 permit ^(110_)+$

ip as-path access-list 10 permit ^(110_)+_[0-9]+$

!…etc to achieve outbound loadsharing

!

ip route 0.0.0.0 0.0.0.0 Serial 1/0 250

ip route 221.10.0.0 255.255.248.0 null0

!
```

Disconnected Backbone

- **Peer with AS120**

  Receive just default route

  Announce /22 address

- **Peer with AS110**

  Receive full routing table – filter with AS-path filter

  Announce /22 address

  Point backup static default – distance 252 – in case AS120 goes down
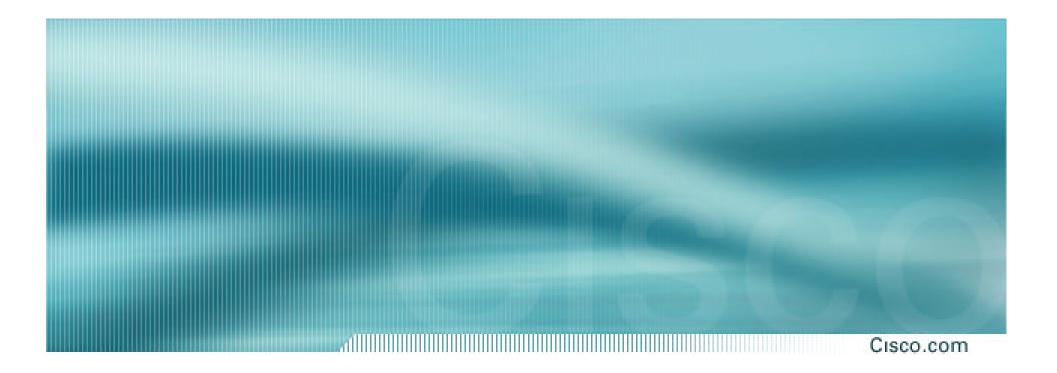
Disconnected Backbone

- **Default ensures that disconnected parts of AS100 are reachable**

  Static route backs up AS120 default

  No BGP loop detection – relying on default route

- **Do not announce /19 aggregate**

  No advantage in announcing /19 and could lead to problems

IDC Multihoming

IDC Multihoming

- **IDCs typically are not registry members so don't get their own address block**

  Situation also true for small ISPs and "Enterprise Networks"
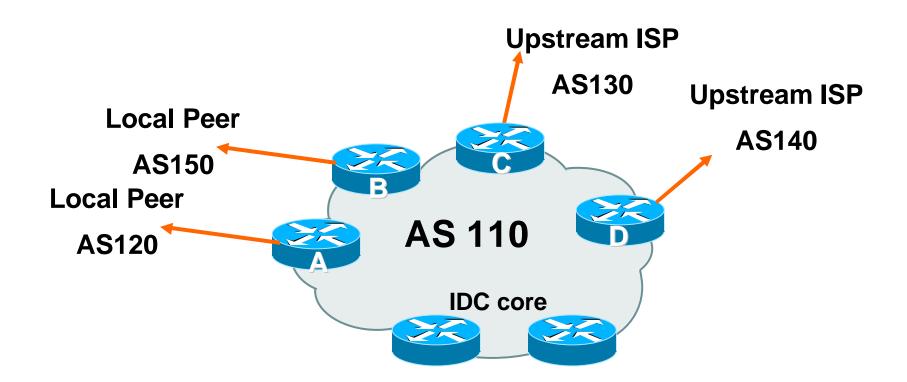
- **Smaller address blocks being announced**

  Address space comes from both upstreams

  Should be apportioned according to size of circuit to upstream

- **Outbound traffic paths matter**

Two Upstreams, Two Local Peers – IDC

Upstream ISP

AS130

Upstream ISP

AS140

Local Peer

AS150

Local Peer

AS120

C

B

D

A

AS 110

IDC core

Assigned /24 from AS130 and /23 from AS140.

Circuit to AS130 is 2Mbps, circuit to AS140 is 4Mbps

IDC Multihoming

- ## Router A and B configuration

    In: Should accept all routes from AS120 and AS150

    Out: Should announce all address space to AS120 and AS150

    Straightforward

IDC Multihoming

- ## Router C configuration

    In: Accept partial routes from AS130

    e.g. ^130_[0-9]+$

    In: Ask for a route to use as default

    set local preference on default to 80

    Out: Send /24, and send /23 with AS-PATH prepend of one AS

IDC Multihoming

- **Router D configuration**

  In: Ask for a route to use as default

  Leave local preference of default at 100

  Out: Send /23, and send /24 with AS-PATH prepend of one AS

IDC Multihoming – Fine Tuning

- ## For local fine tuning, increase circuit capacity

  Local circuits usually are cheap

  Otherwise…

- ## For longer distance fine tuning

  In: Modify as-path filter on Router C

  Out: Modify as-path prepend on Routers C and D

  Outbound traffic flow is usual critical for an IDC so **inbound** policies need to be carefully thought out

IDC Multihoming – Other Details

- **Redundancy**

  Circuits are terminated on separate routers

- **Apply thought to address space use**

  Request from both upstreams

  Utilise address space evenly across IDC

  Don't start with /23 then move to /24 – use both blocks at the same time in the same proportion

  Helps with loadsharing – yes, really!

IDC Multihoming – Other Details

- ## What about failover?

  ### /24 and /23 from upstreams' blocks announced to the Internet routing table all the time
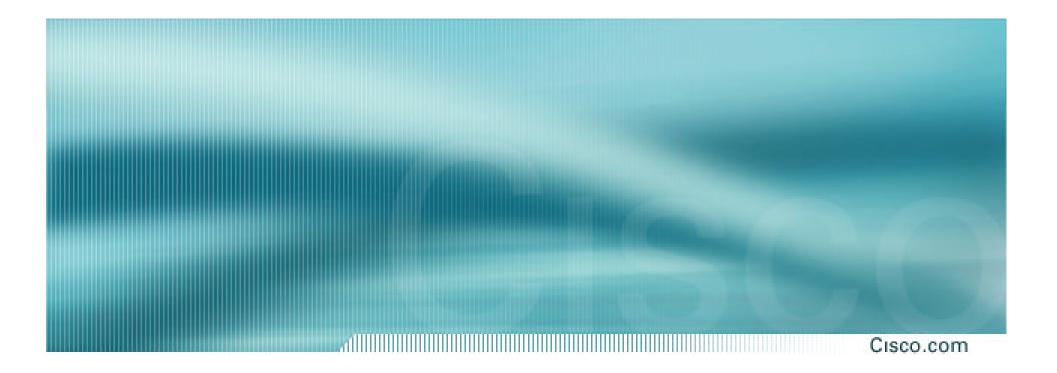
  ### No obvious alternative at the moment

  Conditional advertisement can help in steady state, but subprefixes still need to be announced in failover condition

BGP Multihoming Techniques

- **Definition & Options**

- **Connecting to the same ISP**

- **Connecting to different ISPs**

- **Service Provider Multihoming**

- **Using Communities**

- **Case Study**

Communities

Community usage

- **RFC1998**

- **Examples of SP applications**

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

- ## Sample Customer Router Configuration

```
router bgp 130

 neighbor x.x.x.x remote-as 100

 neighbor x.x.x.x description Backup ISP

 neighbor x.x.x.x route-map config-community out

 neighbor x.x.x.x send-community

!

ip as-path access-list 20 permit ^$

ip as-path access-list 20 deny .*

!

route-map config-community permit 10

 match as-path 20

 set community 100:90
```

RFC1998

- ## Sample ISP Router Configuration

```
! Homed to another ISP

ip community-list 70 permit 100:70

! Homed to another ISP with equal ASPATH length

ip community-list 80 permit 100:80

! Customer backup routes

ip community-list 90 permit 100:90

!

route-map set-customer-local-pref permit 10

 match community 70

 set local-preference 70
```

RFC1998

- ## Sample ISP Router Configuration

```
route-map set-customer-local-pref permit 20
 match community 80
 set local-preference 80
!
route-map set-customer-local-pref permit 30
 match community 90
 set local-preference 90
!
route-map set-customer-local-pref permit 40
 set local-preference 100
```
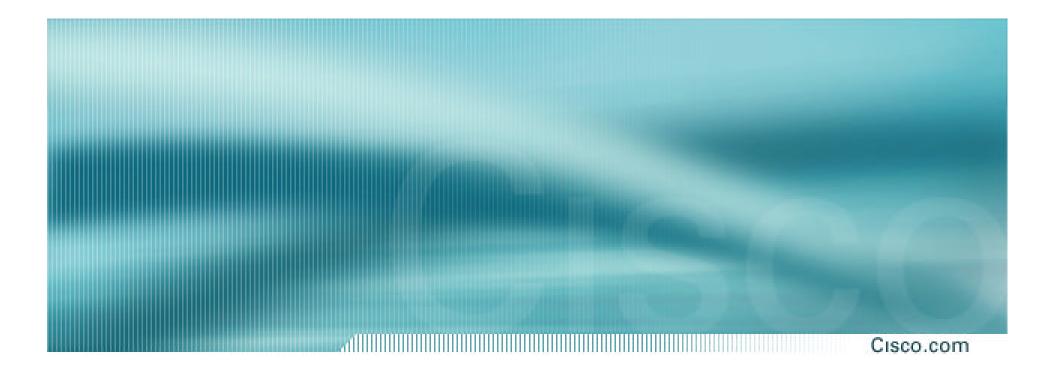
RFC1998

- **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

Two links to the same ISP

One link primary, the other link backup only

Two links to the same ISP

primary

AS 100

AS 65534

backup

- **AS100 proxy aggregates for AS 65534**

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

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

    primary link makes standard announcement

    backup link sends community

- **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 A Configuration**

```
router bgp 65534
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.2 remote-as 100
 neighbor 222.222.10.2 description RouterC
 neighbor 222.222.10.2 prefix-list aggregate out
 neighbor 222.222.10.2 prefix-list default in
!
ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
```

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

- ## Router B Configuration

```
router bgp 65534
  network 221.10.0.0 mask 255.255.224.0
  neighbor 222.222.10.6 remote-as 100
  neighbor 222.222.10.6 description RouterD
  neighbor 222.222.10.6 send-community
  neighbor 222.222.10.6 prefix-list aggregate out
  neighbor 222.222.10.6 route-map routerD-out out
  neighbor 222.222.10.6 prefix-list default in
  neighbor 222.222.10.6 route-map routerD-in in
!
..next slide
```

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

```
ip prefix-list aggregate permit 221.10.0.0/19

ip prefix-list default permit 0.0.0.0/0

!

route-map routerD-out permit 10
 match ip address prefix-list aggregate
 set community 100:90

route-map routerD-out permit 20

!

route-map routerD-in permit 10
 set local-preference 90

!
```

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

- **Router C Configuration (main link)**

```
router bgp 100
  neighbor 222.222.10.1 remote-as 65534
  neighbor 222.222.10.1 default-originate
  neighbor 222.222.10.1 prefix-list Customer in
  neighbor 222.222.10.1 prefix-list default out
!
ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```

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

- **Router D Configuration (backup link)**

```
router bgp 100
  neighbor 222.222.10.5 remote-as 65534
  neighbor 222.222.10.5 default-originate
  neighbor 222.222.10.5 prefix-list Customer in
  neighbor 222.222.10.5 route-map bgp-cust-in in
  neighbor 222.222.10.5 prefix-list default out
!
ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
..next slide
```
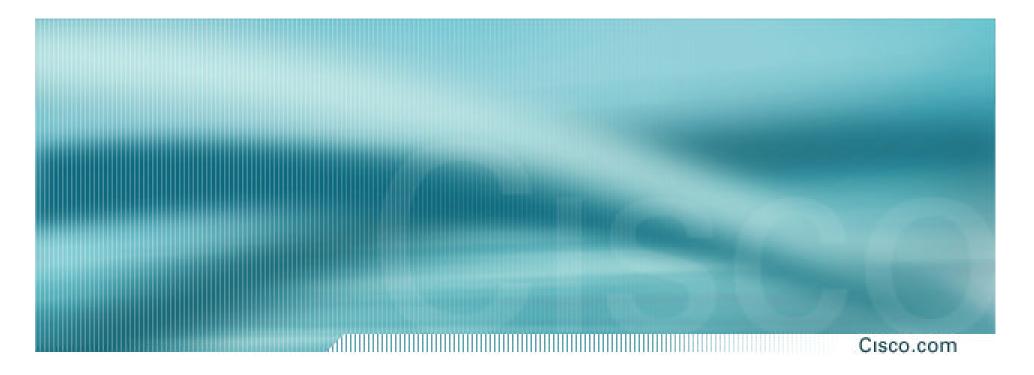
Two links to the same ISP (one as backup only)

```
ip prefix-list Customer permit 221.10.0.0/19

ip prefix-list default permit 0.0.0.0/0

!

ip community-list 90 permit 100:90

!

<snip>

route-map bgp-cust-in permit 30
 match community 90
  set local-preference 90
route-map bgp-cust-in permit 40
  set local-preference 100
```

Service Provider use of Communities

Some working examples

Background

- **RFC1998 is okay for "simple" multihomed customers**

    assumes that upstreams are interconnected

- **ISPs create many other communities to handle more complex situations**

    Simplify ISP BGP configuration

    Give customer more policy control

Some ISP Examples

- ## Public policy is usually listed in the IRR

  ### Following examples are all in the IRR or referenced from the AS Object in the IRR

- ## Consider creating communities to give policy control to customers

  ### Reduces technical support burden

  ### Reduces the amount of router reconfiguration, and the chance of mistakes

Some ISP Examples
Connect.com.au

```
aut-num:      AS2764
as-name:      ASN-CONNECT-NET
descr:        connect.com.au pty ltd
admin-c:      CC89
tech-c:       MP151
remarks:      Community  Definition
remarks:      ----------------------------------------------
remarks:      2764:1 Announce to "domestic" rate ASes only
remarks:      2764:2 Don't announce outside local POP
remarks:      2764:3 Lower local preference by 25
remarks:      2764:4 Lower local preference by 15
remarks:      2764:5 Lower local preference by 5
remarks:      2764:6 Announce to non customers with "no-export"
remarks:      2764:7 Only announce route to customers
remarks:      2764:8 Announce route over satellite link
notify:       routing@connect.com.au
mnt-by:       CONNECT-AU
changed:      mrp@connect.com.au 19990506
source:       CCAIR
```

```
aut-num: AS702
as-name: AS702
descr:   UUNET - Commercial IP service provider in Europe
remarks: ------------------------------------------------------
remarks: UUNET uses the following communities with its customers:
remarks: 702:80  Set Local Pref 80 within AS702
remarks: 702:120 Set Local Pref 120 within AS702
remarks: 702:20  Announce only to UUNET AS'es and UUNET customers
remarks: 702:30  Keep within Europe, don't announce to other UUNET AS's
remarks: 702:1   Prepend AS702 once at edges of UUNET to Peers
remarks: 702:2   Prepend AS702 twice at edges of UUNET to Peers
remarks: 702:3   Prepend AS702 thrice at edges of UUNET to Peers
remarks: Details of UUNET's peering policy and how to get in touch with
remarks: UUNET regarding peering policy matters can be found at:
remarks: http://www.uu.net/peering/
remarks: ------------------------------------------------------
mnt-by:  UUNET-MNT
changed: eric-apps@eu.uu.net 20010928
source:  RIPE
```

Some ISP
BT Ig

```
aut-num:        AS5400
as-name:        CIPCORE
descr:          BT Ignite European Backbone
remarks:        The following BGP communities can be set by BT Ignite
remarks:        BGP customers to affect announcements to major peers.
remarks:
remarks:        Community to                           Community to
remarks:        Not announce        To peer:           AS prepend 5400
remarks:
remarks:        5400:1000 European peers               5400:2000
remarks:        5400:1001 Sprint (AS1239)              5400:2001
remarks:        5400:1003 Unisource (AS3300)           5400:2003
remarks:        5400:1005 UUnet (AS702)                5400:2005
remarks:        5400:1006 Carrier1 (AS8918)            5400:2006
remarks:        5400:1007 SupportNet (8582)            5400:2007
remarks:        5400:1008 AT&T (AS2686)                5400:2008
remarks:        5400:1009 Level 3 (AS9057)             5400:2009
remarks:        5400:1010 RIPE (AS3333)                5400:2010
<snip>
remarks:        5400:1100         US peers             5400:2100
notify:         notify@eu.ignite.net
mnt-by:         CIP-MNT
source:         RIPE
```

And many many more!

Some ISP
Carrier1

```
aut-num:        AS8918
descr:          Carrier1 Autonomous System
<snip>

remarks:        Community Support Definitions:
remarks:        Communities that determine the geographic
remarks:        entry point of routes into the Carrier1 network:
remarks:        *
remarks:        Community      Entry Point
remarks:        -------------------------------
remarks:        8918:10        London
remarks:        8918:15        Hamburg
remarks:        8918:18        Chicago
remarks:        8918:20        Amsterdam
remarks:        8918:25        Milan
remarks:        8918:28        Berlin
remarks:        8918:30        Frankfurt
remarks:        8918:35        Zurich
remarks:        8918:40        Geneva
remarks:        8918:45        Stockholm
<snip>
notify:         inoc@carrier1.net
mnt-by:         CARRIER1-MNT
source:         RIPE
```

And many many more!

Some ISP Level

```
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
remarks:      65002:0   - prepend twice to all peers
remarks:      65002:XXX - prepend twice at peerings to AS XXX
remarks:      65003:0   - prepend 3x to all peers
remarks:      65003:XXX - prepend 3x    at peerings to AS XXX
remarks:      65004:0   - prepend 4x to all peers
remarks:      65004:XXX - prepend 4x    at peerings to AS XXX
<snip>
mnt-by:       LEVEL3-MNT
source:       RIPE
```

And many many more!

BGP Multihoming Techniques

- **Definition & Options**

- **Connecting to the same ISP**

- **Connecting to different ISPs**

- **Service Provider Multihoming**

- **Using Communities**

- **Case Study**

Case Study

First Visit

Case Study – Requirements (1)

- **ISP needs to multihome:**

  To AS5400 in Europe

  To AS2516 in Japan

  /19 allocated by APNIC

  AS 17660 assigned by APNIC

  1Mbps circuits to both upstreams

Case Study – Requirements (2)

- ## ISP wants:

  Symmetric routing and equal link utilisation in and out (as close as possible)
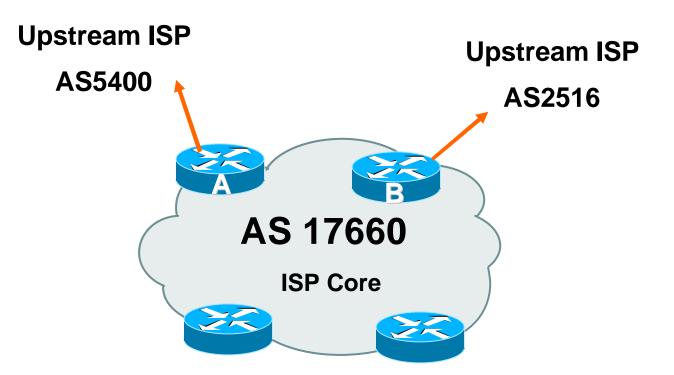
  - international circuits are expensive

  Has two Cisco 2600 border routers with 64Mbytes memory

  - Cannot afford to upgrade memory or hardware on border routers or internal routers

- ## "Philip, make it work, please"

Case Study

Upstream ISP

AS5400

Upstream ISP

AS2516

A

B

AS 17660

ISP Core

Allocated /19 from APNIC

Circuit to AS5400 is 1Mbps, circuit to AS2516 is 1Mbps

Case Study

- **Both providers stated that routers with 128Mbytes memory required for AS17660 to multihome**

    Wrong!

    Full routing table is rarely required or desired

- **Solution:**

    Accept default from one upstream

    Accept partial prefixes from the other

Case Study – Inbound Loadsharing

- **First cut: Went to a few US Looking Glasses**

  Checked the AS path to AS5400

  Checked the AS path to AS2516

  AS2516 was one hop "closer"

  Sent AS-PATH prepend of one AS on AS2516 peering

Case Study – Inbound Loadsharing

- ## Refinement

  ### Did not need any

  ### First cut worked, seeing on average 600kbps inbound on each circuit

  ### Does vary according to time of day, but this is as balanced as it can get, given customer profile

  ### ☺

Case Study – Outbound Loadsharing

- **First cut:**

  Requested default from AS2516

  Requested full routes from AS5400

- **Then looked at my Routing Report**

  Picked the top 5 ASNs and created a filter-list

  If 701, 1, 7018, 1239 or 7046 are in AS-PATH, prefixes are discarded

  Allowed prefixes originated by AS5400 and up to two AS hops away

  Resulted in 32000 prefixes being accepted in AS17660

Case Study – Outbound Loadsharing

- **Refinement**

  32000 prefixes quite a lot, seeing more outbound traffic on the AS5400 path

  Traffic was very asymmetric

  out through AS5400, in through AS2516

  Added the next 3 ASNs from the Top 20 list

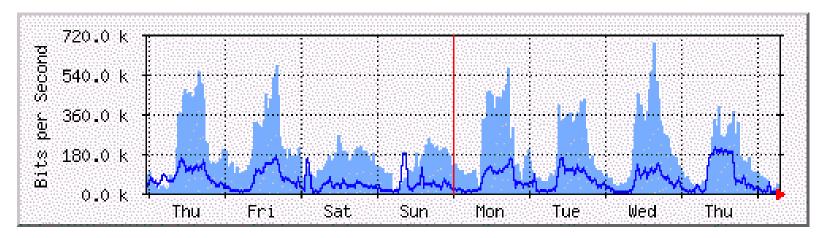 **209, 2914 and 3549**

  Now seeing 14000 prefixes

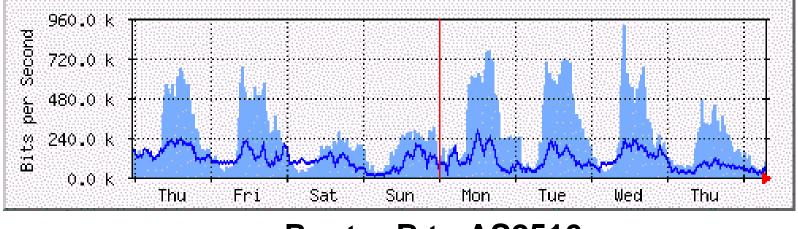 **Traffic is now evenly loadshared outbound**

  Around 200kbps on average

  Mostly symmetric

Case Study
MRTG Graphs

Router A to AS5400

Router B to AS2516

Case Study
Configuration Router A

```
router ospf 100

 log-adjacency-changes

 passive-interface default

 no passive-interface Ethernet0/0

 default-information originate metric 20

!

router bgp 17660

 no synchronization

 no bgp fast-external-fallover

 bgp log-neighbor-changes

 bgp deterministic-med

...next slide
```

Case Study
Configuration Router A

```
neighbor 166.49.165.13 remote-as 5400

neighbor 166.49.165.13 description eBGP multihop to AS5400

neighbor 166.49.165.13 ebgp-multihop 5

neighbor 166.49.165.13 update-source Loopback0

neighbor 166.49.165.13 prefix-list in-filter in

neighbor 166.49.165.13 prefix-list out-filter out

neighbor 166.49.165.13 filter-list 1 in

neighbor 166.49.165.13 filter-list 3 out

!

prefix-list in-filter  deny rfc1918etc in

prefix-list out-filter permit 202.144.128.0/19

!

ip route 0.0.0.0 0.0.0.0 serial 0/0 254

...next slide
```

Case Study
Configuration Router A

```
ip as-path access-list 1 deny _701_

ip as-path access-list 1 deny _1_

ip as-path access-list 1 deny _7018_

ip as-path access-list 1 deny _1239_

ip as-path access-list 1 deny _7046_

ip as-path access-list 1 deny _209_

ip as-path access-list 1 deny _2914_

ip as-path access-list 1 deny _3549_

ip as-path access-list 1 permit _5400$

ip as-path access-list 1 permit _5400_[0-9]+$

ip as-path access-list 1 permit _5400_[0-9]+_[0-9]+$

ip as-path access-list 1 deny .*

ip as-path access-list 3 permit ^$

!
```

Case Study
Configuration Router B

```
router ospf 100
 log-adjacency-changes
 passive-interface default
 no passive-interface Ethernet0/0
 default-information originate
!
router bgp 17660
 no synchronization
 no auto-summary
 no bgp fast-external-fallover
...next slide
```

Case Study
Configuration Router B

```
bgp log-neighbor-changes

bgp deterministic-med

 neighbor 210.132.92.165 remote-as 2516

 neighbor 210.132.92.165 description eBGP peering

 neighbor 210.132.92.165 soft-reconfiguration inbound

 neighbor 210.132.92.165 prefix-list default-route in

 neighbor 210.132.92.165 prefix-list out-filter out

 neighbor 210.132.92.165 route-map as2516-out out

 neighbor 210.132.92.165 maximum-prefix 100

 neighbor 210.132.92.165 filter-list 2 in

 neighbor 210.132.92.165 filter-list 3 out

!

...next slide
```

Case Study
Configuration Router B

```
!
prefix-list default-route permit 0.0.0.0/0
prefix-list out-filter permit 202.144.128.0/19
!
ip as-path access-list 2 permit _2516$
ip as-path access-list 2 deny .*
ip as-path access-list 3 permit ^$
!
route-map as2516-out permit 10
 set as-path prepend 17660
!
```

Configuration Summary

- ## Router A

  Hears full routing table – throws away most of it

  AS5400 BGP options are all or nothing

  Static default pointing to serial interface – if link goes down, OSPF default removed

- ## Router B

  Hears default from AS2516

  If default disappears (BGP goes down or link goes down), OSPF default is removed
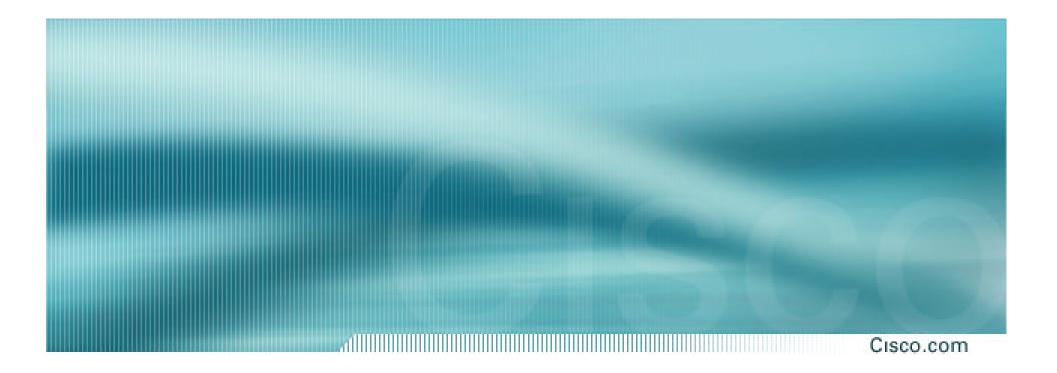
Case Study Summary

- **Multihoming is not hard, really!**

  Needs a bit of thought, a bit of planning

  Use this case study as an example strategy

  Does not require sophisticated equipment, big memory, fast CPUs…

Case Study

Second Visit

Case Study – Current Status

- ## ISP currently multihomes:

  ### To AS5400 in the UK

  ### To AS2516 in Japan

  ### /19 allocated by APNIC

  ### AS 17660 assigned by APNIC

  ### 1Mbps circuits to both upstreams

Case Study – Requirements

- ## ISP wants:

    - To add a new satellite connection, a 640K link to AS22351 in Germany to support the AS5400 link to UK

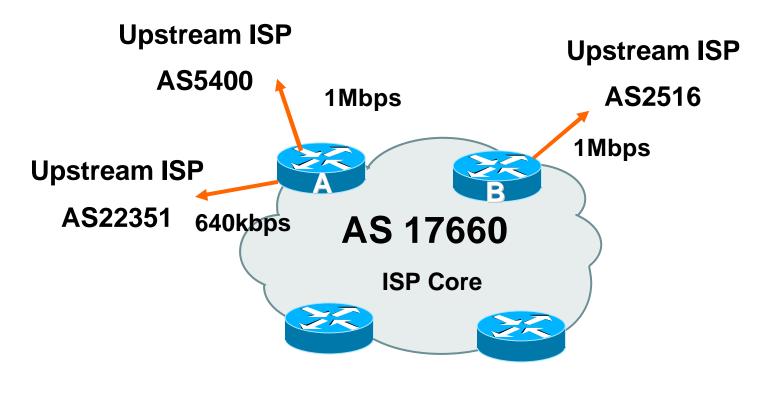   - Still want symmetric routing and equal link utilisation in and out (as close as possible)

        - international circuits are expensive

    - Has upgraded to two Cisco 3725 border routers with plenty of memory

- ## Despite the working previous configuration with "sparse routing table", wanted full prefixes

- ## Talked them out of that, and here is how…

Case Study

Upstream ISP

AS5400

1Mbps

Upstream ISP

AS2516

1Mbps

Upstream ISP

AS22351   **640kbps**

A

B

AS 17660

ISP Core

Allocated /19 from APNIC

Case Study – Inbound Loadsharing

- **First cut: Went to a few US Looking Glasses**

    Checked the AS path to AS5400

    Checked the AS path to AS2516

    Checked the AS path to AS22351

    AS2516 was one hop "closer" than the other two

    Sent AS-PATH prepend of one AS on AS2516 peering

    this is unchanged from two years ago

Case Study – Inbound Loadsharing

- **Refinement**

    Needed some – AS5400 seemed to be always preferred over AS22351

    AS5400 now supports RFC1998 style communities for customer use

    see **whois –h whois.ripe.net AS5400**

    Sent AS5400 some communities to insert prepends towards specific peers

    Now saw some traffic on AS22351 link but not much

    Sent a /23 announcement out AS22351 link

    Now saw more traffic on AS22351 link

Case Study – Inbound Loadsharing

- ## Results:

    Around 600kbps on the AS5400 link

    Around 750kbps on the AS2516 link

    Around 300kbps on the AS22351 link

    Inbound traffic fluctuates quite substantially based on time of day

- ## Status:

    Situation left pending monitoring by the ISP's NOC

Case Study – Outbound Loadsharing

- **First cut:**

  Already receiving default from AS2516

  Receiving full routes from AS5400

  Requested full routes from AS22351 – the only option

- **Retained the AS5400 configuration**

  Discard prefixes which had top 5 ASNs in the path

- **AS22351 configuration uses similar ideas to AS5400 configuration**

  But only accepted prefixes originated from AS22351 or their immediate peers
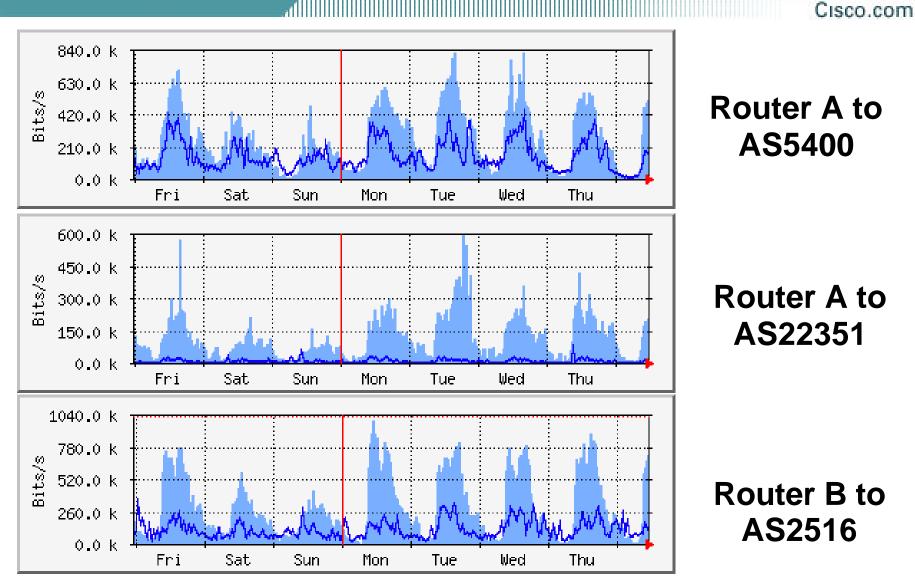
Case Study – Outbound Loadsharing

- **Results:**

  Around 35000 prefixes from AS5400

  Around 2000 prefixes from AS22351

  Around 200kbps on both the AS5400 and AS2516 links

  Around 50kbps on the AS22351 link

  Outbound traffic fluctuates quite substantially based on time of day

- **Status:**

  Situation left pending monitoring by the ISP's NOC

Case Study
MRTG Graphs

Router A to AS5400

Router A to AS22351

Router B to AS2516

Case Study
Configuration Router A

```
router bgp 17660
 no synchronization
 no bgp fast-external-fallover
 bgp log-neighbor-changes
 bgp deterministic-med
 neighbor 80.255.39.241 remote-as 22351
 neighbor 80.255.39.241 description ebgp peer to AS22351
 neighbor 80.255.39.241 send-community
 neighbor 80.255.39.241 prefix-list in-filter in
 neighbor 80.255.39.241 prefix-list out-filter-as22351 out
 neighbor 80.255.39.241 route-map as22351-out out
 neighbor 80.255.39.241 maximum-prefix 120000 95 warning-only
 neighbor 80.255.39.241 filter-list 3 in
 neighbor 80.255.39.241 filter-list 5 out
...next slide
```

Case Study
Configuration Router A

```
    neighbor 166.49.165.13 remote-as 5400

    neighbor 166.49.165.13 description eBGP multihop to AS5400

    neighbor 166.49.165.13 ebgp-multihop 5

    neighbor 166.49.165.13 update-source Loopback0

    neighbor 166.49.165.13 send-community

    neighbor 166.49.165.13 prefix-list in-filter in

    neighbor 166.49.165.13 prefix-list out-filter out

    neighbor 166.49.165.13 route-map as5400-out out

    neighbor 166.49.165.13 filter-list 1 in

    neighbor 166.49.165.13 filter-list 5 out

!

ip prefix-list in-filter  deny rfc1918 prefixes etc

ip prefix-list out-filter permit 202.144.128.0/19

ip prefix-list out-filter-as22351 permit 202.144.128.0/19

ip prefix-list out-filter-as22351 permit 202.144.158.0/23

...next slide
```

Case Study
Configuration Router A

```
ip as-path access-list 1 deny _701_

ip as-path access-list 1 deny _1_

ip as-path access-list 1 deny _7018_

ip as-path access-list 1 deny _1239_

ip as-path access-list 1 deny _7046_

ip as-path access-list 1 permit _5400$

ip as-path access-list 1 permit _5400_[0-9]+$

ip as-path access-list 1 permit _5400_[0-9]+_[0-9]+$

ip as-path access-list 1 deny .*

ip as-path access-list 3 permit _22351$

ip as-path access-list 3 permit _22351_[0-9]+$

ip as-path access-list 3 deny .*

ip as-path access-list 5 permit ^$

!

route-map as5400-out permit 10

 set community 5400:2001 5400:2101 5400:2119 5400:2124 5400:2128

route-map as22351-out permit 10
```

Case Study
Configuration Router B

```
router bgp 17660
 no synchronization
 no auto-summary
 no bgp fast-external-fallover
 bgp log-neighbor-changes
 bgp deterministic-med
 neighbor 210.132.92.165 remote-as 2516
 neighbor 210.132.92.165 description eBGP Peering with AS2516
 neighbor 210.132.92.165 send-community
 neighbor 210.132.92.165 prefix-list default-route in
 neighbor 210.132.92.165 prefix-list out-filter out
 neighbor 210.132.92.165 route-map as2516-out out
 neighbor 210.132.92.165 maximum-prefix 100
 neighbor 210.132.92.165 filter-list 2 in
 neighbor 210.132.92.165 filter-list 5 out
...next slide
```

Case Study
Configuration Router B

```
!
prefix-list default-route permit 0.0.0.0/0
prefix-list out-filter permit 202.144.128.0/19
!
ip as-path access-list 2 permit _2516$
ip as-path access-list 2 deny .*
ip as-path access-list 5 permit ^$
!
route-map as2516-out permit 10
 set as-path prepend 17660
!
```

Interesting Aside

- **Prior to installation of the new 640kbps link, ISP was complaining that both 1Mbps links were saturated inbound**

  Hence the requirement for the new 640kbps circuit

- **Research using NetFlow, cflowd and FlowScan showed that Kazaa was to blame!**

  Kazaa is a peer to peer file sharing utility

  Consumes all available bandwidth

  Found that many customers were using Kazaa for file sharing, saturating the links inbound
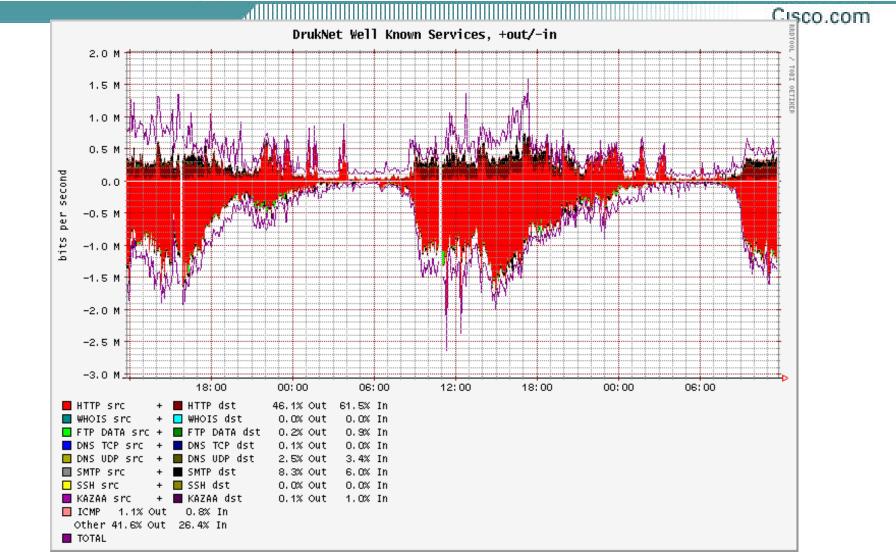
Interesting Aside

- ## Solution

  A time of day filter which blocked Kazaa during working hours, 8am to 8pm

  Inbound and outbound ACLs on border routers had tcp/1214 filters added

  ```
  access-list 100 deny tcp any any eq 1214 time-range OfficeHours
  access-list 101 deny tcp any any eq 1214 time-range OfficeHours
  !
  time-range OfficeHours
   periodic weekdays 8:00 to 20:00
  ```

  The result: inbound traffic on external links dropped by 50%

  And complaints about "the 'net" being slow have reduced

Interesting Aside

Typical FlowScan graph – no longer showing the effects of Kazaa
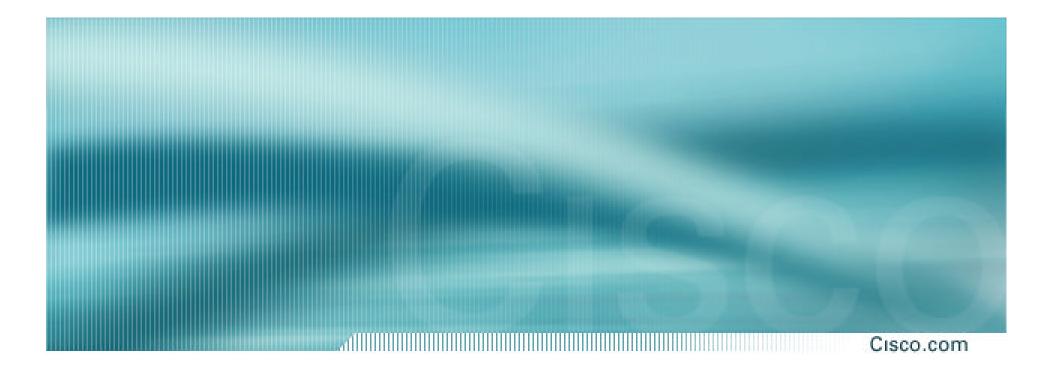
Summary

- **Multihoming solution with three links of different bandwidths works well**

  Fluctuates significantly during the day time, maybe reflecting users browsing habits?

  NOC is monitoring the situation

  NOTE: Full routing table is not required ☺
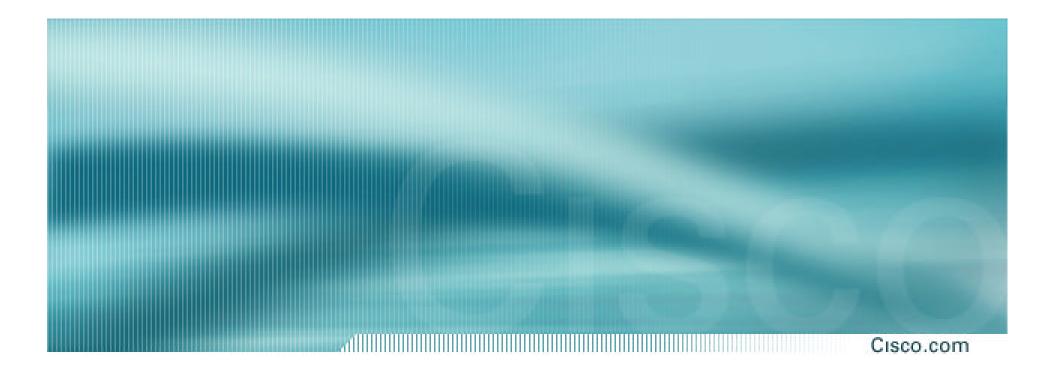
Summary

Summary

- ## Multihoming is not hard, really…

  Keep It Simple! is a very good principle to follow

- ## Full routing table is rarely required

  A default is just as good

  If customers want 120k prefixes, charge them money for it

Cisco.com

BGP Multihoming Techniques

End of Tutorial