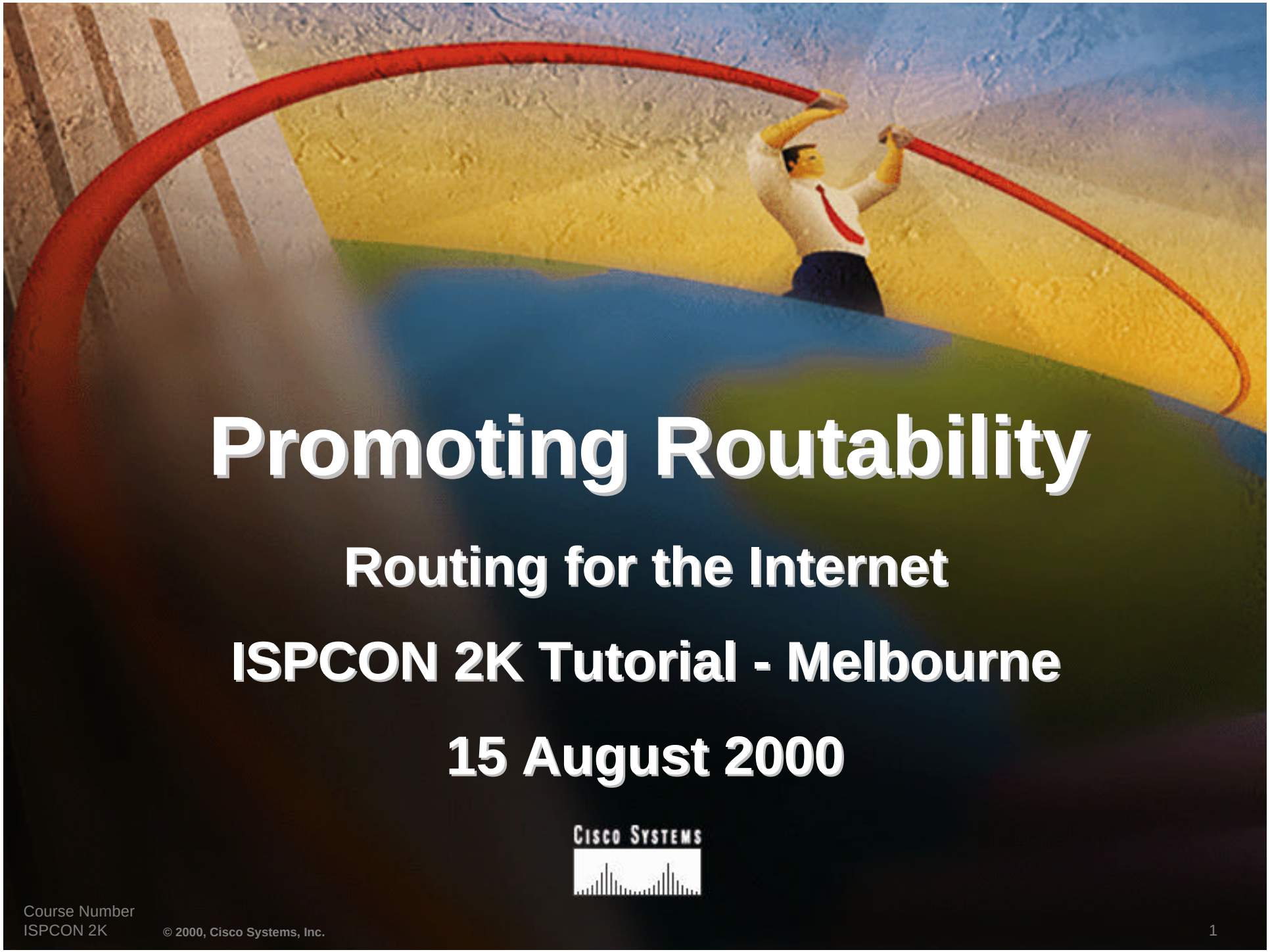


Promoting Routability

Routing for the Internet

ISPCON 2K Tutorial - Melbourne

15 August 2000

Cisco Systems

Introduction

- **Presenter:**

  Philip Smith, Consulting Engineer

  Office of the CTO, Cisco Systems

  e-mail: pfs@cisco.com

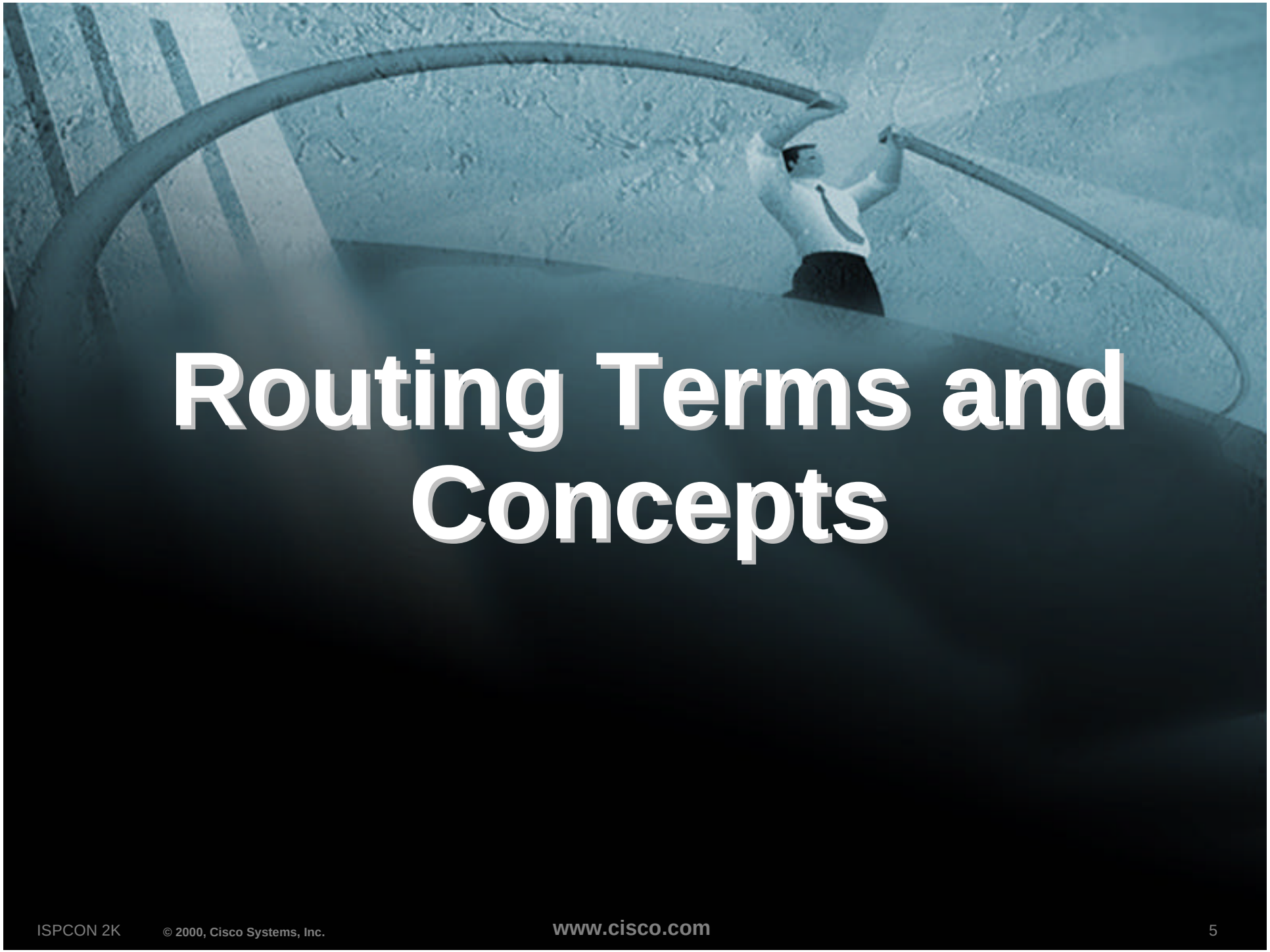
- **Please ask questions**

Agenda

- Routing Terms and Concepts
- Introduction to IGPs
- BGP for ISPs
- Routing Design for ISPs
- Routing Etiquette and the IRR

Goals

- Promoting a healthy Internet
- Efficient and Effective Routing Configuration
- Internet Routing Registry
  - awareness
  - understanding
  - participation



Routing Terms and Concepts

Network Topologies

Routed backbone

- HDLC or PPP links between routers
- Easier routing configuration and debugging

Switched backbone

- Frame Relay/ATM switches in core
- Surrounded by routers
- Complex routing & debugging
- Traffic Engineering

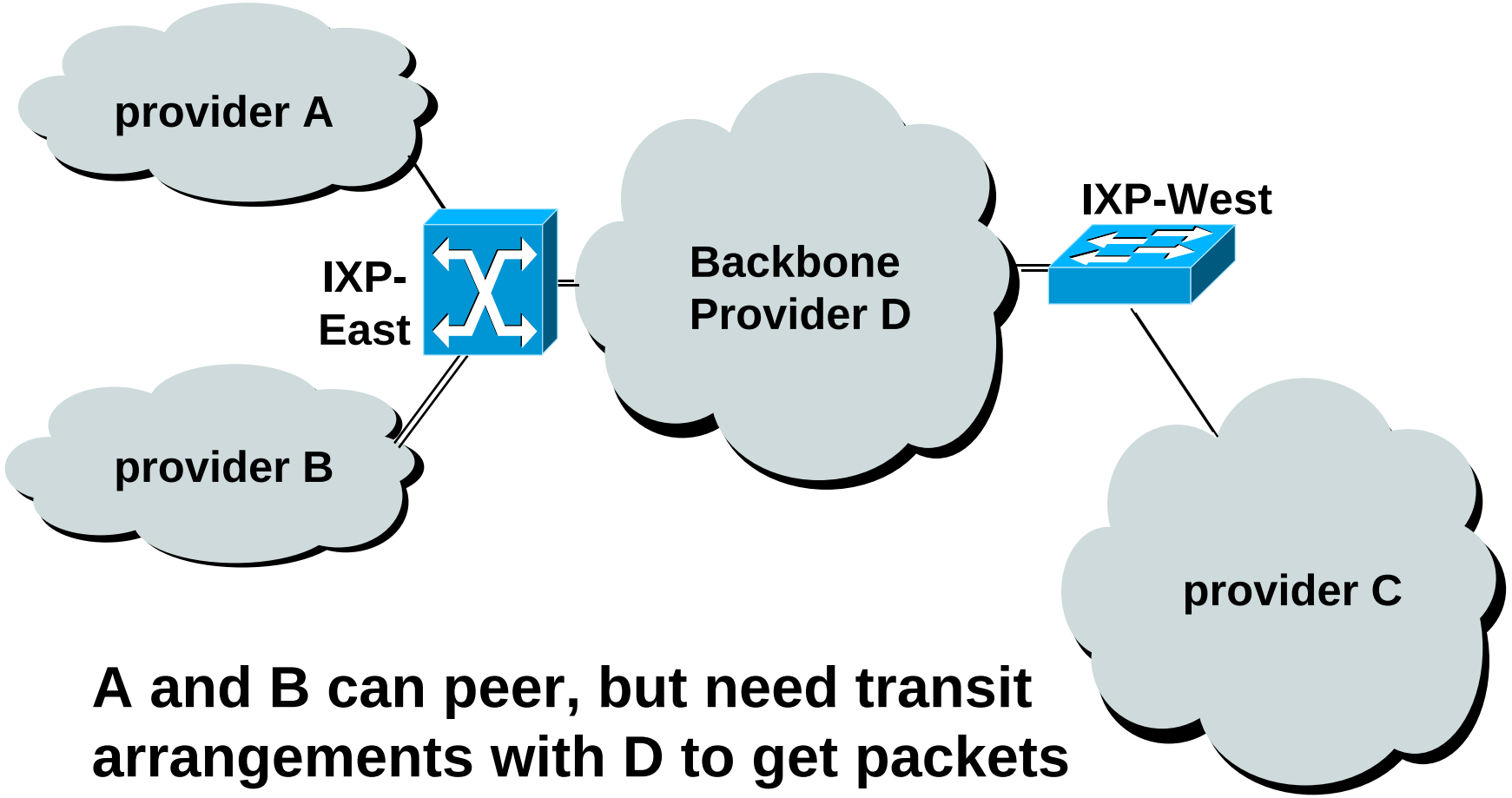
PoP Topologies

- **Core routers - high speed trunk connections**
- **Distribution routers and Access routers - high port density**
- **Border routers - connections to other AS's**
- **Service routers - hosting and servers**
- **Some functions might be handled by a single router**

Transit, Peering and Default

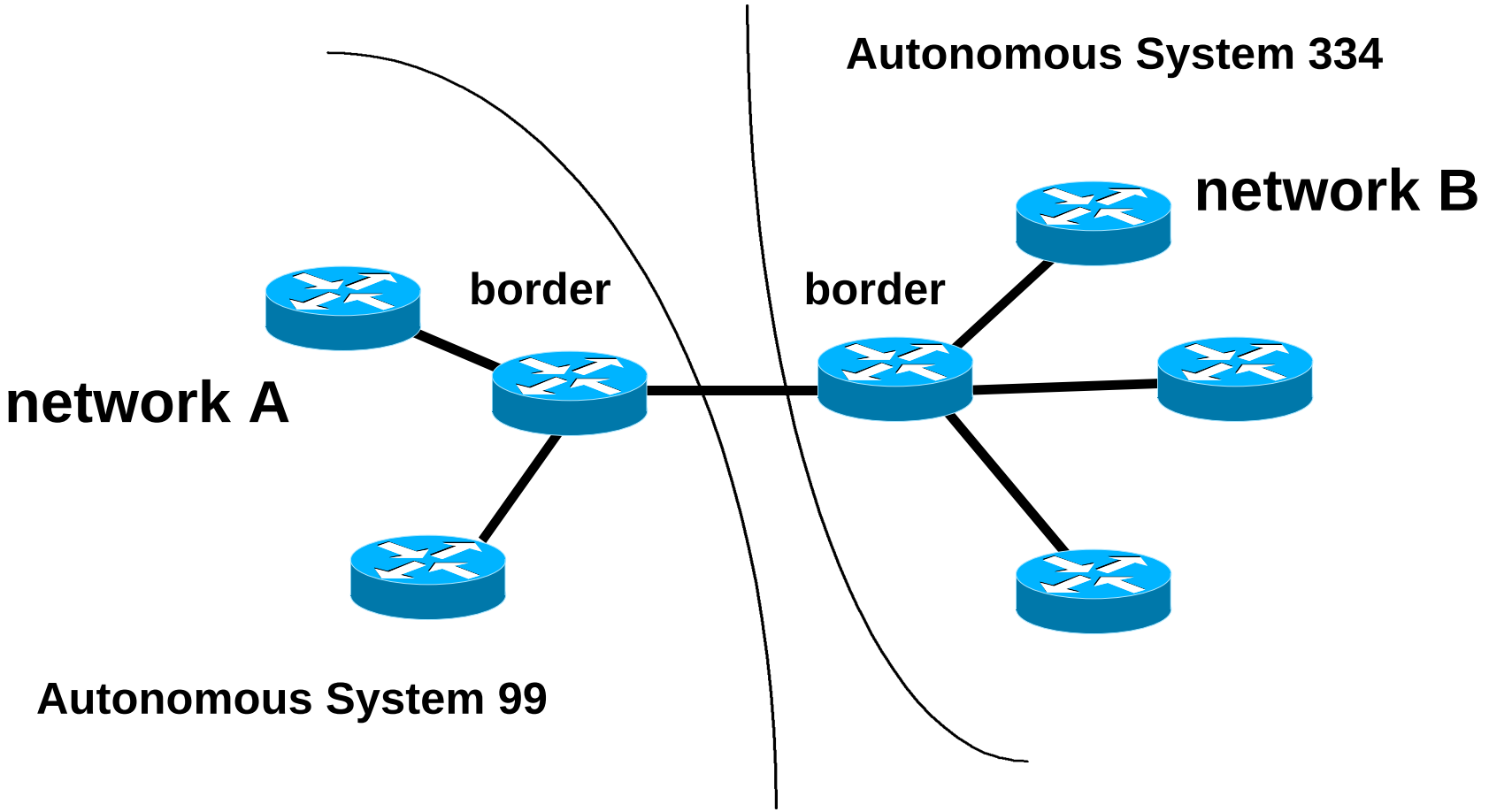
- **Transit - carrying traffic across a network, usually for a fee**
- **Peering - exchanging routing information and traffic**
- **Default - where to send traffic when there is no explicit match is in the routing table**

Peering and Transit example



A and B can peer, but need transit arrangements with D to get packets to/from C

Private Interconnect



Public Interconnect Points

- **IXP - Internet eXchange Point**
- **NAP - Network Access Point**
- **local IXPs**

  peering point for a group of local/regional providers
- **transit IXPs**

  connects local providers to backbone (transit) providers
- **hybrid IXPs**

  combines the function of local and transit

Public Interconnect Point

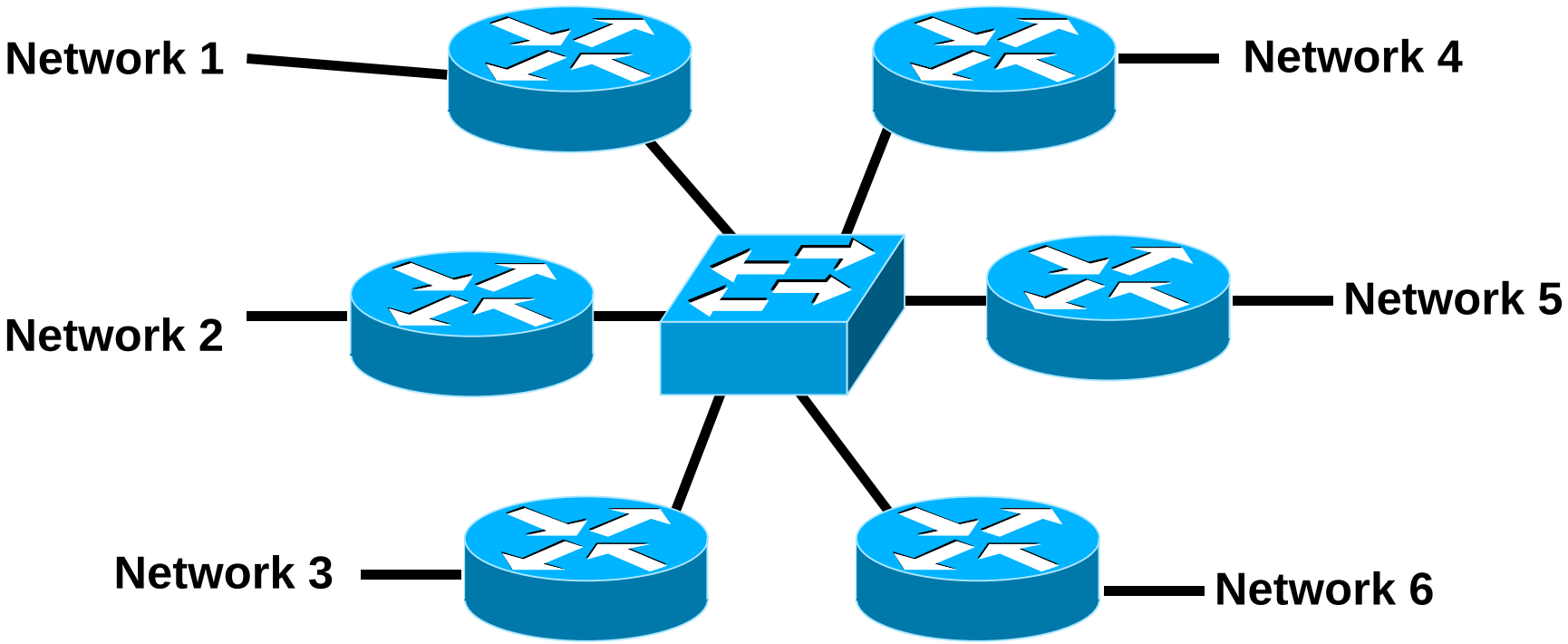
- **Centralised (in one facility)**
- **Distributed (connected via WAN links)**
- **Shared, switched or routed interconnect**

  Router, FDDI, Ethernet, ATM, Frame relay, SMDS, etc.

- **Each provider establishes relationship with other provider at IXP**

  ISP border router peers with all other provider border routers

Public Interconnect



each of these represents a border router in a different autonomous system

Route Server

- **Device which maintains BGP routing table at IXP and forwards it to IXP participants**
- **Advantages:**

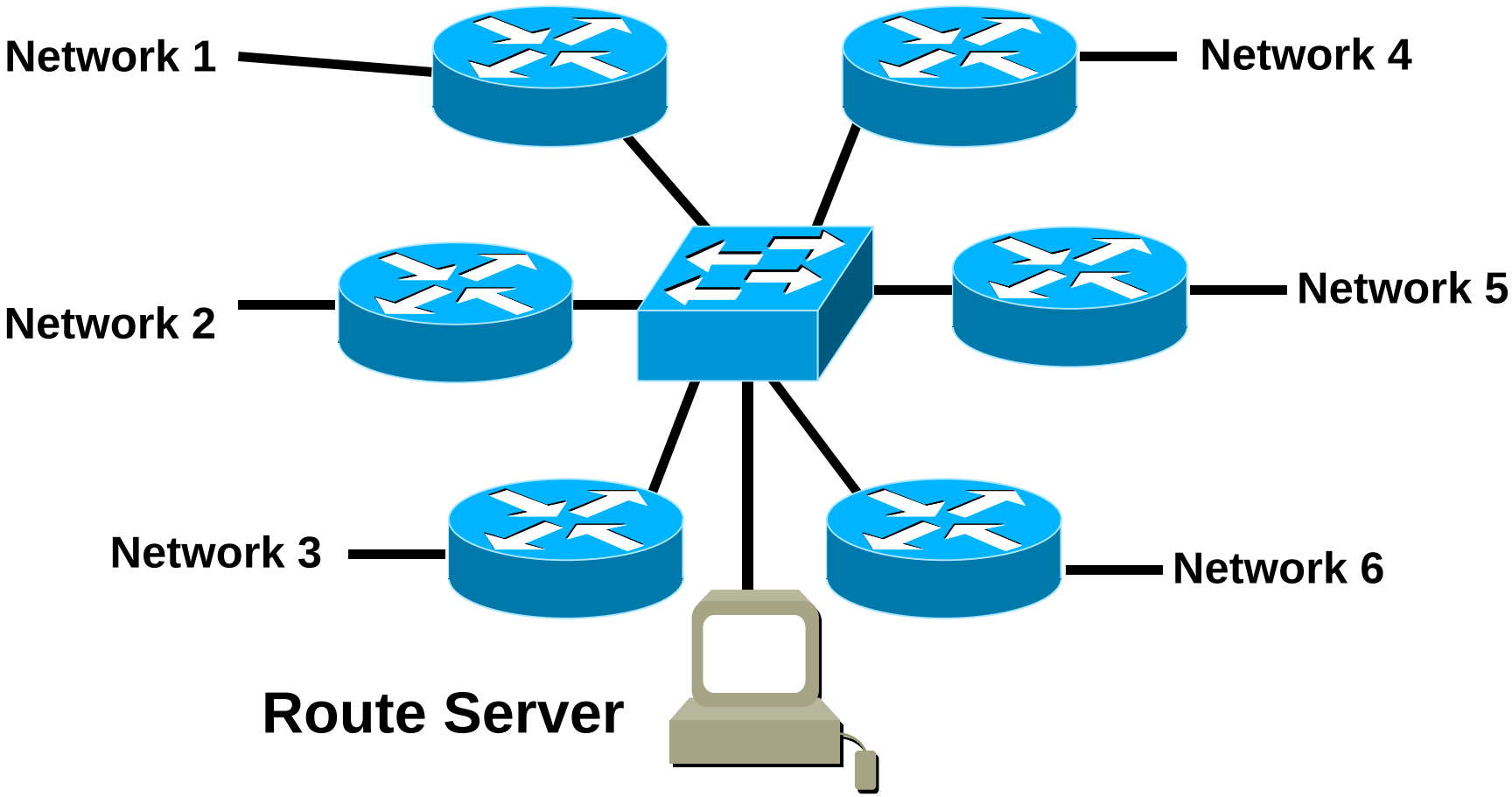
  reduces resource burden on border routers (CPU, memory, configuration complexity)

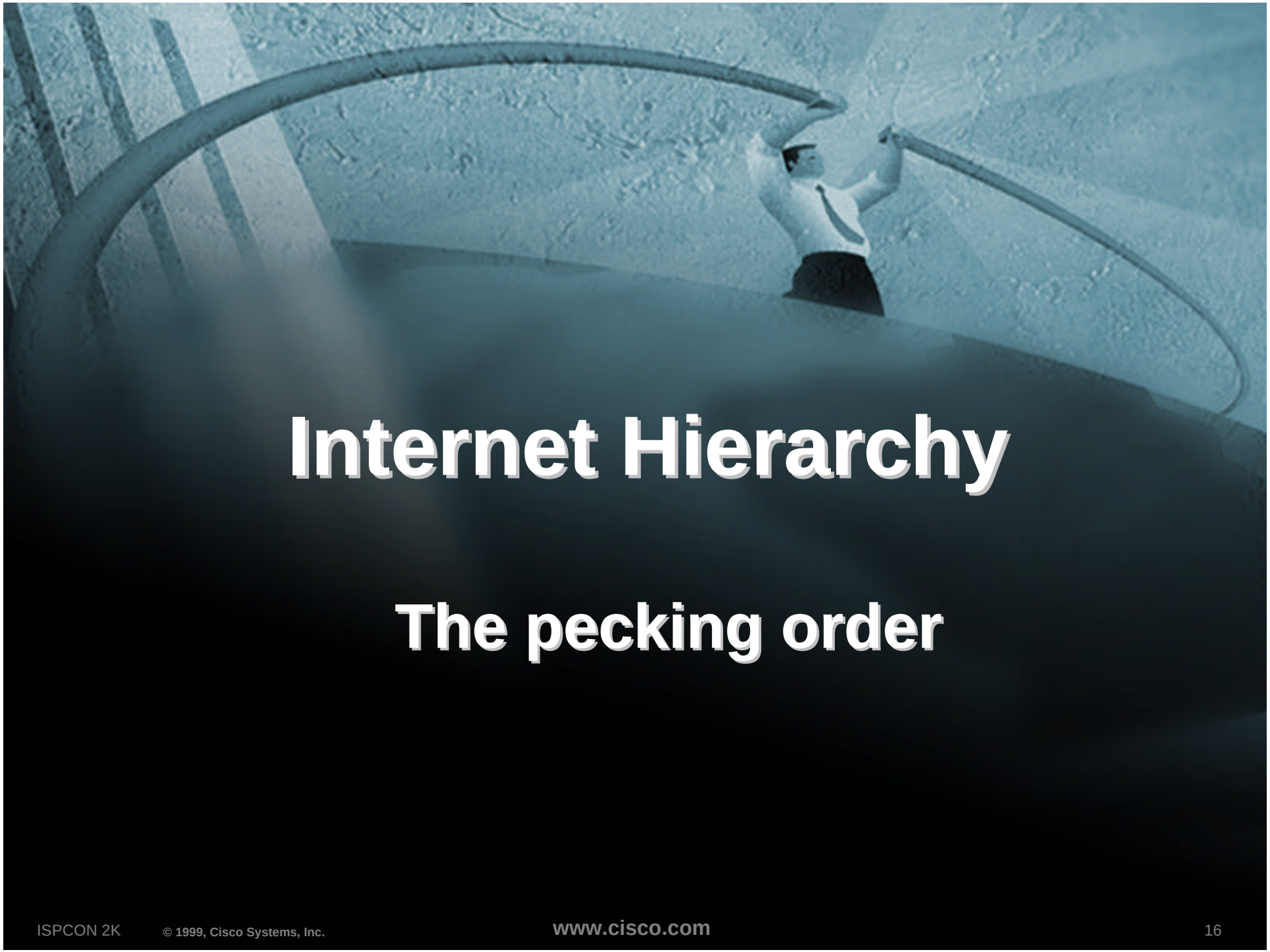
  reduces administrative burden on providers

- **Disadvantages:**

  must rely on a third party (for management, configuration, software updates, maintenance, etc)

Route Server





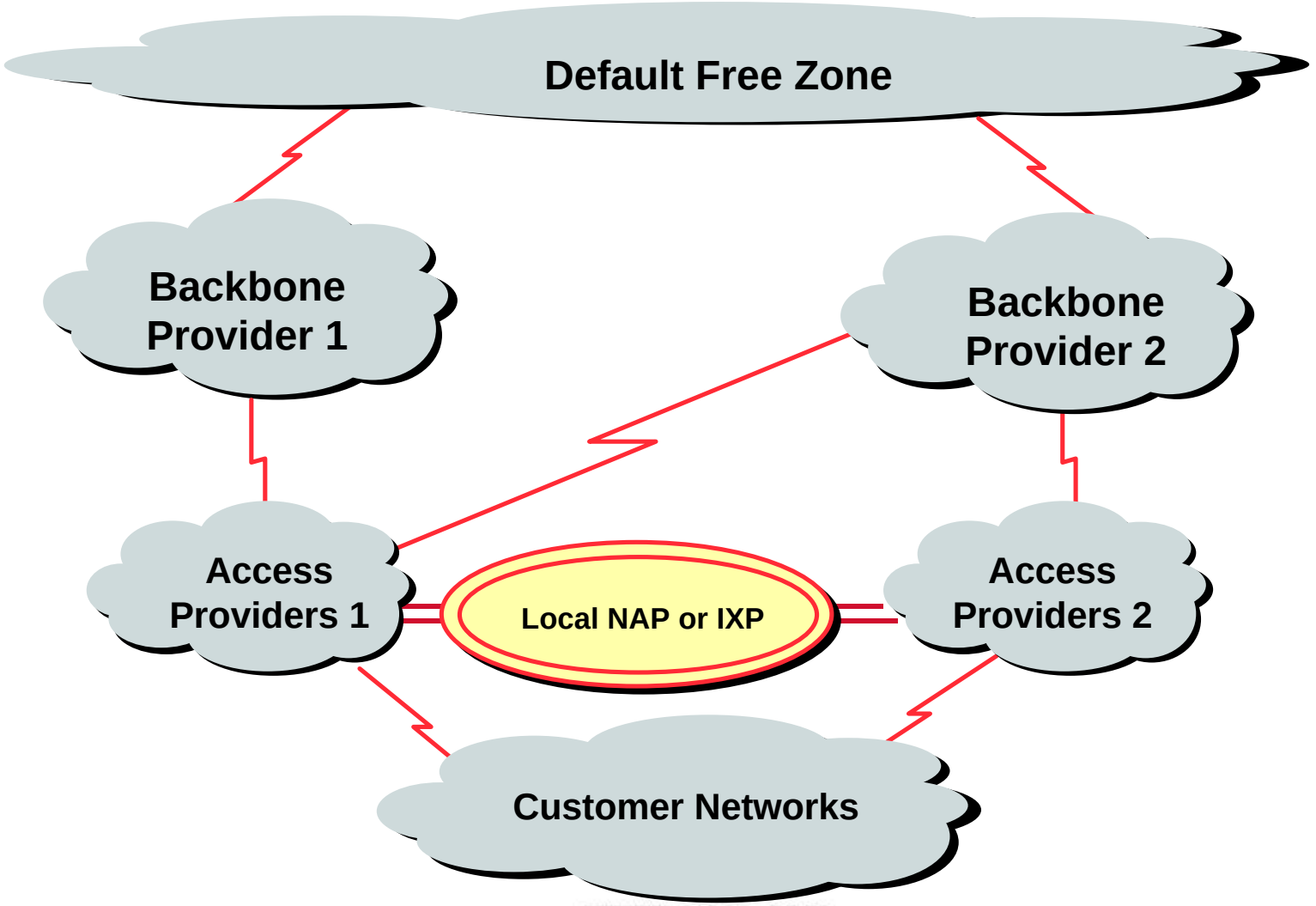
Internet Hierarchy

The pecking order

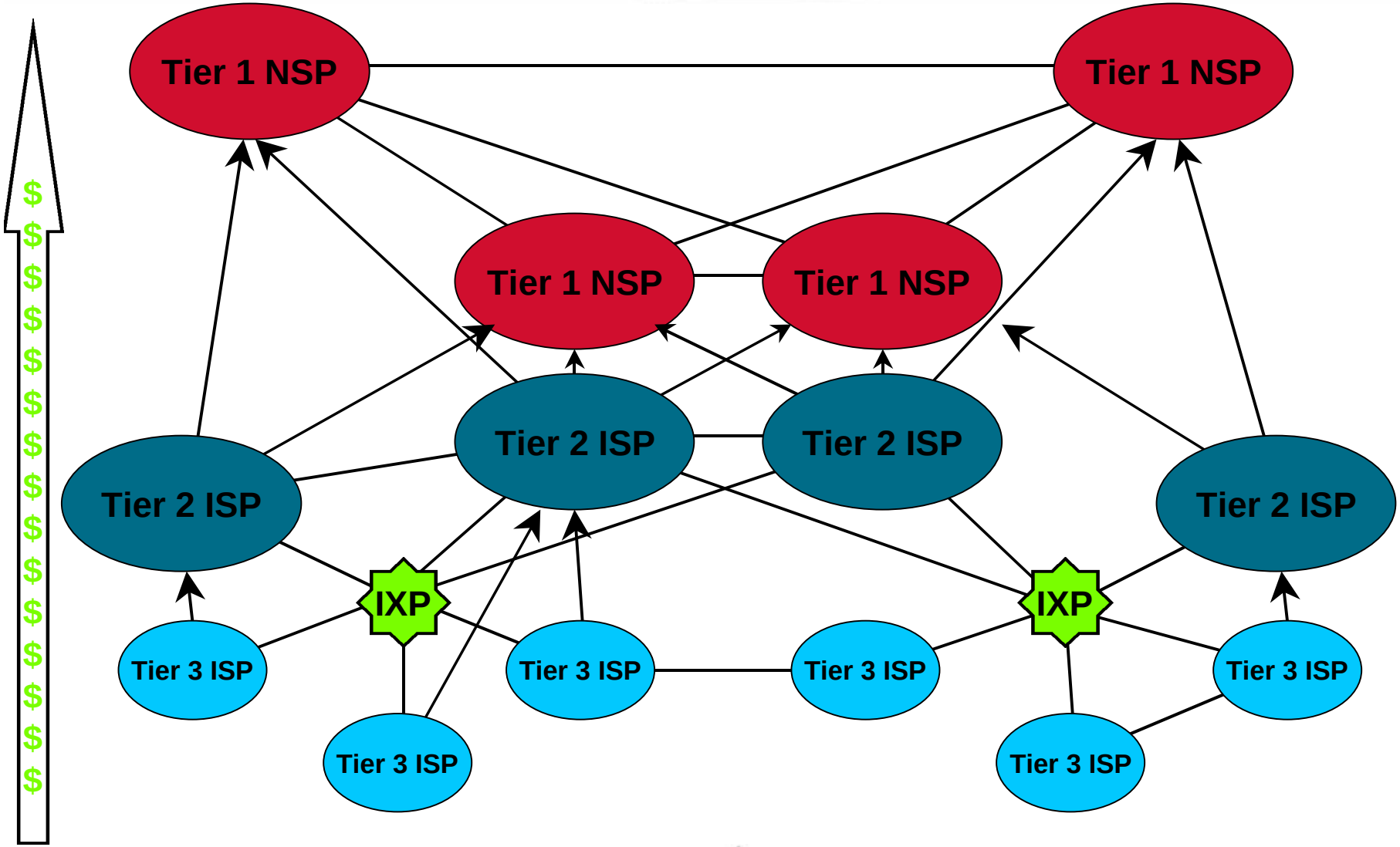
Default Free Zone

The default free zone is made up of Internet routers which have explicit routing information about the rest of the Internet, and therefore do not need to use a default route.

High Level View of the Global Internet



Categorising ISPs



Inter-provider relationships

- **Peering between equivalent sizes of service providers (eg Tier 2 to Tier 2)**

  shared cost private interconnection, equal traffic flows

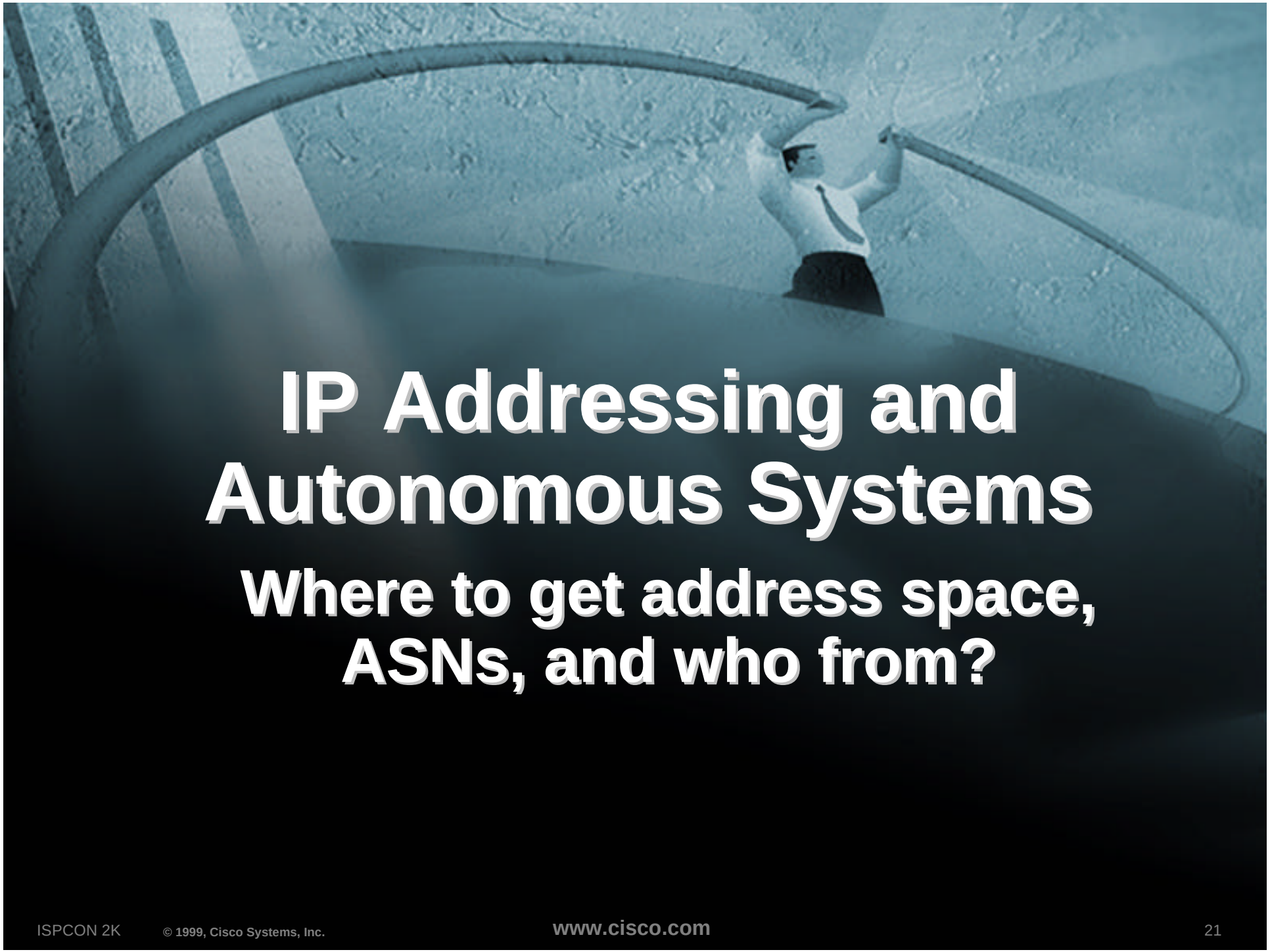
  “no cost peering”

- **Peering across exchange points**

  if convenient, of mutual benefit, technically feasible

- **Fee based peering**

  unequal traffic flows, “market position”



IP Addressing and Autonomous Systems

Where to get address space, ASNs, and who from?

IP Addressing

- Internet is classless
- Concept of Class A, class B or class C is no more

  engineers talk in terms of prefix length, for example the class B 158.43 is now called 158.43/16.

- All routers must be CIDR capable

  Classless InterDomain Routing

  RFC1812 - Router Requirements

IP Addressing

- **Pre-CIDR (<1994)**

  big networks got a class A

  medium networks got a class B

  small networks got a class C

- **Nowadays**

  allocations/assignments made according to demonstrated need - CLASSLESS

  No boundaries, no barriers

IP Addressing

- **IPv4 Address space is a resource shared amongst all Internet users**
  - **Regional Internet Registries delegated allocation responsibility by the IANA**
  - **APNIC, ARIN, RIPE NCC are the three RIRs**
  - **RIRs allocate address space to ISPs and Local Internet Registries**
  - **ISPs/LIRs assign address space to end customers or other ISPs**
- **51% of available IPv4 address space used**

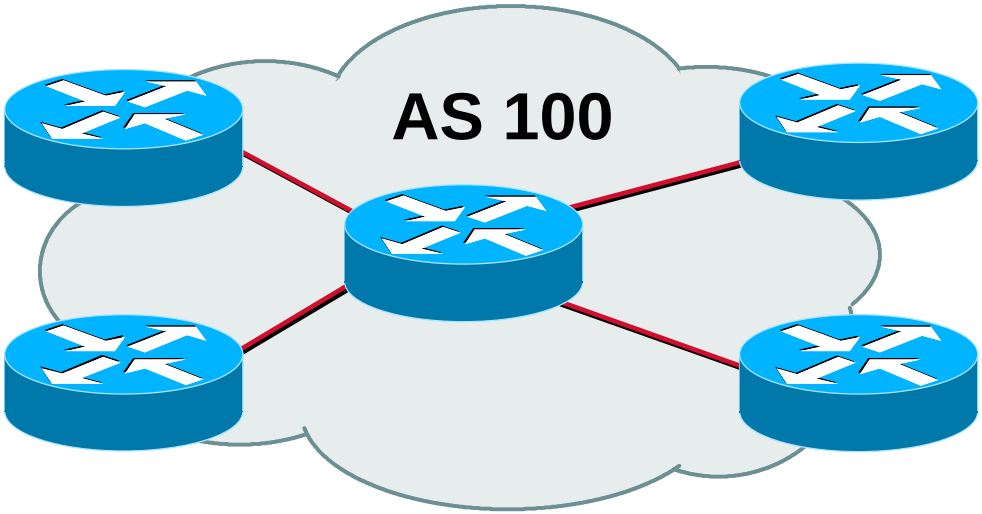
Definitions

- **Non-portable - ‘provider aggregatable’ (PA)**
  - **Customer uses RIR member’s address space while connected to Internet**
  - **Customer has to renumber to change ISP**
  - **Aids control of size of Internet routing table**
  - **May fragment provider block when multihoming**
- **PA space is allocated to the RIR member with the requirement that all assignments are announced as an aggregate**

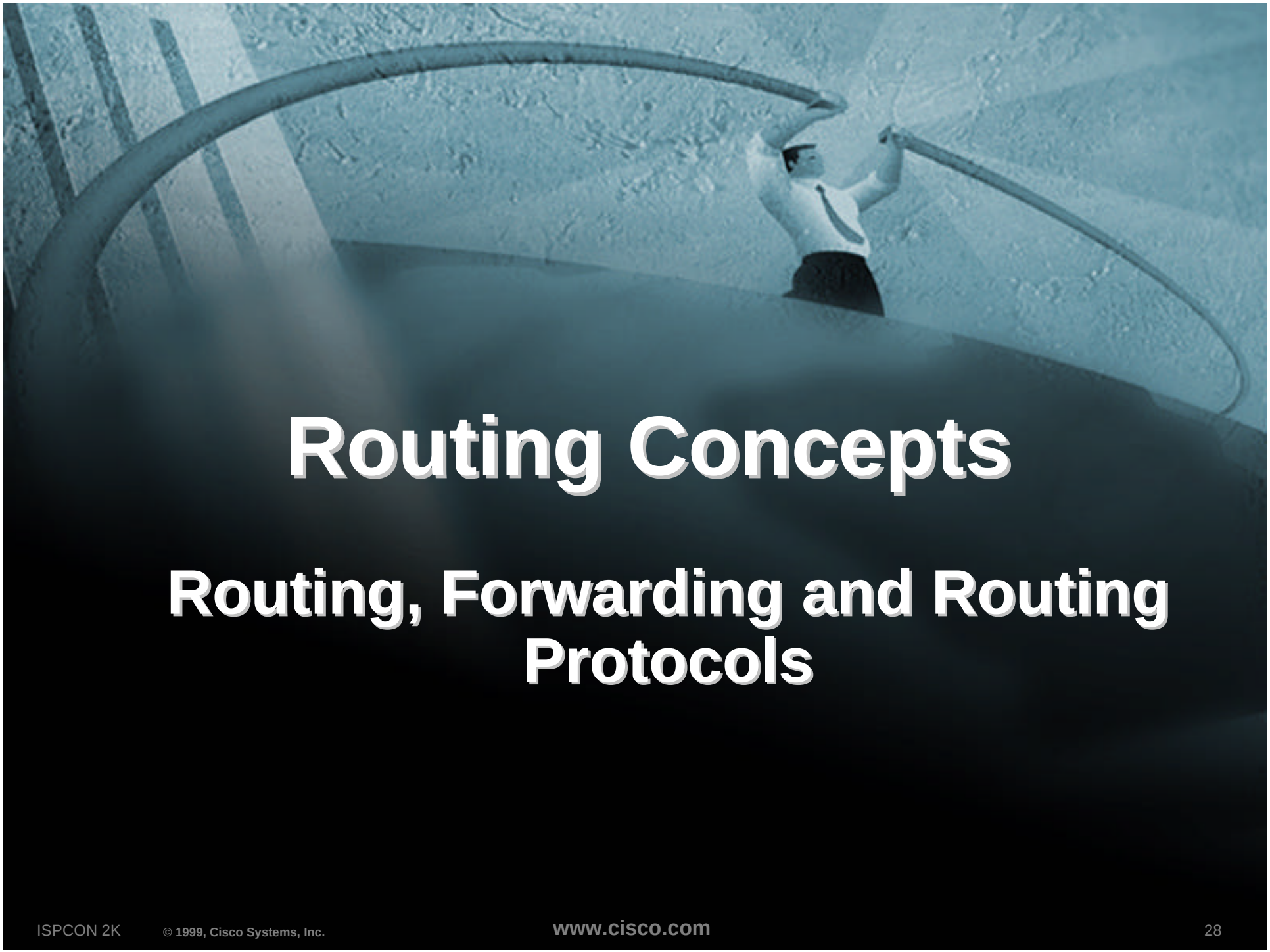
Definitions

- **Portable - ‘provider independent’ (PI)**
  - **Customer gets or has address space independent of ISP**
  - **Customer keeps addresses when changing ISP**
  - **Bad for size of Internet routing table**
  - **PI space is rarely distributed by the RIRs**

Autonomous System (AS)



- Collection of networks with same routing policy
- Single routing protocol
- Usually under single ownership, trust and administrative control
- AS number obtained from RIR or upstream ISP

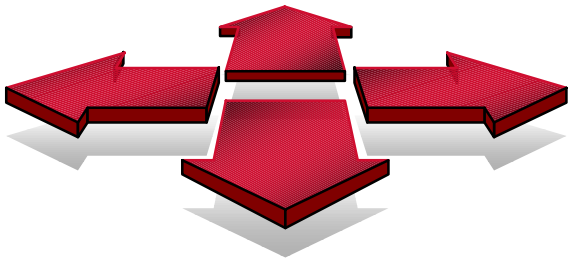


Routing Concepts

Routing, Forwarding and Routing Protocols

Routing versus Forwarding

- **Routing = building maps and giving directions**
- **Forwarding = moving packets between interfaces according to the "directions"**



IP Routing - finding the path

- Path derived from information received from a routing protocol
- Several alternative paths may exist

  best next hop stored in forwarding table
- Decisions are updated periodically or as topology changes (event driven)
- Decisions are based on:

  topology, policies and metrics (hop count, filtering, delay, bandwidth, etc.)

IP route lookup

- **Based on destination IP packet**
- **“longest match” routing**

  more specific prefix preferred over less specific prefix

  example: packet with destination of 10.1.1.1/32 is sent to the router announcing 10.1/16 rather than the router announcing 10/8.

IP Forwarding

- Router makes decision on which interface a packet is sent to
- Forwarding table populated by routing process
- Forwarding decisions:

  destination address

  class of service (fair queuing, precedence, others)

  local requirements (packet filtering)
- Can be aided by special hardware

Explicit versus Default routing

- **Default:**
  - **simple, cheap (cycles, memory, bandwidth)**
  - **low granularity (metric games)**
- **Explicit (default free zone)**
  - **high overhead, complex, high cost, high granularity**
- **Hybrid**
  - **minimise overhead**
  - **provide useful granularity**
  - **requires some filtering knowledge**

Egress Traffic

- How packets leave your network
- Egress traffic depends on:
  - route availability (what others send you)
  - route acceptance (what you accept from others)
  - policy and tuning (what you do with routes from others)
  - Peering and transit agreements

Ingress Traffic

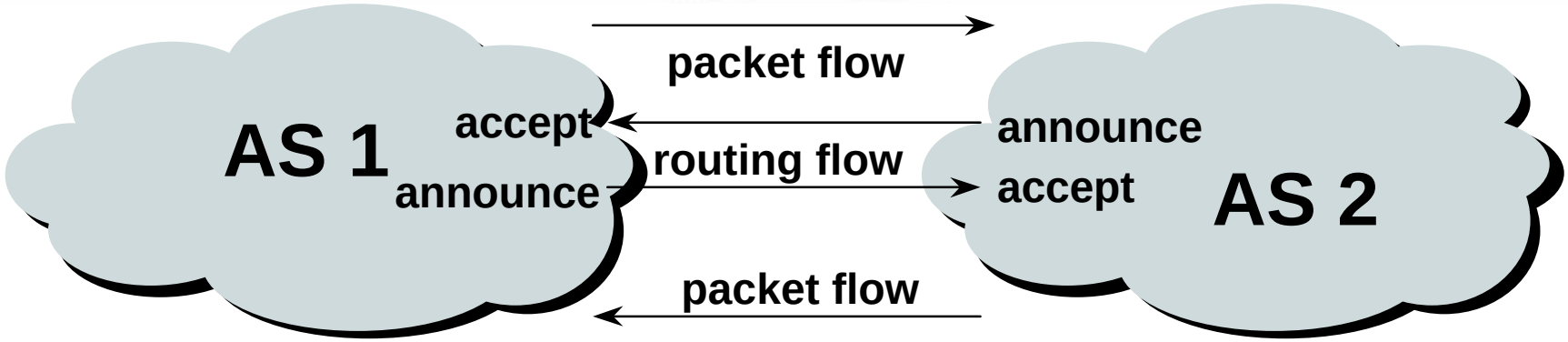
- How packets get to your network and your customers' networks
- Ingress traffic depends on:

  what information you send and to whom

  based on your addressing and AS's

  based on others' policy (what they accept from you and what they do with it)

Routing flow and packet flow



- For networks in AS1 and AS2 to communicate:
  - AS1 must announce to AS2
  - AS2 must accept from AS1
  - AS2 must announce to AS1
  - AS1 must accept from AS2
- Traffic flow is always in the opposite direction of the flow of routing information

What Is an IGP?

- Interior Gateway Protocol
- Within an Autonomous System
- Carries information about internal prefixes
- Examples - OSPF, ISIS, EIGRP…

What Is an EGP?

- Exterior Gateway Protocol
- Used to convey routing information between Autonomous Systems
- De-coupled from the IGP
- Current EGP is BGP4

Why Do We Need an EGP?

- **Scaling to large network**
  - **Hierarchy**
  - **Limit scope of failure**
- **Policy**
  - **Control reachability to prefixes**
  - **Merge separate organizations**
  - **Connect multiple IGPs**

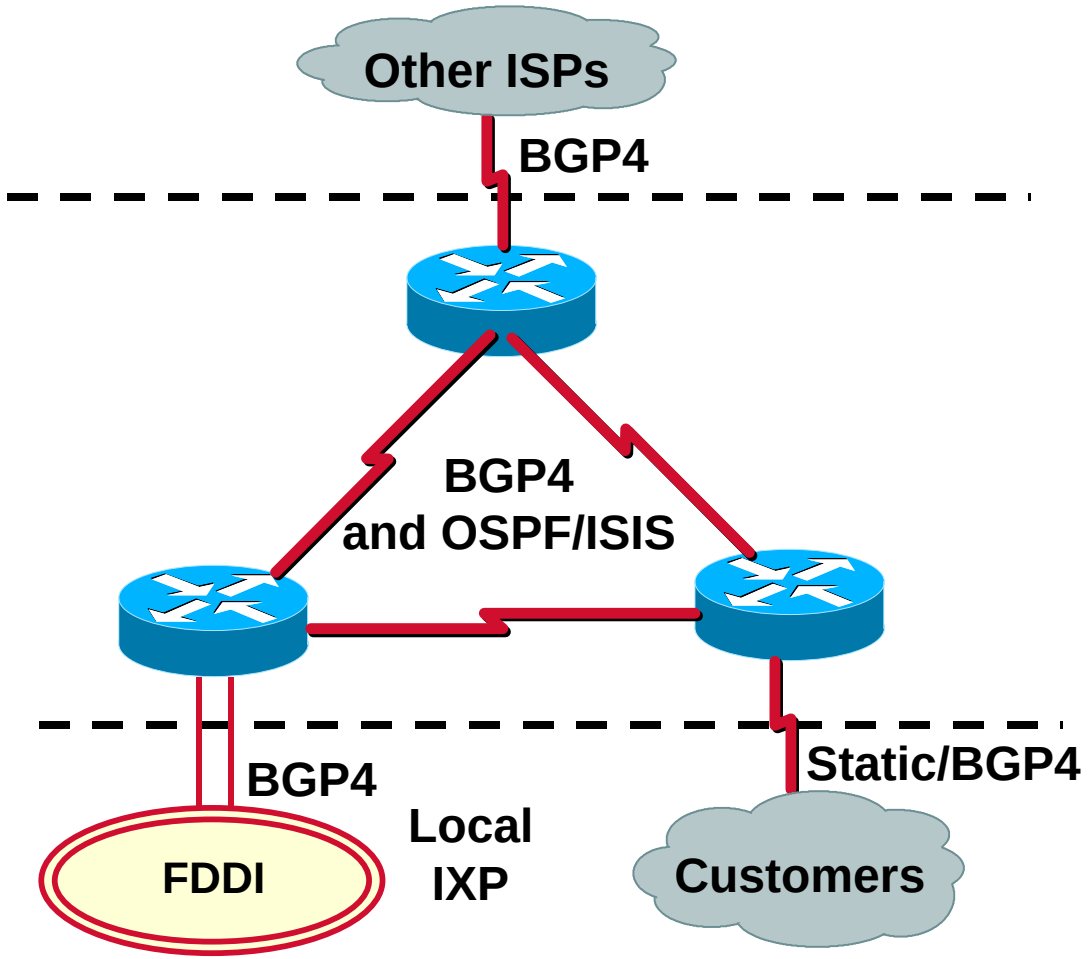
Interior versus Exterior Routing Protocols

- **Interior**
  - automatic neighbour discovery
  - generally trust your IGP routers
  - routes go to all IGP routers
  - binds routers in one AS together
- **Exterior**
  - specifically configured peers
  - connecting with outside networks
  - set administrative boundaries
  - binds AS's together

Interior versus Exterior Routing Protocols

- **Interior**
  - **Carries ISP infrastructure addresses only**
  - **ISPs aim to keep the IGP small for efficiency and scalability**
- **Exterior**
  - **Carries customer prefixes**
  - **Carries Internet prefixes**
  - **EGPs are independent of ISP network topology**

Hierarchy of Routing Protocols



Introduction to IGPs

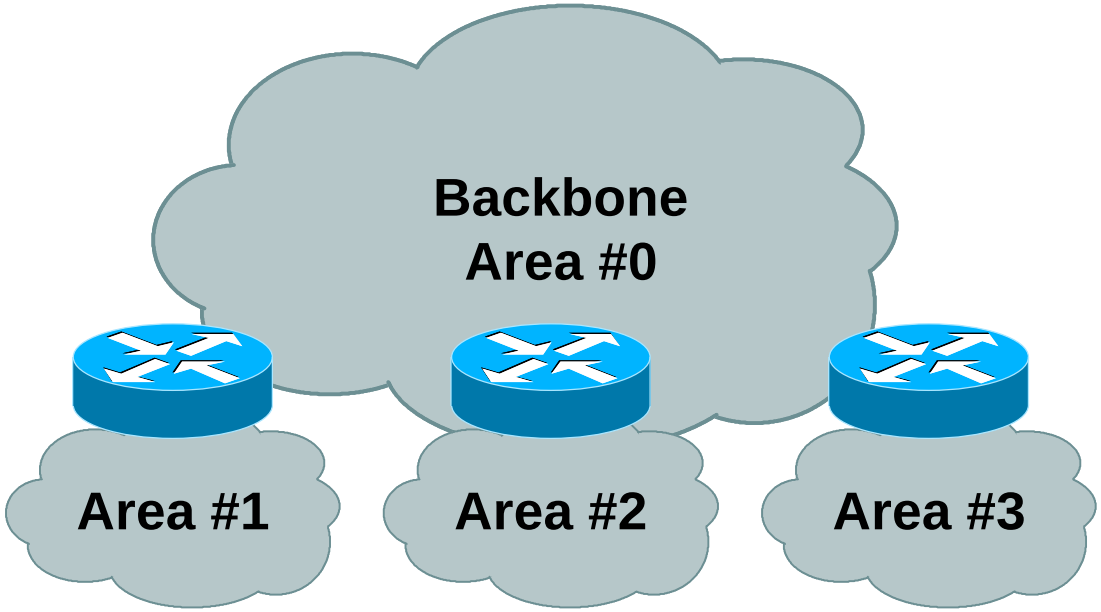
ISIS - Intermediate System to Intermediate System

- Link State Routing Protocol
- OSI development now continued in IETF
- Supports VLSM
- Low bandwidth requirements
- Supports two levels
  - The backbone (level 2) and areas (level 1)
- Route summarisation

OSPF - Open Shortest Path First

- Link State Routing Protocol
- Designed by IETF for TCP/IP - RFC2328
- Supports VLSM
- Low bandwidth requirements
- Supports different types of areas
- Route summarisation and authentication

Why Areas - OSPF Example



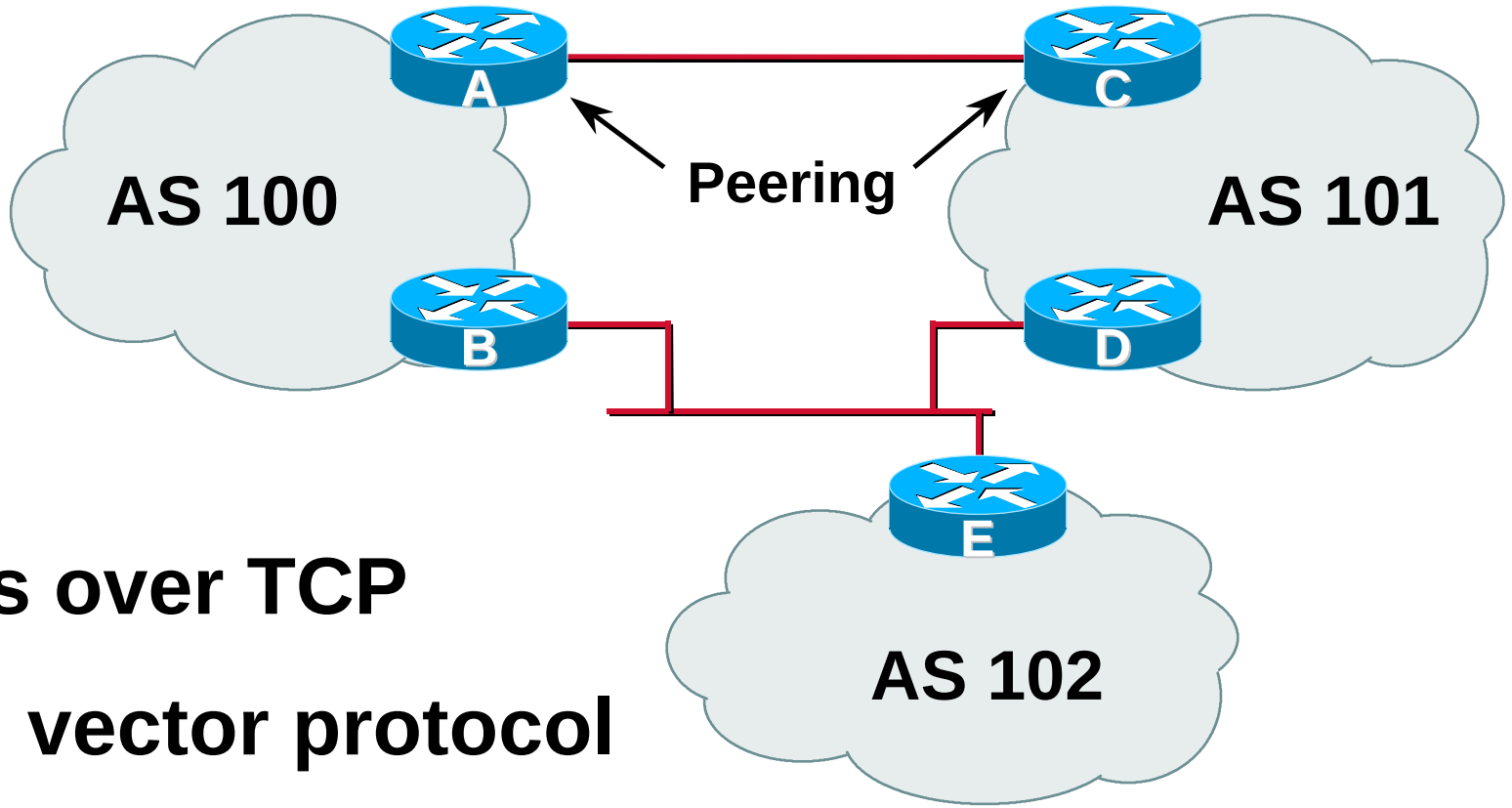
- Topology of an area is invisible from outside of the area
- Results in marked reduction in routing traffic

Scalable Network Design

- **ISIS**
  - **Implement level1 - level 2/level 1 hierarchy for large networks only**
  - **Internet friendly enhanced features**
- **OSPF**
  - **Implement area hierarchy**
  - **Enforces good network design**
- **Requires Addressing Plan**
- **Implement Route Summarisation**

BGP for ISPs

BGP Basics

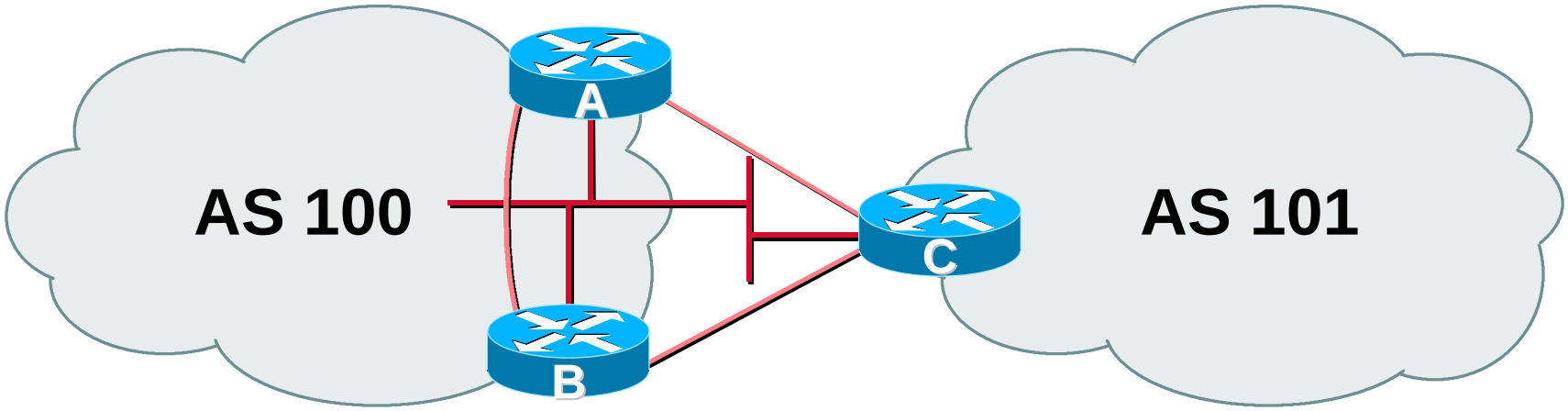


- Runs over TCP
- Path vector protocol
- Incremental update

BGP General Operation

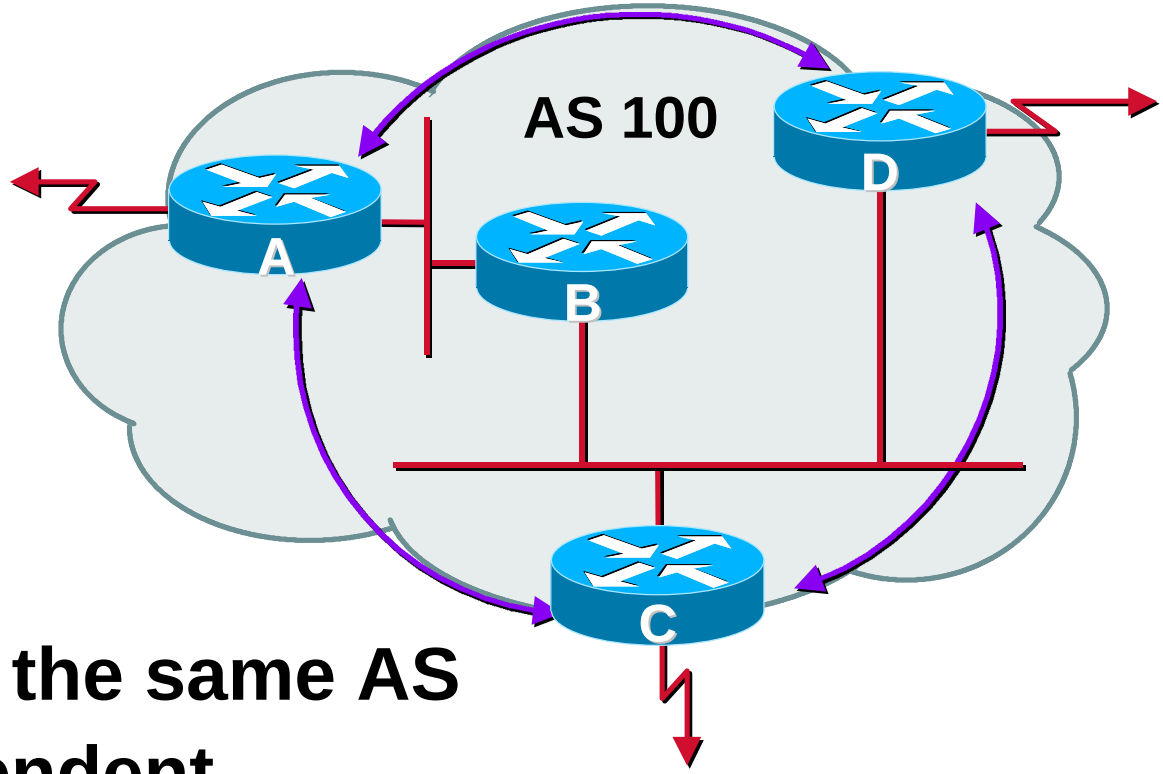
- Learns multiple paths via internal and external BGP speakers
- Picks the best path and installs in the IP forwarding table
- Policies applied by influencing the best path selection

External BGP Peering (eBGP)



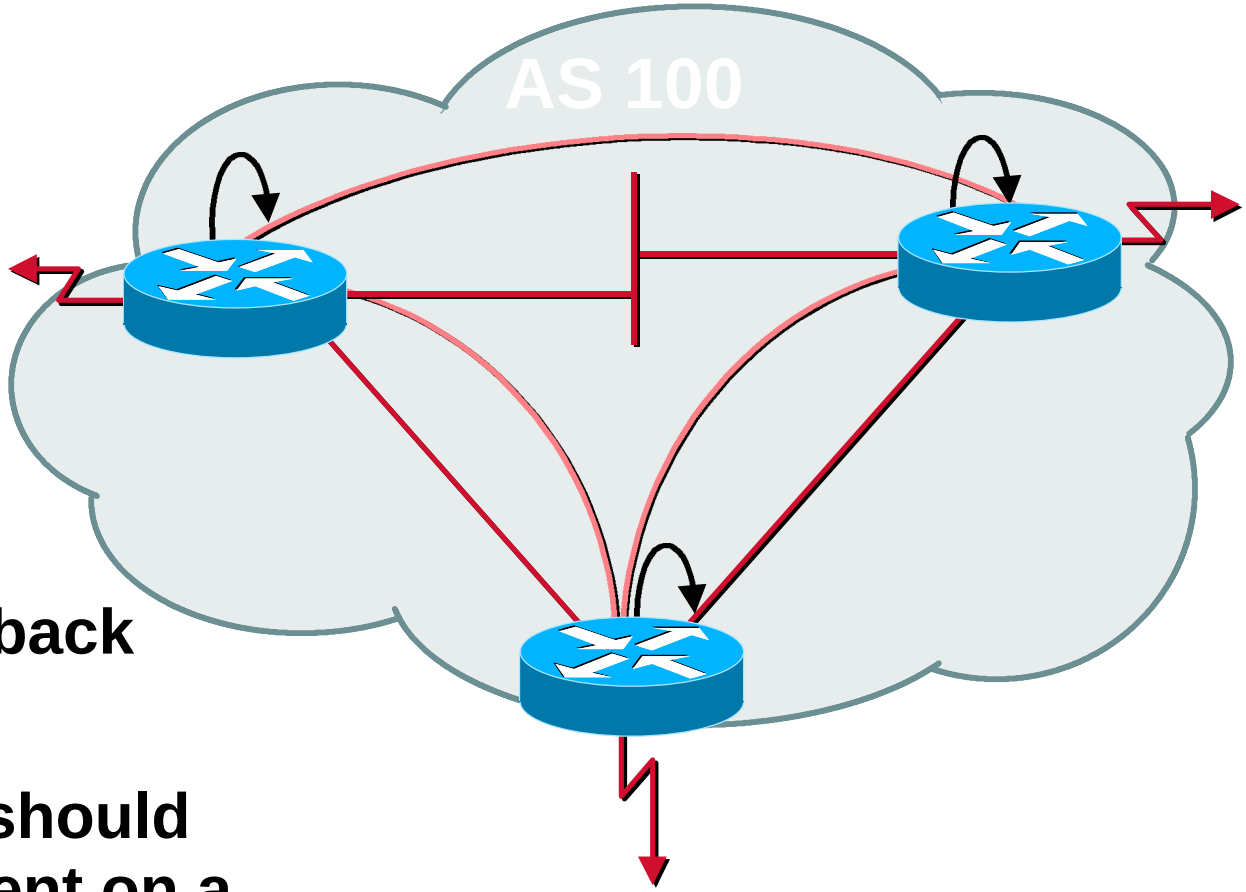
- Between BGP speakers in different AS
- Should be directly connected

Internal BGP Peering (iBGP)



- BGP peer within the same AS
- Topology independent
- Each iBGP speaker must peer with every other iBGP speaker in the AS

Stable iBGP peering - loopback interface



- **Peer with loopback interface**
- **iBGP session should not be dependent on a physical interface**

BGP Attributes

- Describes characteristics of a prefix
- Some BGP attributes:

  AS path, Next hop, Local preference, Multi-Exit Discriminator (MED), Origin, Aggregator and Community.

- Some are mandatory, some are transitive

BGP Path Selection Algorithm

- Do not consider path if no route to next hop
- Highest local preference (global within AS)
- Shortest AS path
- Lowest origin code

  IGP < EGP < incomplete

BGP Path Selection Algorithm (continued)

- **Multi-Exit Discriminator**

  Considered only if paths are from the same AS

- **Prefer eBGP path over iBGP path**
- **Path with shortest next-hop metric wins**
- **Lowest router-id**

BGP in ISP Backbones

- All routers take part in BGP
- BGP carries
  - some or all of the Internet routing table
  - customer prefixes
- IGPs are used to carry next hop and internal network information
  - recursive route lookup
- Routes are never redistributed from BGP into the IGP or from the IGP into BGP

Scaling Techniques

Bigger better networks!

Scaling Techniques

- **Administrative scaling**

  (BGP Communities)

- **Router resource scaling**

  Route Reflectors

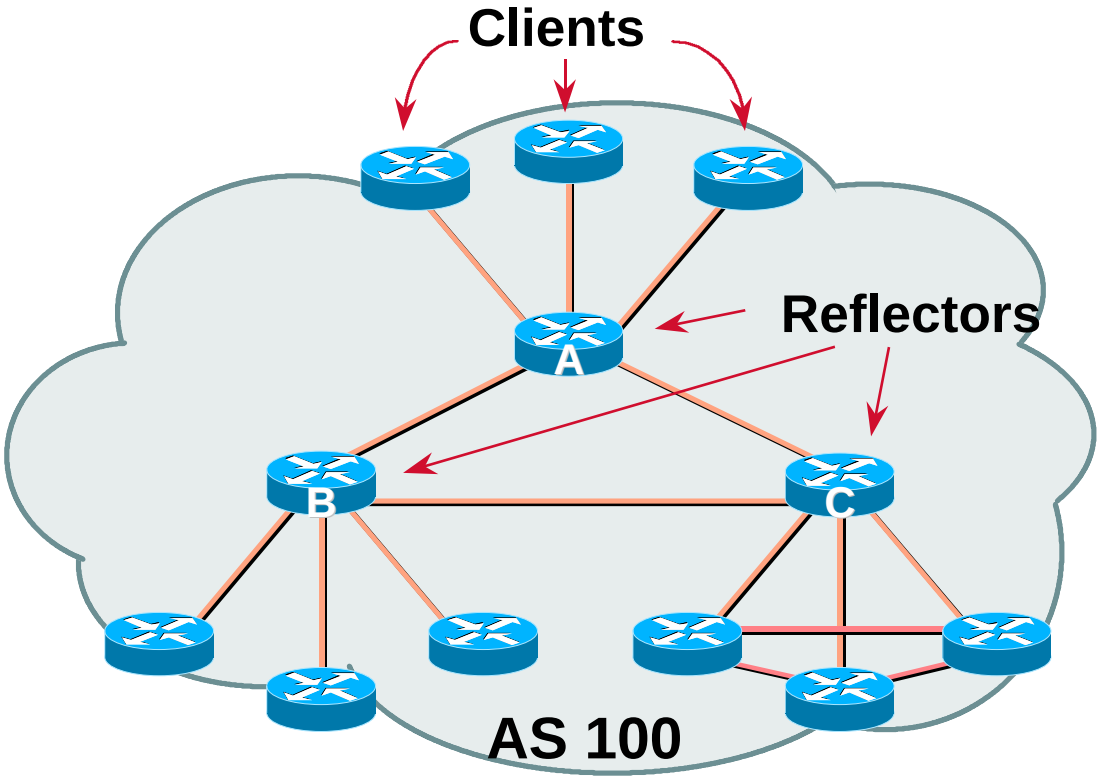
  (Confederations)

  Route Flap Dampening

  Dynamic Reconfiguration

Route Reflector

- Scalable alternative to full iBGP mesh
- Reflector receives path from clients and non-clients
- Selects best path
- Best path is from client—reflect to non-clients
- Best path is from non-client—reflect to clients
- Non-meshed clients



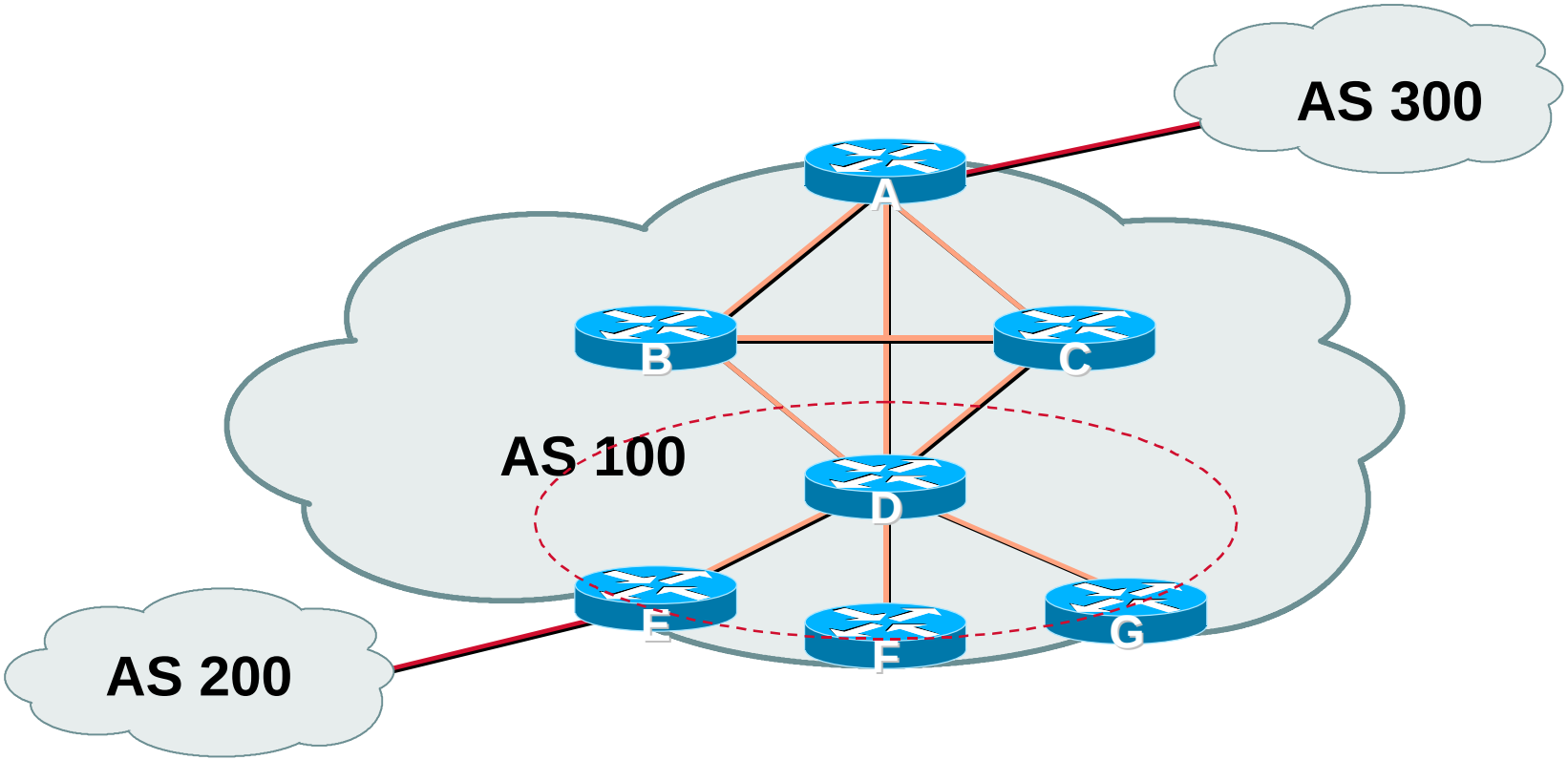
Route Reflector

- Divide the backbone into multiple clusters (hint - build on OSPF/ISIS areas)
- At least one route reflector and few clients per cluster
- Route reflectors are fully meshed
- Clients in a cluster could be fully meshed
- Single IGP to carry next hop and local routes

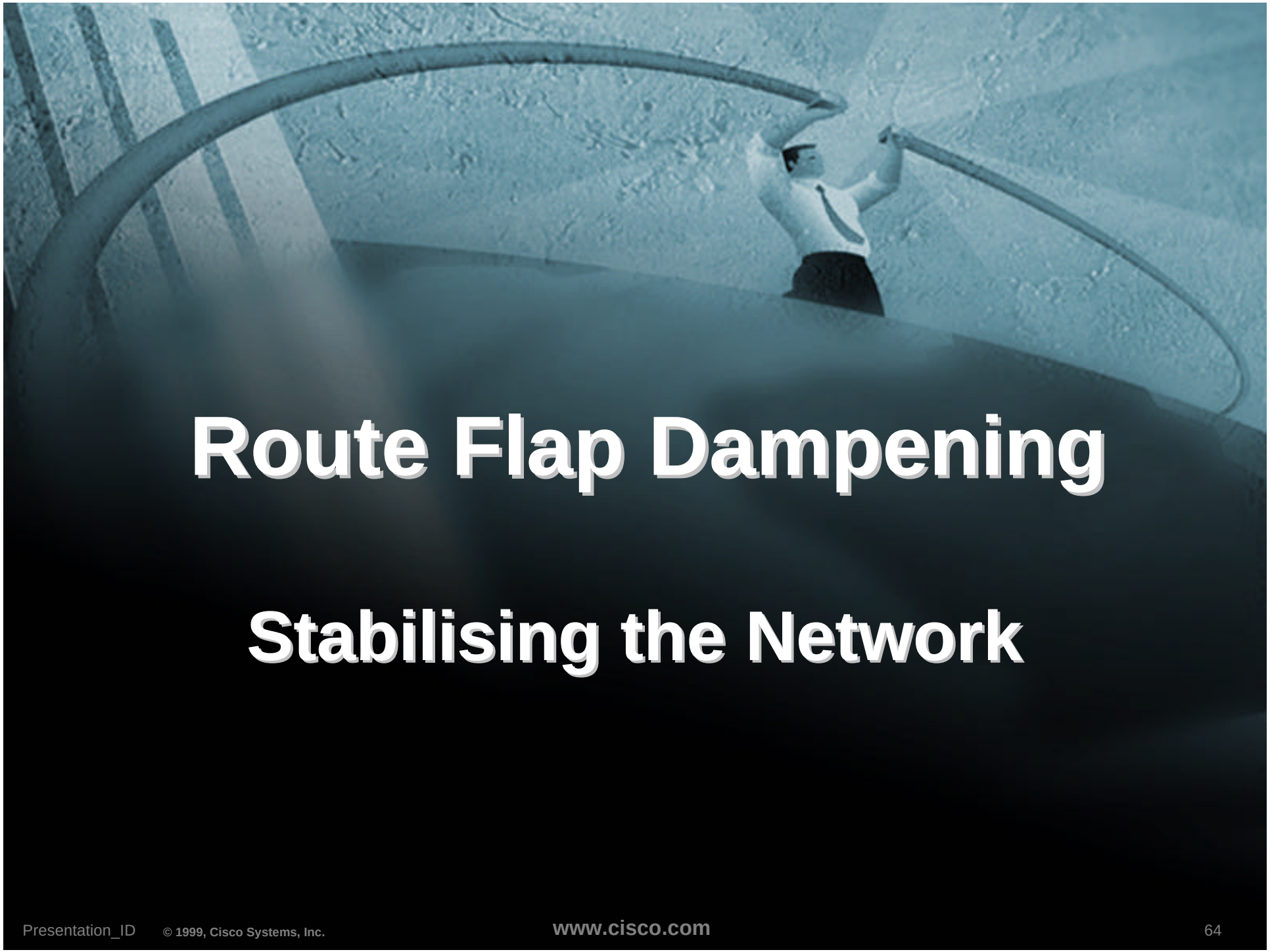
Route Reflector: Benefits

- Solves iBGP mesh problem
- Packet forwarding is not affected
- Normal BGP speakers co-exist
- Multiple reflectors for redundancy
- Easy migration
- Multiple levels of route reflectors

Route Reflector: Migration



- **Migrate small parts of the network, one part at a time.**



Route Flap Dampening

Stabilising the Network

Route Flap Dampening

- **Route flap**
  - **Going up and down of path/change in attribute**
  - **Ripples through the Internet, wastes CPU**
- **Dampening aims to reduce route flap propagation**
  - **Fast convergence for normal route changes**
  - **History predicts future behaviour**
  - **Suppress oscillating routes, advertise stable routes**
- **Described in RFC2439**

Route Flap Dampening - Operation

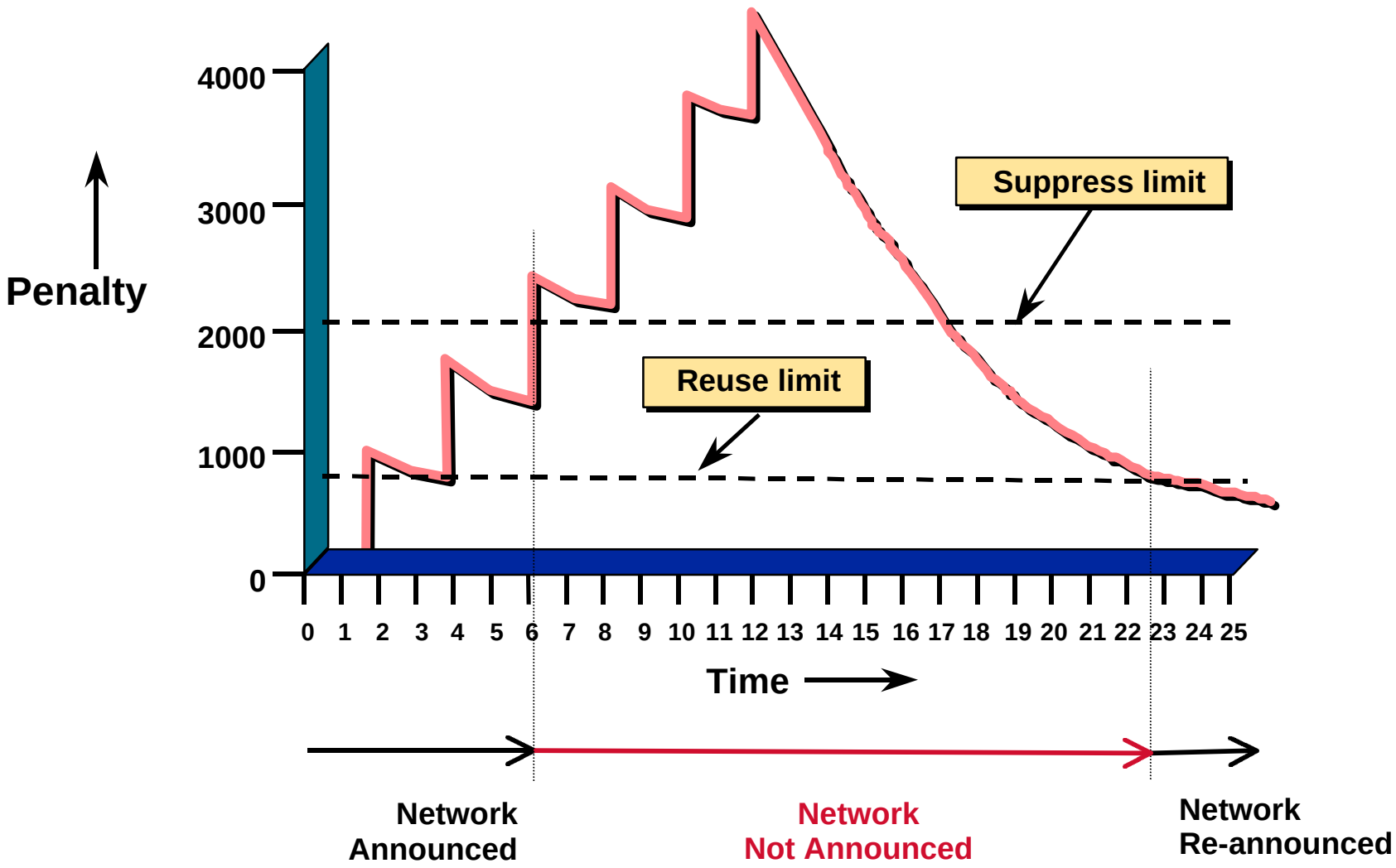
- **Add penalty (1000) for each flap**
- **Exponentially decay penalty**

  half life determines decay rate
- **Penalty above suppress-limit**

  do not advertise route to BGP peers
- **Penalty decayed below reuse-limit**

  re-advertise route to BGP peers

Route Flap Dampening



Route Flap Dampening - Operation

- **Only applied to inbound announcements from eBGP peers**
- **Alternate paths still usable**
- **Controlled by:**
  - **Half-life (default 15 minutes)**
  - **reuse-limit (default 750)**
  - **suppress-limit (default 2000)**
  - **maximum suppress time (default 30 minutes)**

Flap Dampening: Enhancements

- **Selective dampening based on**

  AS-path, Community, Prefix

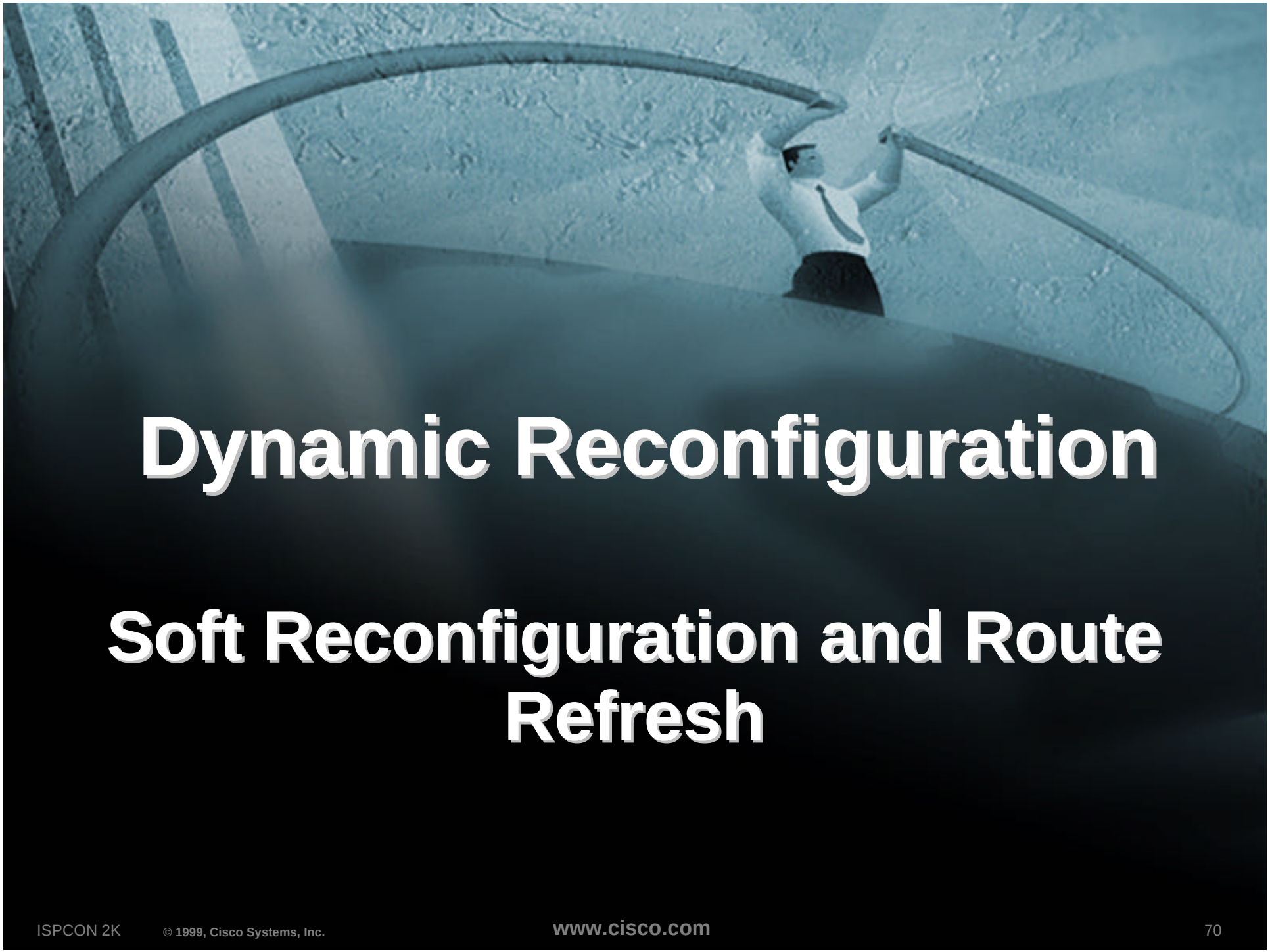
- **Variable dampening**

  recommendations for ISPs

  http://www.ripe.net/docs/ripe-210.html

- **Flap statistics**

  ```
  show ip bgp neighbor <x.x.x.x> [dampened-routes |
  flap-statistics]
  ```



Dynamic Reconfiguration

Soft Reconfiguration and Route Refresh

Soft Reconfiguration

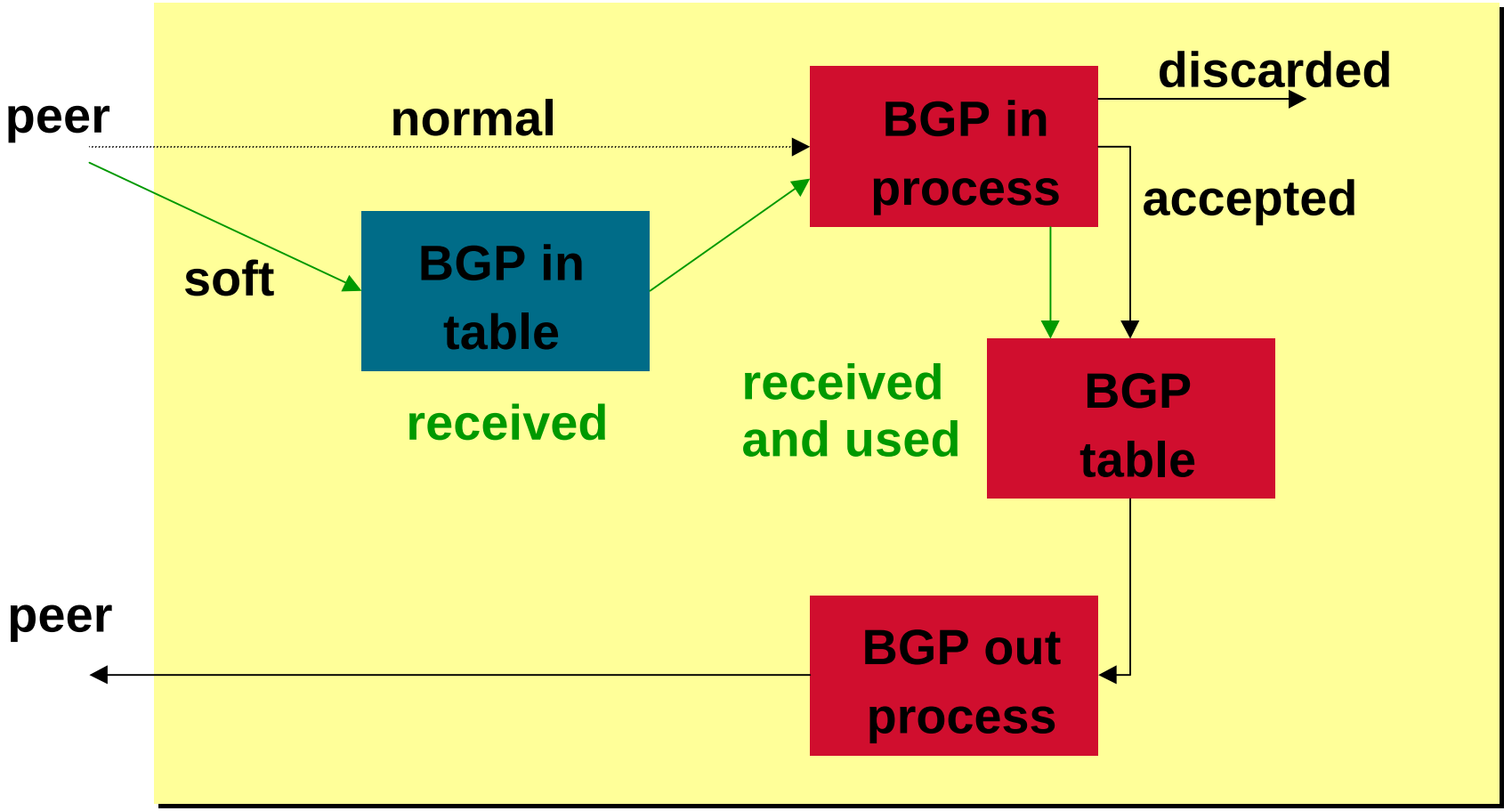
Problem:

- **Hard BGP peer clear required after every policy change because the router does not store prefixes that are denied by a filter**
- **Hard BGP peer clearing consumes CPU and affects connectivity for all networks**

Solution:

- **Soft-reconfiguration**

Soft Reconfiguration



Soft Reconfiguration

- New policy is activated without tearing down and restarting the peering session
- Per-neighbour basis
- Use more memory to keep prefixes whose attributes have been changed or have not been accepted

Configuring Soft reconfiguration

```
router bgp 100
 neighbor 1.1.1.1 remote-as 101
 neighbor 1.1.1.1 route-map infilter in
 neighbor 1.1.1.1 soft-reconfiguration inbound
```

! ***Outbound does not need to be configured*** !

Then when we change the policy, we issue an exec command

```
clear ip bgp 1.1.1.1 soft [in | out]
```

Route Refresh Capability

- Facilitates non-disruptive policy changes
- No configuration is needed
- No additional memory is used
- Requires peering routers to support “route refresh capability” - RFC2842
- clear ip bgp x.x.x.x in tells peer to resend full BGP announcement

Soft Reconfiguration vs Route Refresh

- **Use Route Refresh capability if supported**
  - **find out from “show ip bgp neighbor”**
  - **does not require additional memory**
- **Otherwise use Soft Reconfiguration**

Routing Design for ISPs

Network Design

- **Aim for simplicity, scalability and reliability**
- **Plan the network coverage**
- **Estimate growth over the next year**
- **Design the network**

Network Coverage

- Where will you start and how?
- Where will it grow?
  - One year is a long time in the Internet
  - Future PoP sites
- How big will it grow?
  - Inter-site bandwidth availability
- Does it match the business plan?

Network Design

- **Start as you mean to continue**
- **Design scalability from day one**
  - **hierarchy**
  - **separate functions**
- **Choose your IGP carefully**
  - **scalability, standards**
  - **knowledge and expertise**

Designed in Redundancy

- **Design goal should be two of everything**

  Each site should have at least two backbone WAN connections

  Consider two core routers for each backbone site

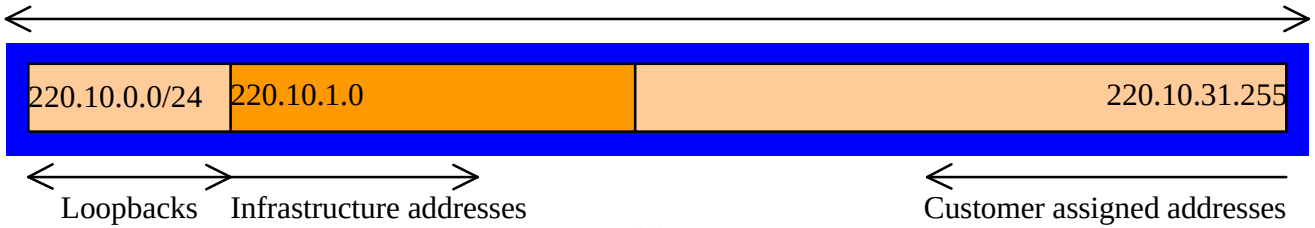
- **Out of Band management network**
- **Test lab/network**
- **Documentation!**

Address Space

- **Approach upstream ISP or consider RIR membership for address space**
- **Supply addressing plan when requested**

  remember Internet is classless

  addresses assigned according to need not want
- **Assign addresses to backbone and other network layers - remember scalability!**



Deploying IGP

- **Keep IGP small!**

  Smaller IGP, faster convergence in case of link problems

  Use BGP for customer prefixes, dial pools, and other networks

- **Use summarisation between areas of network hierarchy**
- **Use ip unnumbered where possible**

External Connections

- **Don’t need BGP from day one**

  apply for an AS and deploy BGP only when it is needed i.e. when multihoming

- **When deploying BGP**

  iBGP carries customer networks only

  IGP carries network link information only

  Do not distribute BGP routes into IGP and vice-versa

Routing Etiquette

“Problems on the Internet”

- **Concern about rate of Internet growth**

  http://www.isc.org/ds/

- **Large number of routes**

  http://www.employees.org/~tbates/cidr.plot.html

- **Routing instability**

  http://www.merit.edu/ipma/reports

- **Difficulties diagnosing problems**
- **Quality of Service??**

Effects of CIDR on Internet

- Currently around 85000 routes
- If Internet were unaggregated
  - Would be over 250000 networks
  - May have run out of IPv4 addresses
  - What size of routers required?
  - How stable would the Internet be?

CIDR - Examples

- **Must** announce network block assigned by RIR or upstream ISP
- Do **not** announce subnets of network block, or subnets of other ISPs' network blocks unless exceptional circumstances
- On Cisco routers use

  redistribute static, or aggregate-address, or network/mask pair

CIDR – Examples

Redistribute static

```
router bgp 1849
 network 194.216.0.0
 redistribute static
! Must have a matching IGP route
ip route 194.216.0.0 255.255.0.0 null0
```

Aggregate address

```
router bgp 1849
 network 194.216.0.0
 aggregate-address 194.216.0.0 255.255.0.0
! More specific route must exist in BGP table
```

Network/mask pair

```
router bgp 1849
 network 194.216.0.0 mask 255.255.0.0
! Must have a matching IGP route
ip route 194.216.0.0 255.255.0.0 null0
```

CIDR - Positive Efforts

- Most ISPs now filter all prefixes longer than /24
- Some ISPs filter according to policy registered in the Internet Routing Registry
- No aggregation or bad aggregation could result in no connectivity

Aggregation

- Announce aggregate to rest of Internet
- Put it into Routing Registry (route object)
- Keep more specifics internal to network
  - Use iBGP for carrying customer networks
  - Use IGP for carrying backbone addresses
  - Aggregate internally when possible

Aggregation - Good Example

- **Customer link goes down**
  - **their /26 network becomes unreachable**
- **/19 aggregate is still being announced**
  - **no BGP hold down problems**
  - **no BGP propagation delays**
  - **no dampening by other ISPs**

Aggregation - Good Example

- Customer link returns
- Their /26 network is visible again
- The whole Internet becomes visible immediately
- Quality of Service perception

Aggregation - Bad Example

- **Customer link goes down**
  - **Their /23 network becomes unreachable**
- **Their ISP doesn't aggregate their /19 network block**
  - **/23 network withdrawal announced to peers**
  - **starts rippling through the Internet**
  - **added load on all Internet backbone routers as network is removed from routing table**

Aggregation - Bad Example

- **Customer link returns**
  - **Their /23 network is now visible to their ISP**
  - **Their /23 network is re-advertised to peers**
  - **Starts rippling through Internet**
  - **Load on Internet backbone routers as network is reinserted into routing table**
  - **Some ISP's dampen flaps**
  - **Internet may take 10-20 min or longer to be visible**
  - **Quality of Service???**

Aggregation - Summary

- **Good example is what everyone should do!**
  - **Adds to Internet stability**
  - **Reduces size of routing table**
  - **Reduces routing churn**
  - **Improves Internet QoS for everyone**
- **Bad example is what many still do!**
  - **Laziness? Lack of knowledge?**

“The New Swamp”

- Areas of poor aggregation
- 192/3 space contributes 69000 networks - rest of Internet contributes 16000 networks

Block	Networks	Block	Networks	Block	Networks	Block	Networks
192/8	6352	200/8	2436	208/8	4804	12/8	1047
193/8	2746	201/8	0	209/8	4755	24/8	1122
194/8	2963	202/8	3712	210/8	1375	61/8	80
195/8	1689	203/8	5494	211/8	532	62/8	428
196/8	525	204/8	4694	212/8	1859	63/8	2198
197/8	0	205/8	3210	213/8	635	64/8	1439
198/8	4481	206/8	4206	214/7	14		
199/8	4084	207/8	3943	216/8	4177		

Original Swamp Cause

- Early growth of Internet
- Classful network allocation
- Small number of connected networks
- Lack of foresight by all

New Swamp Persists

- **Lazy or technically naïve ISPs**

  announcing 32 /24s rather than /19 aggregate block

  announcing customer prefixes as they connect rather than aggregate block only

- **Poorly thought out multihoming**
- **Technical solutions keep ahead of problem so far:**

  faster routers, more memory, CIDR

Solutions

- Don’t route other ISP’s address space unless in failure mode during multihoming
- Aggregate!
- Don’t announce subprefixes of your assigned block
- Be prudent when announcing small prefixes out of former A and B space

Solutions

- **Encourage other ISPs to be good citizens**

  don’t route their bad citizenship

- **Multihoming**

  fragments address space

  think carefully about set up and requirements

  load balancing versus resilience

  http://infopage.cw.net/Routing

Efforts

- **Tony Bates' CIDR report**
  - **sent to nanog, apops and eof mail lists**
- **Routing Report**
  - **sent to apops and RIPE routing-wg**
- **Regional Internet Registries**
- **Many ISPs**
- **Peer pressure**
- **YOU!**

Renumbering - motivation

- **Same as motivation for aggregation**

  holes are bad, using swamp space

- **First time Internet connection**

  legal address space, practical addressing scheme

- **New Provider**

  renumber into new provider's block

  reduces fragmentation and improves routability

Renumbering - how to?

- **PIER - Procedures for Internet and Enterprise Renumbering**

  http://www.isi.edu/div7/pier/papers.html

- **Be aware of effect on essential services**

  e.g. DNS ttl requires lowering, router filters

- **Use DHCP, secondary addressing**
- **Not difficult but needs planning**

Route Flap Dampening

- **Route Flap**

  technical description earlier

- **Many ISPs now suppress route flaps at network borders**

- **Cisco BGP Case Study at**

  http://www.cisco.com/warp/public/459/16.html

- **Recommended parameters are at**

  http://www.ripe.net/docs/ripe-210.html

Route Flap Dampening - Caution

- Be aware of potential problems
- Unreachability could be due to dampening, not disconnection
- Border routers need more memory and CPU
- Train your staff!

Filtering Policies

- **Filter announcements by peers**
  - **AS list, prefixes**
- **Only accept what is listed in routing registry**
  - **avoids configuration errors and routing problems**
  - **authorisation?**
- **Only announce what you list in routing registry**
- **Keep routing registry and filters up to date**

“Documenting Special Use Addresses” - DSUA

- **Private and Special Use addresses must be blocked on all BGP peerings, in and out:**

  http://www.ietf.org/internet-drafts/draft-manning-dsua-03.txt

```
ip prefix-list rfc1918-dsua deny 0.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny 10.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny 127.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny 169.254.0.0/16 le 32
ip prefix-list rfc1918-dsua deny 172.16.0.0/12 le 32
ip prefix-list rfc1918-dsua deny 192.0.2.0/24 le 32
ip prefix-list rfc1918-dsua deny 192.168.0.0/16 le 32
ip prefix-list rfc1918-dsua deny 224.0.0.0/3 le 32
ip prefix-list rfc1918-dsua deny 0.0.0.0/0 ge 25
ip prefix-list rfc1918-dsua permit 0.0.0.0/0 le 32
```

The Internet Routing Registry

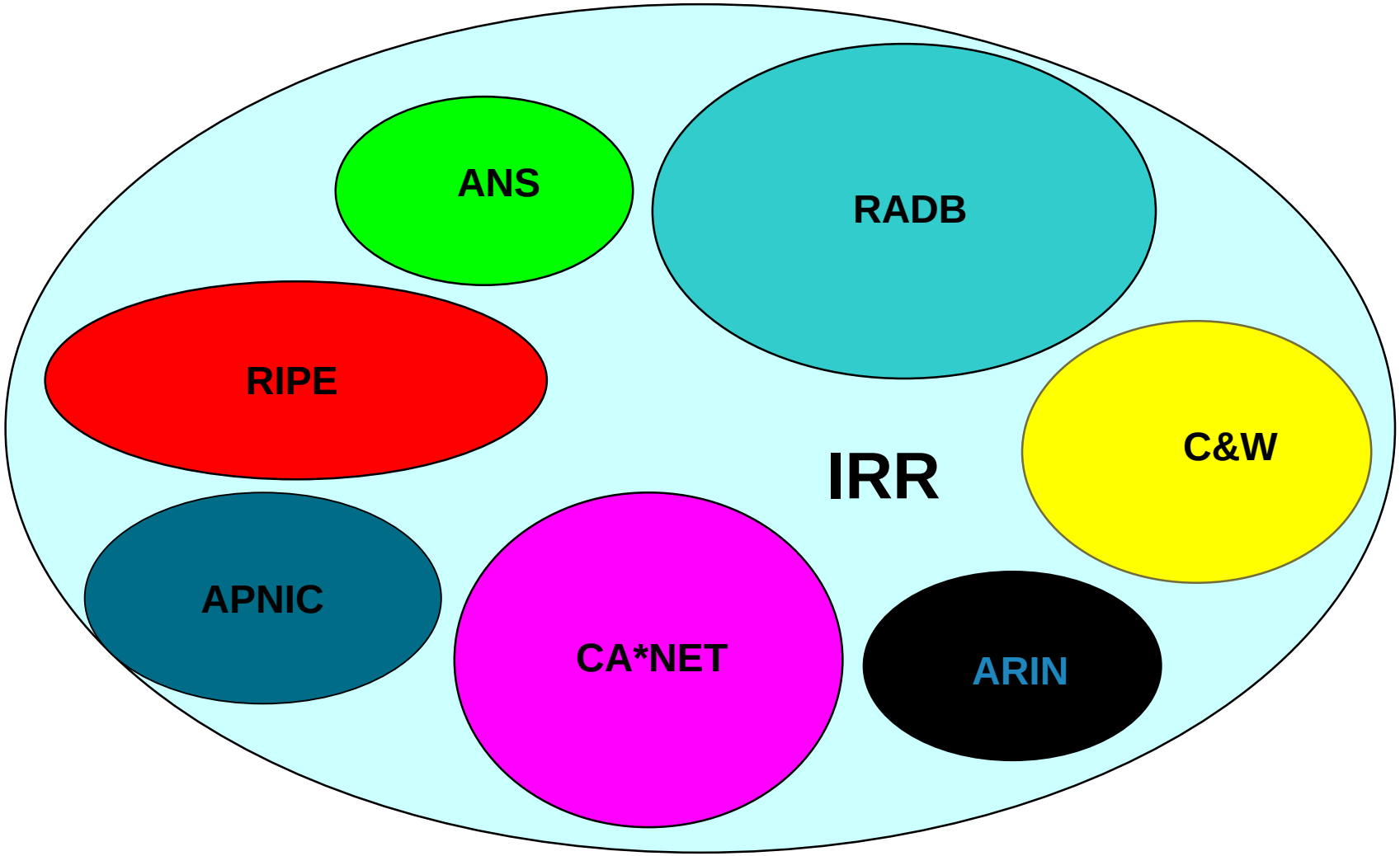
Definition

- **“A public authoritative distributed repository of routing information”**
  - **Public databases**
  - **Distributed repository of information**
  - **Have authoritative data**
  - **Vendor independent**

Composition

- Routing Policy Details
- Routes and their aggregates
- Topology Linking AS's
- Network components such as routers
- Is separate from other information such as domains and networks

Entities of the IRR



Relationship Table

Registry	Routing Policy	Routes	Networks	Domains
APNIC	**Yes**	No	Yes	No
RIPE	**Yes**	**Yes**	Yes	Yes
RADB	**Yes**	**Yes**	No	No
C&W	**Yes**	**Yes**	No	No
ANS	**Yes**	**Yes**	No	No
CA*NET	**Yes**	**Yes**	No	No
ARIN	**Yes**	**Yes**	Yes	No
“InterNIC”	No	No	No	Yes

Relationships

- C&W, ANS and CA*Net - provider run RRs
- RIPE RR - European providers
- ARIN RR - launched 8 February 1999
- RADB - Default RR for rest of world
- APNIC - plans to be full member of IRR very soon.

Benefits of an IRR

- Operational Support
- Information
- Configuration
- Problem diagnosis
- Improved Service Quality
- Tools for consistency checking

Information

- Routing policy repository
- “Map of global routing topology”
- Routing policy between neighbouring AS’s
- Device independent description of routing policy

Configuration

- Supports network filtering
- Configures routers and policies
- Revision control
- Sanity checking
- Simulation

Improved Quality of Service

All this adds up to improved quality of service

Participation is essential!

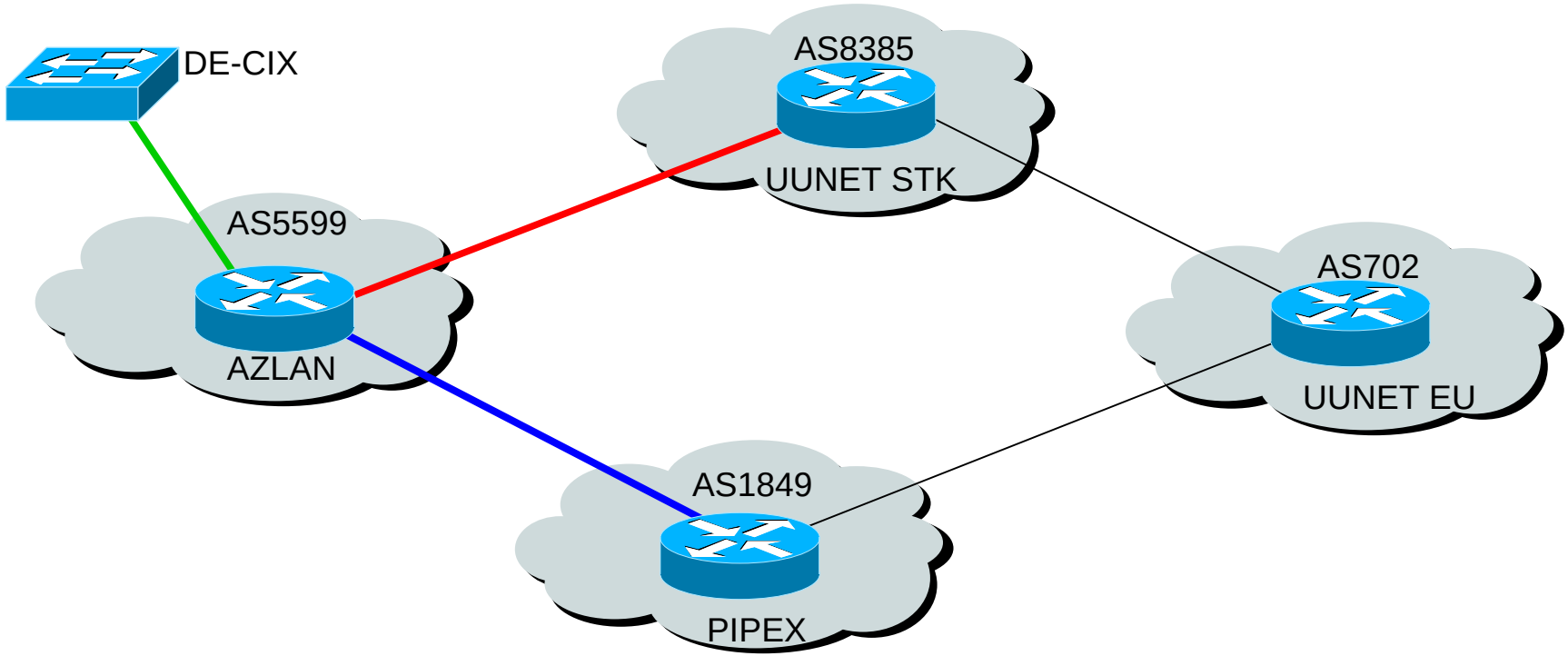
Key Objects and Syntax of RIPE-181

- Representation
- AS Object
- AS Macro
- Route Object
- Authorisation - Maintainer Object

Representation

- ASCII printable
- Attributes by tag:value lines
- Objects separated by empty lines
- RIPE-181
- RPSL (not covered)

Real World Example!



AS-Object

```
aut-num:    AS5599
descr:      Azlan Scandinavia
descr:      Internet Business Unit
descr:      Glostrup NOC
as-in:      from AS1849 100 accept AS-PIPEXEURO
as-in:      from AS1835 100 accept AS1835
as-in:      from AS2863 100 accept AS2863
as-in:      from AS3292 100 accept AS-DKNET AS3292
as-in:      from AS3308 100 accept AS3308
as-in:      from AS5492 100 accept AS5492
as-in:      from AS5509 100 accept AS5509
as-in:      from AS6785 100 accept AS6785
as-in:      from AS6834 100 accept AS6834
as-in:      from AS8526 100 accept AS8526
as-in:      from AS8385 100 accept {146.188.0.0/16}
```

```
as-out:     to AS1849 announce AS5599
as-out:     to AS1835 announce AS5599
as-out:     to AS2863 announce AS5599
as-out:     to AS3292 announce AS5599
as-out:     to AS3308 announce AS5599
as-out:     to AS5492 announce AS5599
as-out:     to AS5509 announce AS5599
as-out:     to AS6785 announce AS5599
as-out:     to AS6834 announce AS5599
as-out:     to AS8526 announce AS5599
as-out:     to AS8385 announce AS5599
default:    AS8385 100
admin-c:    MW89-RIPE
tech-c:     KE30-RIPE
mnt-by:     AS5599-MNT
changed:    klaus@azlan.net 970207
changed:    klaus@azlan.net 971209
source:     RIPE
```

Connection to exchange point
Connection transit provider
Connection to backup provider

Syntax for AS Object

- **Can represent policy using**
  - **Boolean expressions (AND, OR, NOT)**
  - **Keyword ANY - means “everything”**
  - **Communities and AS Macros**
  - **Route lists - {prefixes}**
  - **Cost to indicate preference**
  - **Attribute DEFAULT - accept 0.0.0.0**

Fields in AS Object

- **Mandatory Fields**

  aut-num, descr, admin-c, tech-c, mnt-by, changed, source, as-in, as-out

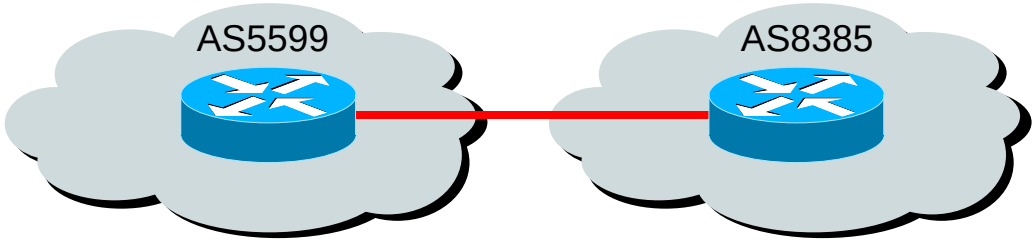
- **Optional Fields**

  as-name, interas-in, interas-out, as-exclude, default, guardian, remarks, notify

IP Routing Policy

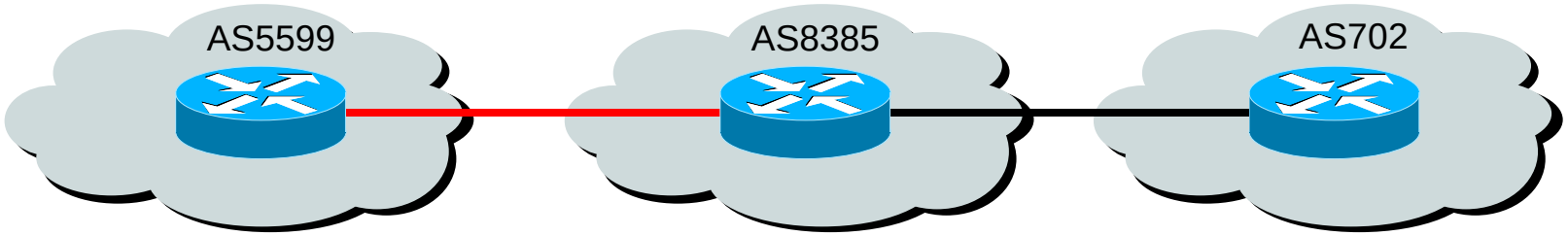
- Relationship between AS's
- What to announce to each neighbour
- What to accept from each neighbour
- Selection between multiple paths
- Preferred paths
- Use default route?

Basic Policy Example




```
aut-num:    AS5599
as-in:      from AS8385 100 accept {146.188.0.0/16}
as-out:     to AS8385 announce AS5599

aut-num:    AS8385
as-in:      from AS5599 100 accept AS5599
as-out:     to AS5599 announce {146.188.0.0/16}
```

Transit Policy Example



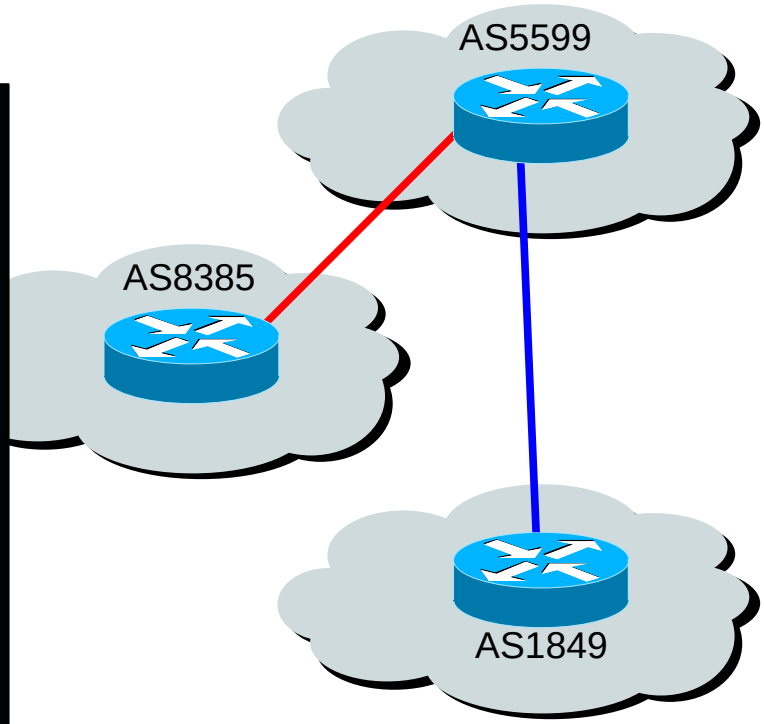

```
aut-num: AS8385
as-in:     from AS702 100 accept ANY
as-in:     from AS5599 100 accept AS5599
as-out:    to AS702 announce AS8385 AS5599 AS8473 AND NOT {0.0.0.0/0}
as-out:    to AS5599 announce {146.188.0.0/16}
default:   AS702     50 {146.188.0.0/16}

aut-num: AS702
as-in:     from AS8385 100 accept AS8385 AS5599 AS8473
as-out:    to AS8385 announce ANY
```

Multihoming Policy Example

aut-num: AS5599
as-in: from AS1849 100 accept AS-PIPEXEURO
as-in: from AS8385 100 accept {146.188.0.0/16}
as-out: to AS8385 announce AS5599
as-out: to AS1849 announce AS5599

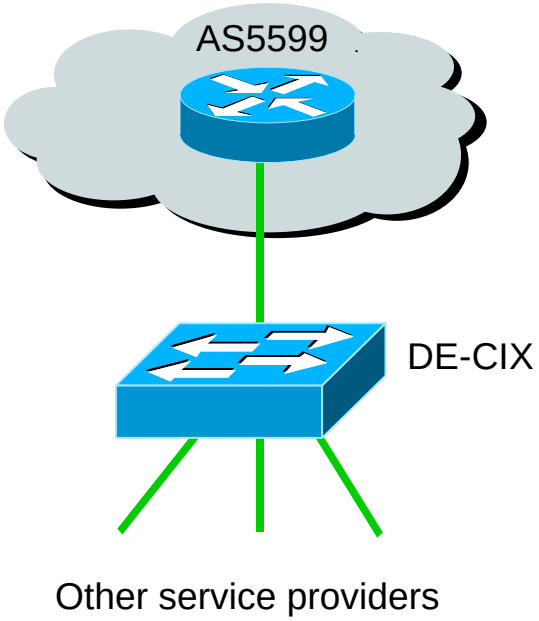
aut-num: AS1849
as-in: from AS5599 100 accept AS5599
as-out: to AS5599 announce AS-PIPEXEURO

aut-num: AS8385
as-out: to AS5599 announce {146.188.0.0/16}
as-in: from AS5599 100 accept AS5599



Exchange Point Policy Example

```
aut-num:    AS5599
as-out:     to AS1835 announce AS5599
as-out:     to AS2863 announce AS5599
as-out:     to AS3292 announce AS5599
as-out:     to AS3308 announce AS5599
as-out:     to AS5492 announce AS5599
as-out:     to AS5509 announce AS5599
as-out:     to AS6785 announce AS5599
as-out:     to AS6834 announce AS5599
as-out:     to AS8526 announce AS5599
```



AS Macro

- Collection of AS's or other AS macros
- Describes membership of a set
- Contains no policy info
- Scales better
- Can differentiate between customer and peer routes

Fields in AS Macro

- **Mandatory Fields**

  as-macro, descr, as-list, tech-c, admin-c, mnt-by, changed, source

- **Optional Fields**

  guardian, remarks, notify

AS Macro

as-macro:	AS-UUNETSTK
descr:	UUNET customer routes in Stockholm
as-list:	AS-TAIDE
as-list:	AS-KOLUMBUS
as-list:	AS1759
as-list:	AS8385
as-list:	AS702
tech-c:	KCH251
admin-c:	ES199
remarks:	AS702 Stockholm routes are community tagged
notify:	intl-net-eng@uu.net
mnt-by:	UUNET-MNT
changed:	annel@uu.net 971113
source:	RIPE

Used in

aut-num:	AS702
as-out:	to AS1759 announce AS-UUNETSTK

Route Object

- Represents a route in the Internet
- Contains all membership information
- Only one origin possible
- Classless (should be aggregated)
- Can support holes and withdrawn

Fields in Route Object

- **Mandatory Fields**

  route, descr, origin, mnt-by, changed, source

- **Optional Fields**

  hole, withdrawn, comm-list, remarks, notify

- **Example:**

```
route:      195.129.0.0/19
descr:      UUNET-NET
origin:     AS702
remarks:    UUNET filter inbound on prefixes longer than /24
notify:     intl-net-eng@uu.net
mnt-by:     UUNET-MNT
changed:    annel@uu.net 970501
source:     RIPE
```

Route Object

route:	**194.216.0.0/16**
descr:	**PIPEX-BLOCK194216**
origin:	**AS1849**
hole:	**194.216.59.0/24**
remarks:	**UUNET UK filter inbound on prefixes longer than /24**
mnt-by:	**AS1849-MNT**
changed:	**philip@uk.uu.net 19980107**
source:	**RIPE**

```
stk-gw1>show ip bgp 194.216.0.0 255.255.0.0 longer-prefixes
BGP table version is 53607058, local router ID is 195.242.36.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 194.216.0.0/16   146.188.30.162                       0    702 1849 i
*> 194.216.59.0     146.188.30.162                       0    702 701 3491 5557 i
```

How to register and update information in the IRR

- **Frequently used objects**
- **Update procedures**
  - **Modifying Objects**
  - **Deleting Objects**
  - **Submitting Objects**
  - **Authorisation/Notification**
  - **Errors and Warnings**
  - **NIC handles**

Frequently Used Objects

- Person - contact person
- Maintainer - authorisation of objects
- Inetnum - address assignment
- Aut-num - autonomous systems
- AS-macro - set of AS's
- Route - announced routes

Unique Keys

- Uniquely identifies an object
- Updating object overwrites old entry - need unique key
- Used in querying whois
- Web based full text searches available now, e.g.

  http://whois.apnic.net/apnic-bin/whois.pl

Unique Keys

- Person - name plus NIC handle
- Maintainer - maintainer name
- Inetnum - network number
- Aut-num - AS number
- AS-macro - AS macro name
- Route - route value plus origin

Modifying an Object

Before

person:	**Philip F. Smith**
address:	UUNET UK
address:	Internet House
address:	332 Science Park
address:	Milton Road
address:	Cambridge CB4 4BZ
address:	England, UK
phone:	+44 1223 250100
fax-no:	+44 1223 250101
e-mail:	philip@uk.uu.net
nic-hdl:	**PFS2-RIPE**
notify:	philip@uk.uu.net
changed:	philip@uk.uu.net 19971202
source:	RIPE

Submitted and After

person:	**Philip F. Smith**
address:	Cisco Systems Australia
address:	Level 8, 80 Albert Street
address:	Brisbane 4000
address:	QLD
address:	Australia
phone:	+61 7 3238 8200
fax-no:	+61 7 3211 3889
e-mail:	pfs@cisco.com
e-mail:	philip@dial.pipex.com
nic-hdl:	**PFS2-RIPE**
notify:	philip@dial.pipex.com
changed:	pfs@cisco.com 19980209
source:	RIPE

- **Unique keys must stay the same**
- **Remember to use current date**
- **NIC handle mandatory**

Deleting an Object

```
person:    Philip F. Smith
address:   UUNET UK
address:   332 Science Park
address:   Milton Road
address:   Cambridge
address:   England, UK
phone:     +44 1223 250100
fax-no:    +44 1223 250101
e-mail:    philip@uk.uu.net
nic-hdl:   PFS2-RIPE
notify:    philip@uk.uu.net
changed:   philip@uk.uu.net 19971202
source:    RIPE
delete:    philip@dial.pipex.com left company
```

- **delete deletes object from database**
- **current object must be submitted exactly as is, only with extra delete line**
- **If there is a mnt-by line, need the password!**

Submitting Objects

- **Email Interface - eg APNIC**

  auto-dbm@apnic.net

  Robot mail box

  Send all database updates to this mailbox

  Can use LONGACK and HELP in the subject line

  apnic-dbm@apnic.net

  human mailbox

  questions on the database process

Authorisation/Notification

```
route:      194.216.0.0/16
descr:      PIPEX-BLOCK194216
origin:     AS1849
hole:       194.216.59.0/24
remarks:    UUNET UK filter inbound on prefixes longer than /24
mnt-by:     AS1849-MNT
notify:     support@uk.uu.net
changed:    philip@uk.uu.net 19980107
source:     RIPE
```

- **mnt-by the maintainer object**
- **notify who is notified of changes**

Maintainer Object

- **Who is authorised**
- **Authorisation Method**

  email-from and crypt-pw

- **Mandatory Fields**

  mntner, descr, admin-c, tech-c, upd-to, auth, mnt-by

- **Optional Fields**

  mnt-nfy, changed, notify, source

Maintainer Object

Maintainer Object AS1849-MNT

```
mntner:   AS1849-MNT
descr:    AS 1849 Maintainer - PIPEX UK
admin-c:  PFS2-RIPE
tech-c:   PFS2-RIPE
upd-to:   philip@uk.uu.net
mnt-nfy:  netdev@uk.uu.net
auth:     CRYPT-PW fjOImdmwKsx
mnt-by:   AS1849-MNT
changed:  philip@uk.uu.net 19980109
source:   RIPE
```

Object has to be registered manually

Authorisation/Notification

```
route:      194.216.0.0/16
descr:      PIPEX-BLOCK194216
origin:     AS1849
hole:       194.216.59.0/24
hole:       194.216.136.0/23
remarks:    UUNET UK filter inbound on prefixes longer than /24
mnt-by:     AS1849-MNT
passwd:     c4Ange5
notify:     support@uk.uu.net
changed:    philip@uk.uu.net 19980109
source:     RIPE
```

- New hole to be added.
- passwd field to allow change
- <support@uk.uu.net> will be notified of this change
- updated changed field

Warnings and Errors

- **Warnings**
  - **Object corrected then accepted**
  - **Notification of action taken sent in acknowledgement**
- **Errors**
  - **Object not corrected and not accepted**
  - **Diagnostics in acknowledgement**
- **Syntax checking is very strict**

NIC Handles

```
mntner:     AS1849-MNT
descr:      AS 1849 Maintainer - PIPEX UK
admin-c:    PFS2-RIPE
tech-c:     PFS2-RIPE
upd-to:     philip@uk.uu.net
mnt-nfy:    netdev@uk.uu.net
auth:       CRYPT-PW fjOImdmwKsx
mnt-by:     AS1849-MNT
changed:    philip@uk.uu.net 19980109
source:     RIPE
```

- **PFS2-RIPE is the NIC Handle of the person**
- **Only way of avoiding ambiguity in person objects**
- **Mandatory**
- **Format: <initials><number>-<regional registry>**
- **Local differences for obtaining NIC Handles.**

What tools and resources?

- **RAToolset**

  www.isi.edu/ra/RAToolSet

- **RIPE whois**

  ftp.ripe.net/ripe/tools

- **Looking Glasses**

  nitrous.digex.net

RAToolSet

- Runs on most Unix platforms
- Requires g++, tcl and tk
- Excellent for housekeeping, debugging and configuration

RAToolSet Tools

- **RTconfig**

  Generate router configurations

- **AOE - aut-num object editor**

  update aut-num, as-macro objects

- **ROE - route-object editor**

  update route-object

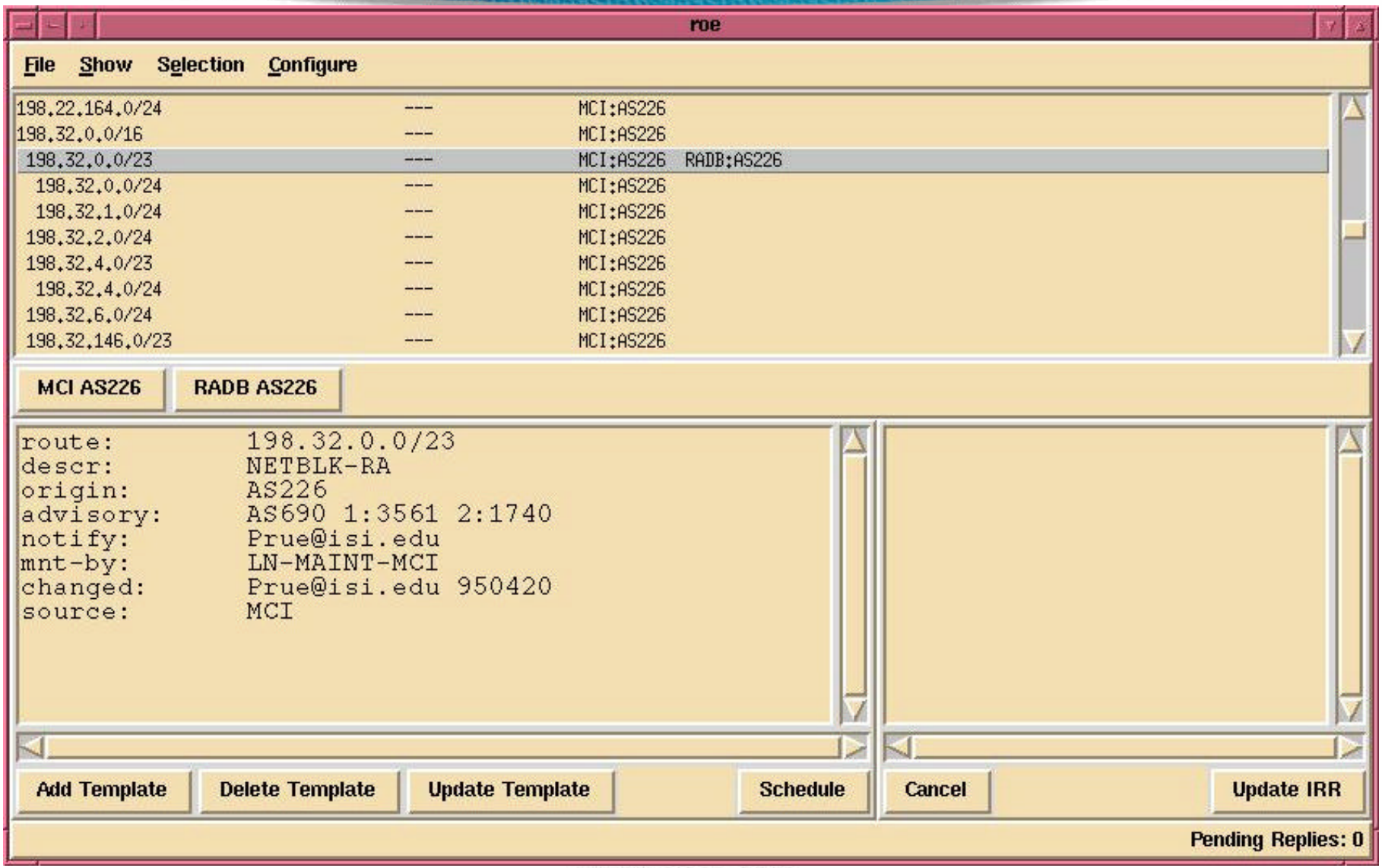
- **CIDRadvisor**

  advice on CIDRisation

ROE Uses

- **Route object editor used to:**
  - **check for consistency of route objects in IRRs**
  - **synchronise route object entries in different IRRs**
  - **detect missing or unwanted route objects**

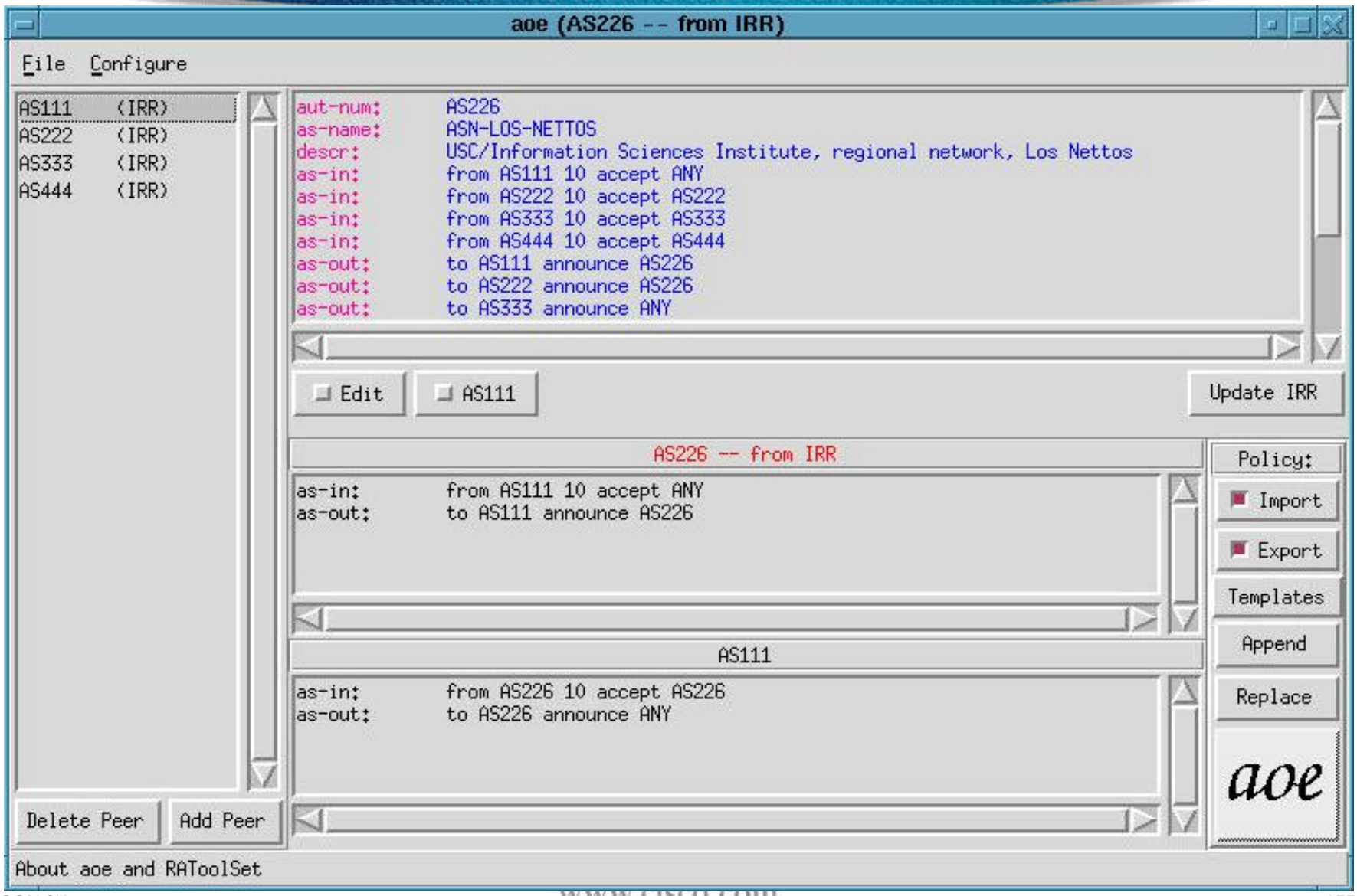
ROE example



AOE Uses

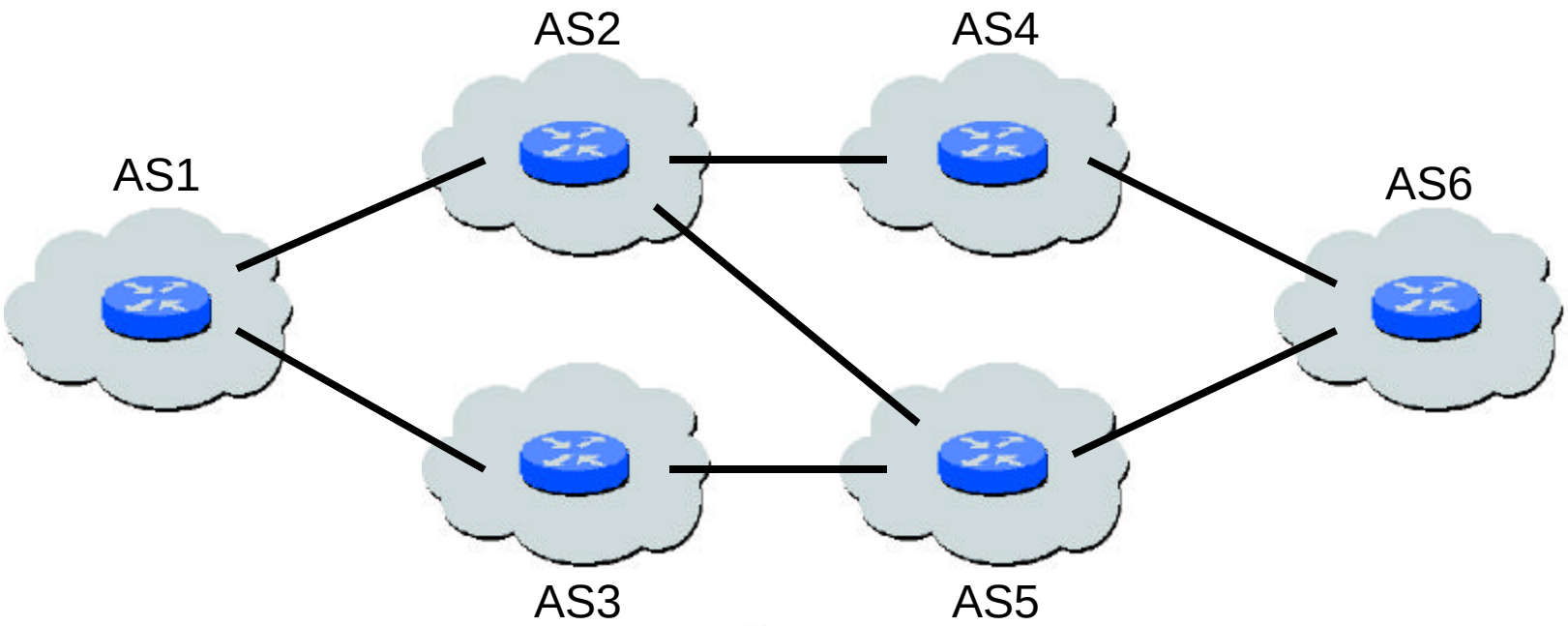
- **AS Object editor used to:**
  - **generate AS objects and policies as-in and as-out**
  - **check policies listed in AS object on the IRRs**
  - **check policies according to BGP dump**

AOE example



PRtraceroute

- **PRIDE modified traceroute which includes AS information and a comparison between the real route and the route according to the IRR.**
- **Cisco IOS trace command refers to BGP table**



PRtraceroute Example

```
% prtraceroute -lv collegepk-cr9.bbnplanet.net
traceroute with AS and policy additions [Jan 13 20:21:19 UTC]

               from  AS109 lovefm.cisco.com (171.68.228.35)
               to     AS86 collegepk-cr9.bbnplanet.net (192.239.103.9)

 1  AS109  al.cisco.com                    171.68.228.3     [I] 4 1 1 ms
 2  AS109  acorn.cisco.com                 171.68.0.134     [I] 2 1 1 ms
 3  AS109  gaza-gw2.cisco.com              171.68.0.91      [I] 2 1 1 ms
 4  AS109  sj-wall-2.cisco.com             198.92.1.138     [I] 3 3 2 ms
 5  AS109  barrnet-gw.cisco.com            192.31.7.37      [I] 4 3 2 ms
 6  AS200  paloalto-cisco.bbnplanet.net    131.119.26.9     [?] 4 4 3 ms
 7  AS200  paloalto-br1.bbnplanet.net      131.119.0.193    [I] 7 8 7 ms
 8    AS1  chicago2-br1.bbnplanet.net      4.0.1.2          [E1] 58 59 58 ms
 9    AS1  collegepk-br1.bbnplanet.net     4.0.1.6          [I] 82 73 75 ms
10   AS86  collegepk-cr9.bbnplanet.net     128.167.252.9    [E1] 86 81 ms

AS Path followed: AS109 AS200 AS1 AS86

 AS109 = Cisco Systems
 AS200 = BBN Planet Western Region
   AS1 = BBN Planet backbone
  AS86 = SURAnet Northern AS
```

ERROR	hop should not have been taken
NH ASx	possible NEXT_HOP followed
I	intra AS hop
En	nth choice inter AS hop
Dn	nth choice default hop
C	connected hop
?	No information in IRR

RIPE whois client

- Runs on most (UNIX) platforms
- Easy to install
- Can use to query all other IRR's
- Expanded whois functionality
- Good for housekeeping, debugging, operations
- RECOMMENDED!

Open Issues

- **Why isn't the IRR used more today?**
  - **Ignorance?**
  - **Education?**
  - **Security fears?**
  - **No local routing registry?**
- **What tools are missing?**

Tool Availability

- **Should software be available as a commercial package?**
  - **Better bundled/supported/debugged?**
  - **Better integration/training?**
- **Most tools are freely available public efforts for the good of the “community”**

Routing Registries

- **Belief that the Internet works with out the IRR.**
  - **It does but for how much longer?**
  - **Many ISPs rely on the data kept in the registry**
  - **Subset of tools available are being used on a daily basis**

Awareness & Training

- Is there enough awareness about Internet routability?
- Is there enough training on the promotion of routability
- Headcount requirement
  - depends on organisation
  - too easy and cheaper to be irresponsible
- Overall organisational awareness of the issues ® overall efficiency, quality of service and support

Ways forward

- **Routing Registry enhancements**
  - **RPSL matches today's BGP capabilities**
- **Feedback on tool enhancements**
- **Feedback to vendors on equipment configuration enhancements**
- **More training, more education, more feedback!**

Summary

- ISP networks and terminology
- The application of IGPs and BGP in an Internet network
- Shown tools which help diagnose and solve routing problems more easily
- Application of routing registries

Summary

- Made you more aware of the issues facing the Internet today
- Showed you how to make a positive contribution to the functioning of the Internet
- Promoted Routability!
- Any questions?

Useful URL's & Reading

1. CIDR
   - ftp://ftp.isi.edu/in-notes/rfc{1517,1518,1519}.txt
   - http://www.ibm.net.il/~hank/cidr.html
   - ftp://ftp.uninett.no/pub/misc/eidnes-cidr.ps.Z
   - Network addressing when using CIDR
2. AS numbers
   - ftp://ftp.isi.edu/in-notes/rfc1930.txt
   - Guidelines for creation, selection, and registration of an AS
3. Address Allocation and Private Internets
   - ftp://ftp.isi.edu/in-notes/rfc1918.txt
4. BGP Dampening
   - http://www.cisco.com/warp/public/459/16.html
   - ftp://ftp.ripe.net/ripe/docs/ripe-210.txt
   - European recommendations for route flap dampening
   - ftp://engr.ans.net/pub/slides/nanog/feb-1995/route-dampen.ps
5. Routing Discussion
   - http://www.ripe.net/wg/routing/index.html

Useful URL's & Reading

6. Traceroute server repository
   - http://www.boardwatch.com/isp/trace.htm
   - http://nitrous.digex.net
   - Internet Looking Glass
7. ISP Tips
   - http://www.amazing.com/internet/faq.html
   - http://www.cisco.com/public/cons/isp/
8. BGP Table
   - http://www.telstra.net/ops/bgptable.html
   - http://www.employees.org/~tbates/cidr.hist.plot.html
   - http://www.merit.edu/ipma/reports
   - http://www.apnic.net/stats/bgp
9. Route server views
   - http://www.caida.org
10. NANOG archive
    - http://www.merit.edu/mail.archives/html/nanog/maillist.htm

IRR Reading List

1. RFC1786 “Representation of IP Routing Policies in a Routing Registry”
   ftp://ftp.isi.edu/in-notes/rfc1786.txt
2. RATools and RSPL
   ftp://ftp.apnic.net/ietf/rfc/rfc2280.txt
   Tools http://www.isi.edu/ra/*
   Mailing List <ratoolset@isi.edu>
3. PRIDE
   Slides ftp://ftp.ripe.net/pride/docs/course-slides
   Guide ftp://ftp.ripe.net/pride/docs/guide-2.0txt.{ps}.tar.gz
   Tools ftp://ftp.ripe.net/pride/tools/*
4. IRR authorisation/notification
   ftp://ftp.ripe.net/ripe/docs/ripe-120.txt
5. RADB pointers
   http://www.ra.net
   http://www.ra.net/.faq.htm
6. ISP run RR User documents
   http://infopage.cw.net/Routing