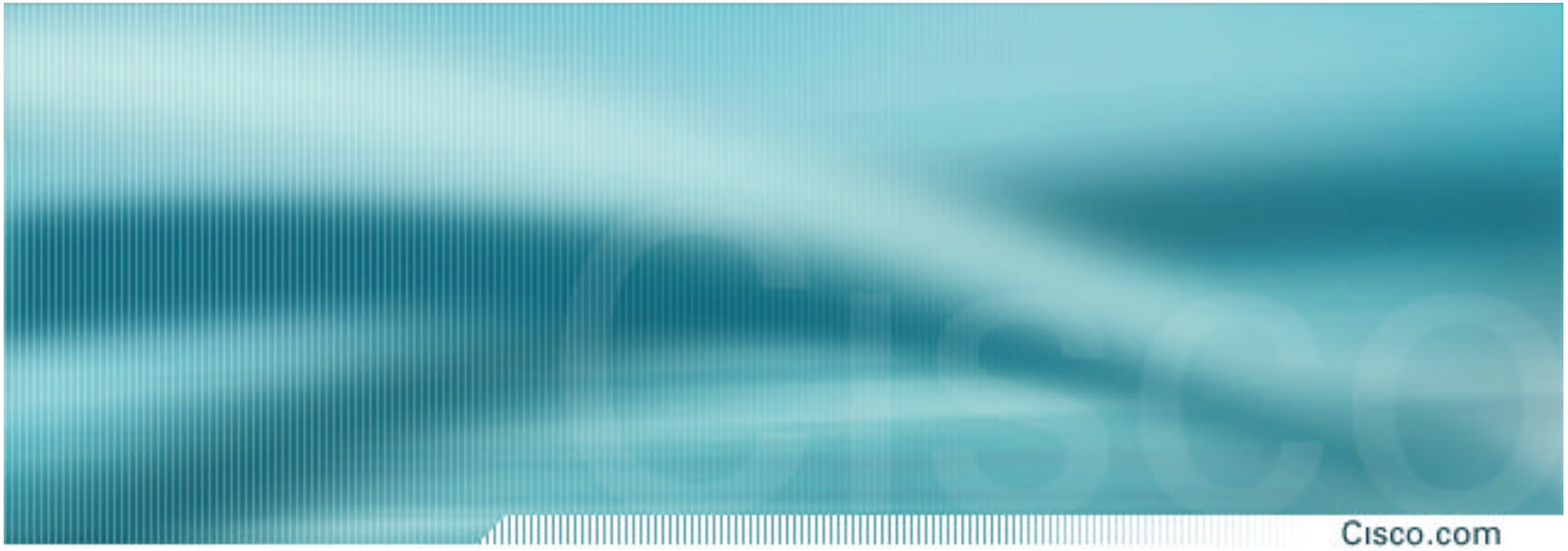


BGP Multihoming Techniques

Philip Smith <pfs@cisco.com>

APNIC 2002, Kitakyushu, Japan

Presentation Slides

- **Available at**

  www.apnic.net/meetings/14/programme/docs/bgp-tut-slides-pfs.pdf

  www.cisco.com/public/cons/seminars/APNIC2002

Introduction

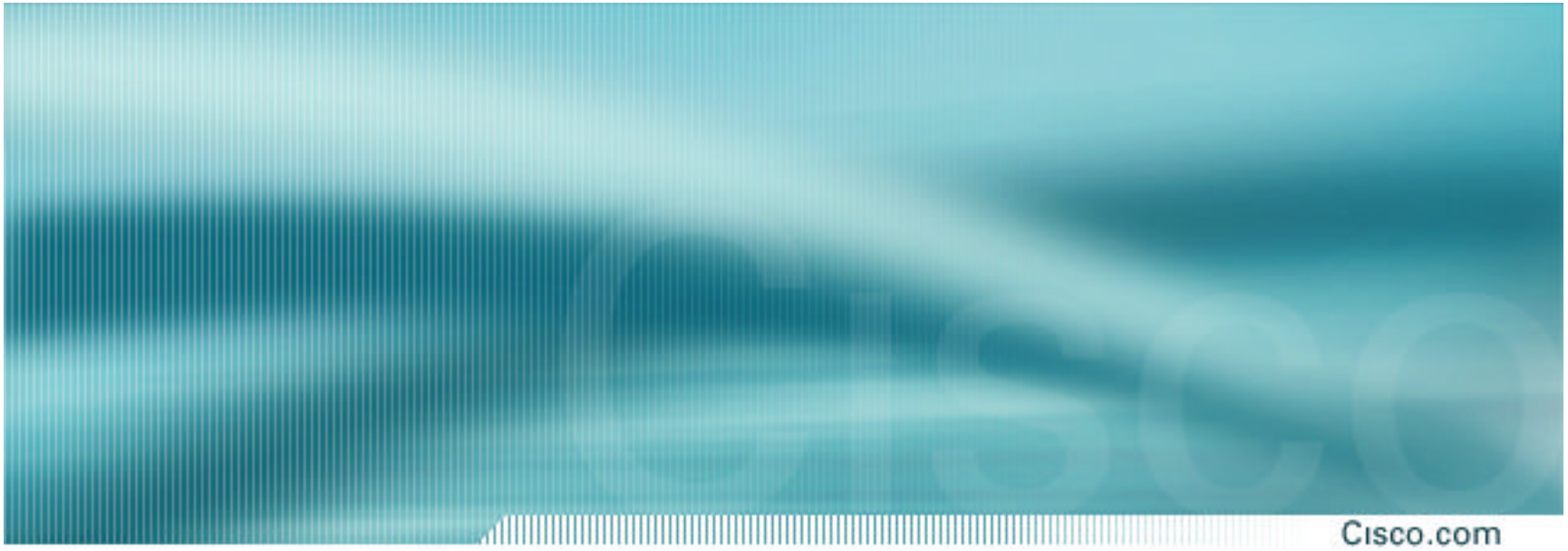
Cisco.com

- 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

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- Using Communities
- Case Study



Preparations

Before we begin…

Preparations

- Definition of Multihoming
- ASN and how to get one
- Aggregation/BGP Best Practices
- Policy Options on the Router
- Multihoming Options
- Private ASNs

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

Obtaining an AS Number

- **Existing Local Internet Registry (LIR)**

  (this means APNIC member)

  Fill up the application form – http://ftp.apnic.net/apnic/docs/asn-request

- **Customer of an LIR**

  Ask them to apply on your behalf

  They may charge you a fee for this service

Obtaining an AS Number (cont)

- **Become an LIR**

  Apply for APNIC membership

  http://www.apnic.net/member/application.html

- **Apply for a one-off allocation as non-member**

  http://www.apnic.net/member/non-member-application.html

  Costs US$500, with US$50 yearly maintenance

APNIC's AS application form

Cisco.com

- **Fields in the Autnum Template**

as-name	Choose a name
descr	Give description
country	ISO country code
import	list what you will receive from upstreams
export	list what you will send to upstreams
default	if applicable, list a default
remarks	any comments
admin-c	your admin contact (for bills etc)
tech-c	your tech contact (for technical issues)
notify	who to notify of changes
mnt-by	Maintainer object to protect the autnum

APNIC's AS application form – Example

aut-num	*MY-AS*
as-name	PHILIP-ISP
descr	Philip's own ISP
descr	Brisbane
country	AU
import	from AS703
	accept any
import	from AS1221
	accept AS1221
export	to AS703
	announce MY-AS
export	to AS1221
	announce MY-AS
default	to AS703
	networks ANY
admin-c	PFS1-AP
tech-c	PFS1-AP
notify	pfs@cisco.com
mnt-by	MAINT-PFS1-AP

Placeholder for the number APNIC will assign.

AS703 is the upstream provider, providing the full routing table.

AS1221 is a local peer, providing only AS1221's prefixes.

The default route is to AS703.

This aut-num describes how PHILIP-ISP is multihomed between AS703 and AS1221.

Import/Export

- **The aut-num lists the import/export policies**
  - **Here the ISP applying for the AS lists the ASNs it will be multihoming with**
  - **The ISP lists the policy – see the APNIC supporting documentation for examples of these**
  - **Need at least two ASNs in the import/export policy section before the ISP will assigned an ASN**

Announcing Prefixes

- **ISPs receive address block from Regional Registry or upstream provider**
- **Aggregation means announcing the address block only, not subprefixes**
  - **Subprefixes should only be announced in special cases – see later.**
- **Aggregate should be generated internally**
  - **Not on the network borders!**

Configuring Aggregation

Cisco.com

- ISP has 221.10.0.0/19 address block
- To put into BGP as an aggregate:

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
ip route 221.10.0.0 255.255.224.0 null0
```

- The static route is a “pull up” route
  - more specific prefixes within this address block ensure connectivity to ISP’s customers
  - “longest match lookup”

Announcing Aggregate: Cisco IOS

- **Configuration Example**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 110
 neighbor 222.222.10.1 prefix-list out-filter out
!
ip route 221.10.0.0 255.255.224.0 null0
!
ip prefix-list out-filter permit 221.10.0.0/19
```

Announcing an Aggregate

- ISPs who don't and won't aggregate are held in poor regard by community
- Registries' minimum allocation size is now a /20

  Very few reasons to see subprefixes of allocated blocks in the Internet

  BUT there are currently >63000 /24s!

- Several ISPs filter based on the RIRs' minimum allocation size

  Called the "Net Police Filters" by some

The Internet Today

- **Current Internet Routing Table Statistics**

BGP Routing Table Entries	114514
Prefixes after maximum aggregation	74191
Unique prefixes in Internet	55037
Prefixes larger than registry alloc	46614
/24s announced	63202
only 5506 /24s are from 192.0.0.0/8	
ASes in use	13566

Part of the “Net Police” prefix list


```
!! 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 200.0.0.0/8 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 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
```

“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

Receiving Prefixes: From Downstreams

- ISPs should only accept prefixes which have been assigned or allocated to their downstream customer
- For example

  downstream has 220.50.0.0/20 block

  should only announce this to peers

  peers should only accept this from them

Receiving Prefixes: Cisco IOS

- **Configuration Example on upstream**

```
router bgp 100
 neighbor 222.222.10.1 remote-as 110
 neighbor 222.222.10.1 prefix-list customer in
!
ip prefix-list customer permit 220.50.0.0/20
```

Receiving Prefixes: From Upstreams

- **Not desirable unless really necessary**
  - **special circumstances – see later**
- **Ask upstream to either:**
  - **originate a default-route**
    - **-or-**
  - **announce one prefix you can use as default**

Receiving Prefixes: From Upstreams

- **Downstream Router Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 110
 neighbor 221.5.7.1 prefix-list infilter in
 neighbor 221.5.7.1 prefix-list outfilter out
!
ip prefix-list infilter permit 0.0.0.0/0
!
ip prefix-list outfilter permit 221.10.0.0/19
```

Receiving Prefixes: From Upstreams

- **Upstream Router Configuration**

```
router bgp 110
 neighbor 221.5.7.2 remote-as 100
 neighbor 221.5.7.2 default-originate
 neighbor 221.5.7.2 prefix-list cust-in in
 neighbor 221.5.7.2 prefix-list cust-out out
!
ip prefix-list cust-in permit 221.10.0.0/19
!
ip prefix-list cust-out permit 0.0.0.0/0
```

Receiving Prefixes: From Peers and Upstreams

- If necessary to receive prefixes from any provider, care is required
  - don't accept RFC1918 etc prefixes
    - http://www.ietf.org/internet-drafts/draft-manning-dsua-08.txt
  - don't accept your own prefix
  - don't accept default (unless you need it)
  - don't accept prefixes longer than /24
- Check Rob Thomas' list of "bogons"
  - http://www.cymru.org/Documents/bogon-list.html

Receiving Prefixes


```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 110
 neighbor 221.5.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0           ! Block default
ip prefix-list in-filter deny 0.0.0.0/8 le 32
ip prefix-list in-filter deny 10.0.0.0/8 le 32
ip prefix-list in-filter deny 127.0.0.0/8 le 32
ip prefix-list in-filter deny 169.254.0.0/16 le 32
ip prefix-list in-filter deny 172.16.0.0/12 le 32
ip prefix-list in-filter deny 192.0.2.0/24 le 32
ip prefix-list in-filter deny 192.168.0.0/16 le 32
ip prefix-list in-filter deny 221.10.0.0/19 le 32 ! Block local prefix
ip prefix-list in-filter deny 224.0.0.0/3 le 32   ! Block multicast
ip prefix-list in-filter deny 0.0.0.0/0 ge 25     ! Block prefixes >/24
ip prefix-list in-filter permit 0.0.0.0/0 le 32
```

Configuring Policy on the Router

Cisco.com

- **Three BASIC Principles**

  prefix-lists to filter prefixes

  filter-lists to filter ASNs

  route-maps to apply policy

- **Avoids confusion!**

Policy Tools

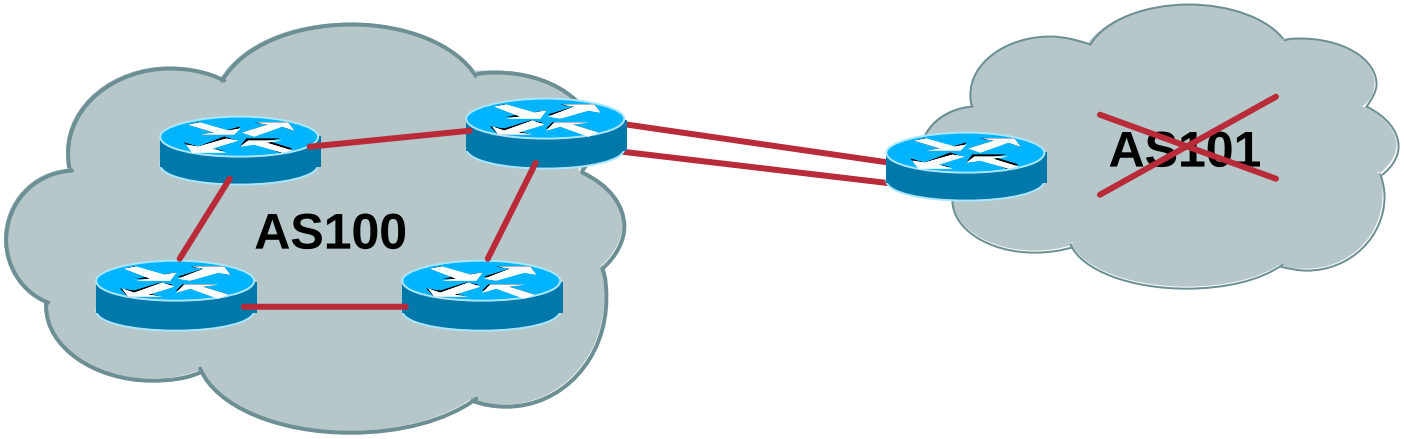
- **Local preference**
  - **outbound traffic flows**
- **Metric (MED)**
  - **inbound traffic flows (local scope)**
- **AS-PATH prepend**
  - **inbound traffic flows (Internet scope)**
- **Communities**
  - **specific inter-provider peering**

Multihoming Scenarios

Cisco.com

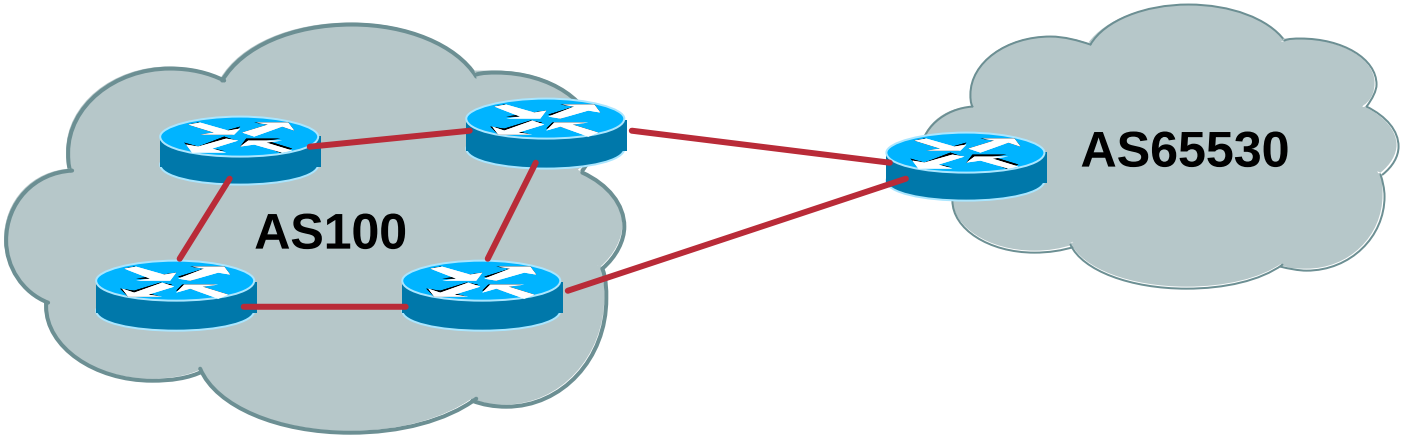
- **Stub network**
- **Multi-homed stub network**
- **Multi-homed network**
- **Configuration Options**

Stub Network



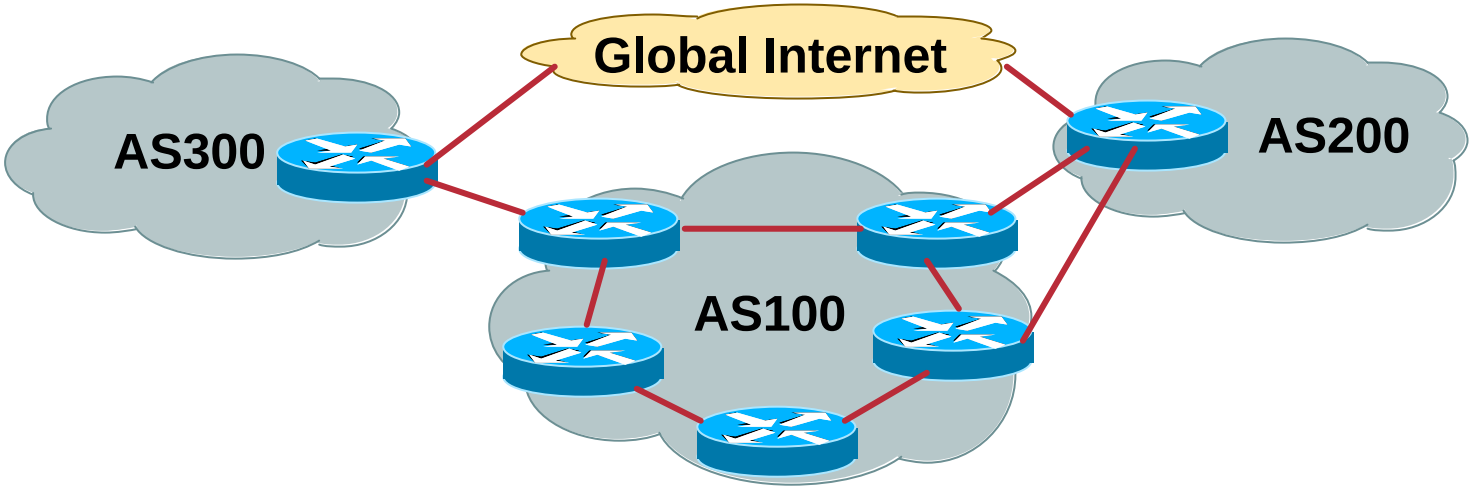
- No need for BGP
- Point static default to upstream ISP
- Upstream ISP advertises stub network
- Policy confined within upstream ISP's policy

Multi-homed Stub Network



- 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

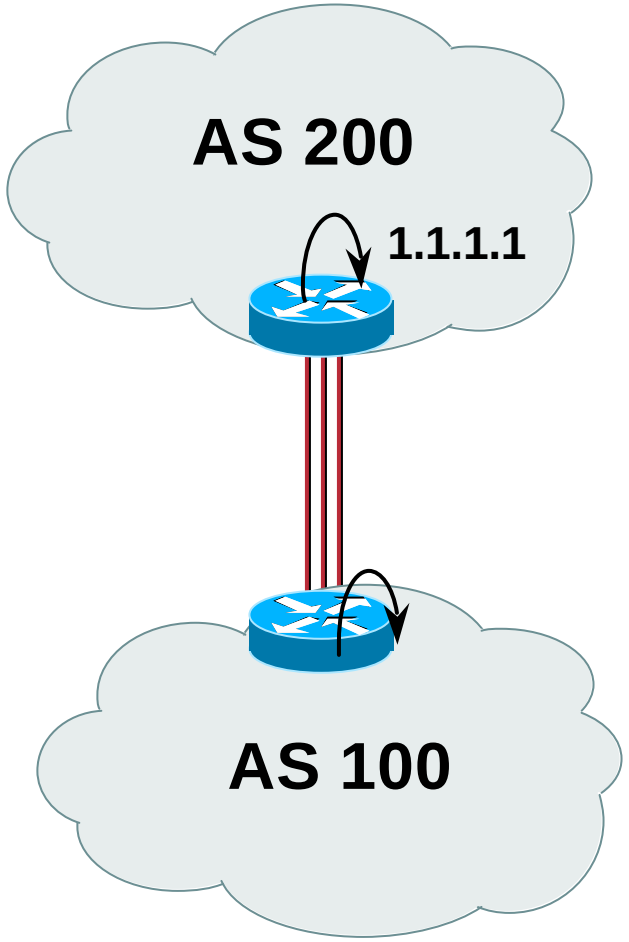


- **Many situations possible**
  - **multiple sessions to same ISP**
  - **secondary for backup only**
  - **load-share between primary and secondary**
  - **selectively use different ISPs**

Multiple Sessions to an ISP – Example One

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

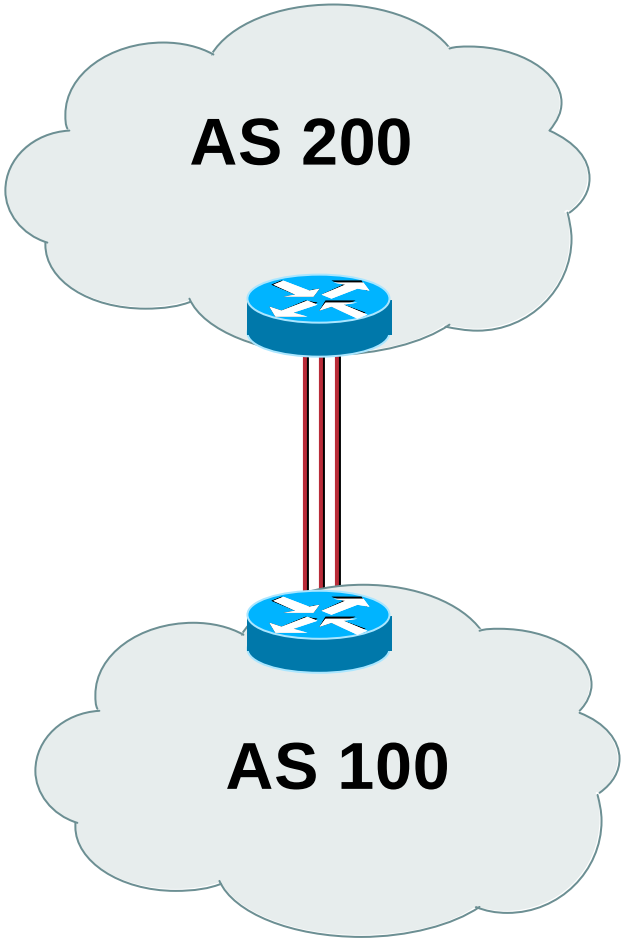
```
router bgp 100
 neighbor 1.1.1.1 remote-as 200
 neighbor 1.1.1.1 ebgp-multihop 5
!
ip route 1.1.1.1 255.255.255.255 serial 1/0
ip route 1.1.1.1 255.255.255.255 serial 1/1
ip route 1.1.1.1 255.255.255.255 serial 1/2
```



Multiple Sessions to an ISP – Example Two

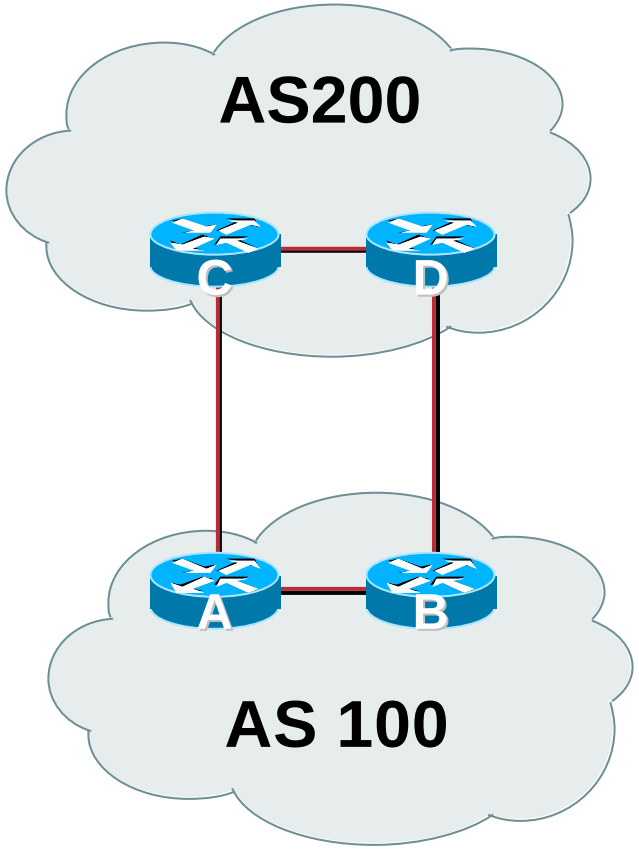
- BGP multi-path
- Three BGP sessions required
- limit of 6 parallel paths

```
router bgp 100
 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
```



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

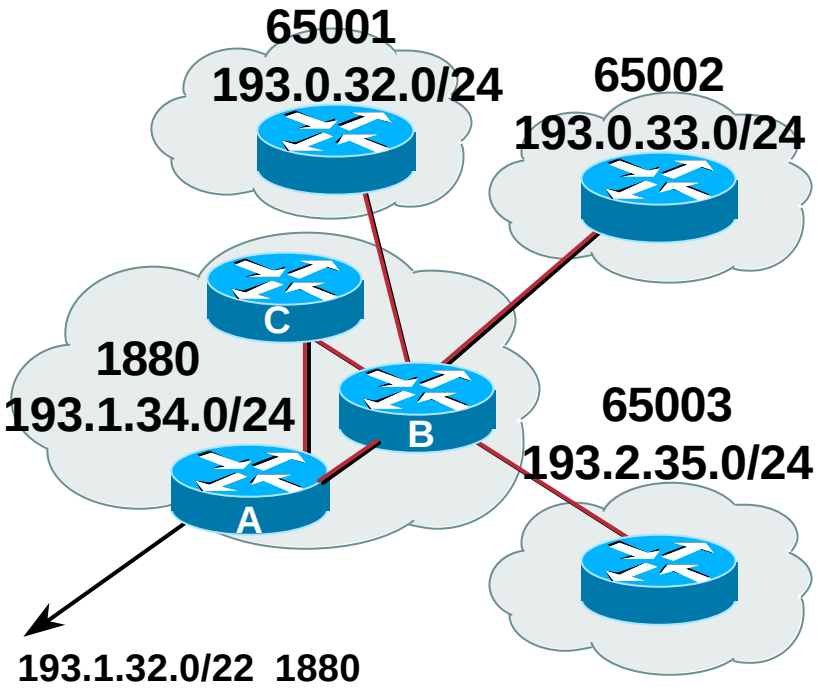


Private-AS – Application

- **Applications**

  ISP with single-homed customers (RFC2270)

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



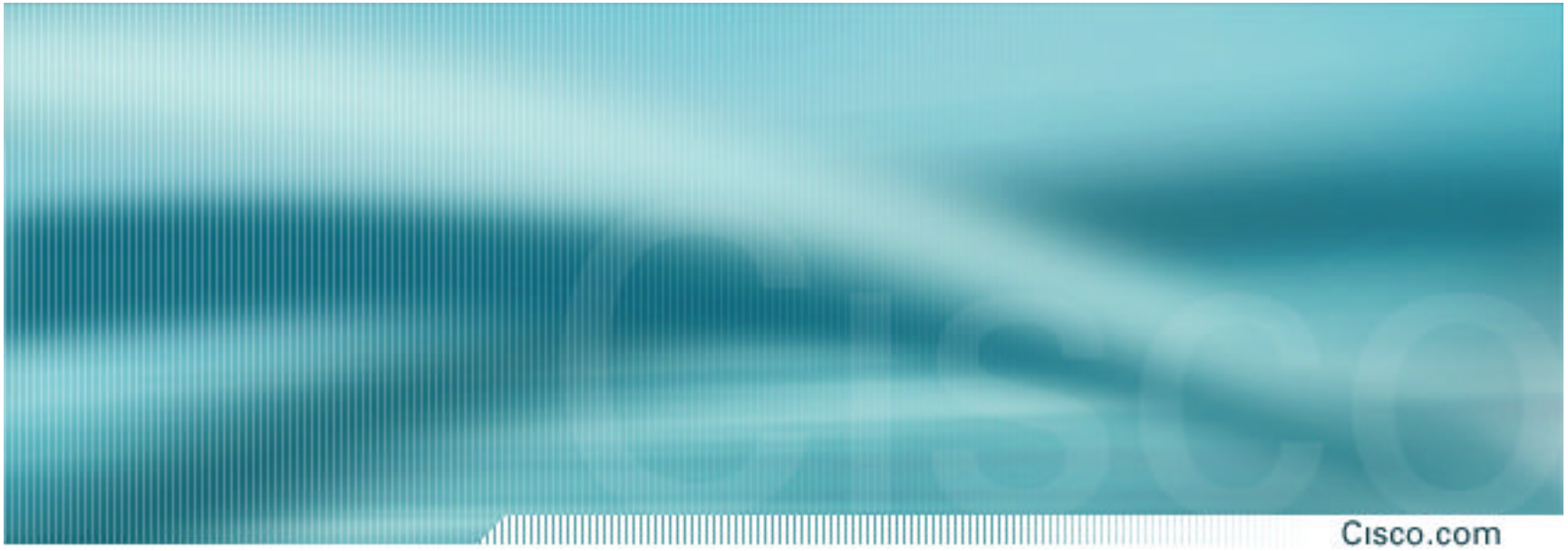
Private-AS Removal

- **neighbor x.x.x.x remove-private-AS**
- **Rules:**
  - **available for eBGP neighbors 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**
- **This command should be MANDATORY on all ISP eBGP configurations**

BGP Multihoming Techniques

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- 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)

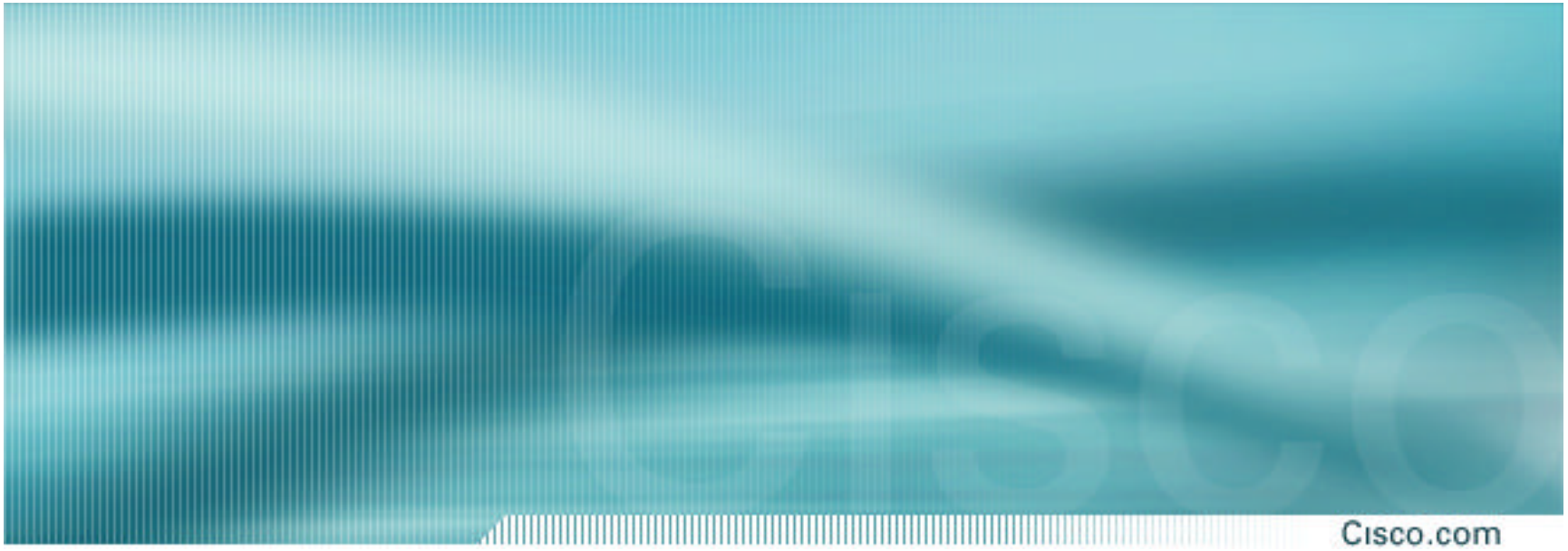
  public AS not required as any routing policy is only local to the upstream ISP

- **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)

- **Assumptions**

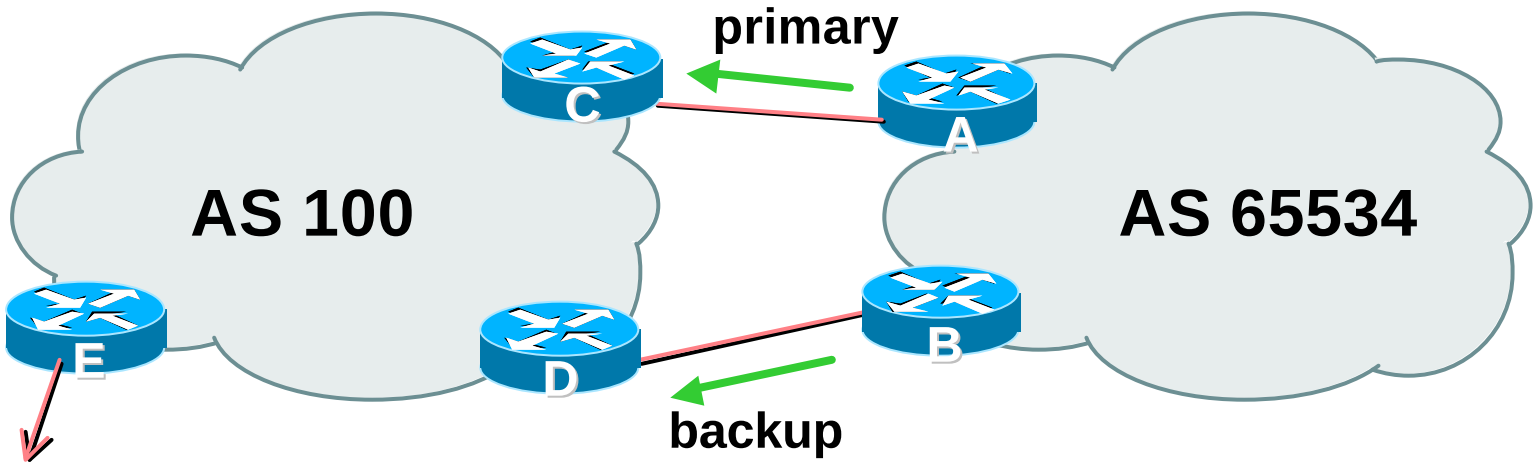
  Customer ISP uses AS65534 and has a /19 address block assigned to them



Two links to the same ISP

One link primary, the other link backup only

Two links to the same ISP



- **AS109 removes private AS and any customer subprefixes from Internet announcement**

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

- **AS65534 announces /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)

Cisco.com

```
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 – Example One**

```
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 (one as backup only)

- **Router E Configuration – Example Two**

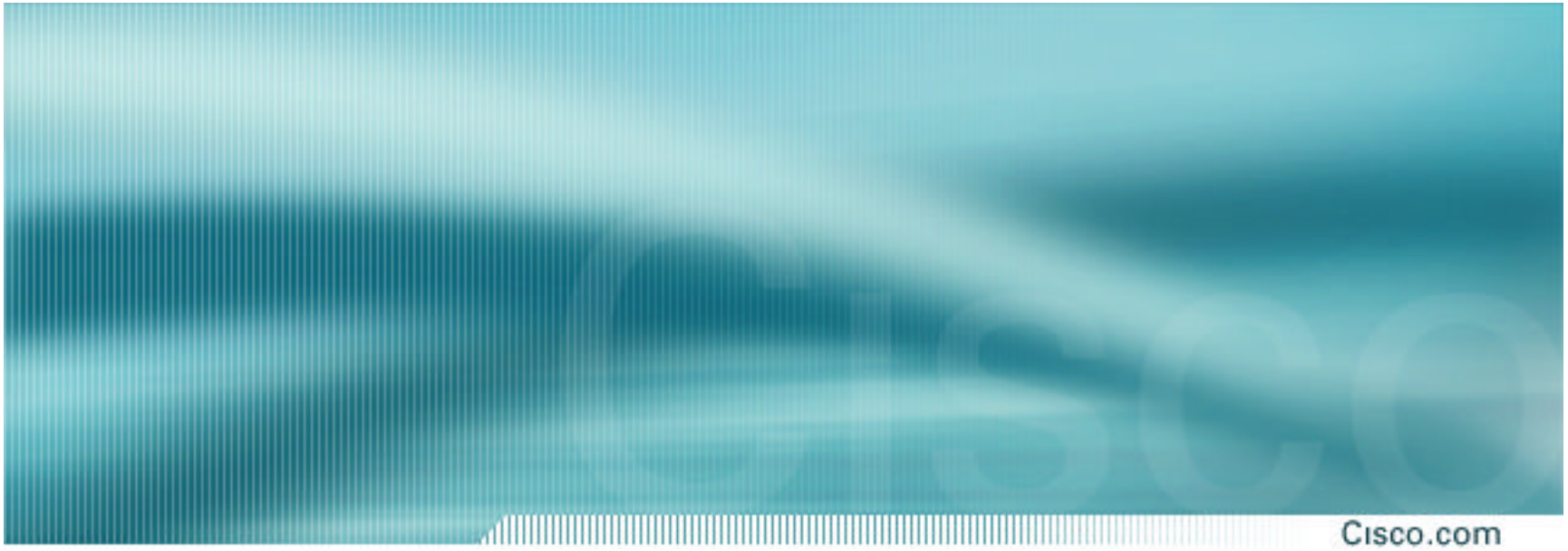
  alternative, and no longer recommended

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.17 remote-as 110
 neighbor 222.222.10.17 filter-list 10 out
 neighbor 222.222.10.17 prefix-list Customer out
!
ip prefix-list Customer permit 221.10.0.0/19
ip as-path access-list 10 deny ^(65534_)+$
ip as-path access-list 10 permit .*
ip route 221.10.0.0 255.255.224.0 null0
```

- **Router E proxy aggregates for AS65534**

Router E configuration

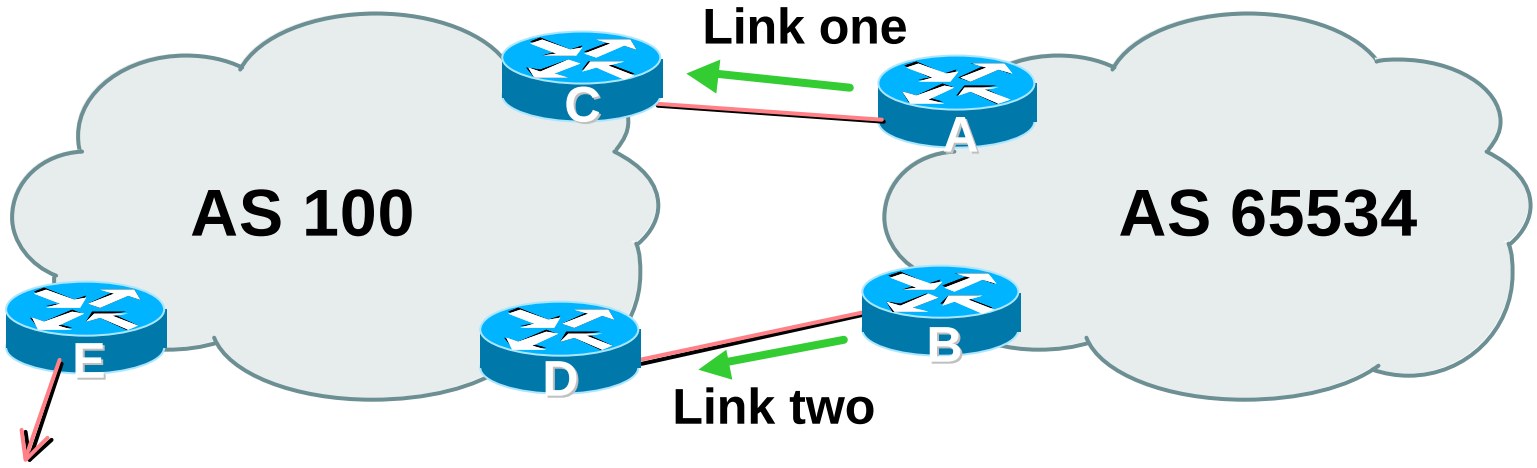
- Example One is the recommended way to do this now
  - Private AS is automatically stripped at the network edge
  - AS65534's generated aggregate is transited by AS100
- Example Two was the method prior to the "remove-private-AS" BGP option
  - More complex, more likelihood for mistakes to be made
  - Called proxy aggregation – Router E generates the aggregate for AS65534's network block



Two links to the same ISP

With Loadsharing

Loadsharing to the same ISP



- **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

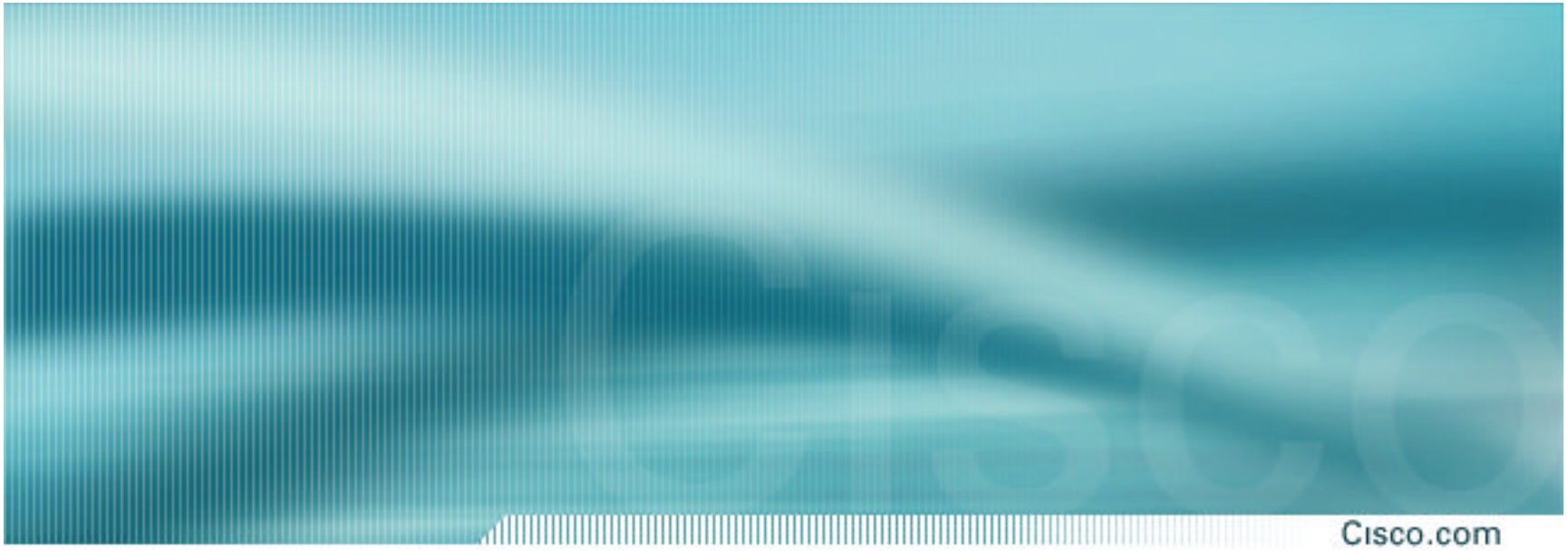
- **Router C Configuration**

```
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

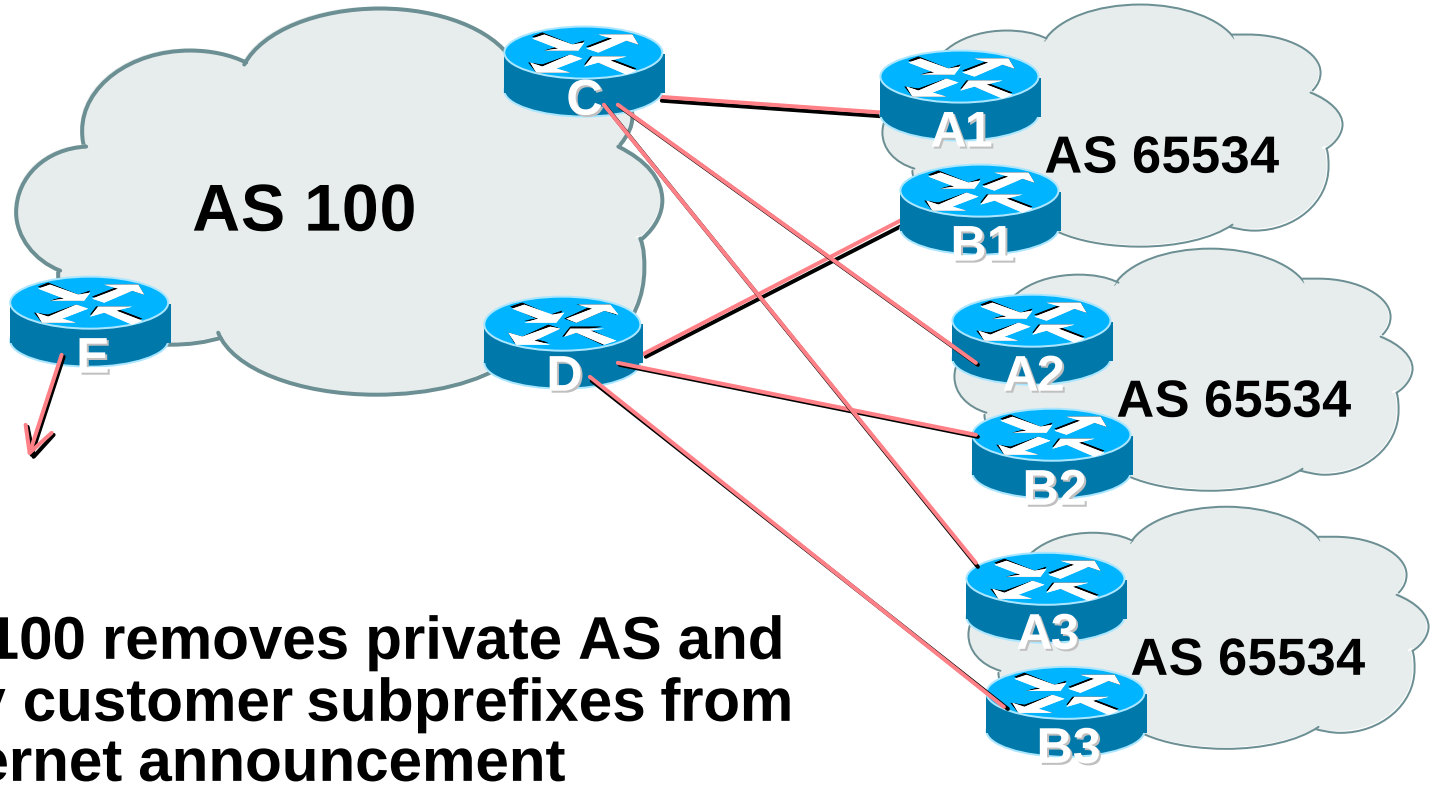


Two links to the same ISP

Multiple Dualhomed Customers

(RFC2270)

Multiple Dualhomed Customers (RFC2270)



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

Multiple Dualhomed Customers

Cisco.com

- **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

Cisco.com

- **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

Cisco.com

```
 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

Cisco.com

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

  do NOT announce them to the Internet

  announce ISP aggregate only

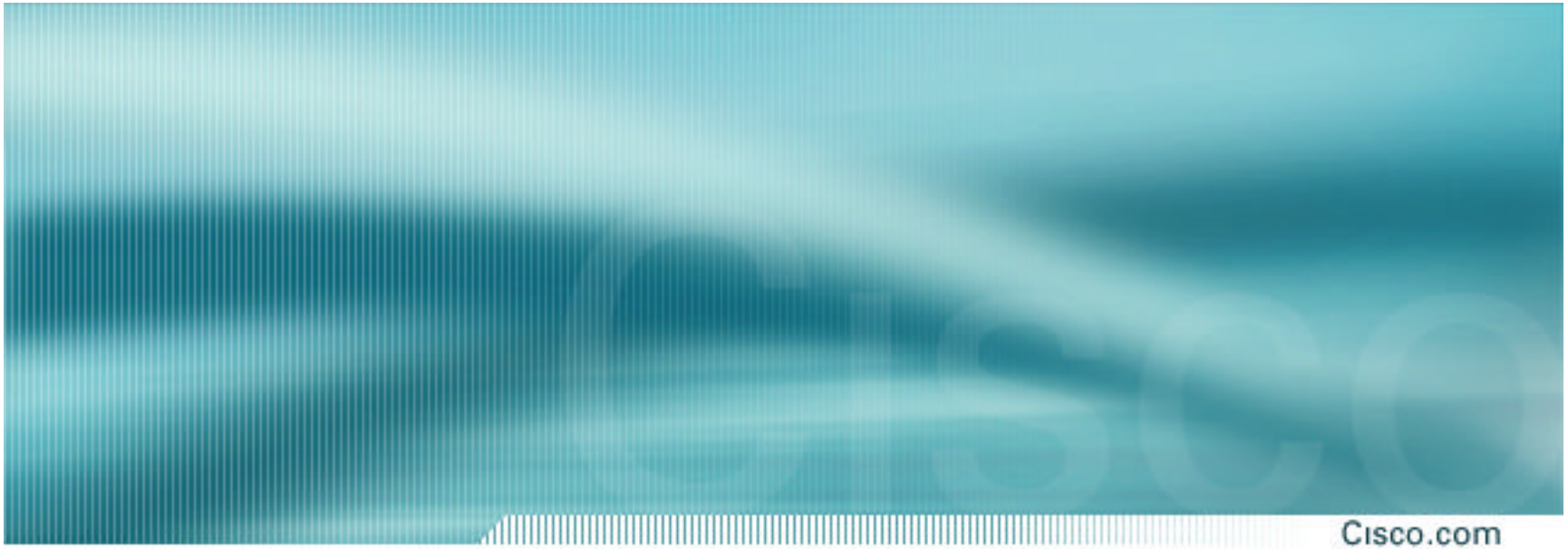
- **Router E configuration:**

```
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

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- Using Communities
- Case Study



Multihoming to different ISPs

Two links to different ISPs

- **Use a Public AS**

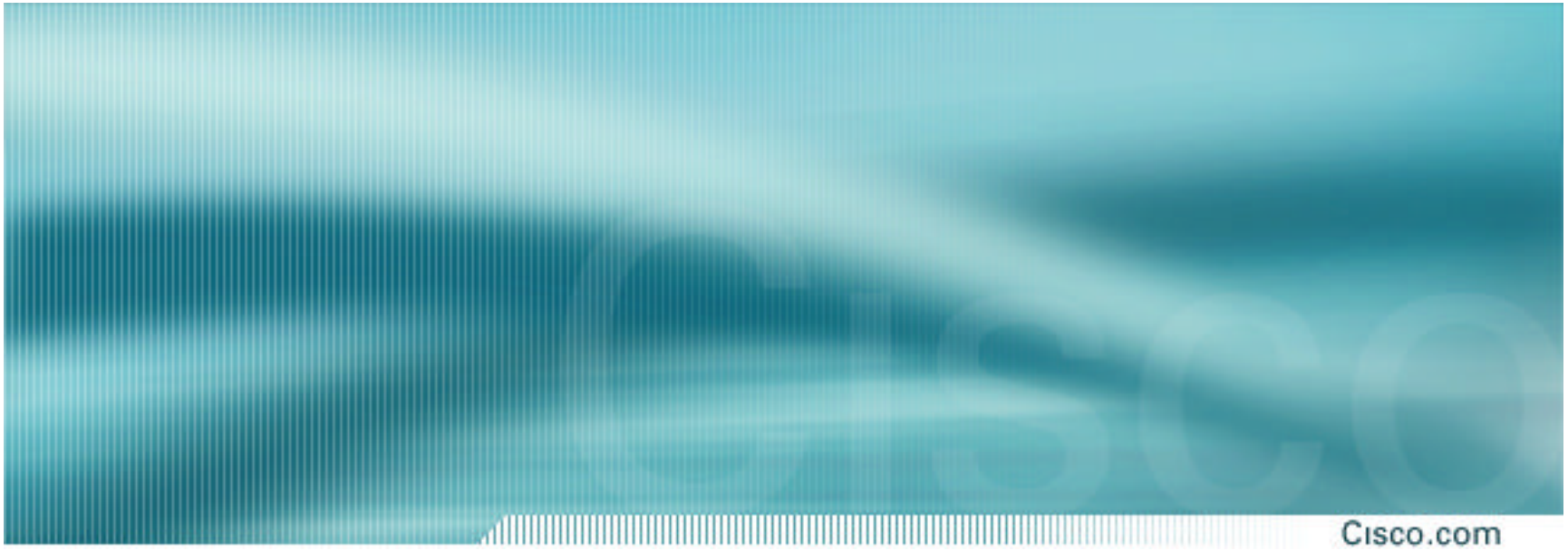
  or use private AS if agreed with the other ISP

- **Address space comes from**

  both upstreams or

  Regional Internet Registry

- **Configuration concepts very similar**



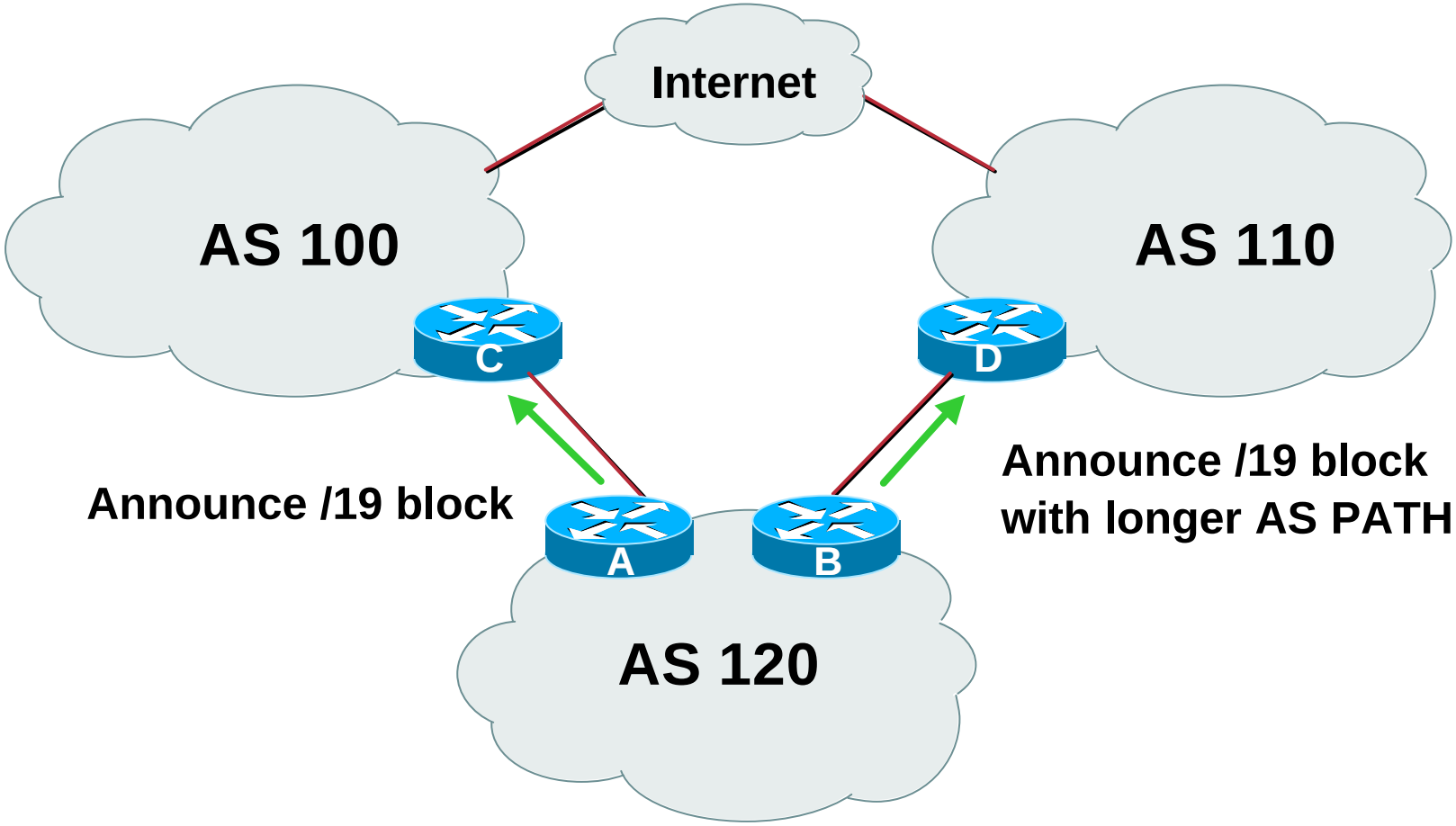
Two links to different ISPs

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)



Two links to different ISPs (one as backup only)

- **Router A Configuration**

```
router bgp 120
 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**

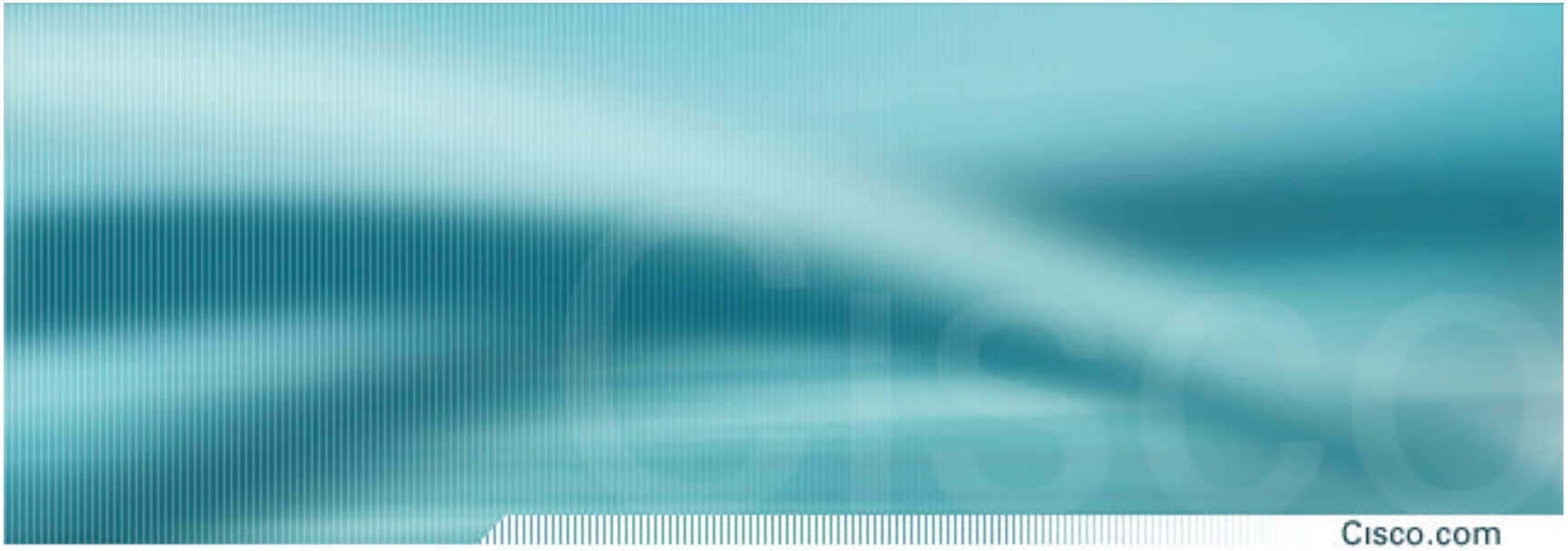
```
router bgp 120
 network 221.10.0.0 mask 255.255.224.0
 neighbor 220.1.5.1 remote-as 110
 neighbor 220.1.5.1 prefix-list aggregate out
 neighbor 220.1.5.1 route-map routerD-out out
 neighbor 220.1.5.1 prefix-list default in
 neighbor 220.1.5.1 route-map routerD-in in
!
ip prefix-list aggregate permit 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 120 120 120
!
route-map routerD-in permit 10
 set local-preference 80
```

Two links to different ISPs (one as backup only)

- **Router C Configuration**

```
router bgp 100
 neighbor 222.222.10.2 remote-as 120
 neighbor 222.222.10.2 default-originate
 neighbor 222.222.10.2 prefix-list Customer in
 neighbor 222.222.10.2 prefix-list default out
!
ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```

- **Router C only allows in /19 prefixes from customer block**
- **Router D configuration is almost identical**



Two links to different ISPs

With Loadsharing

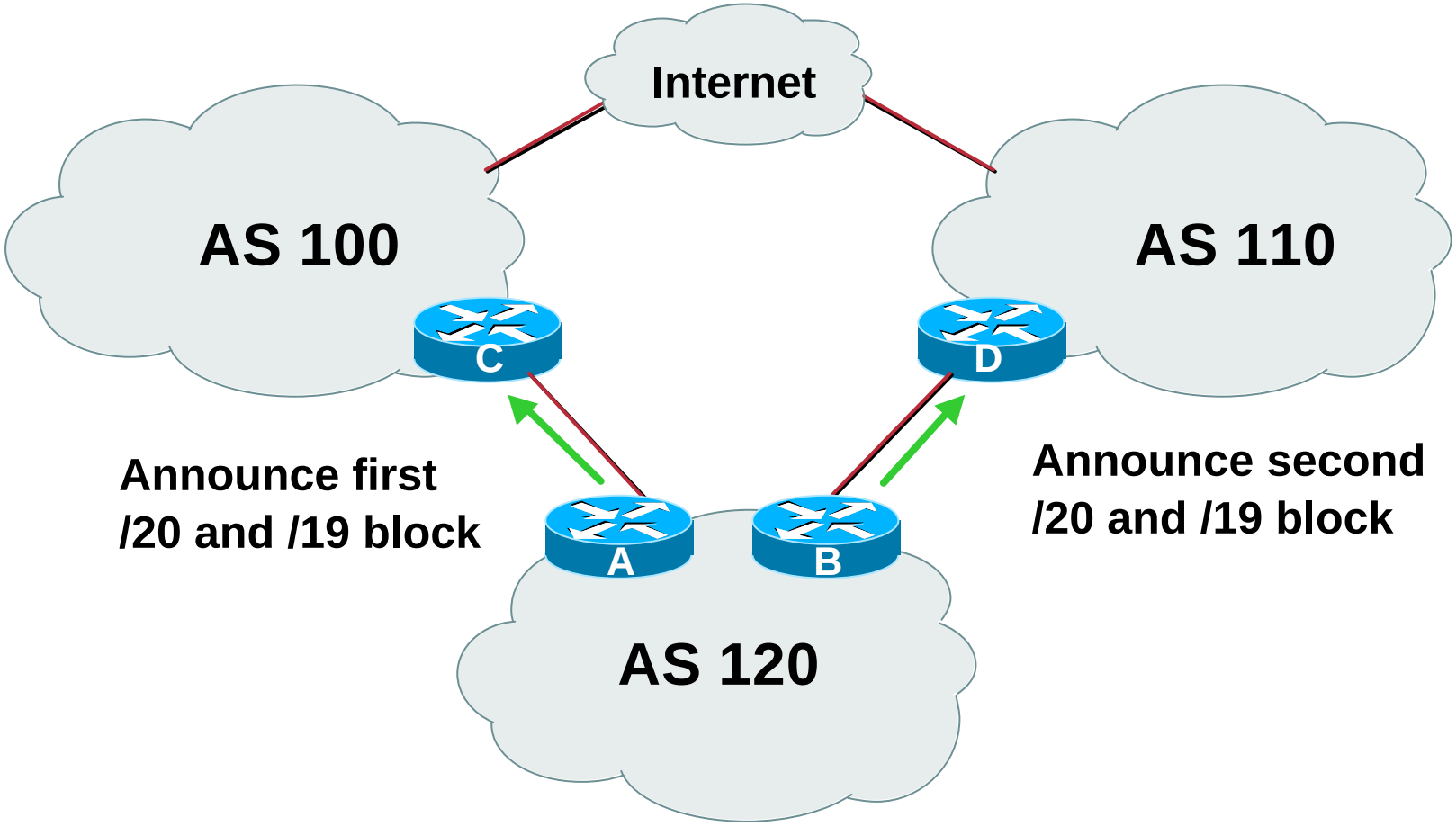
Two links to different ISPs (with loadsharing)

- 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)



Two links to different ISPs (with loadsharing)

- **Router A Configuration**

```
router bgp 120
 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)

Cisco.com

- **Router B Configuration**

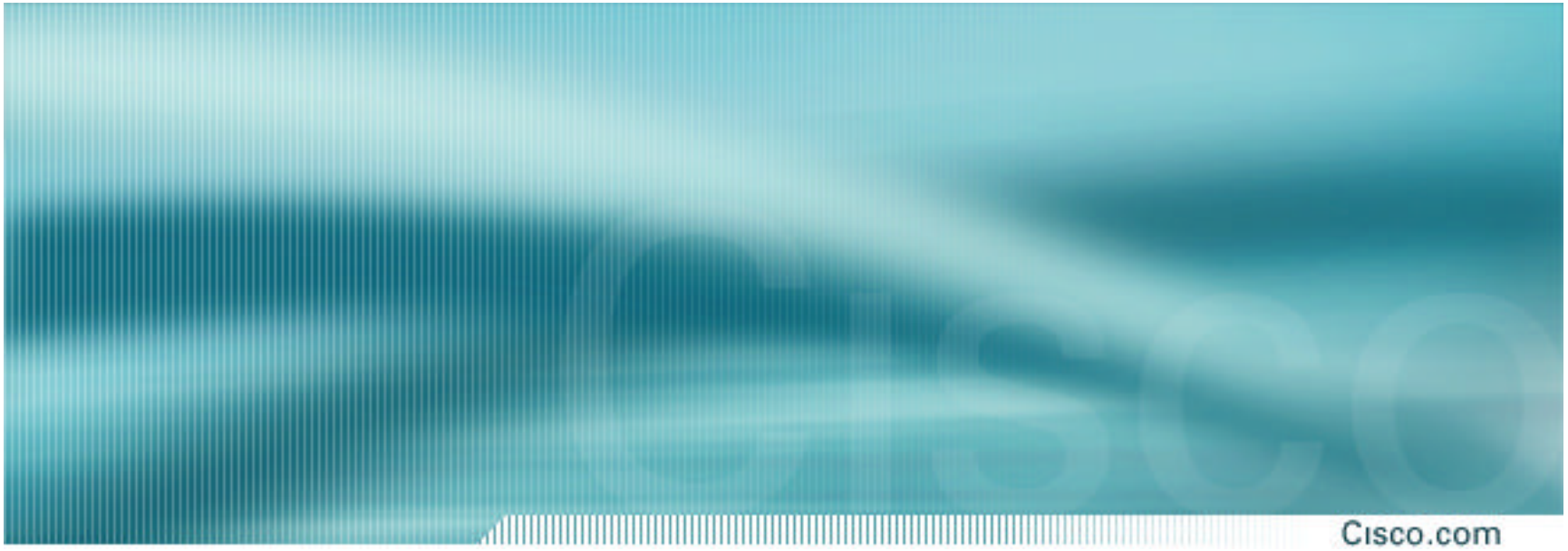
```
router bgp 120
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.16.0 mask 255.255.240.0
 neighbor 220.1.5.1 remote-as 110
 neighbor 220.1.5.1 prefix-list secondblock out
 neighbor 220.1.5.1 prefix-list default in
!
ip prefix-list default permit 0.0.0.0/0
!
ip prefix-list secondblock permit 221.10.16.0/20
ip prefix-list secondblock permit 221.10.0.0/19
```

Two links to different ISPs (with loadsharing)

- **Router C Configuration**

```
router bgp 100
 neighbor 222.222.10.2 remote-as 120
 neighbor 222.222.10.2 default-originate
 neighbor 222.222.10.2 prefix-list Customer in
 neighbor 222.222.10.2 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 almost identical**
- **Note that upstream keeps configuration simple to allow flexibility of changes by AS120**



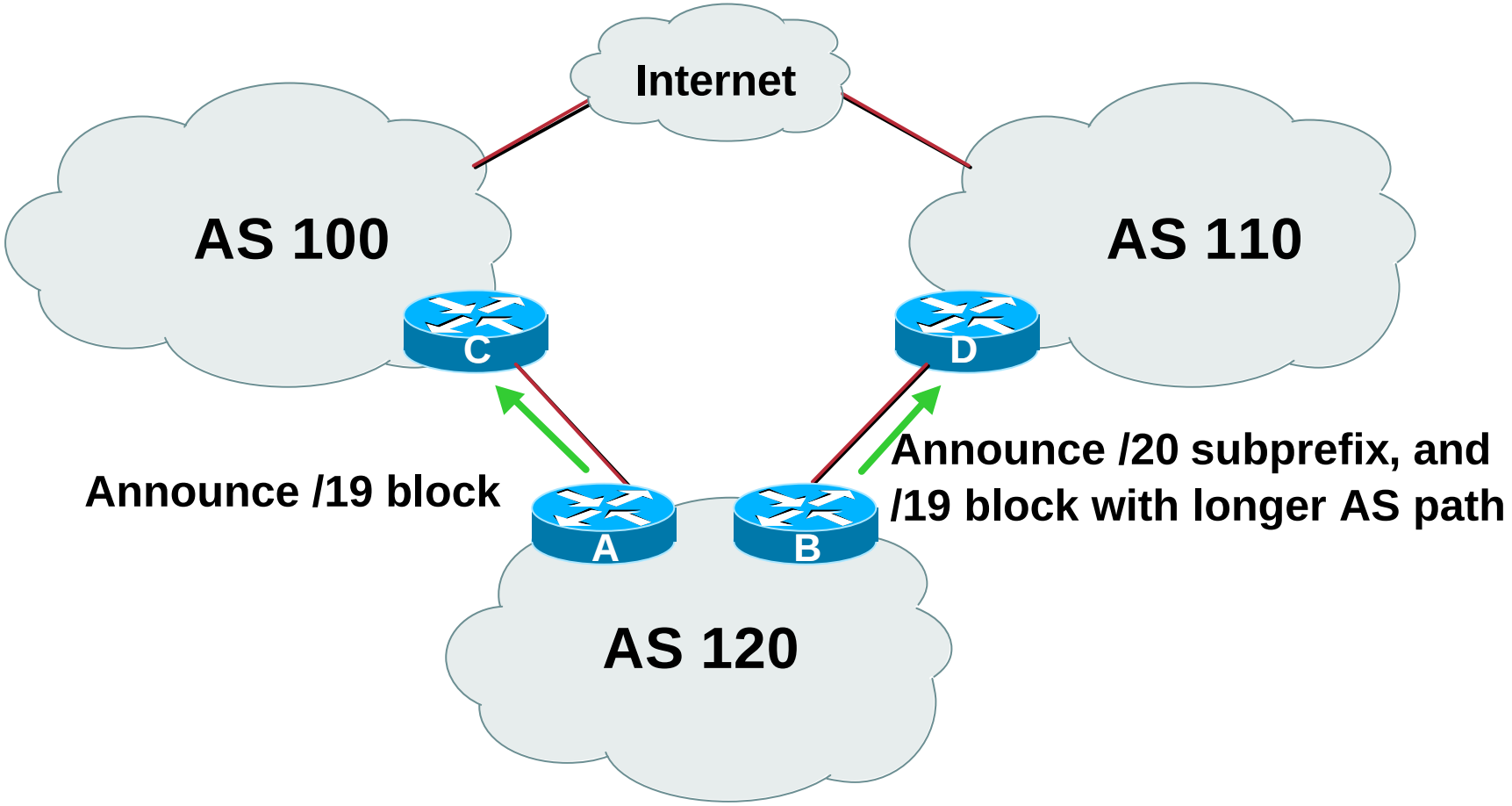
Two links to different ISPs

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



Loadsharing with different ISPs

Cisco.com

- **Router A Configuration**

```
router bgp 120
 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

Cisco.com

- **Router B Configuration**

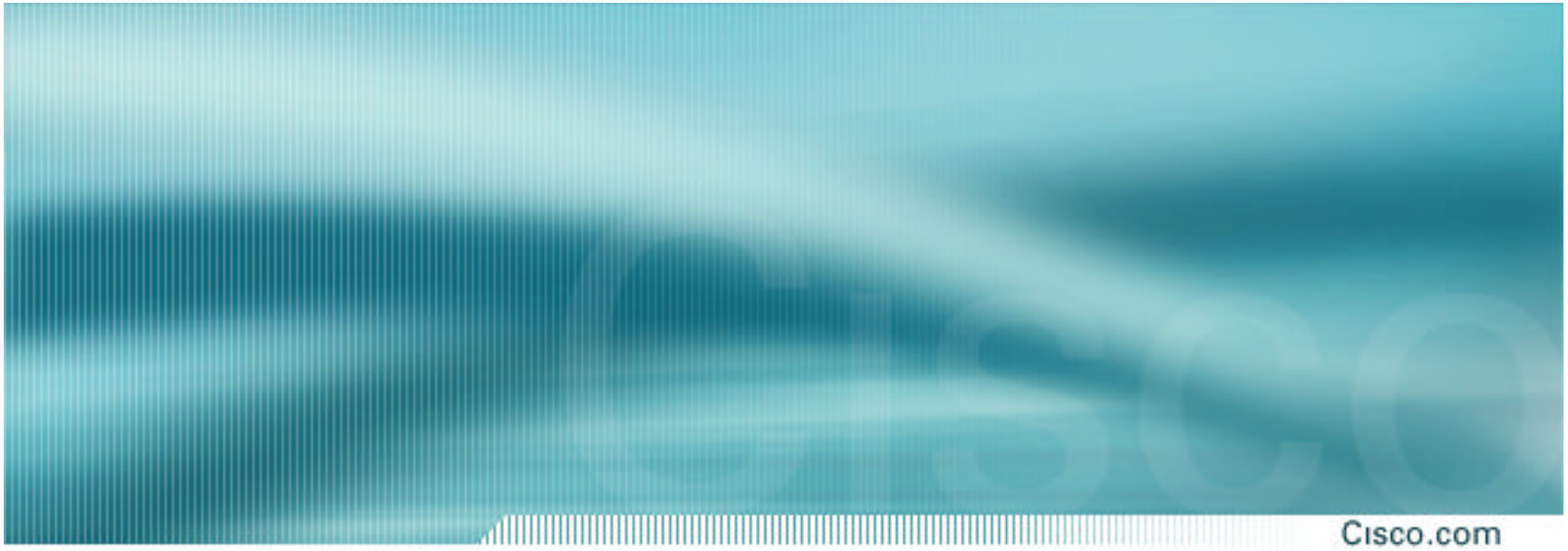
```
router bgp 120
 network 221.10.0.0 mask 255.255.224.0
 network 221.10.16.0 mask 255.255.240.0
 neighbor 220.1.5.1 remote-as 110
 neighbor 220.1.5.1 prefix-list default in
 neighbor 220.1.5.1 prefix-list subblocks out
 neighbor 220.1.5.1 route-map routerD out
!
route-map routerD permit 10
 match ip address prefix-list aggregate
 set as-path prepend 120 120
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
```

Loadsharing with different ISPs

- **Router C Configuration**

```
router bgp 100
 neighbor 222.222.10.2 remote-as 120
 neighbor 222.222.10.2 default-originate
 neighbor 222.222.10.2 prefix-list Customer in
 neighbor 222.222.10.2 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 almost identical**
- **Note that upstream keeps configuration simple to allow flexibility of changes by AS120**



Loadsharing with different ISPs

More Complexity

Loadsharing with different ISPs

- So far, we've seen simple examples showing configuration tips
- What about more complex scenarios?
- The tool available is subdividing the /19 address block further, maybe to /21s or /22s
  - ALWAYS keep announcing the /19, but use AS-PATH prepend if required
  - Metric has little value unless there is more than one link to the same upstream
  - DO NOT split the /19 into /24s and announce them – /24s contribute to the increasing size of the Internet routing table, and are often filtered by many ISPs

Loadsharing with different ISPs

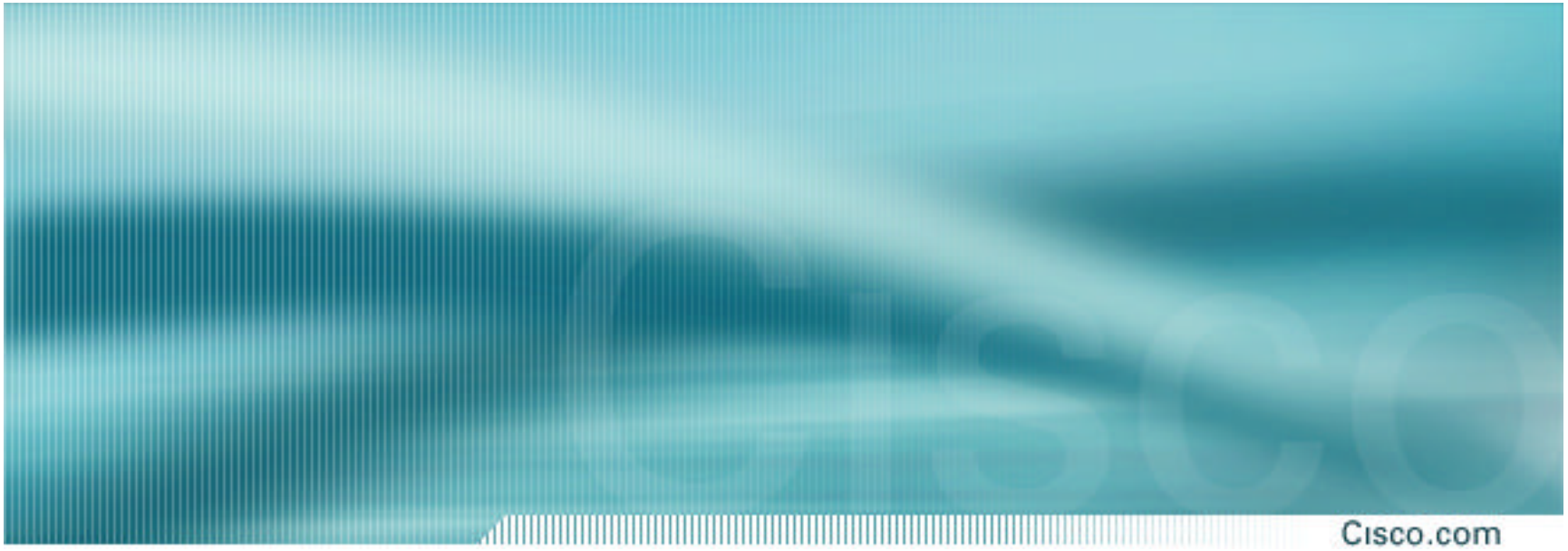
Cisco.com

- **Next section on Service Provider Multihoming looks at more realistic examples**
  - **Details loadsharing both inbound and outbound…**
  - **…and the configuration tips required**

BGP Multihoming Techniques

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- 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**

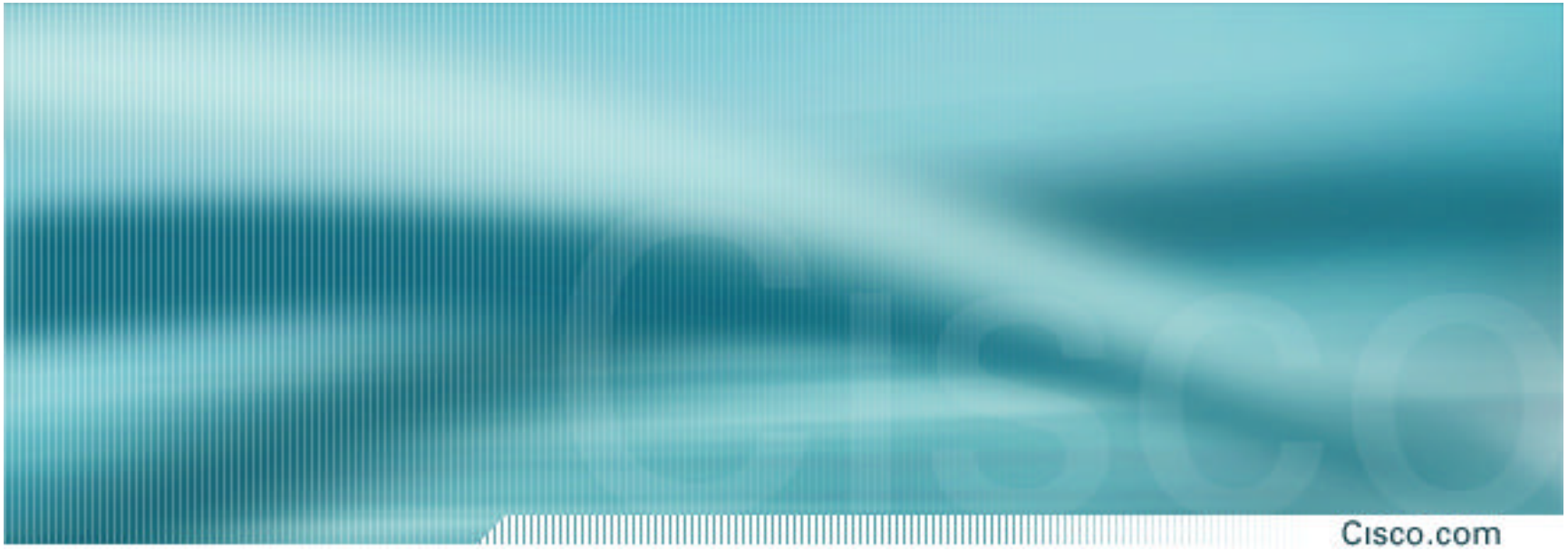
Service Provider Multihoming

- **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

Cisco.com

- **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

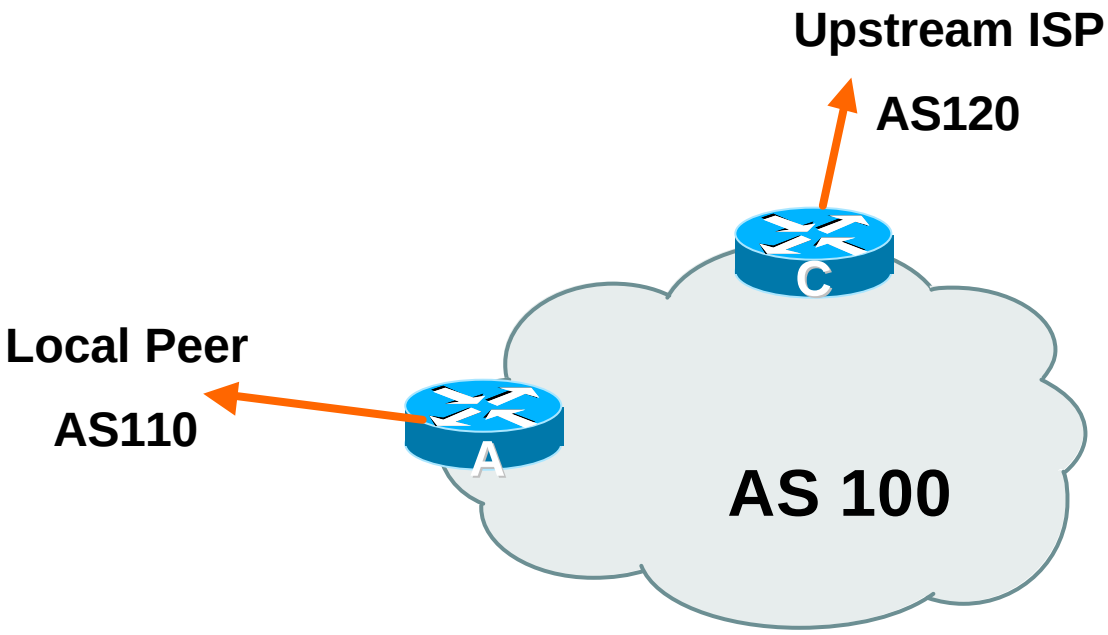
Cisco.com

- 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



One Upstream, One Local Peer

- **Router A Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.2 remote-as 110
 neighbor 222.222.10.2 prefix-list my-block out
 neighbor 222.222.10.2 prefix-list AS110-peer in
!
ip prefix-list AS110-peer permit 222.5.16.0/19
ip prefix-list AS110-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 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.2 remote-as 110
 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 ^(110_)+$
!
ip prefix-list my-block permit 221.10.0.0/19
!
ip route 221.10.0.0 255.255.224.0 null0
```

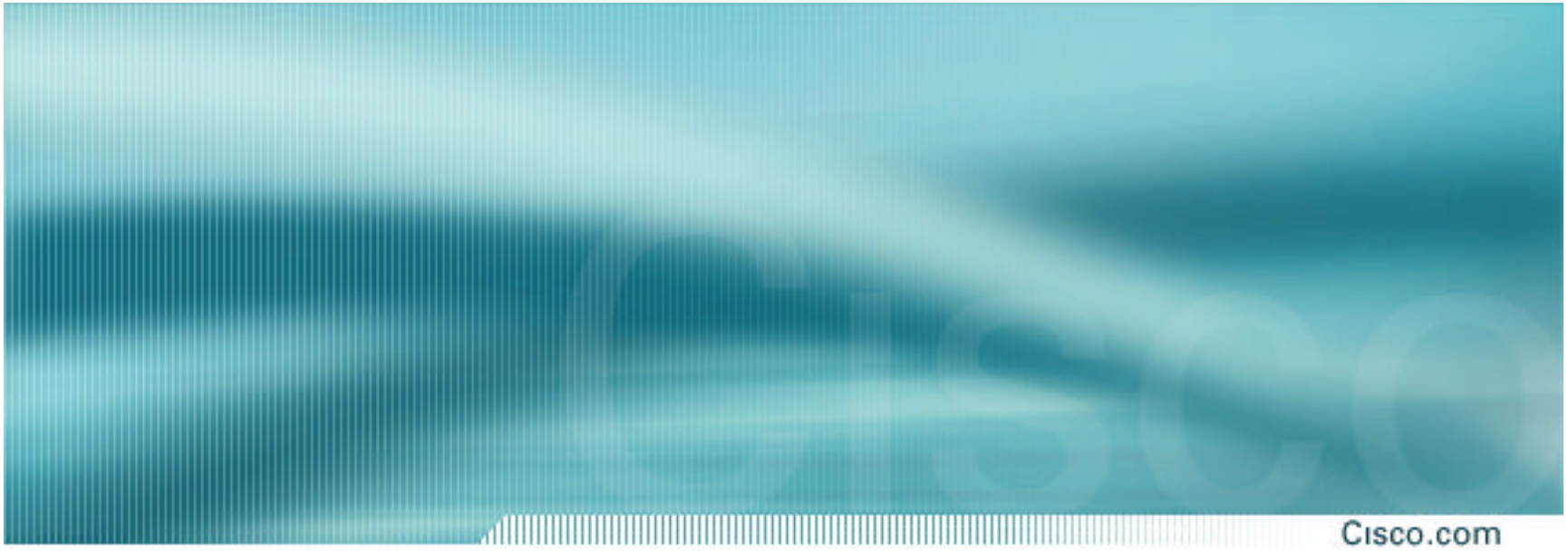
One Upstream, One Local Peer

- **Router C Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 120
 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

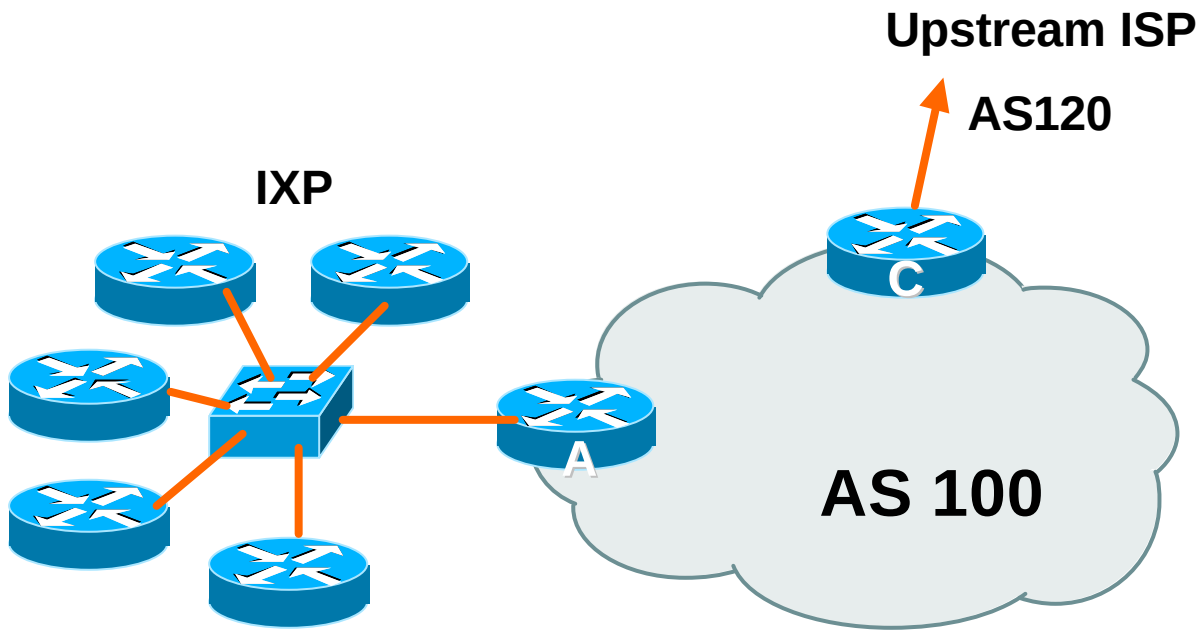
One Upstream, Local Exchange Point

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



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 100
 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

Cisco.com

```
neighbor 220.5.10.2 remote-as 1000
neighbor 222.5.10.2 peer-group ixp-peers
neighbor 222.5.10.2 prefix-list peer1000 in
neighbor 220.5.10.3 remote-as 1010
neighbor 222.5.10.3 peer-group ixp-peers
neighbor 222.5.10.3 prefix-list peer1010 in
neighbor 220.5.10.4 remote-as 1020
neighbor 222.5.10.4 peer-group ixp-peers
neighbor 222.5.10.4 prefix-list peer1020 in
neighbor 220.5.10.5 remote-as 1030
neighbor 222.5.10.5 peer-group ixp-peers
neighbor 222.5.10.5 prefix-list peer1030 in
..next slide
```

One Upstream, Local Exchange Point

Cisco.com

```
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 peer1000 permit 222.0.0.0/19
ip prefix-list peer1010 permit 222.30.0.0/19
ip prefix-list peer1020 permit 222.12.0.0/19
ip prefix-list peer1030 permit 222.18.128.0/19
!
```

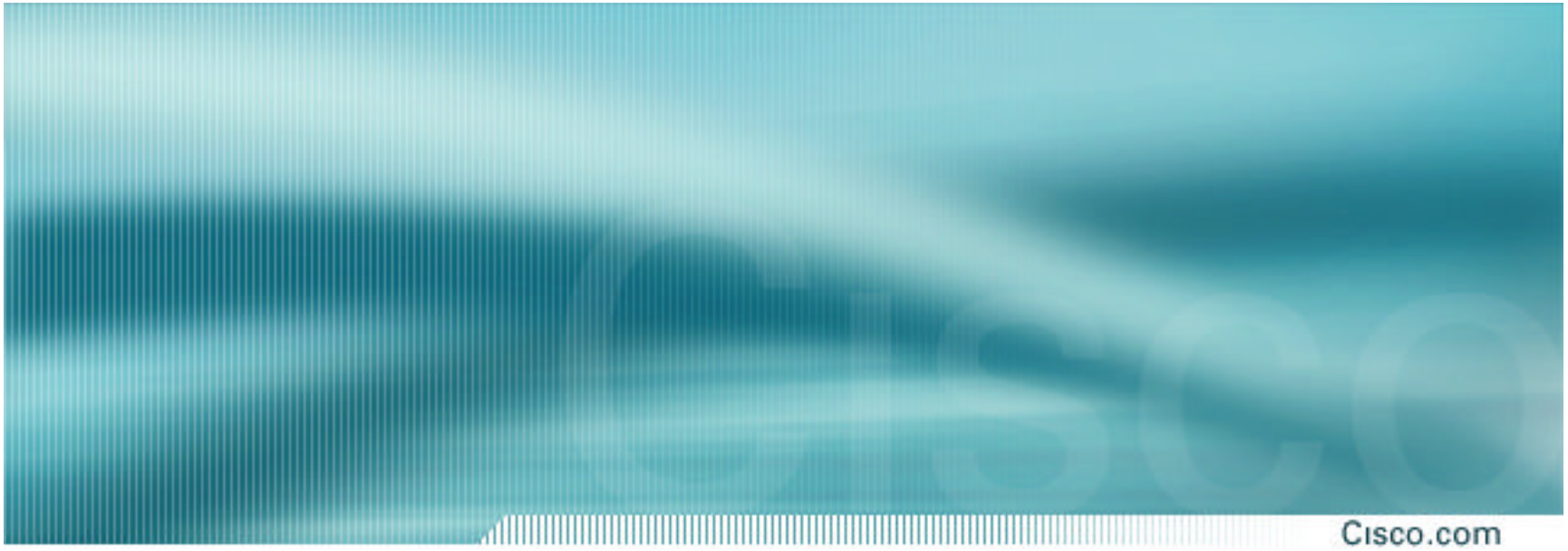
One Upstream, Local Exchange Point

- **Router C Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 120
 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**



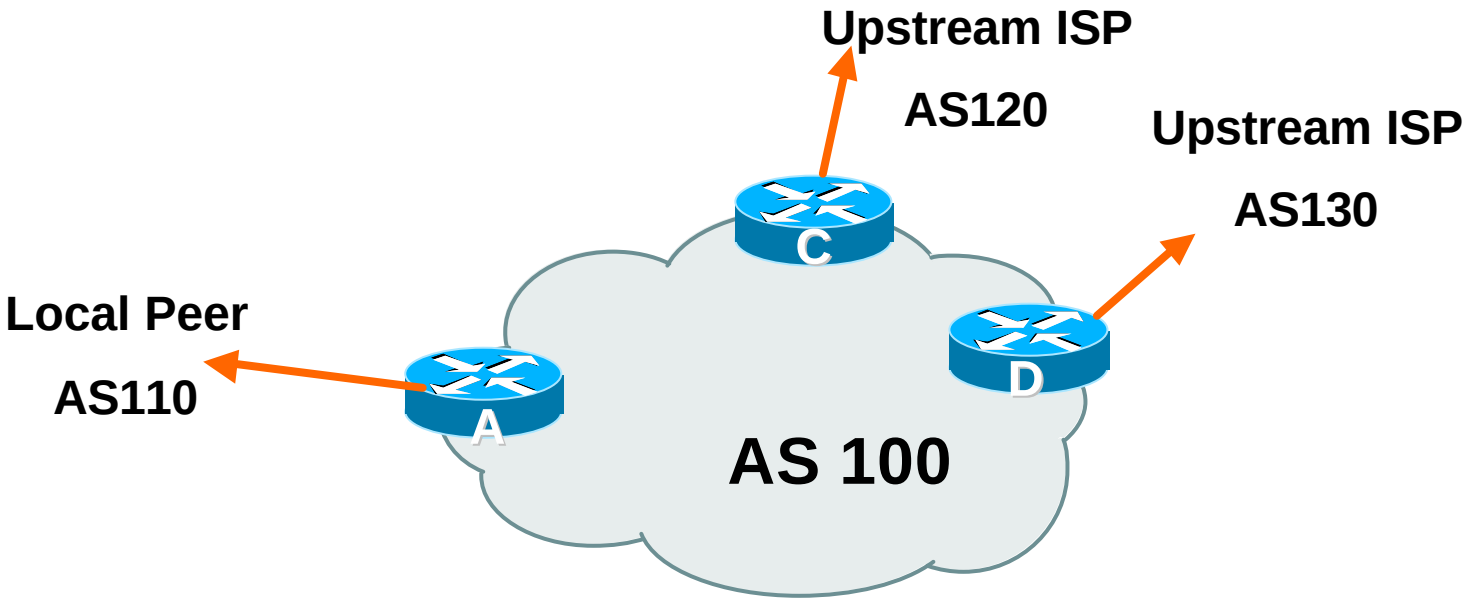
Service Provider Multihoming

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



Two Upstreams, One Local Peer

Cisco.com

- **Router A**
  - **Same routing configuration as in example with one upstream and one local peer**
  - **Same hardware configuration**

Two Upstreams, One Local Peer

Cisco.com

- **Router C Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 120
 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

Cisco.com

- **Router D Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.5 remote-as 130
 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 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 120
 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 AS120-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 ^(120_)+$
ip as-path access-list 10 permit ^(120_)+_[0-9]+$
!
route-map AS120-loadshare permit 10
 match ip as-path 10
 set local-preference 120
route-map AS120-loadshare permit 20
 set local-preference 80
!
```

Two Upstreams, One Local Peer: Full Routes

Cisco.com

- **Router D Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.5 remote-as 130
 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 AS120**
  - **Tag prefixes originated by AS120 and AS120's neighbouring ASes with local preference 120**
    - **Traffic to those ASes will go over AS120 link**
  - **Remaining prefixes tagged with local preference of 80**
    - **Traffic to other all other ASes will go over the link to AS130**
- **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**
    - **Slows convergence rate on local network**
  - **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

Cisco.com

- **Router C Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 120
 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 ^(120_)+$
ip as-path access-list 10 permit ^(120_)+_[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

Cisco.com

- **Router D Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.5 remote-as 130
 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 AS120**
    - **(or get them to send less)**
  - **Filter ASNs so only AS120 and AS120's neighbouring ASes are accepted**
  - **Allow default, and set it to local preference 80**
  - **Traffic to those ASes will go over AS120 link**
  - **Traffic to other all other ASes will go over the link to AS130**
  - **If AS130 link fails, backup via AS120 – and vice-versa**

Two Upstreams, One Local Peer: Partial Routes

- Partial routes from upstreams
  - Not expensive
    - only carry the routes necessary for loadsharing
    - Much quicker convergence in local network
  - 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 100
 default-information originate metric 30
 passive-interface Serial 0/0
!
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 120
 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 ^(120_)+$
ip as-path access-list 10 permit ^(120_)+_[0-9]+$
!
```

Two Upstreams, One Local Peer: Partial Routes

- **Router D Configuration**

```
router ospf 100
 default-information originate metric 10
 passive-interface Serial 0/0
!
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.5 remote-as 130
 neighbor 222.222.10.5 prefix-list deny-all in
 neighbor 222.222.10.5 prefix-list my-block out
!
..next slide
```

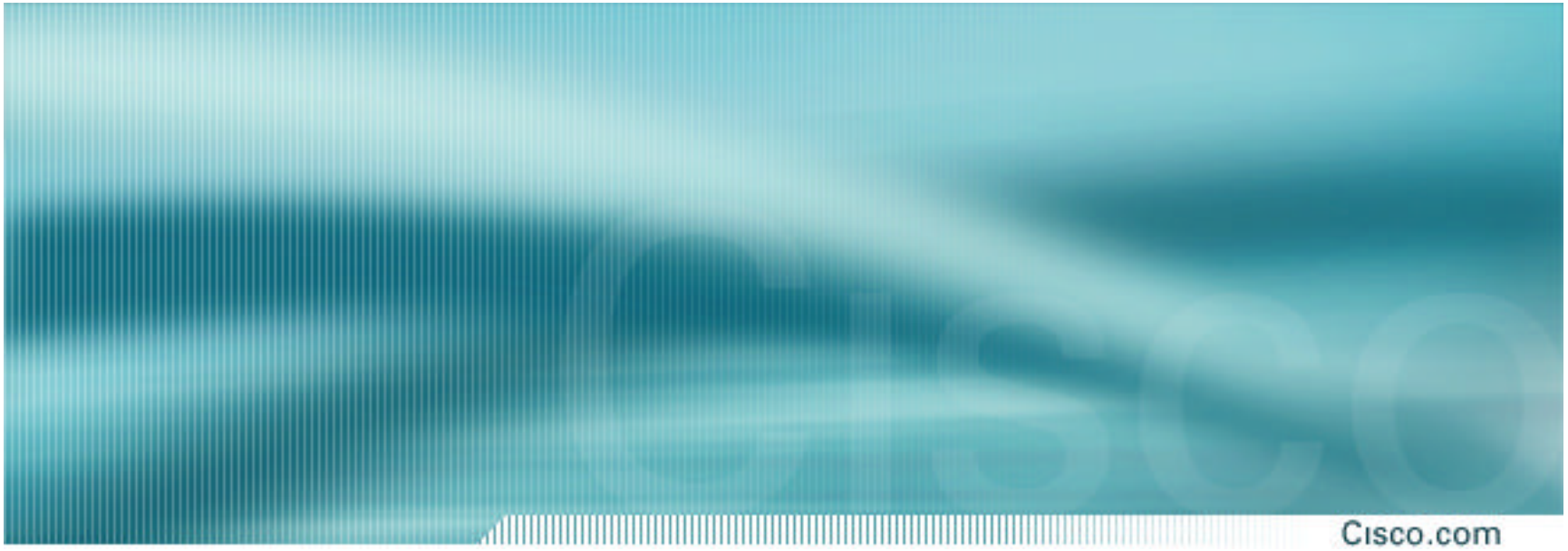
Two Upstreams, One Local Peer: Partial Routes

Cisco.com

```
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**



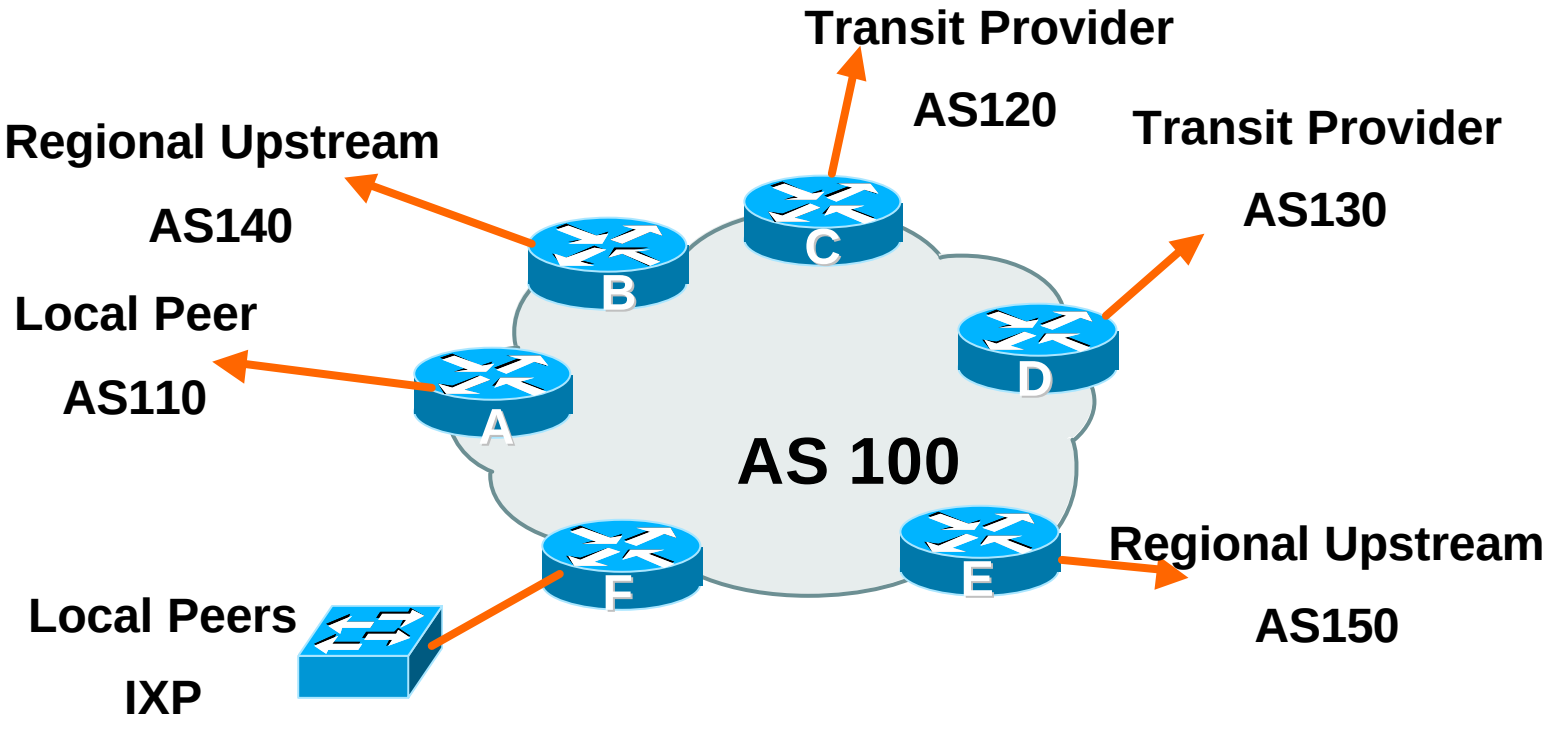
Service Provider Multihoming

Two Transit upstreams, two regional upstreams, and local peers

Transit & 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

Transit & Regional Upstreams, Local Peers



Detail 1

- **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**

Detail 2

- **Router B – regional upstream**
  - **They provide transit to Internet, but longer AS path than the Transit Providers**
  - **Accept all regional routes from them**
    - **e.g. ^140_[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 Transit Provider links go down**

Detail 3

- **Router E – regional upstream**
  - **They provide transit to Internet, but longer AS path than Transit Providers**
  - **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 70**
  - **Will provide backup to Internet only when direct Transit Provider links go down**

Detail 4

- **Router C – first Transit Provider**
  - **Accept all their customer and AS neighbour routes from them**
    - **e.g. ^120_[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 Transit Provider goes down**

Detail 5

- **Router D – second Transit Provider**
  - **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**

Summary

Cisco.com

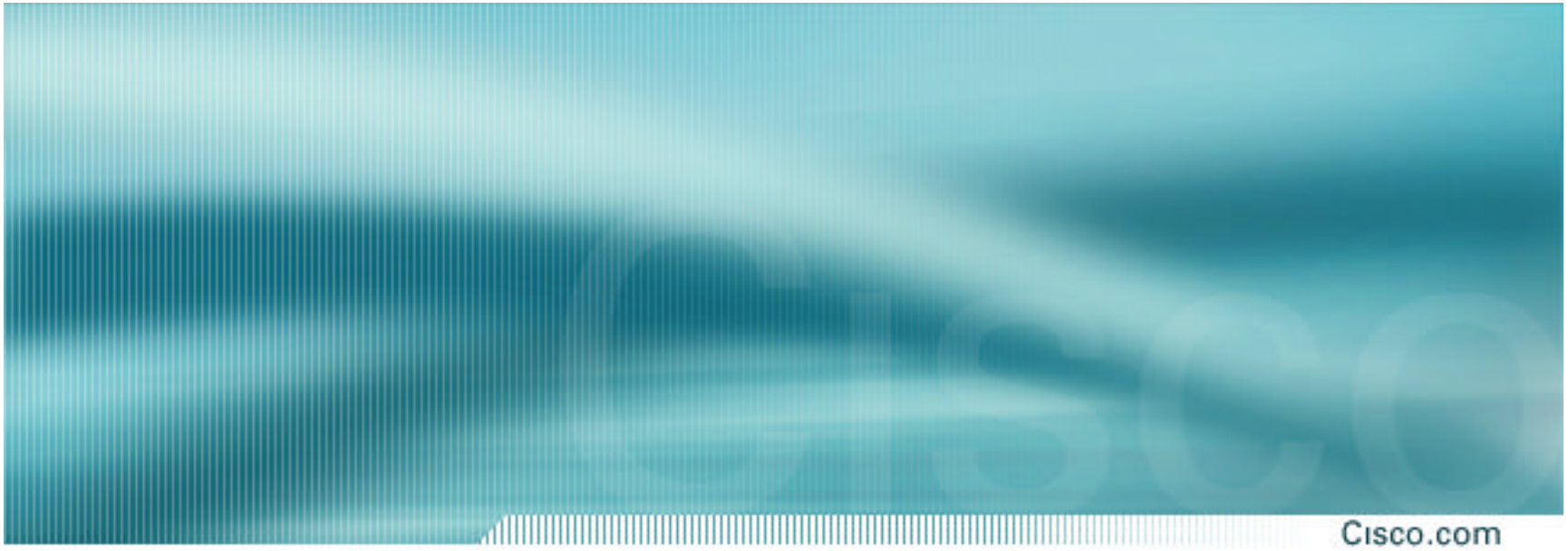
- Local traffic goes to local peer and IXP
- Regional traffic goes to two regional upstreams
- Everything else is shared between the two Transit Providers
- To modify loadsharing tweak what is heard from the two regionals and the first Transit Provider
  - Best way is through modifying the AS-path filter

Outbound?

- **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**

Unequal Circuit Capacity?

- **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

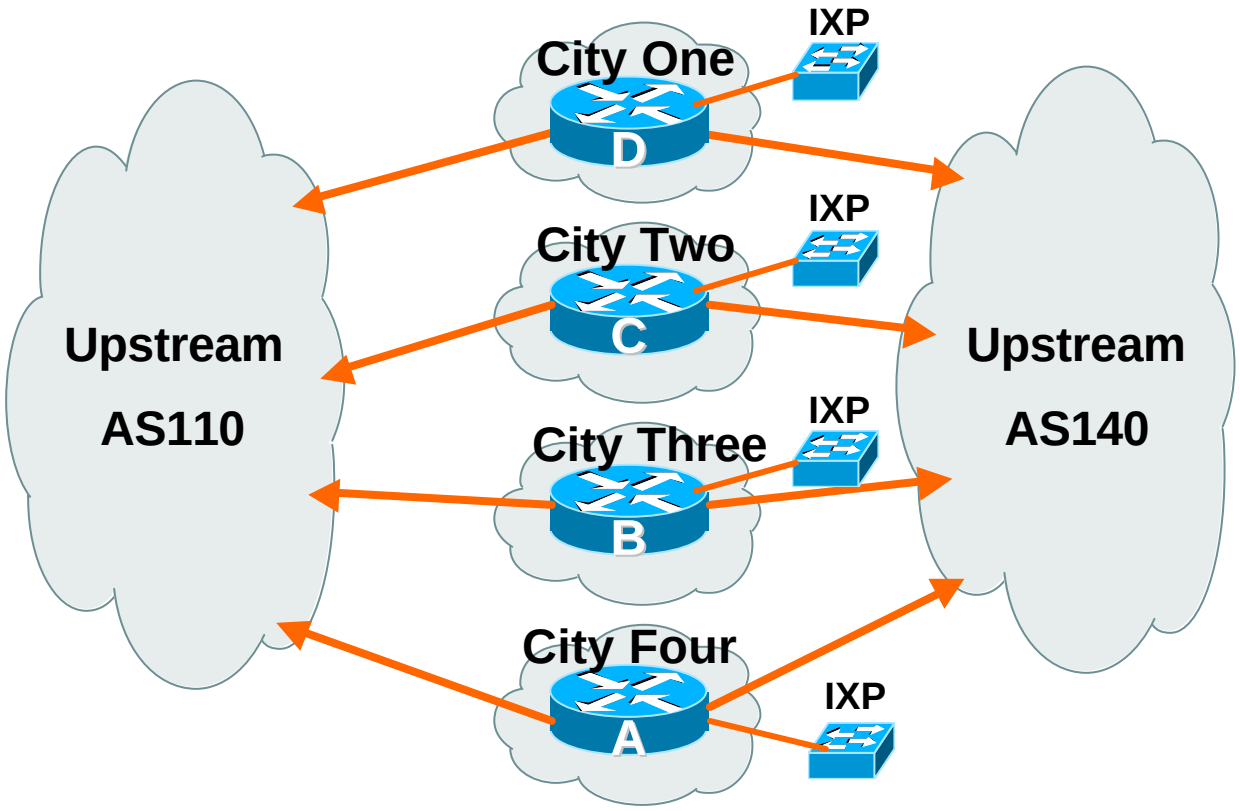
Disconnected Backbone

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



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

Cisco.com

- **Router A Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.248.0
 neighbor 222.200.0.1 remote-as 110
 neighbor 222.200.0.1 description AS110 – 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 140
 neighbor 222.222.10.1 description AS140 – 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

Cisco.com

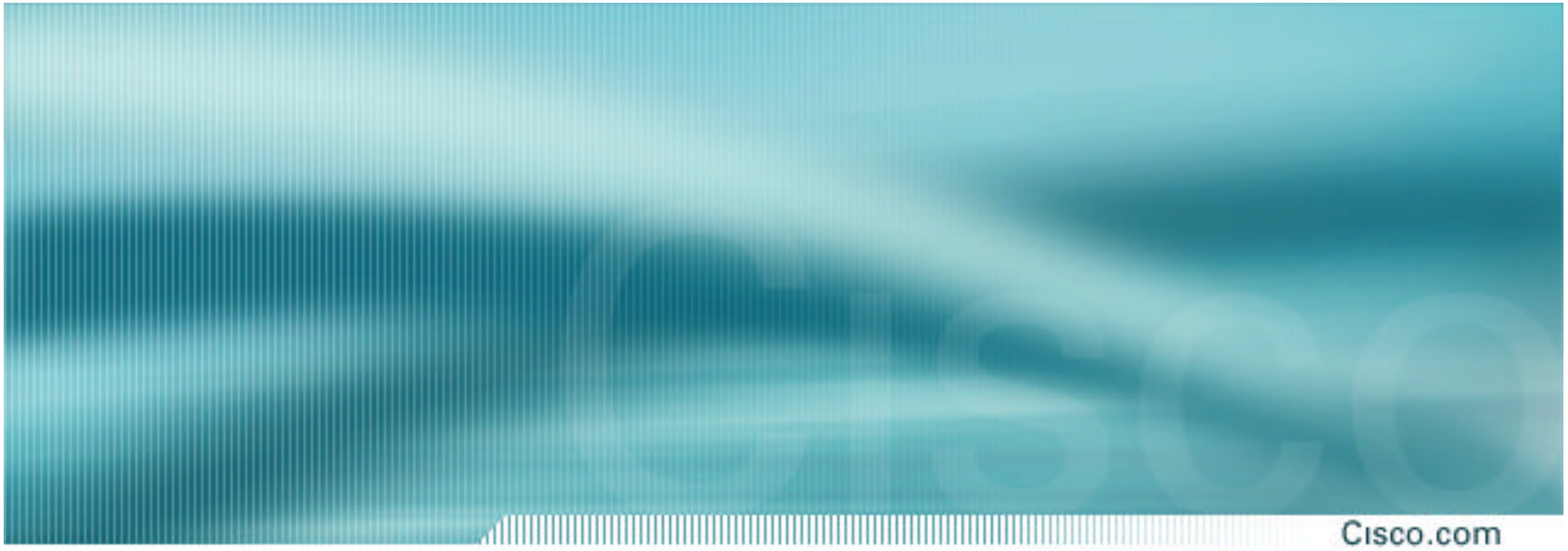
```
ip prefix-list my-block permit 221.10.0.0/21
ip prefix-list default permit 0.0.0.0/0
!
ip as-path access-list 10 permit ^(140_)+$
ip as-path access-list 10 permit ^(140_)+_[0-9]+$
!…etc to achieve outbound loadsharing
!
ip route 0.0.0.0 0.0.0.0 Serial 1/0 250
ip route 221.10.0.0 255.255.248.0 null0
!
```

Disconnected Backbone

- **Peer with AS110**
  - **Receive just default route**
  - **Announce /22 address**
- **Peer with AS140**
  - **Receive full routing table – filter with AS-path filter**
  - **Announce /22 address**
  - **Point backup static default – distance 252 – in case AS110 goes down**

Disconnected Backbone

- **Default ensures that disconnected parts of AS100 are reachable**
  - **Static route backs up AS110 default**
  - **No BGP loop detection – relying on default route**
- **Do not announce /19 aggregate**
  - **No advantage in announcing /19 and could lead to problems**

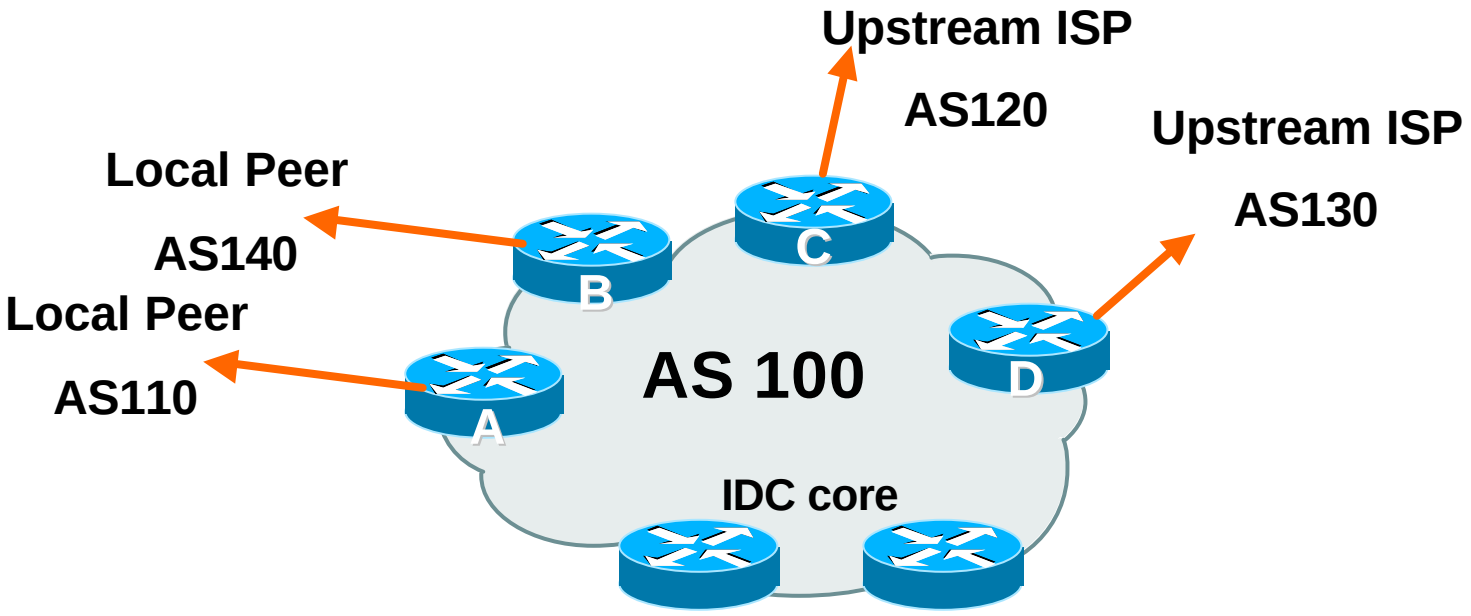


IDC Multihoming

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

IDC Multihoming



Assigned /24 from AS120 and /23 from AS130.

Circuit to AS120 is 2Mbps, circuit to AS130 is 4Mbps

IDC Multihoming

- **Router A and B configuration**

  In: Should accept all routes from AS110 and AS140

  Out: Should announce all address space to AS110 and AS140

  Straightforward

IDC Multihoming

Cisco.com

- **Router C configuration**
  - **In: Accept partial routes from AS120**
    e.g. ^120_[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

Cisco.com

- **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

Cisco.com

- **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

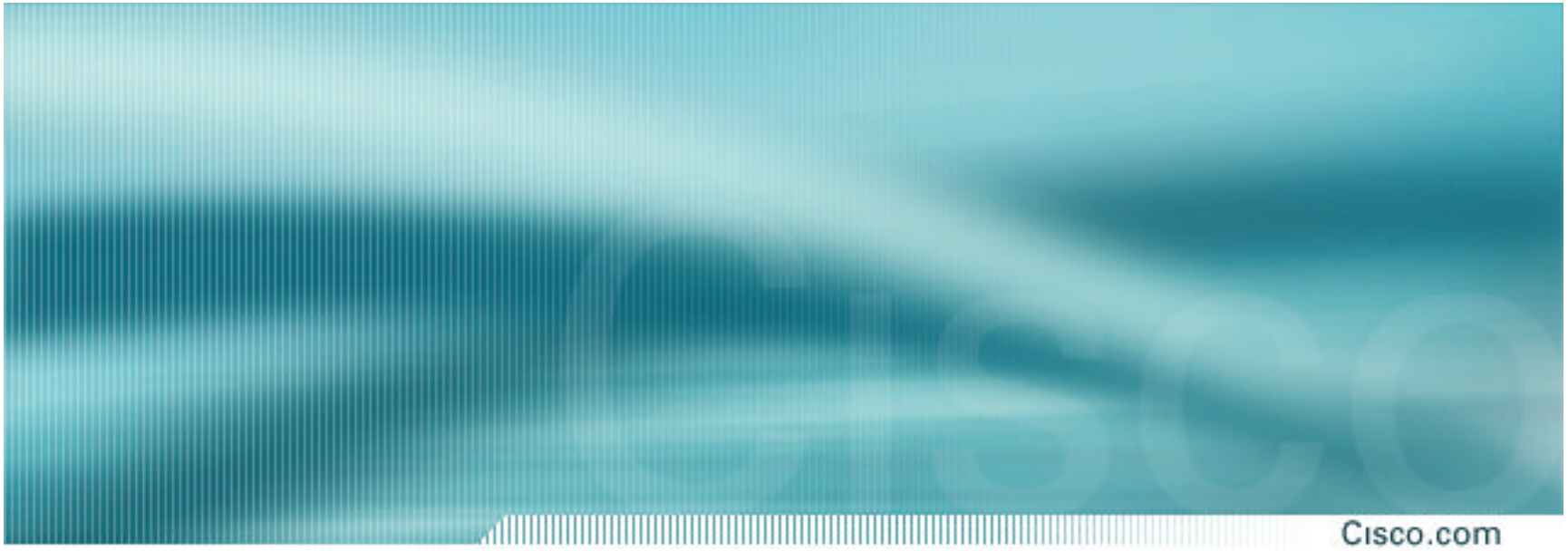
Cisco.com

- **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

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- Using Communities
- Case Study



Internet Exchange Points

Keeping Local Traffic Local

Internet Exchange Point

- IXP is designed to keep local traffic local
- Common Interconnect Point, usually an ethernet switch in a co-location facility
- ISPs bring one router, provide connections from their backbone to the router
- The ISP router connects to the ethernet and peers with other ISP routers at the exchange

Why an Internet Exchange Point?

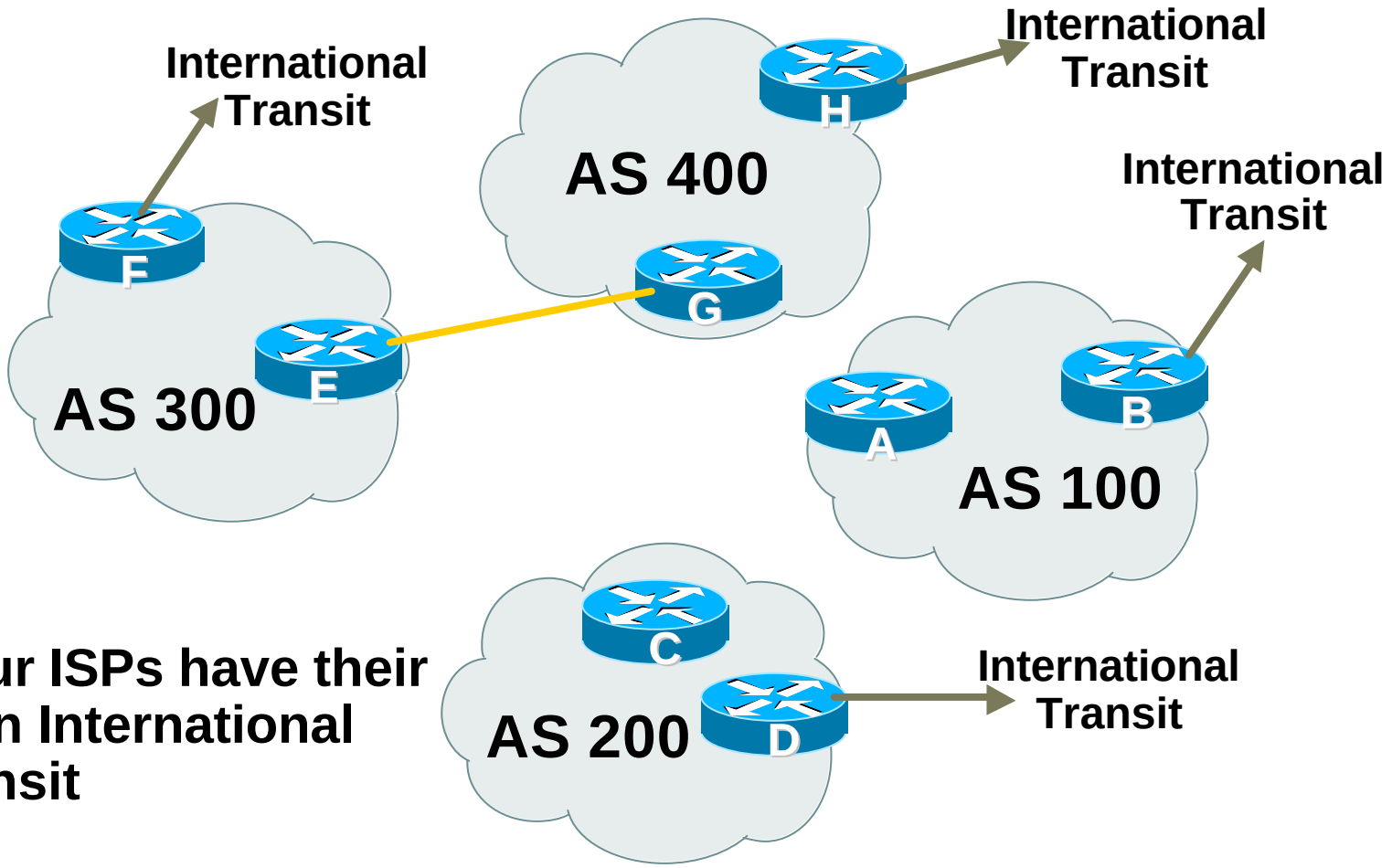
- **Keep local traffic local**
  - **International circuits cost $$$ compared with domestic circuits**
  - **Round trip times substantially reduced for local traffic**
- **Save money**
  - **Why pay overseas ISPs and cable companies for traffic which should stay in country?**
  - **Save yourself money, save your country money, save your customers money**
  - **Customers with more money to spend will buy other services from you ® you become more successful**
- **Help develop your country's Internet economy**

IXP Example

Cisco.com

- **Four ISPs in one country**
  - **Each has own international transit**
  - **Only two ISPs have a connection to each other**
  - **The other two access the network of the others via their international links**

Internet Exchange Point – Example



Four ISPs have their own International transit

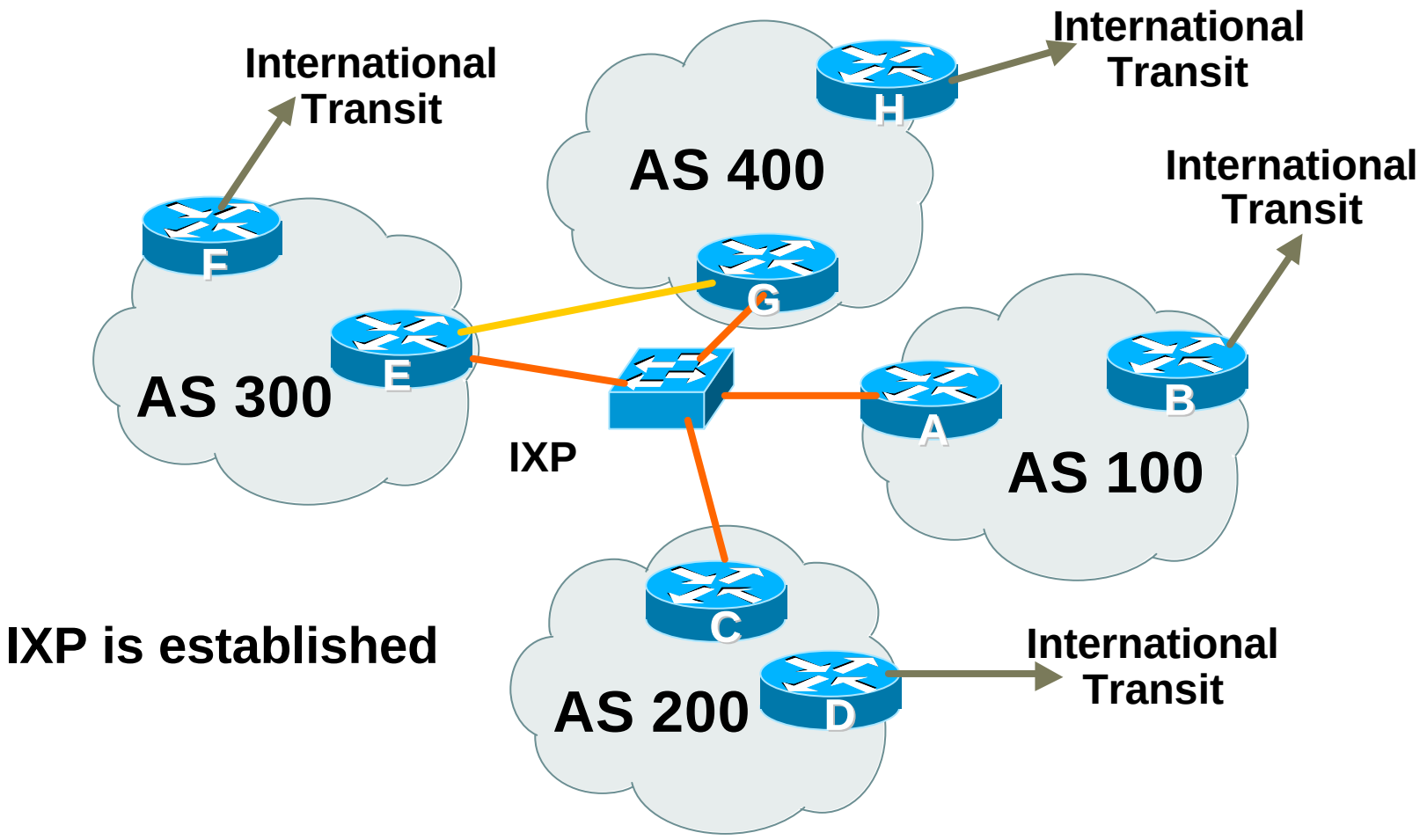
Internet Exchange Point – Example

- Traffic from AS100 to the other networks goes via their international transit
- Traffic from AS200 to the other networks goes via their international transit
- Traffic from AS300 to AS400 goes over the local link – the rest goes over their international transit
- Traffic from AS400 to AS300 goes over the local link – the rest goes over their international transit
- Result:

  AS100 and AS200 are wasting money on international links

  AS300 and AS400 save some money, but could to better

Internet Exchange Point – Example



IXP is established

Post IXP installation

- **AS300 and AS400 retain private interconnect**
  - **“backup” for the IXP**
  - **They could choose to discontinue it**
- **Local traffic between each of the networks stays local – it only crosses the IXP fabric**
- **Configuration is not hard**

Internet 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 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor ixp-peers peer-group
 neighbor ixp-peers prefix-list my-block out
..next slide
```

Internet Exchange Point

Cisco.com

```
neighbor 220.5.10.2 remote-as 200
neighbor 220.5.10.2 description AS200 ISP
neighbor 222.5.10.2 peer-group ixp-peers
neighbor 222.5.10.2 prefix-list peer200 in
neighbor 220.5.10.3 remote-as 300
neighbor 220.5.10.3 description AS300 ISP
neighbor 222.5.10.3 peer-group ixp-peers
neighbor 222.5.10.3 prefix-list peer300 in
neighbor 220.5.10.4 remote-as 400
neighbor 220.5.10.4 description AS400 ISP
neighbor 222.5.10.4 peer-group ixp-peers
neighbor 222.5.10.4 prefix-list peer400 in
..next slide
```

Internet 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 peer200 permit 222.0.0.0/19
ip prefix-list peer300 permit 222.30.0.0/19
ip prefix-list peer400 permit 222.12.0.0/19
!
```

Internet Exchange Point

- **Router B Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 2000
 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
```

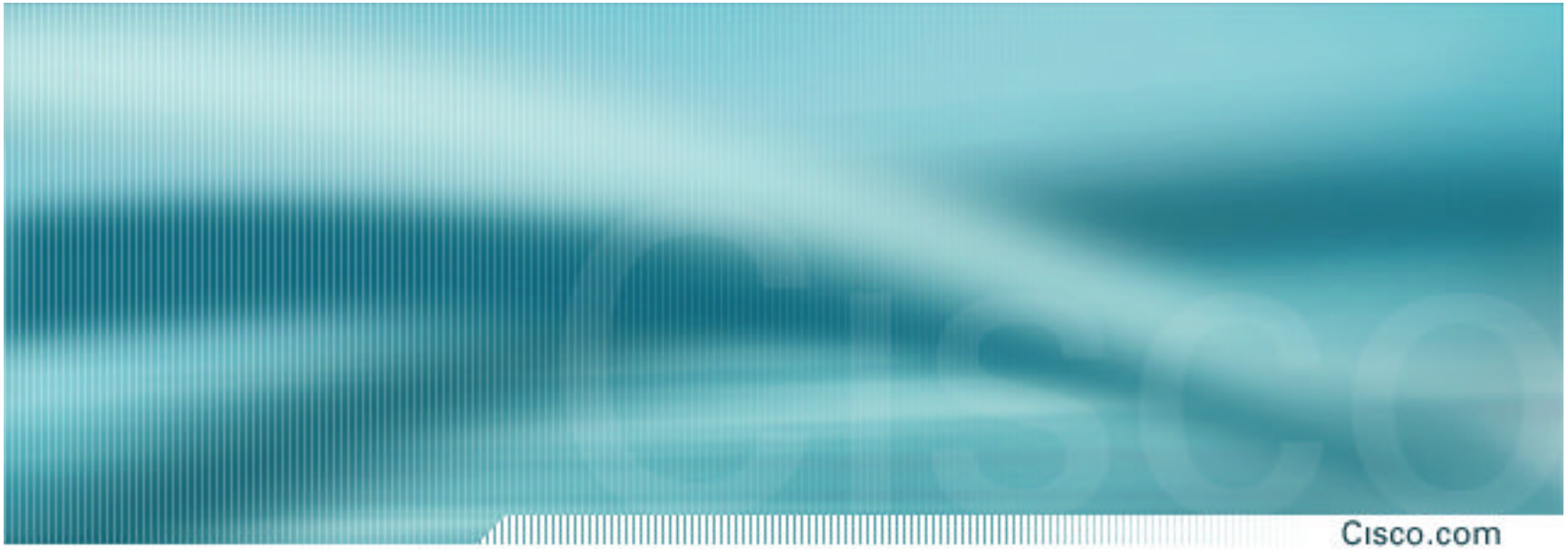
Internet Exchange Point

- **Router A configuration**
  - **Prefix-list higher maintenance, but safer**
  - **uRPF on the FastEthernet interface**
- **Router B configuration**
  - **Standard BGP transit provider upstream – accept default, send them just the local aggregate**
- **IXP traffic goes to and from local IXP, everything else goes to upstream**
- **Not hard to configure, not hard to set up, not hard to run**
  - **Benefits are almost immediately tangible**

BGP Multihoming Techniques

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- Using Communities
- Case Study



Communities

Community usage

Cisco.com

- **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 120
 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
```

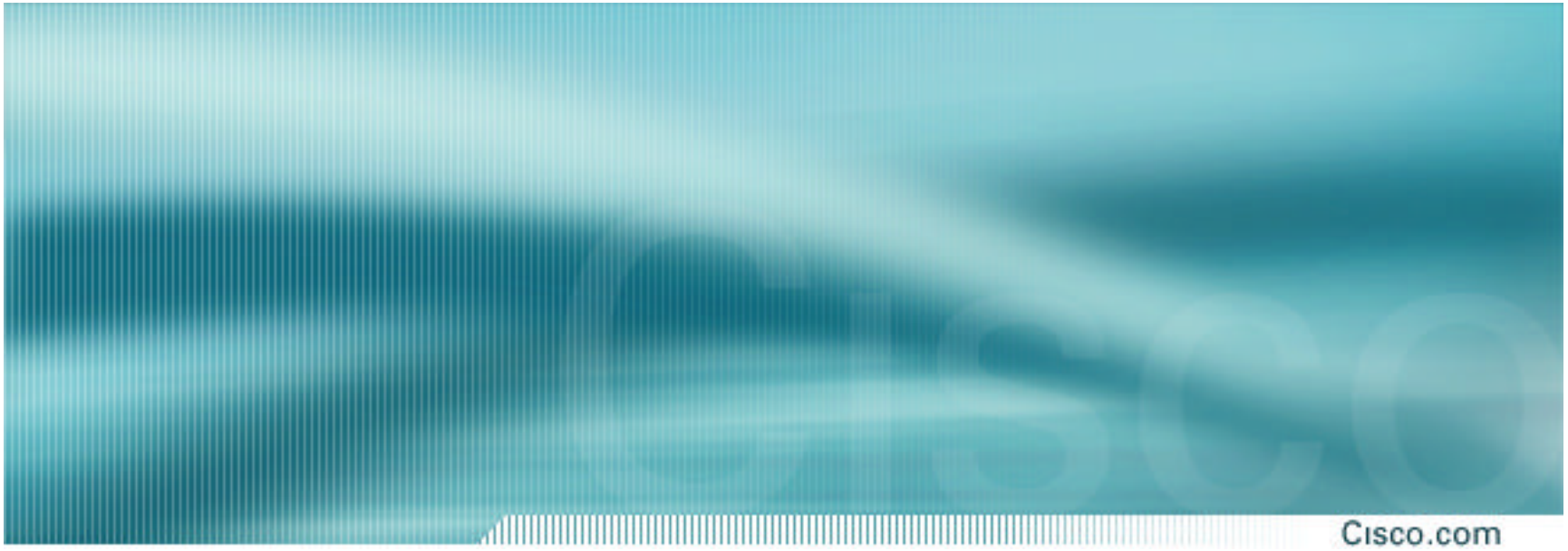
RFC1998

- **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

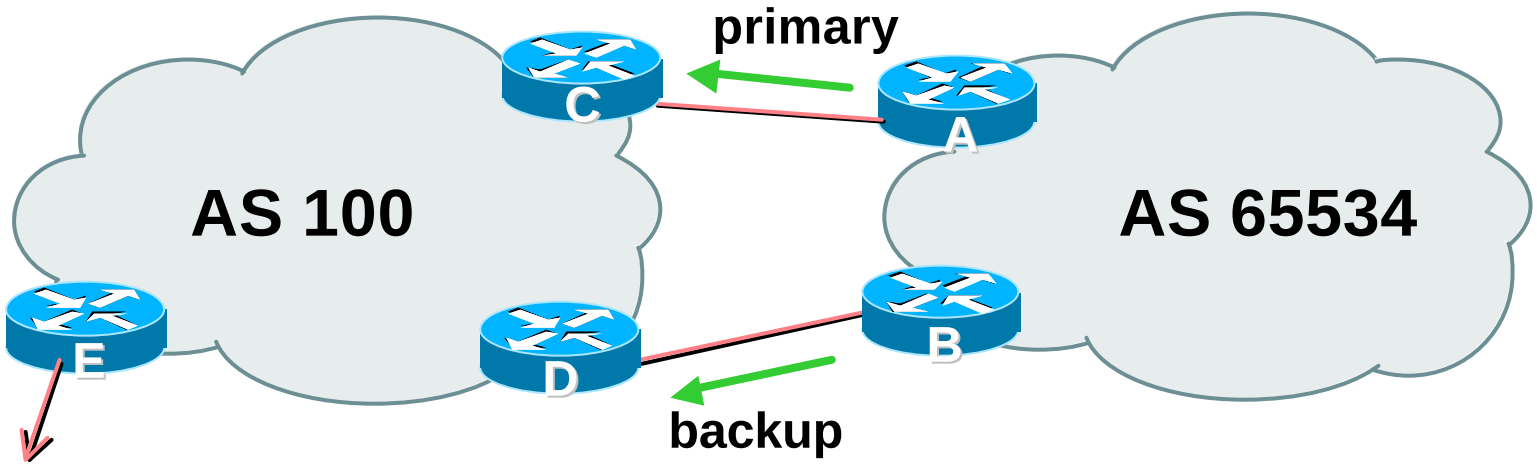
- **Supporting RFC1998**
  - **many ISPs do, more should**
  - **check AS object in the Internet Routing Registry**
  - **if you do, insert comment in AS object in the IRR**



Two links to the same ISP

One link primary, the other link backup only

Two links to the same ISP



- **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)

Cisco.com

```
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)

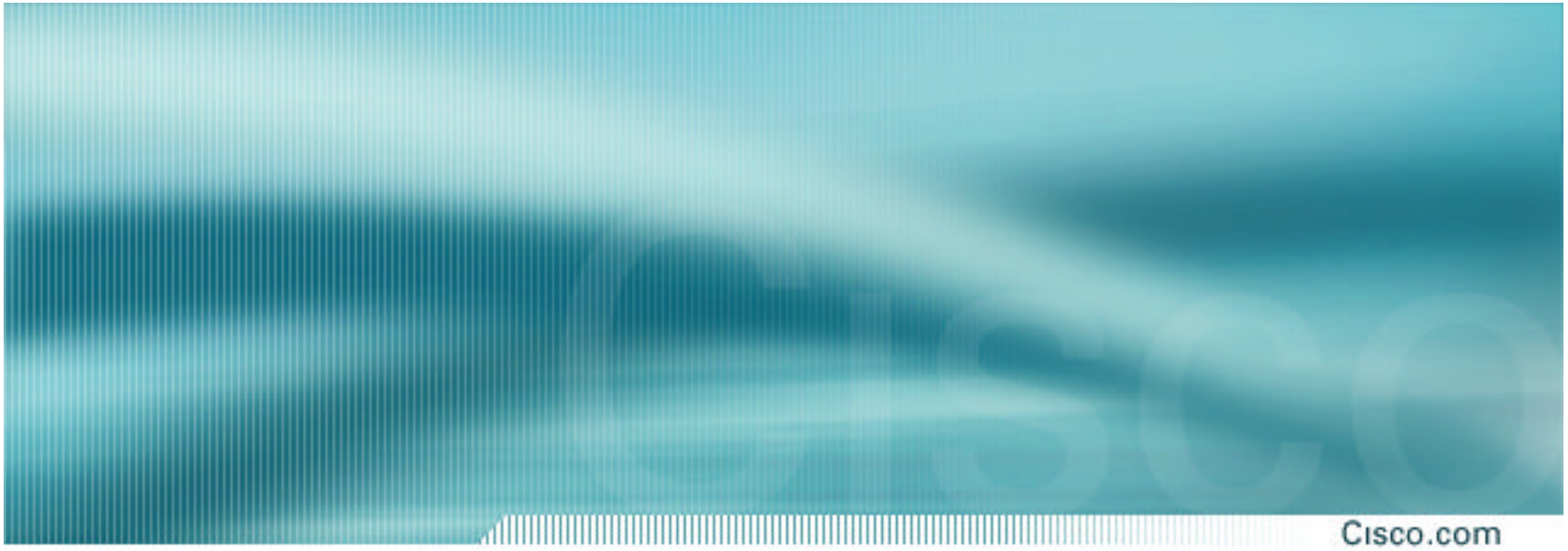
- **Router D Configuration (backup link)**

```
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)

Cisco.com

```
   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

Cisco.com

- **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

Cisco.com

```
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
```

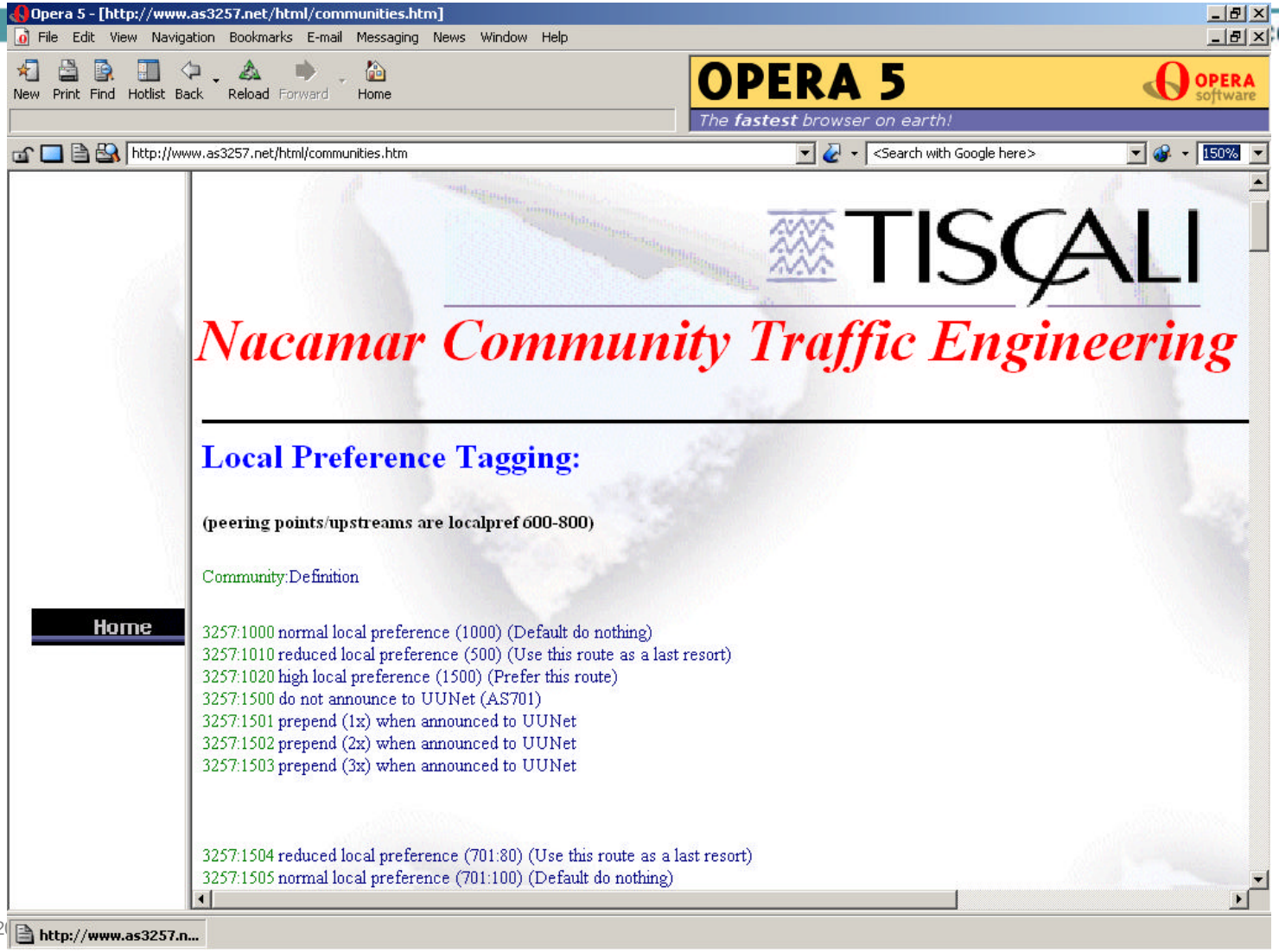
Some ISP Examples
UUNET Europe


```
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 Examples
Concert Europe


```
aut-num:      AS5400
as-name:      CIPCORE
descr:        Concert European Core Network
remarks:      Communities scheme:
remarks:      The following BGP communities can be set by Concert BGP
remarks:      customers to affect announcements to major peerings.
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         Ebone (AS1755)          5400:2001
remarks:      5400:1002         Eunet (AS286)           5400:2002
remarks:      5400:1003         Unisource (AS3300)      5400:2003
<snip>
remarks:      5400:1100         US peers                5400:2100
notify:       peertech@concert.net
mnt-by:       CIP-MNT
source:       RIPE
```

Some ISP Examples
Tiscali/Nacamar



ISP Examples

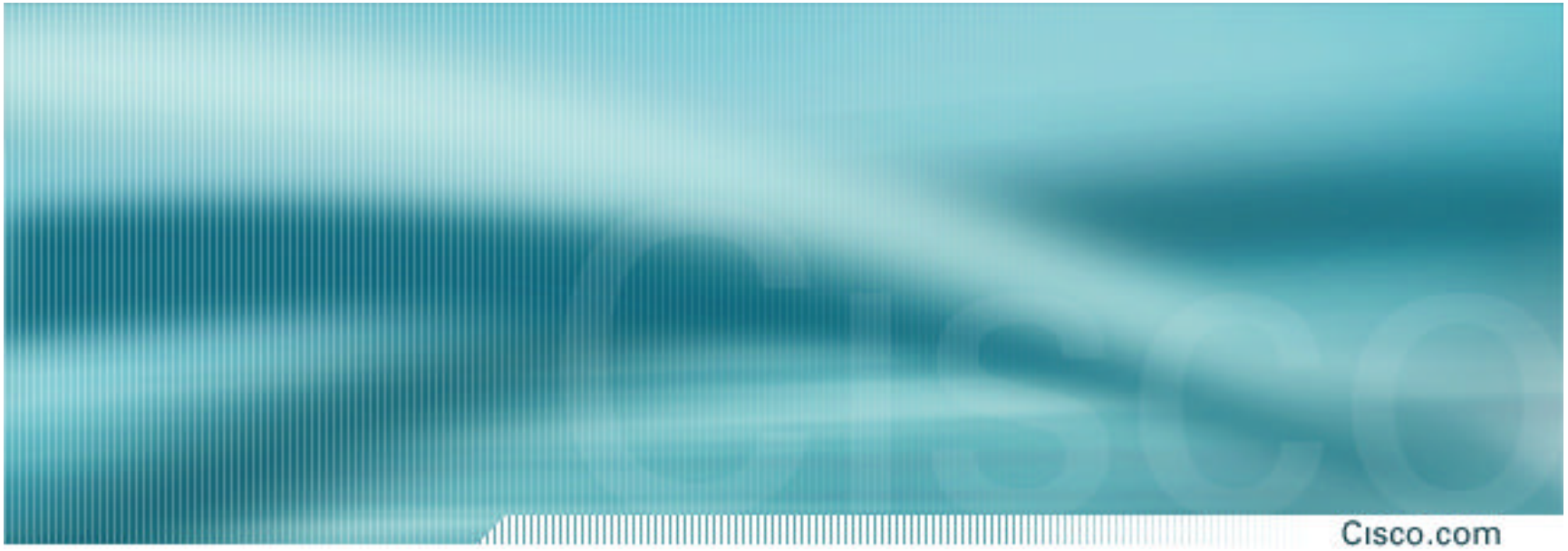
Cisco.com

- Several more…
- Tiscali is very detailed
  - Concept used by others, such as Concert
  - Includes IOS configuration examples
- Many ISP support communities for multihoming preferences

BGP Multihoming Techniques

Cisco.com

- Preparations
- Connecting to the same ISP
- Connecting to different ISPs
- Service Provider Multihoming
- Internet Exchange Points
- Using Communities
- Case Study



Case Study

Case Study – Requirements (1)

Cisco.com

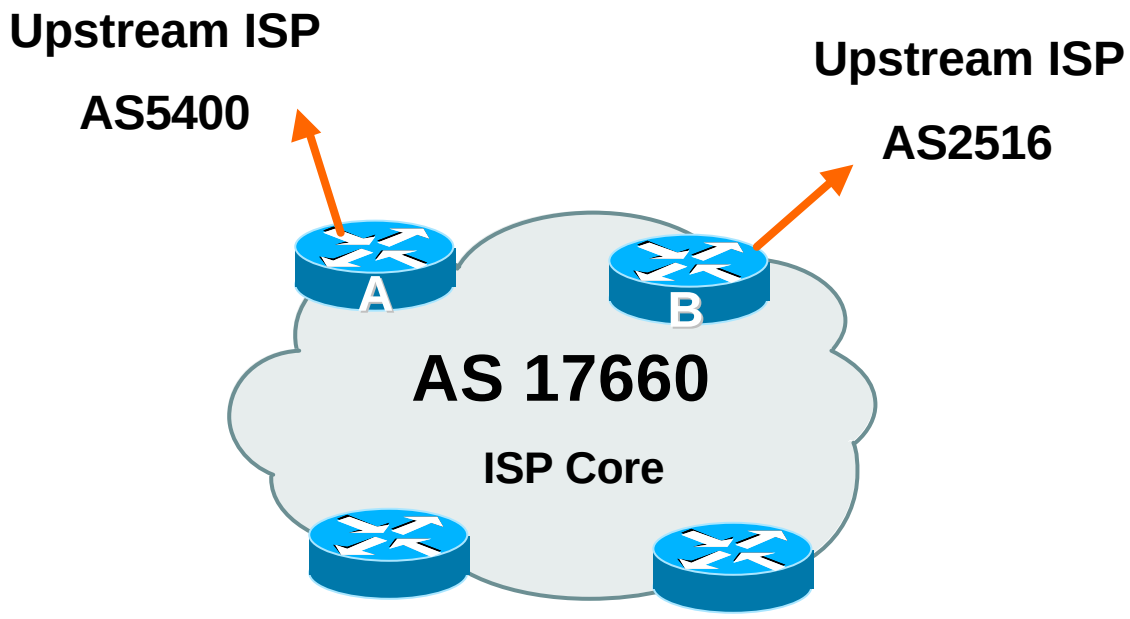
- **ISP needs to multihome:**
  - **To AS5400 in Europe**
  - **To AS2516 in Japan**
  - **/19 allocated by APNIC**
  - **AS 17660 assigned by APNIC**
  - **1Mbps circuits to both upstreams**

Case Study – Requirements (2)

Cisco.com

- **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



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
Configuration Router A

Cisco.com

```
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

Cisco.com

```
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

Cisco.com

```
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

Cisco.com

```
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

Cisco.com

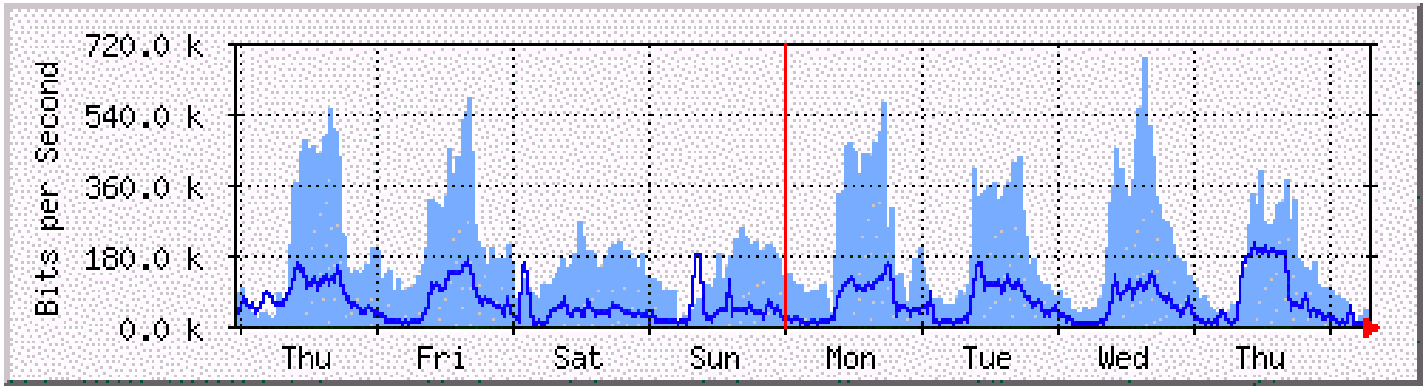
```
!
prefix-list default-route permit 0.0.0.0/0
prefix-list out-filter permit 202.144.128.0/19
!
ip as-path access-list 2 permit _2516$
ip as-path access-list 2 deny .*
ip as-path access-list 3 permit ^$
!
route-map as2516-out permit 10
 set as-path prepend 17660
!
```

Configuration Summary

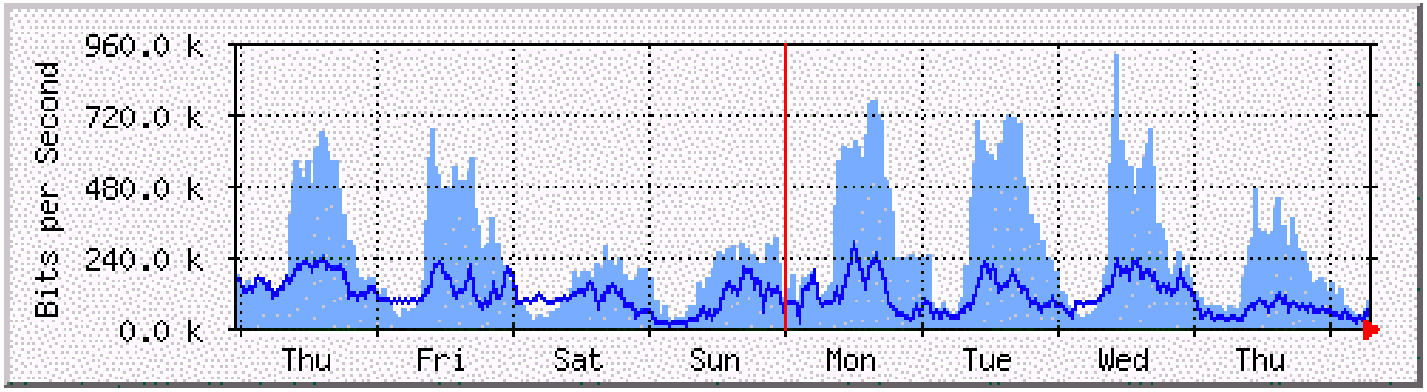
Cisco.com

- **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
MRTG Graphs



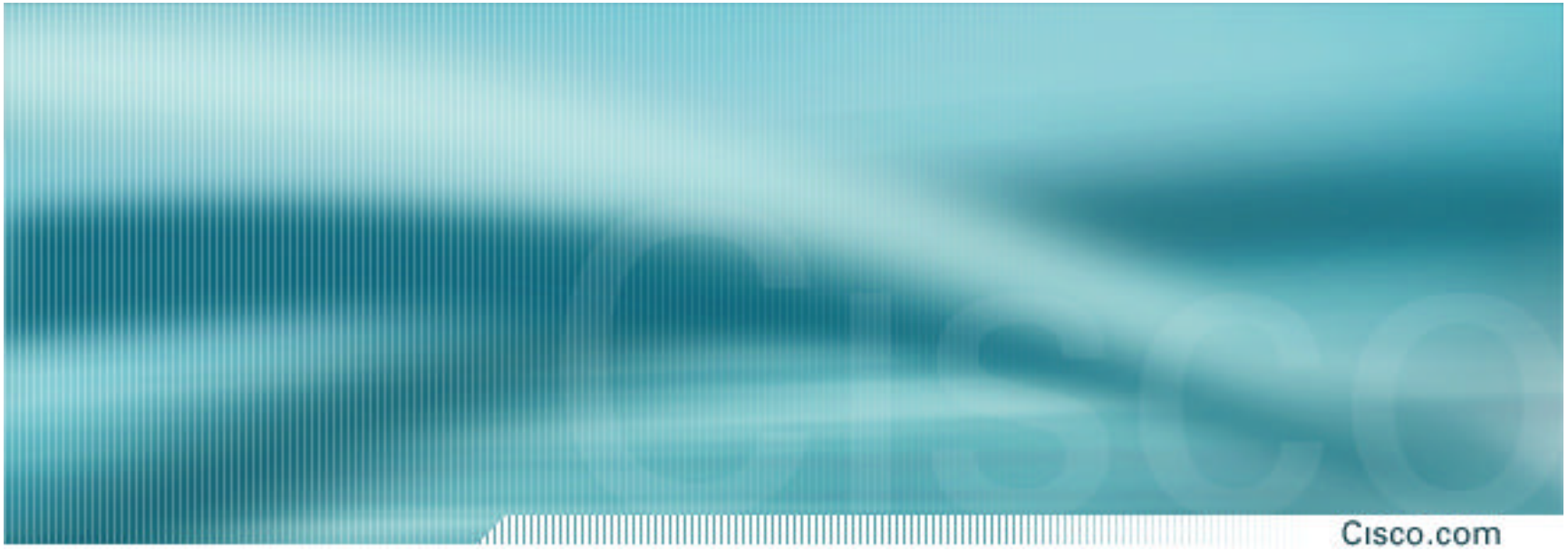
Router A to AS5400



Router B to AS2516

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…**

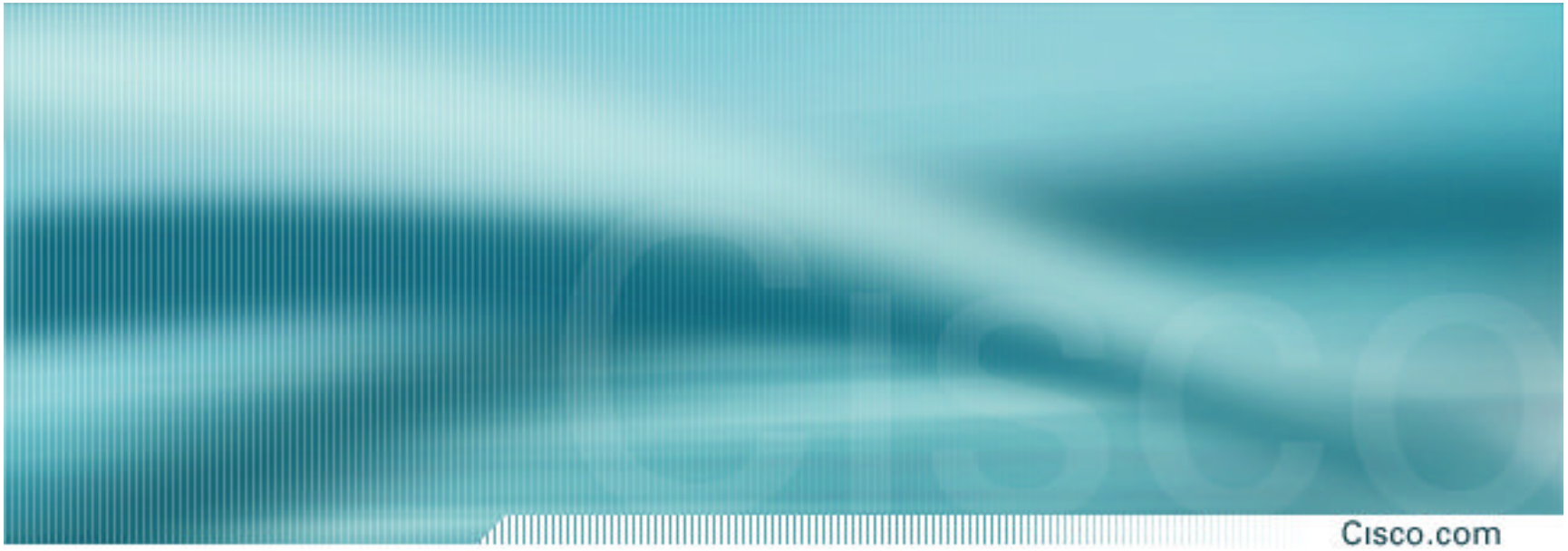


Summary

Summary

Cisco.com

- **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 109k prefixes, charge them money for it**



BGP Multihoming Techniques

End of Tutorial