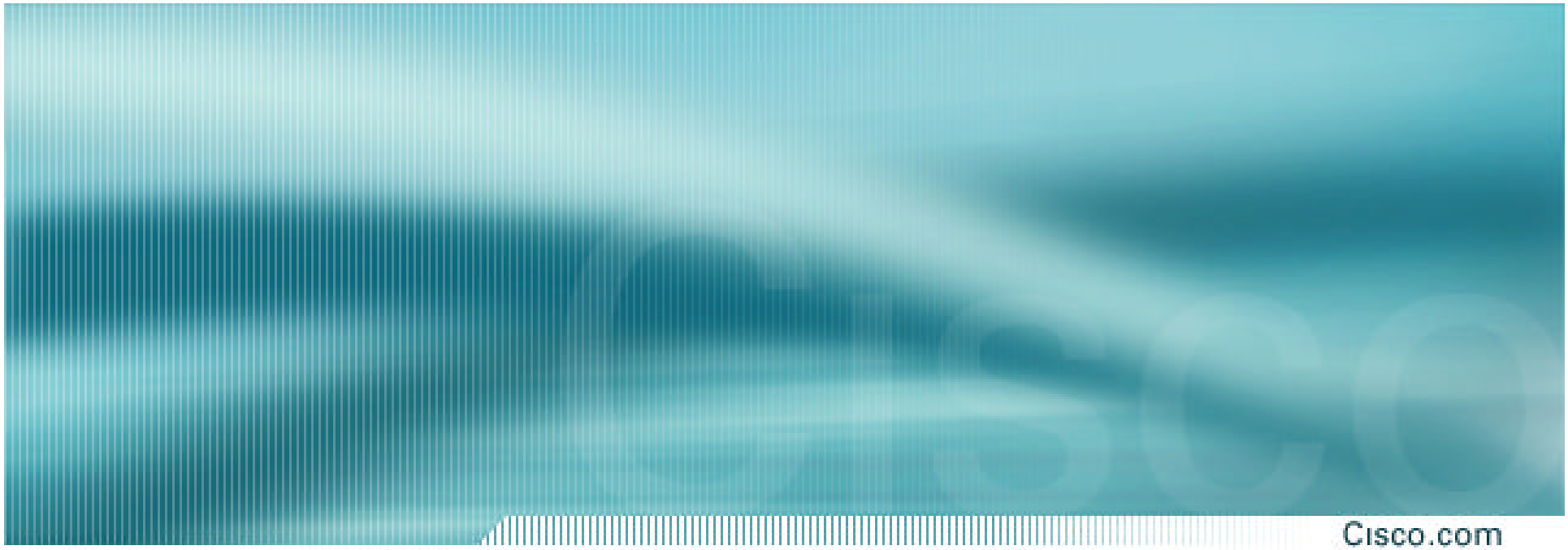
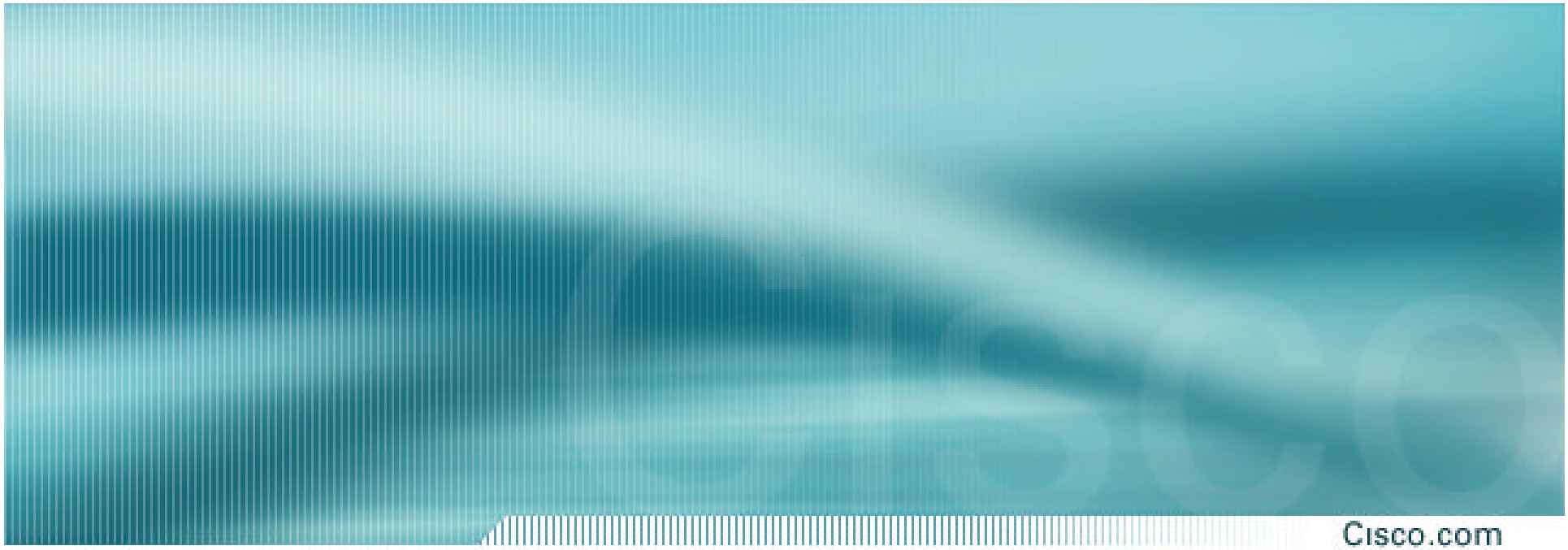


BGP in the Internet

Best Current Practices



Recommended IOS Releases

Which IOS??

Which IOS?

- IOS is a feature rich and highly complex router control system
- ISPs should choose the IOS variant which is most appropriate for the intended application
- There is an exclusive service provider train in IOS
  - This is 12.0S, supporting 7200, 7500, 10000 and 12000
  - Images also available for 2500, 2600, 3600 and 4500, but are completely unsupported
- There is a service provider image in most IOS releases
  - This is the image with –p– in its name, for example:
    - c7200-p-mz.122-8.T1 and c2600-p-mz.121-14
  - The –p– image is IP-only plus ISIS/CLNS

Which IOS?

Cisco.com

- **12.*n* – for example 12.2**
  - **This means the IOS is a mainline image**
  - **NO new features**
  - **ONLY bug fixes**
  - **The aim is stability!**
- **12.*n*T – for example 12.2T**
  - **This means the IOS is the technology release**
  - **NEW features**
  - **Bug fixes**
  - **Avoid unless you need the feature!**

12.0 IOS release images

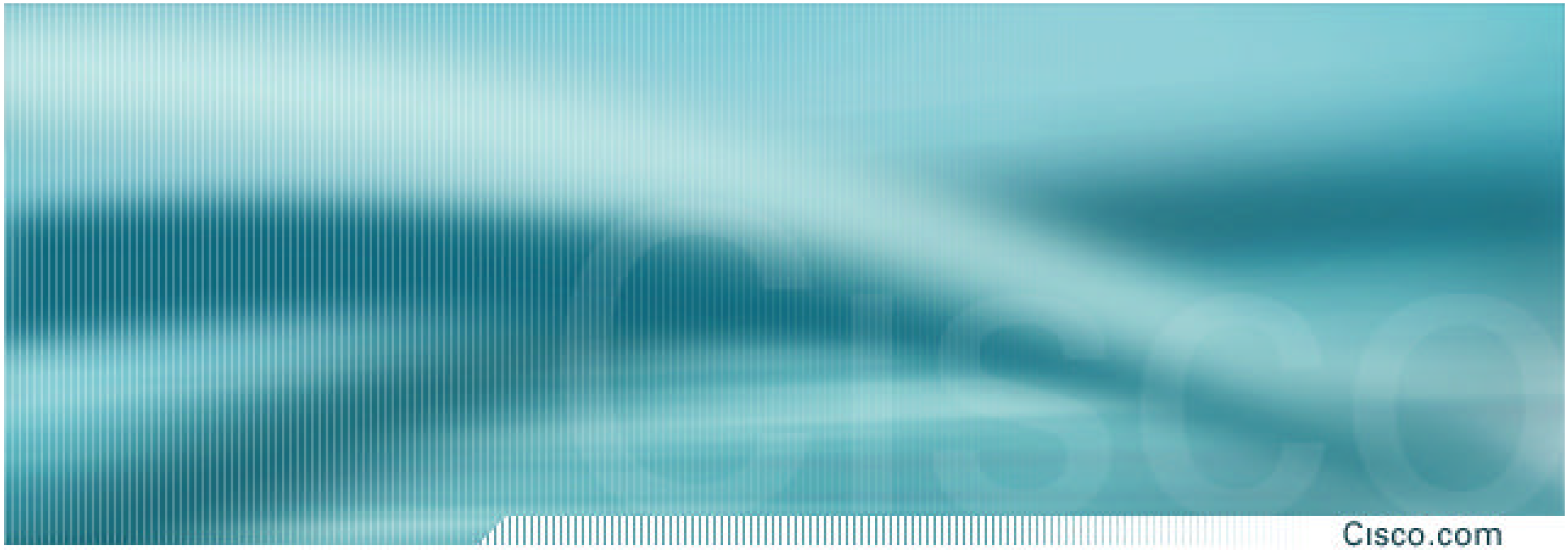
- **12.0S is the release for all ISPs**
  - **for 7200, 7500, 10000 and GSR/12000**
  - **replaces 11.1CC and 11.2GS**
  - **currently at 12.0(23)S1**
- **12.0 is the “mainline” train**
  - **for all other older platforms**
  - **replaces 11.2P and 11.3T**
  - **currently at 12.0(25)**
- **Available on CCO, supported by TAC**

12.1 IOS release images

- **12.1 is the more recent “mainline” train**
  - **Comes from 12.0T, currently at 12.1(18)**
  - **Bug fix release only – aiming for stability**
  - **Supports more platforms and has more features than 12.0**
- **12.1E is the enterprise train**
  - **Started off as the 7600/Cat6500 train**
  - **Has many of the features from 12.0S**
  - **Currently at 12.1(14)E**
- **Available on CCO, supported by TAC**

12.2 IOS release images

- 12.2 is the current “mainline” train
  - Originated from 12.1T, currently at 12.2(13)
  - Bug fix release only – aiming for stability
  - Supports more platforms and has more features than 12.1
- 12.2T is the “technology train”
  - new features introduced in IOS 12.2
  - Includes IPv6 – currently at 12.2(13)T
- Available on CCO, supported by TAC



What is BGP for??

What is an IGP not for?

BGP versus OSPF/ISIS

- **Internal Routing Protocols (IGPs)**
  - **examples are ISIS and OSPF**
  - **used for carrying infrastructure addresses**
  - **NOT used for carrying Internet prefixes or customer prefixes**

BGP versus OSPF/ISIS

- BGP used internally (iBGP) and externally (eBGP)
- iBGP used to carry
  - some/all Internet prefixes across backbone
  - customer prefixes
- eBGP used to
  - exchange prefixes with other ASes
  - implement routing policy

BGP versus OSPF/ISIS

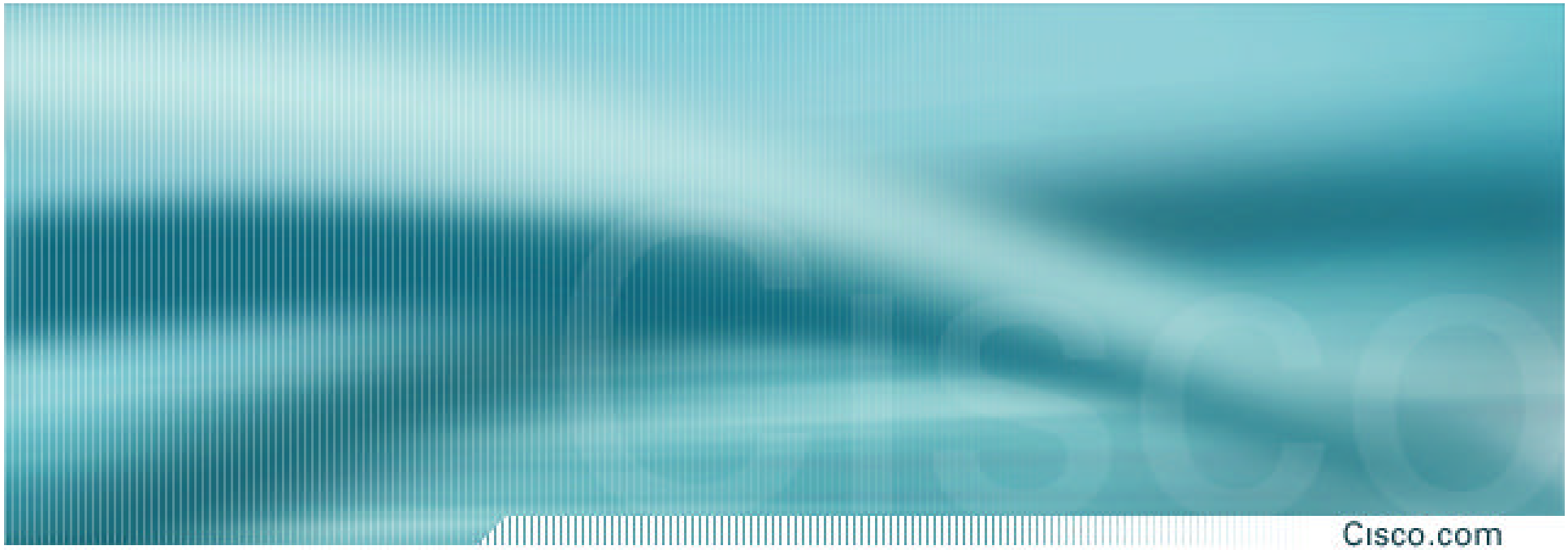
- **DO NOT:**

  distribute BGP prefixes into an IGP

  distribute IGP routes into BGP

  use an IGP to carry customer prefixes

- **YOUR NETWORK WILL NOT SCALE**



Aggregation

Aggregation

- **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 – such as fine-tuning multihoming**
- **Aggregate should be generated internally**
  - **Not on the network borders!**

Configuring Aggregation – Cisco IOS

- 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”

Aggregation

- **Address block should be announced to the Internet as an aggregate**
- **Subprefixes of address block should NOT be announced to Internet unless special circumstances (more later)**

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 101
 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
ip prefix-list out-filter deny 0.0.0.0/0 le 32
```

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**

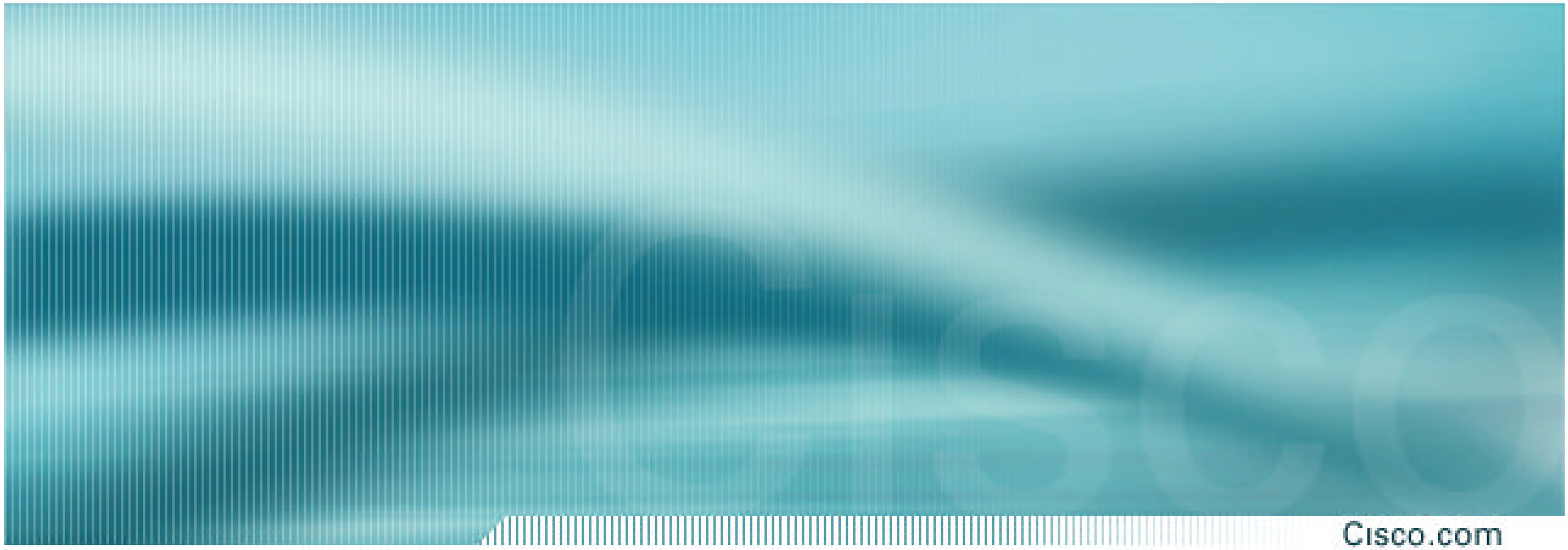
  no real reason to see anything longer than a /21 prefix in the Internet

  BUT there are currently >67000 /24s!

The Internet Today (January 2003)

- **Current Internet Routing Table Statistics**

BGP Routing Table Entries	119544
Prefixes after maximum aggregation	76260
Unique prefixes in Internet	57040
Prefixes smaller than registry alloc	55563
/24s announced	66125
only 5406 /24s are from 192.0.0.0/8	
ASes in use	14361



Receiving Prefixes

Receiving Prefixes from: peers or customers

- ISPs should only accept prefixes which have been assigned or allocated to their peer or 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 101
 neighbor 222.222.10.1 prefix-list customer in
!
ip prefix-list customer permit 220.50.0.0/20
ip prefix-list customer deny 0.0.0.0/0 le 32
```

Receiving Prefixes from: upstream peers

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

Receiving Prefixes from: upstream peers

- **Downstream Router Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 101
 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 infilter deny 0.0.0.0/0 le 32
!
ip prefix-list outfilter permit 221.10.0.0/19
ip prefix-list outfilter deny 0.0.0.0/0 le 32
```

Receiving Prefixes from: upstream peers

- **Upstream Router Configuration**

```
router bgp 101
 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-in deny 0.0.0.0/0 le 32
!
ip prefix-list cust-out permit 0.0.0.0/0
ip prefix-list cust-out deny 0.0.0.0/0 le 32
```

Receiving Prefixes from: upstream peers

- **If necessary to receive prefixes from upstream provider, care is required**
  - **don't accept RFC1918 etc prefixes**
    - **http://www.ietf.org/internet-drafts/draft-manning-dsua-08.txt**
    - **ftp://ftp.rfc-editor.org/in-notes/rfc3330.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

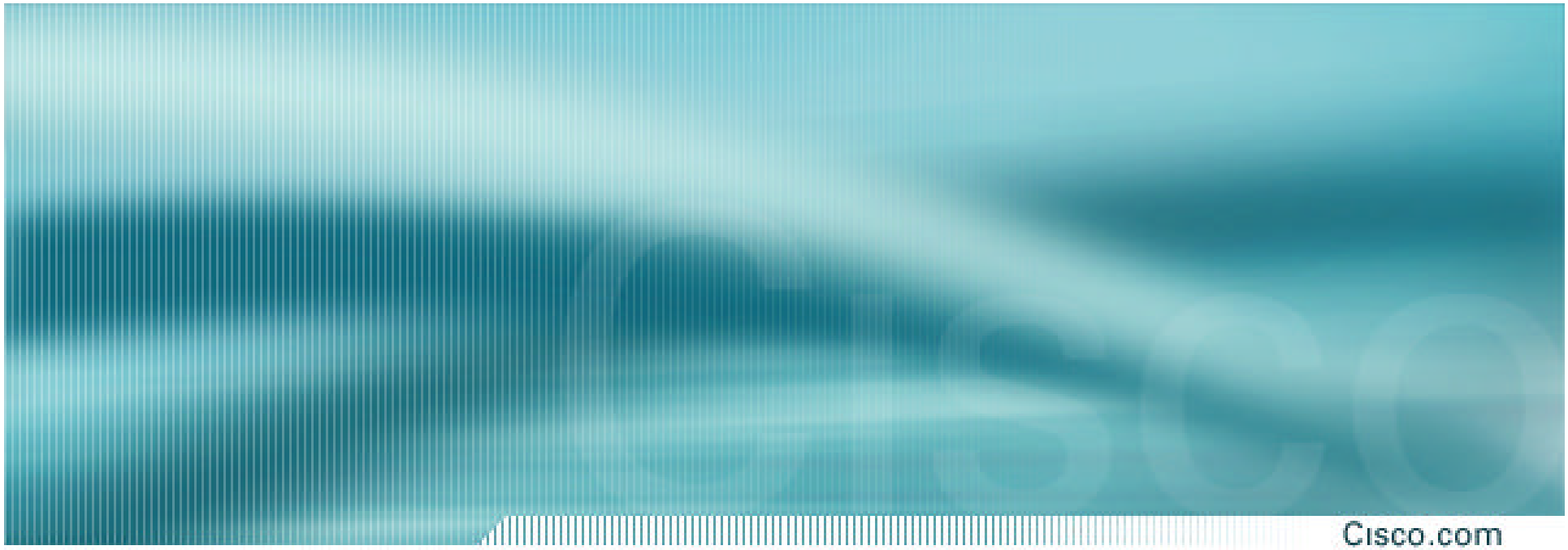

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 101
 neighbor 221.5.7.1 prefix-list bogons in
!
ip prefix-list bogons deny 0.0.0.0/0                    ! Block default
ip prefix-list bogons deny 0.0.0.0/8 le 32
ip prefix-list bogons deny 10.0.0.0/8 le 32
ip prefix-list bogons deny 127.0.0.0/8 le 32
ip prefix-list bogons deny 169.254.0.0/16 le 32
ip prefix-list bogons deny 172.16.0.0/12 le 32
ip prefix-list bogons deny 192.0.2.0/24 le 32
ip prefix-list bogons deny 192.168.0.0/16 le 32
ip prefix-list bogons deny 221.10.0.0/19 le 32          ! Block local prefix
ip prefix-list bogons deny 224.0.0.0/3 le 32            ! Block multicast
ip prefix-list bogons deny 0.0.0.0/0 ge 25              ! Block prefixes >/24
ip prefix-list bogons permit 0.0.0.0/0 le 32
```

Receiving Prefixes

- **Always filter what your neighbour sends you**

  Check in registry database that the ISP is entitled to announce the prefix it is sending

- **Always apply the previous filter when receiving the full routing table or a sizeable number of prefixes from a peer**



Prefixes into iBGP

Injecting prefixes into iBGP

- Use iBGP to carry customer prefixes

  don't use IGP

- Point static route to customer interface
- Use BGP network statement
- As long as static route exists (interface active), prefix will be in BGP

Router Configuration: network statement

- **Example:**

```
interface loopback 0
 ip address 215.17.3.1 255.255.255.255
!
interface Serial 5/0
 ip unnumbered loopback 0
 ip verify unicast reverse-path
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 network 215.34.10.0 mask 255.255.252.0
```

Injecting prefixes into iBGP

- interface flap will result in prefix withdraw and reannounce
  - use “ip route…permanent”
- many ISPs use redistribute static rather than network statement
  - only use this if you understand why

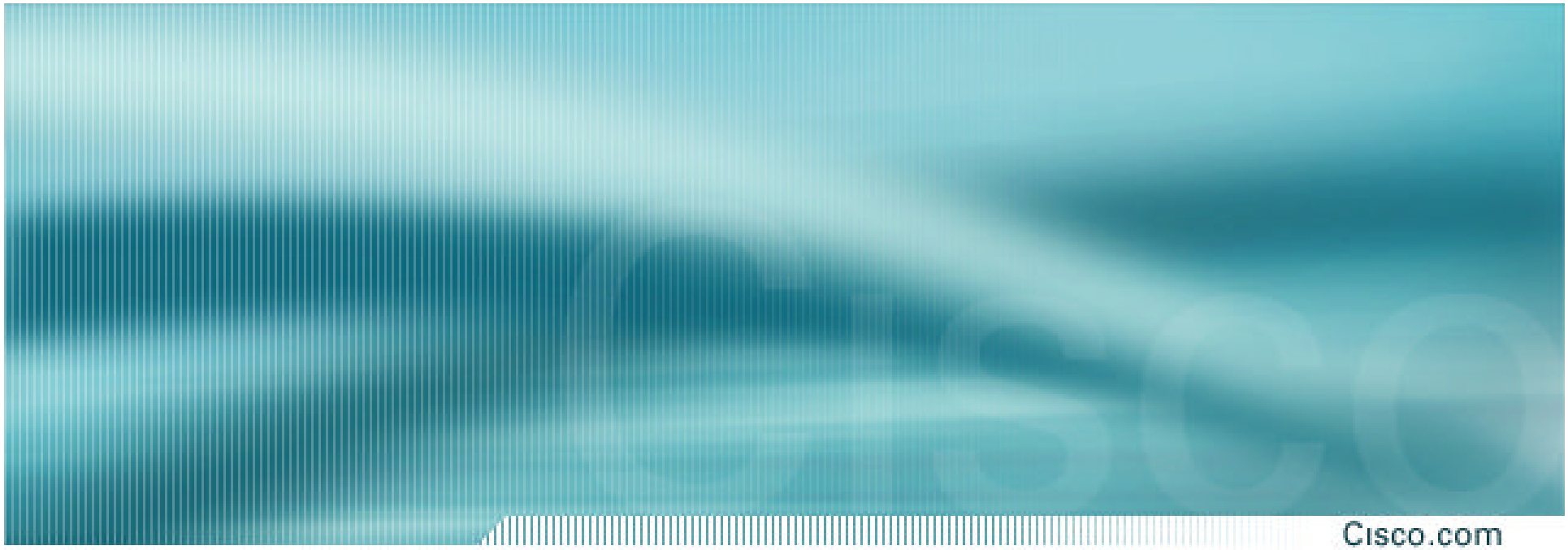
Router Configuration: redistribute static

- **Example:**

```
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 redistribute static route-map static-to-bgp
<snip>
!
route-map static-to-bgp permit 10
 match ip address prefix-list ISP-block
 set origin igp
<snip>
!
ip prefix-list ISP-block permit 215.34.10.0/22 le 30
!
```

Injecting prefixes into iBGP

- **Route-map ISP-block can be used for many things:**
  - **setting communities and other attributes**
  - **setting origin code to IGP, etc**
- **Be careful with prefix-lists and route-maps**
  - **absence of either/both means all statically routed prefixes go into iBGP**



Scaling the network

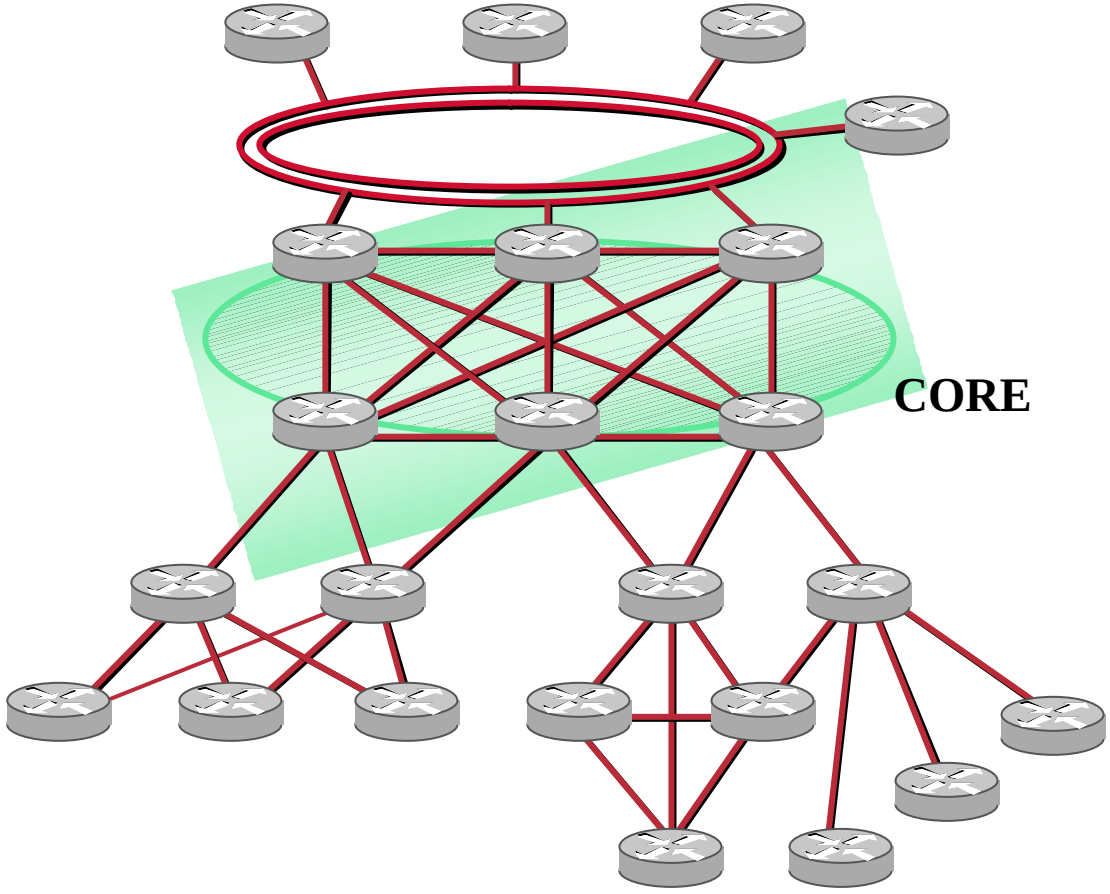
How to get out of carrying all prefixes in IGP

IGP Limitations

- **Amount of routing information in the network**
  - **Periodic updates/flooding**
  - **Long convergence times**
  - **Affects the core first**
- **Policy definition**
  - **Not easy to do**

BGP Cores
Sample Network

- Geographically distributed
- Hierarchical
- Redundant
- Media independent
- A clearly identifiable core



iBGP Core: Migration Plan

- Configure BGP on all the core routers
  - Transit path
  - Turn synchronisation off
  - Turn auto-summarisation off
- Check network borders
  - Ensure eBGP peerings only announce aggregates and won't leak specifics
- Route generation
  - Use static routes to generate summaries if required
  - Redistribution from the IGP is NOT recommended as it will cause instability

iBGP Core
Migration Plan (Cont.)

- **Route Generation – Example:**

```
!
router bgp 109
 network 200.200.200.0
 network 201.201.0.0 mask 255.255.0.0
!
ip route 200.200.200.0 255.255.255.0 null0
ip route 201.201.0.0 255.255.0.0 null0
!
```

iBGP Core Migration Plan (Cont.)

- **Verify consistency of routing information**

  Compare the IGP routing table against the BGP table – they must match!

- **Change the distance parameters so that the BGP routes are preferred**

  distance bgp 20 20 20

  All IGPs have a higher administrative distance

iBGP Core Migration Plan (Cont.)

- **Filter “non-core” IGP routes**
  - **Method will depend on the IGP used**
  - **May require the use of a different IGP process in the core if using a link state protocol**
  - **The routes to reach all the core links plus the BGP peering addresses must be carried by the IGP**

iBGP Core Migration Plan (Cont.)

- **Once iBGP carrying prefixes...**

  apply route-map to IGP redistribute commands so that only infrastructure addresses are in IGP

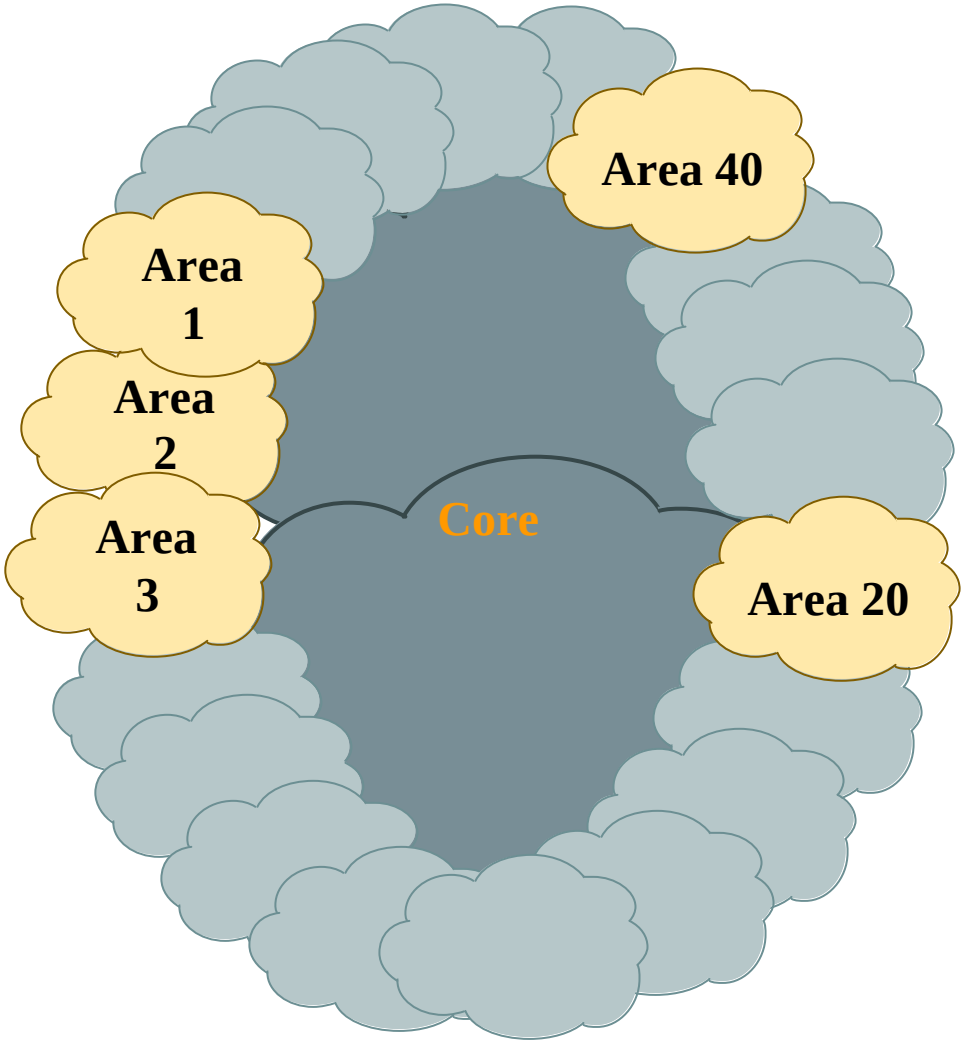
  check that customer routes in IGP have disappeared

  change BGP distance back to default

  no distance bgp 20 20 20

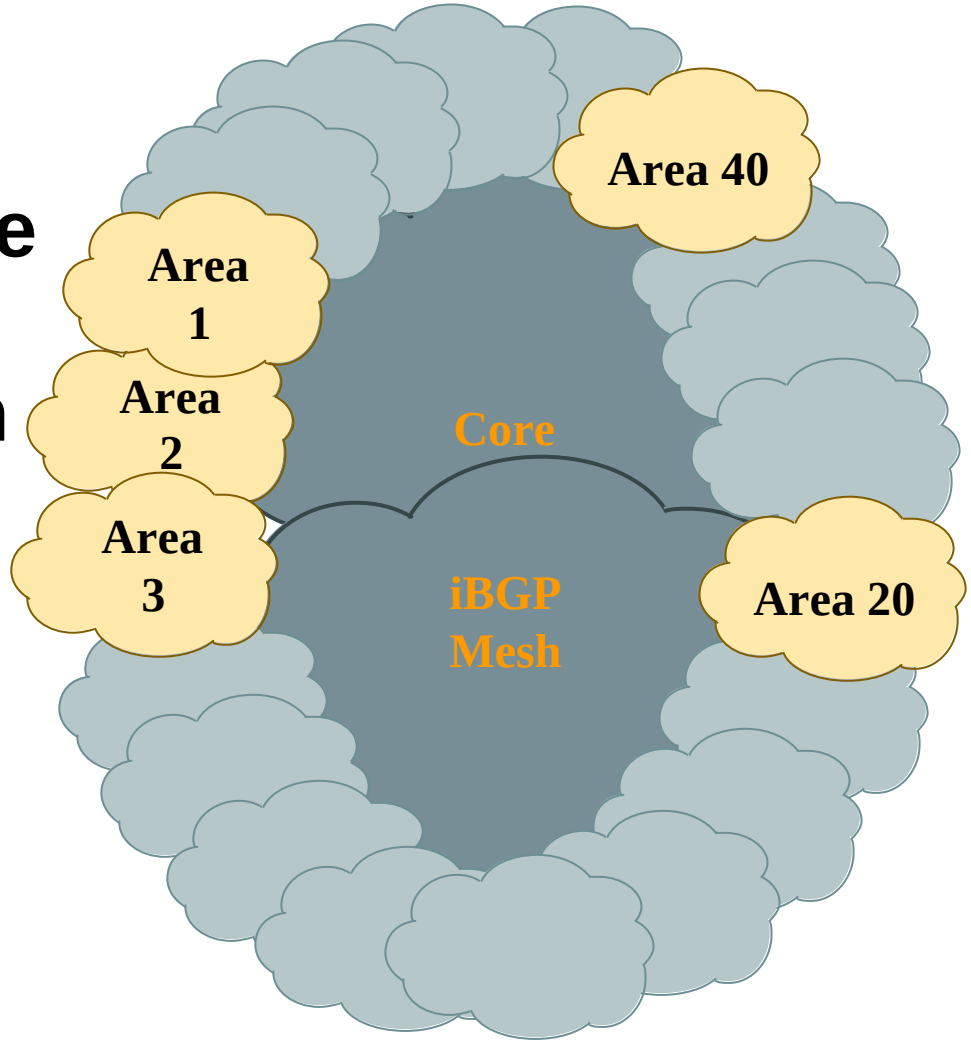
iBGP Core Before...

- IGP carries all the routes
- The core routers may be stressed due to the large number of routes



iBGP Core After...

- IGP carries only core links plus peering address information
- BGP carries all the routes
- Increased Stability!



iBGP Core Results

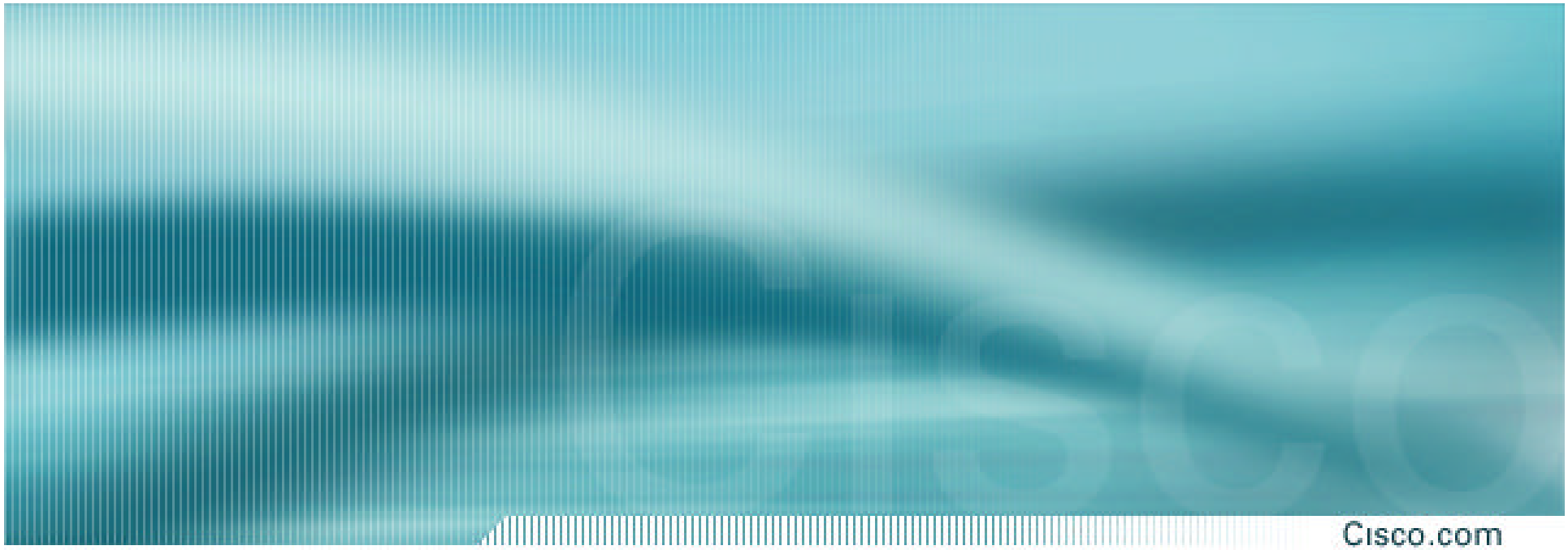
- **The routes from the core cannot be redistributed back into the IGP**
  - **Non-core areas need a default route**
  - **Amount of routing information in non-core areas has been reduced!**
- **Full logical iBGP mesh**
- **External connections must be located in the core**

Scaling Issues

- **Full mesh core**
  - **High number of neighbors**
  - **Update generation**
- **Complex topologies**
  - **Not a “simple” hierarchical network**
  - **Multiple external and/or inter-region connections**
  - **Policy definition and enforcement**

Scaling Issues: Solutions

- **Reduce the number of updates**
  - **Peer groups**
- **Reduce the number of neighbors**
  - **Confederations**
  - **Route reflectors**
- **Use additional information to effectively apply policies**
  - **eBGP provides extra granularity**
  - **Confederations**



BGP in the Internet

Best Current Practices