Internet Exchange Point Design

ISP/IXP Workshops

© 2009 Cisco Systems, Inc. All rights reserved.

IXP Design

- Background
- Why set up an IXP?
- Layer 2 Exchange Point
- Layer 3 Exchange Point
- Design Considerations
- Route Collectors & Servers
- What can go wrong?

© 2009 Cisco Systems, Inc. All rights reserved.

A bit of history

In a time long gone…

© 2009 Cisco Systems, Inc. All rights reserved.

A Bit of History…

- End of NSFnet – one major backbone
- move towards commercial Internet
  - private companies selling their bandwidth
- need for coordination of routing exchange between providers
  - Traffic from ISP A needs to get to ISP B
- Routing Arbiter project created to facilitate this

© 2009 Cisco Systems, Inc. All rights reserved.

What is an Exchange Point

- Network Access Points (NAPs) established at end of NSFnet
  - The original “exchange points”
- Major providers connect their networks and exchange traffic
- High-speed network or ethernet switch
- Simple concept – any place where providers come together to exchange traffic

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Points

- Layer 2 exchange point
  - Ethernet (1000/100Mbps)
  - Older technologies include ATM, Frame Relay, SRP, FDDI and SMDS
- Layer 3 exchange point
  - Router based
  - Historical status

© 2009 Cisco Systems, Inc. All rights reserved.

Why an Internet Exchange Point?

Saving money, improving QoS,

Generating a local Internet economy

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point
Why peer?

- Consider a region with one ISP
  - They provide internet connectivity to their customers
  - They have one or two international connections
- Internet grows, another ISP sets up in competition
  - They provide internet connectivity to their customers
  - They have one or two international connections
- How does traffic from customer of one ISP get to customer of the other ISP?
  - Via the international connections

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point
Why peer?

- Yes, International Connections…
  - If satellite, RTT is around 550ms per hop
  - So local traffic takes over 1s round trip
- International bandwidth
  - Costs significantly more than domestic bandwidth
  - Congested with local traffic
  - Wastes money, harms performance

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point
Why peer?

- Solution:
  - Two competing ISPs peer with each other
- Result:
  - Both save money
  - Local traffic stays local
  - Better network performance, better QoS,…
  - More international bandwidth for expensive international traffic
  - Everyone is happy

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point
Why peer?

- A third ISP enters the equation
  - Becomes a significant player in the region
  - Local and international traffic goes over their international connections
- They agree to peer with the two other ISPs
  - To save money
  - To keep local traffic local
  - To improve network performance, QoS,…

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point
Why peer?

- Peering means that the three ISPs have to buy circuits between each other
  - Works for three ISPs, but adding a fourth or a fifth means this does not scale
- Solution:
  - Internet Exchange Point

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point

- Every participant has to buy just one whole circuit
  - From their premises to the IXP
- Rather than N-1 half circuits to connect to the N-1 other ISPs
  - 5 ISPs have to buy 4 half circuits = 2 whole circuits → already twice the cost of the IXP connection

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point

- Solution
  - Every ISP participates in the IXP
  - Cost is minimal – one local circuit covers all domestic traffic
  - International circuits are used for just international traffic – and backing up domestic links in case the IXP fails
- Result:
  - Local traffic stays local
  - QoS considerations for local traffic is not an issue
  - RTTs are typically sub 10ms
  - Customers enjoy the Internet experience
  - Local Internet economy grows rapidly

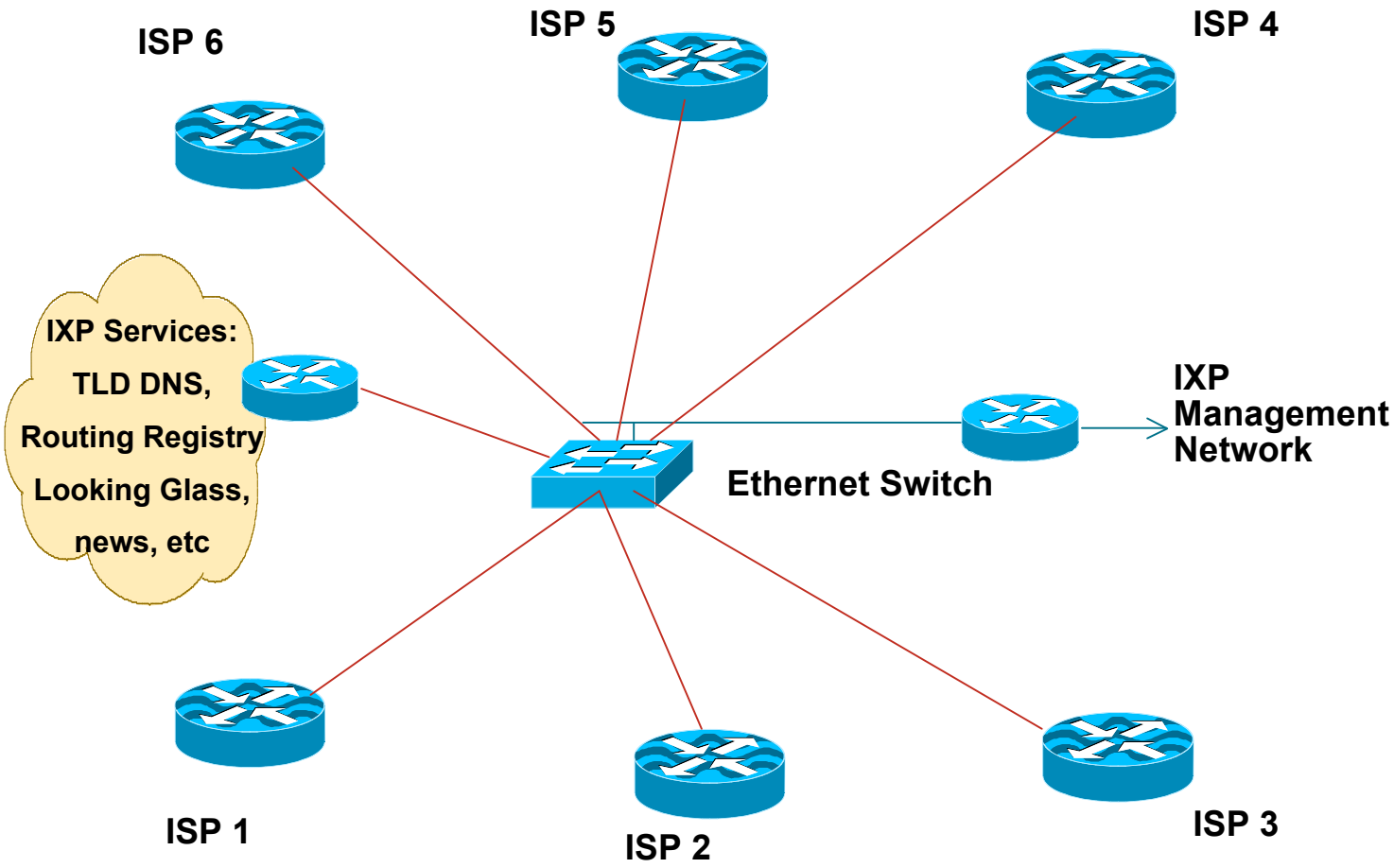
© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 Exchange

The traditional IXP

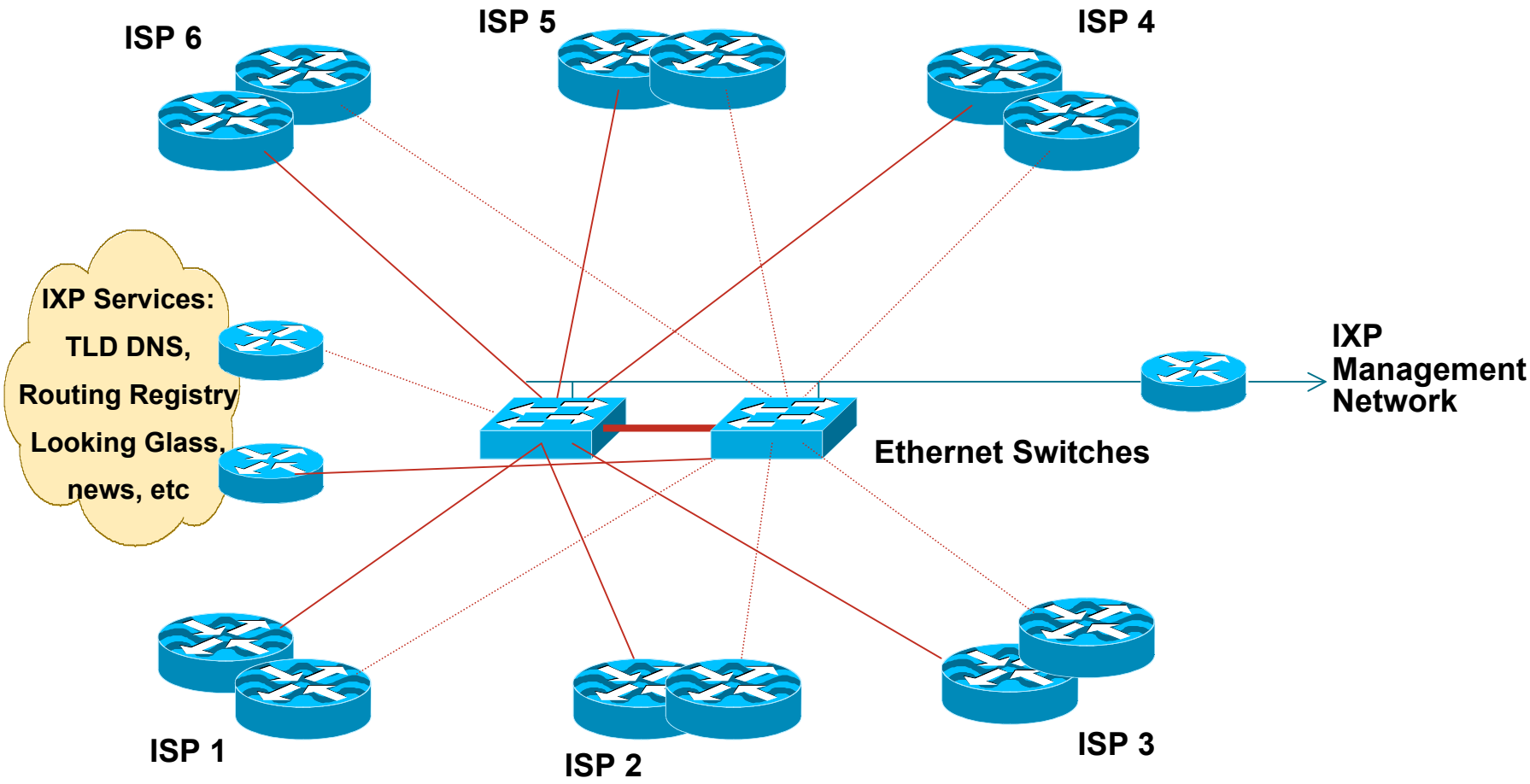
© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 Exchange



© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 Exchange



© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 Exchange

- Two switches for redundancy
- ISPs use dual routers for redundancy or loadsharing
- Offer services for the “common good”
  - Internet portals and search engines
  - DNS TLD, News, NTP servers
  - Routing Registry and Looking Glass

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 Exchange

- Requires neutral IXP management
  - usually funded equally by IXP participants
  - 24x7 cover, support, value add services
- Secure and neutral location
- Configuration
  - private address space if non-transit and no value add services
  - ISPs require AS, basic IXP does not

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 Exchange

- Network Security Considerations
  - LAN switch needs to be securely configured
  - Management routers require TACACS+ authentication, vty security
  - IXP services must be behind router(s) with strong filters

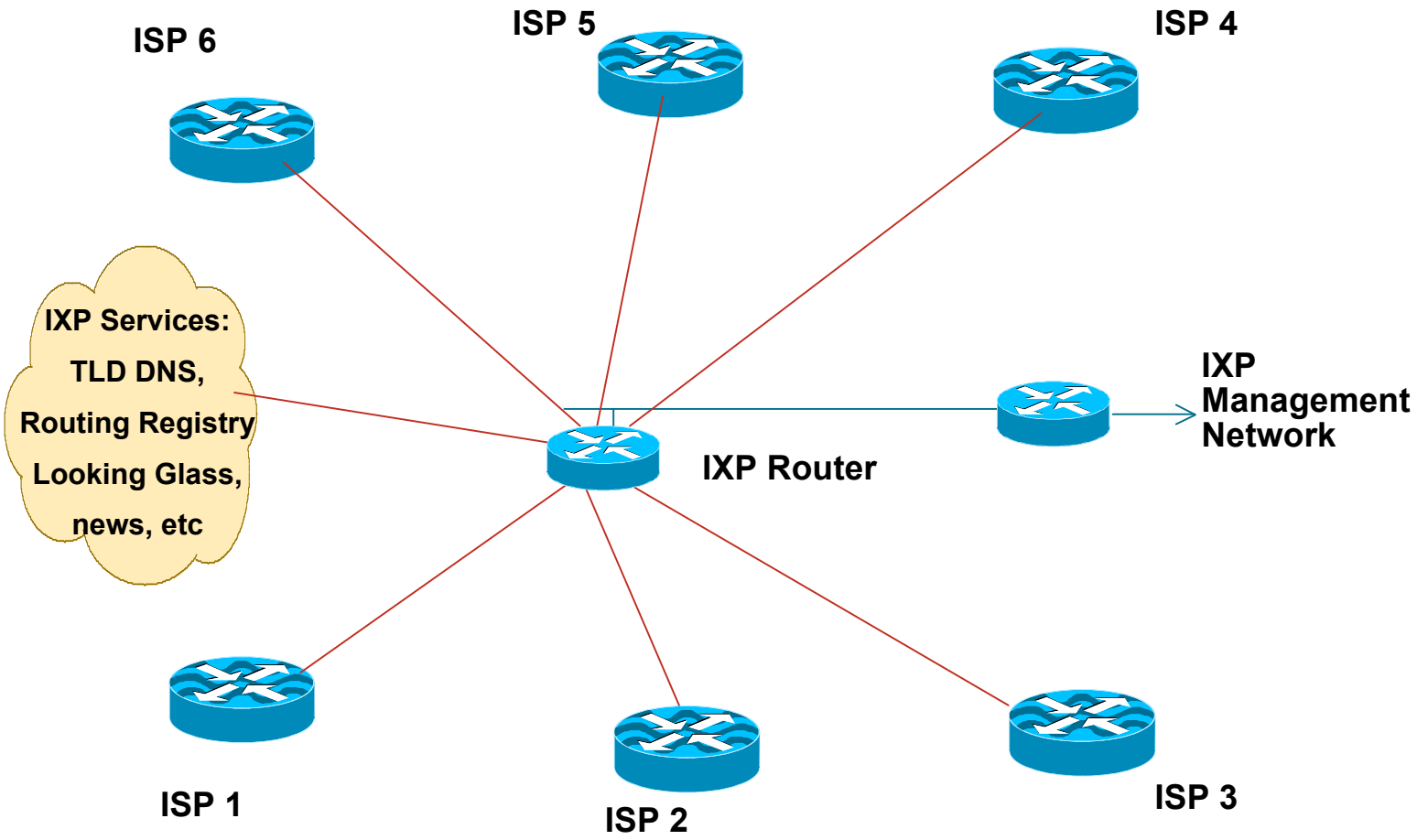
© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Exchange

Aka: The wholesale transit ISP

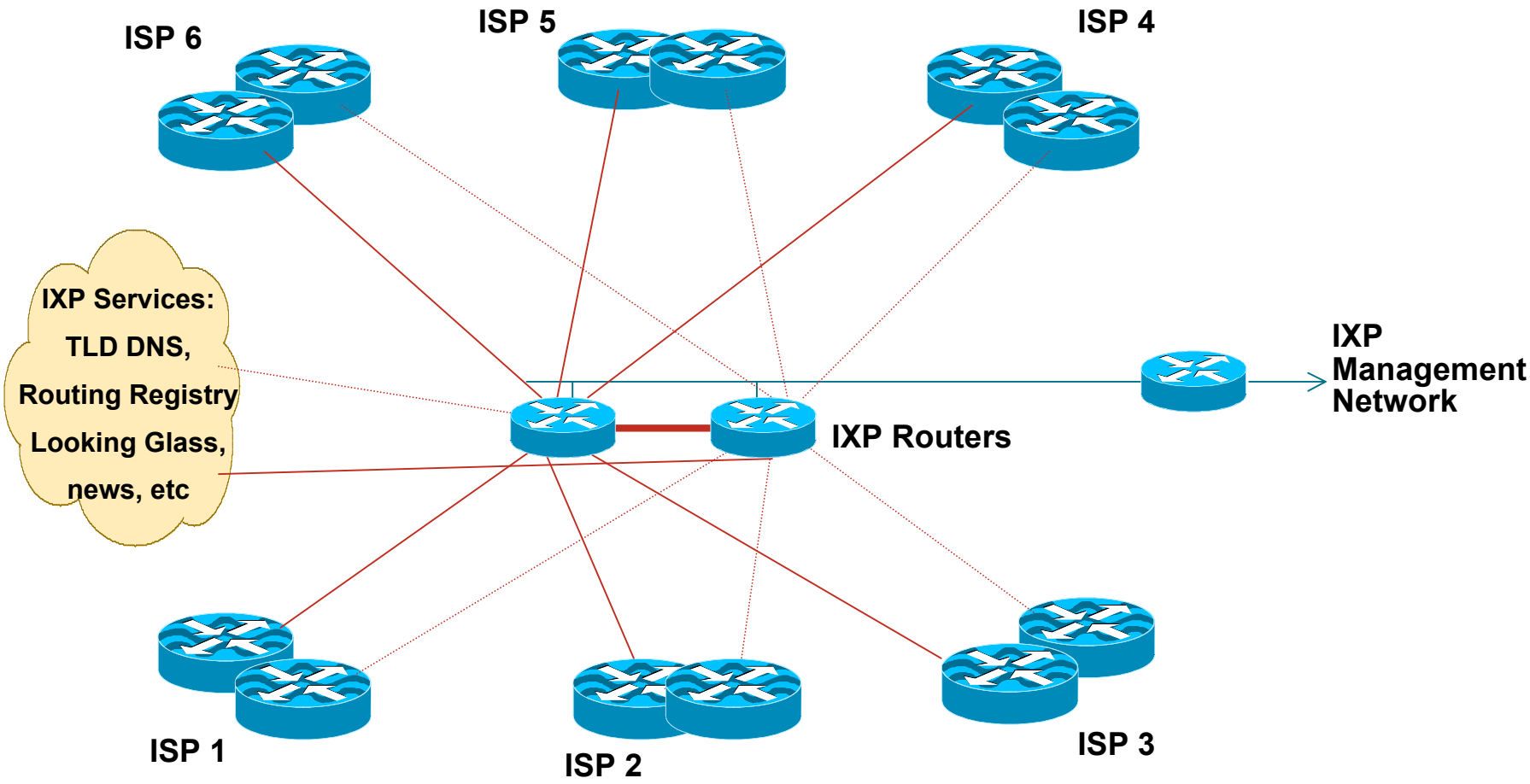
© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Exchange/Wholesale Transit ISP



© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Exchange/Wholesale Transit ISP



© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Exchange/Wholesale Transit ISP

- Two routers for redundancy
- ISPs use dual routers for redundancy or loadsharing
- Offer services for the “common good”
  - Internet portals and search engines
  - DNS TLD, News, NTP servers
  - Routing Registry and Looking Glass

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Exchange/Wholesale Transit ISP

- Requires neutral management
  - Usually funded equally by participants
  - 24x7 cover, support, value add services
  - BGP configuration skills essential
- Secure and neutral location
- Configuration
  - private address space if non-transit and no value add services
  - ISPs and IXP require AS

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Exchange/Wholesale Transit ISP

- Network Security Considerations
  - Core IXP router(s) require strong security, preferably with BGP neighbour authentication
  - Management routers require TACACS+ authentication, vty security
  - IXP services must be behind router(s) with strong filters

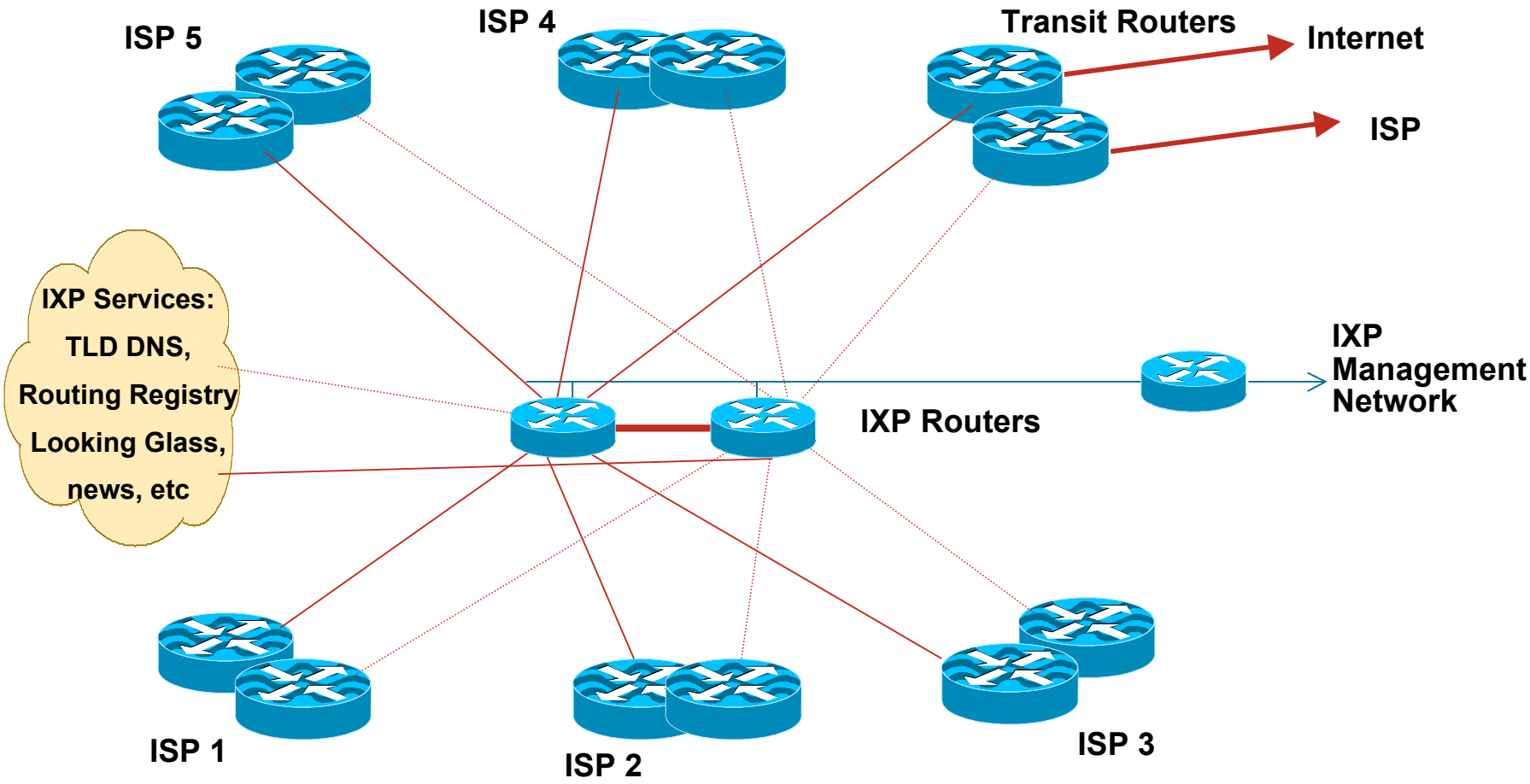
© 2009 Cisco Systems, Inc. All rights reserved.

Transit IXPs/Wholesale Transit ISP

- Provides local Internet exchange facility to members
- Also provides transit to Internet or upstream ISP
- Usually operated as a commercial service
- Usually layer 3 design

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 3 Transit Exchange/Transit ISP



© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 versus Layer 3

- Layer 3
  - IXP team requires good BGP knowledge
  - Rely on 3rd party for BGP configuration
  - Less freedom on who peers with whom
  - Usually competes with IXP membership
  - Tends to be distributed over wide area

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 versus Layer 3

- Layer 2
  - IXP team does not need routing knowledge
  - Easy to get started
  - More complicated to distribute over wide area
  - ISPs free to set up peering agreements with each other as they wish

© 2009 Cisco Systems, Inc. All rights reserved.

Layer 2 versus Layer 3 Summary

- Layer 2 is a **REAL** internet exchange point
- **Layer 3 is marketing concept used by Transit ISPs**
  - Is NOT a real IXP

© 2009 Cisco Systems, Inc. All rights reserved.

IXP Design Considerations

© 2009 Cisco Systems, Inc. All rights reserved.

Exchange Point Design

- The IXP Core is an Ethernet switch
- Has superseded all other types of network devices for an IXP
  - From the cheapest and smallest 12 or 24 port 10/100 switch
  - To the largest 32 port 10GigEthernet switch

© 2009 Cisco Systems, Inc. All rights reserved.

Exchange Point Design

- Each ISP participating in the IXP brings a router to the IXP location
- Router needs:
  - One Ethernet port to connect to IXP switch
  - One WAN port to connect to the WAN media leading back to the ISP backbone
  - To be able to run BGP

© 2009 Cisco Systems, Inc. All rights reserved.

Exchange Point Design

- IXP switch located in one equipment rack dedicated to IXP
  - Also includes other IXP operational equipment
- Routers from participant ISPs located in neighbouring/adjacent rack(s)
- Copper (UTP) connections made for 10Mbps, 100Mbps or 1Gbps connections
- Fibre used for 10Gbps and 40Gbps

© 2009 Cisco Systems, Inc. All rights reserved.

Peering

- Each participant needs to run BGP
  - They need their own AS number
  - **Public** ASN, **NOT** private ASN
- Each participant configures external BGP directly with the other participants in the IXP
  - Peering with all participants
    - or
  - Peering with a subset of participants

© 2009 Cisco Systems, Inc. All rights reserved.

Peering (more)

- Mandatory Multi-Lateral Peering (MMLP)

  Each participant is required to peer with every other participant as part of their IXP membership

  Has no history of success — the practice is **discouraged**

- Multi-Lateral Peering (MLP)

  Each participant peers with every other participant

- Bi-Lateral Peering

  Participants set up peering with each other according to their own requirements and business relationships

  This is the most common situation at IXPs today

© 2009 Cisco Systems, Inc. All rights reserved.

Routing

- ISP border routers at the IXP generally should NOT be configured with a default route or carry the full Internet routing table
  - Carrying default or full table means that this router and the ISP network is open to abuse by non-peering IXP members
  - Correct configuration is only to carry routes offered to IXP peers on the IXP peering router
- Note: Some ISPs offer transit across IX fabrics
  - They do so at their own risk – see above

© 2009 Cisco Systems, Inc. All rights reserved.

Routing (more)

- ISP border routers at the IXP should not be configured to carry the IXP LAN network within the IGP or iBGP
  - Use next-hop-self BGP concept
- Don’t generate ISP prefix aggregates on IXP peering router
  - If connection from backbone to IXP router goes down, normal BGP failover will then be successful

© 2009 Cisco Systems, Inc. All rights reserved.

Address Space

- Some IXPs use private addresses for the IX LAN
  - Public address space means IXP network could be leaked to Internet which may be undesirable
  - Because most ISPs filter RFC1918 address space, this avoids the problem
- Some IXPs use public addresses for the IX LAN
  - Address space available from the RIRs
  - IXP terms of participation often forbid the IX LAN to be carried in the ISP member backbone

© 2009 Cisco Systems, Inc. All rights reserved.

Hardware

- Try not to mix port speeds
  - if 10Mbps and 100Mbps connections available, terminate on different switches (L2 IXP)
- Don’t mix transports
  - if terminating ATM PVCs and G/F/Ethernet, terminate on different devices
- Insist that IXP participants bring their own router
  - moves buffering problem off the IXP
  - security is responsibility of the ISP, not the IXP

© 2009 Cisco Systems, Inc. All rights reserved.

Services Offered

- Services offered should not compete with member ISPs (basic IXP)
  - e.g. web hosting at an IXP is a bad idea unless all members agree to it
- IXP operations should make performance and throughput statistics available to members
  - Use tools such as MRTG to produce IX throughput graphs for member (or public) information

© 2009 Cisco Systems, Inc. All rights reserved.

Services to Offer

- ccTLD DNS
  - the country IXP could host the country's top level DNS
  - e.g. "SE." TLD is hosted at Netnod IXes in Sweden
  - Offer back up of other country ccTLD DNS
- Root server
  - Anycast instances of I.root-servers.net, F.root-servers.net etc are present at many IXes
- Usenet News
  - Usenet News is high volume
  - could save bandwidth to all IXP members

© 2009 Cisco Systems, Inc. All rights reserved.

Services to Offer

- Route Collector
  - Route collector shows the reachability information available at the exchange
  - Technical detail covered later on
- Looking Glass
  - One way of making the Route Collector routes available for global view (e.g. **www.traceroute.org**)
  - Public or members only access

© 2009 Cisco Systems, Inc. All rights reserved.

Services to Offer

- Content Redistribution/Caching

  For example, Akamised update distribution service

- Network Time Protocol

  Locate a stratum 1 time source (GPS receiver, atomic clock, etc) at IXP

- Routing Registry

  Used to register the routing policy of the IXP membership (more later)

© 2009 Cisco Systems, Inc. All rights reserved.

Introduction to Route Collectors

What routes are available at the IXP?

© 2009 Cisco Systems, Inc. All rights reserved.

What is a Route Collector?

- Usually a router or Unix system running BGP
- Gathers routing information from service provider routers at an IXP
  - Peers with each ISP using BGP
- Does not forward packets
- Does not announce any prefixes to ISPs

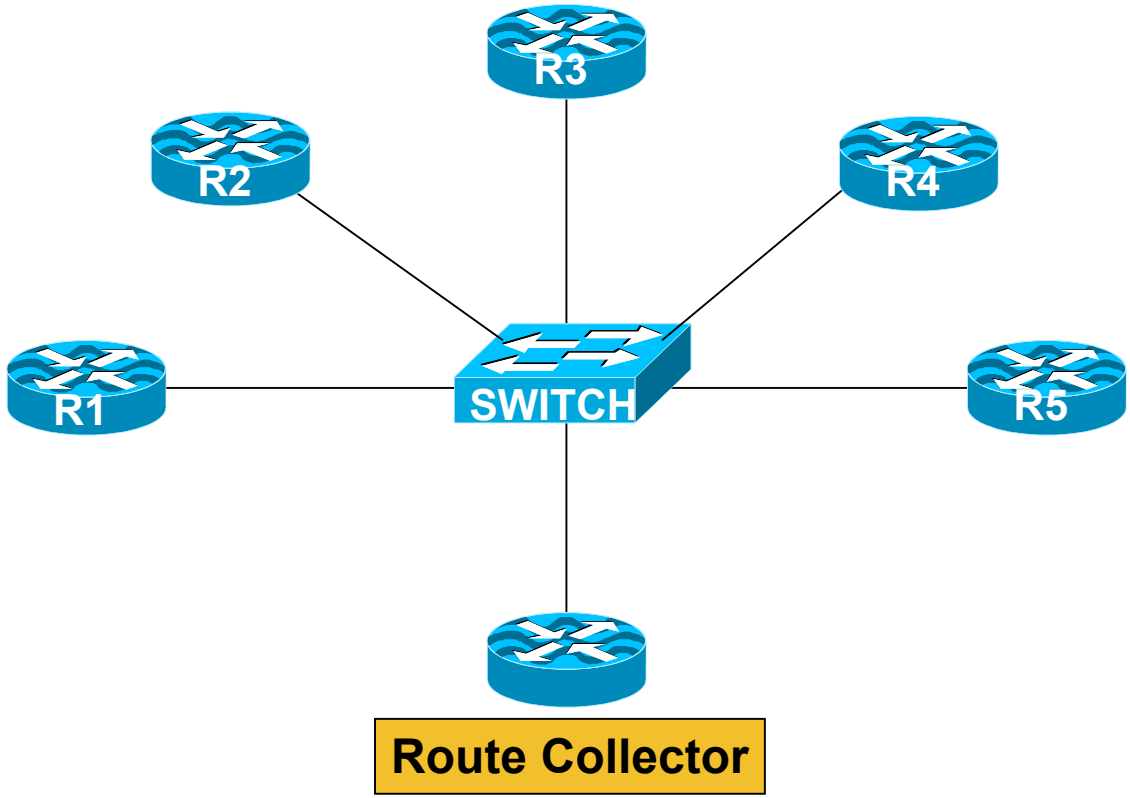
© 2009 Cisco Systems, Inc. All rights reserved.

Purpose of a Route Collector

- To provide a public view of the Routing Information available at the IXP
  - Useful for existing members to check functionality of BGP filters
  - Useful for prospective members to check value of joining the IXP
  - Useful for the Internet Operations community for troubleshooting purposes
    - E.g. **www.traceroute.org**

© 2009 Cisco Systems, Inc. All rights reserved.

Route Collector at an IXP



© 2009 Cisco Systems, Inc. All rights reserved.

Route Collector Requirements

- Router or Unix system running BGP
  - Minimal memory requirements – only holds IXP routes
  - Minimal packet forwarding requirements – doesn't forward any packets
- Peers eBGP with every IXP member
  - Accepts everything; Gives nothing
  - Uses a private ASN
  - Connects to IXP Transit LAN
- "Back end" connection
  - Second Ethernet globally routed
  - Connection to IXP Website for public access

© 2009 Cisco Systems, Inc. All rights reserved.

Route Collector Implementation

- Most IXPs now implement some form of Route Collector
- Benefits already mentioned
- Great public relations tool
- Unsophisticated requirements
  - Just runs BGP

© 2009 Cisco Systems, Inc. All rights reserved.

Introduction to Route Servers

How to scale very large IXPs

© 2009 Cisco Systems, Inc. All rights reserved.

What is a Route Server?

- Has all the features of a Route Collector
- But also:

  Announces routes to participating IXP members according to their routing policy definitions
- Implemented using the same specification as for a Route Collector

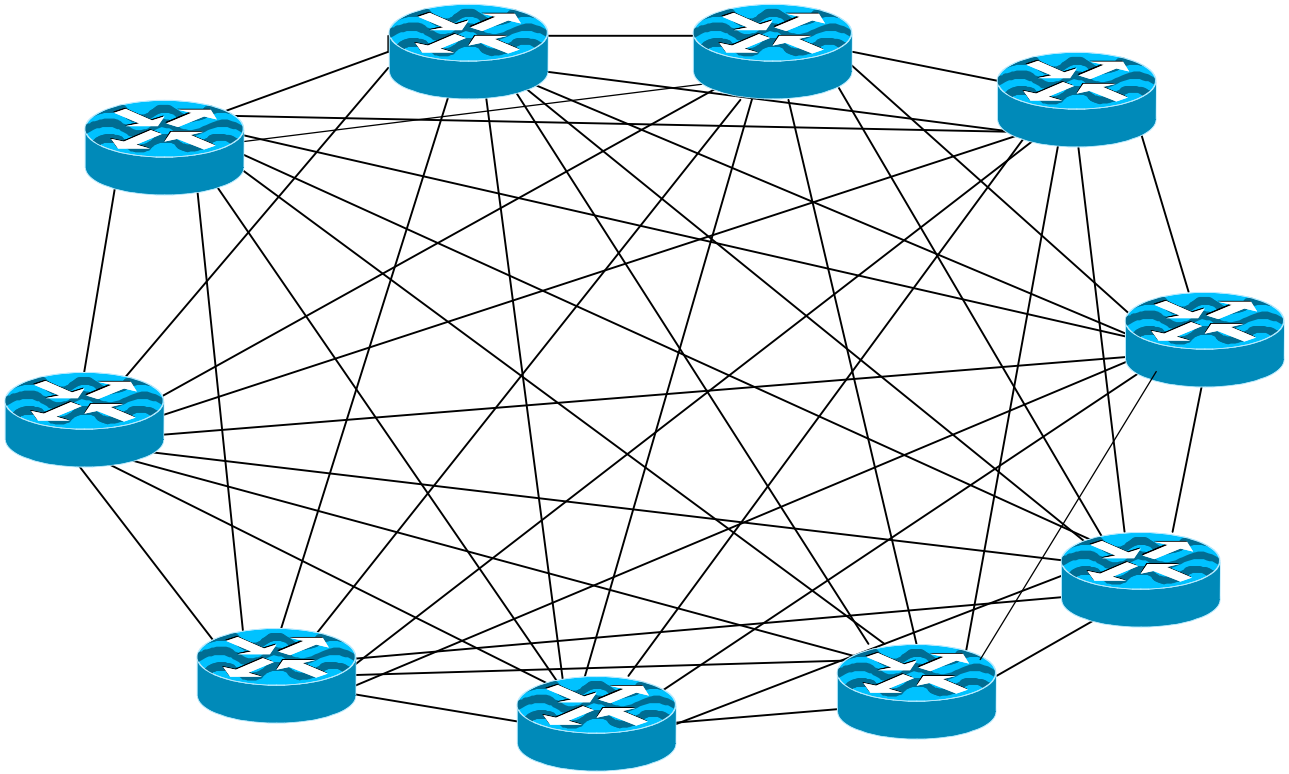
© 2009 Cisco Systems, Inc. All rights reserved.

Features of a Route Server

- Helps scale routing for large IXPs
- Simplifies Routing Processes on ISP Routers
- Optional participation
  - Provided as service, is NOT mandatory
- Does result in insertion of RS Autonomous System Number in the Routing Path
- Optionally uses Policy registered in IRR

© 2009 Cisco Systems, Inc. All rights reserved.

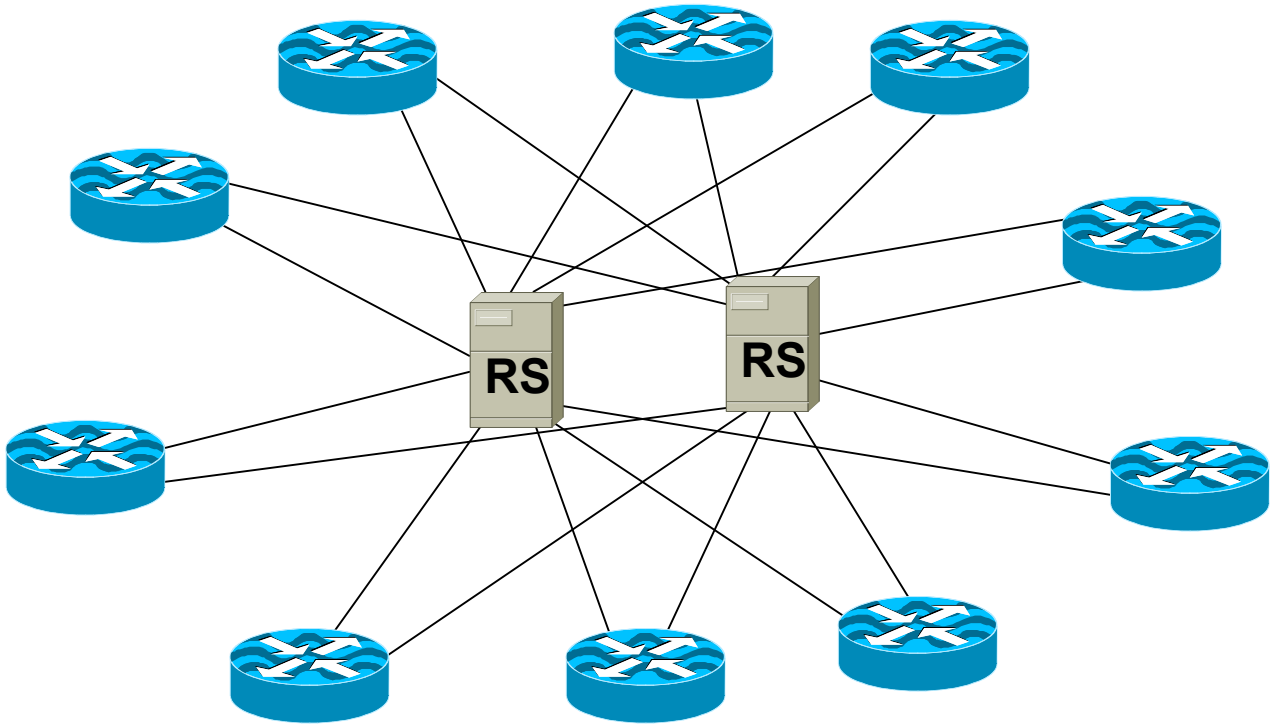
Diagram of N-squared Peering Mesh



- For large IXPs (dozens for participants) maintaining a larger peering mesh becomes cumbersome and often too hard

© 2009 Cisco Systems, Inc. All rights reserved.

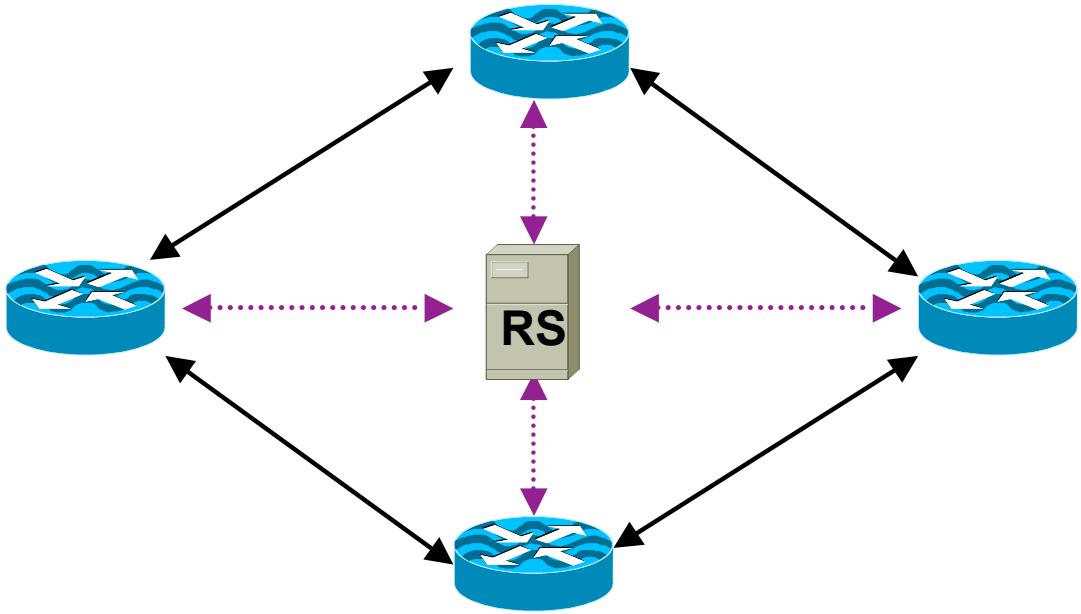
Peering Mesh with Route Servers



- ISP routers peer with the Route Servers
  - Only need to have two eBGP sessions rather than N

© 2009 Cisco Systems, Inc. All rights reserved.

RS based Exchange Point Routing Flow



TRAFFIC FLOW

ROUTING INFORMATION FLOW

© 2009 Cisco Systems, Inc. All rights reserved.

Advantages of Using a Route Server

- Helps scale Routing for very large IXPs
- Separation of Routing and Forwarding
- Simplify Routing Configuration Management on ISPs routers

© 2009 Cisco Systems, Inc. All rights reserved.

Disadvantages of using a Route Server

- ISPs can lose direct policy control
  - If RS is only peer, ISPs have no control over who their prefixes are distributed to
- Completely dependent on 3rd party
  - Configuration, troubleshooting, etc…
- Insertion of RS ASN into routing path
  - Traffic engineering/multihoming needs more care
- These are major disadvantages
  - Usually out-weigh the advantages

© 2009 Cisco Systems, Inc. All rights reserved.

Typical usage of a Route Server

- Route Servers may be provided as an **OPTIONAL** service
  - Most common at large IXPs (>50 participants)
  - Examples: LINX, TorIX, AMS-IX, etc
- ISPs peer:
  - Directly with significant peers
  - With Route Server for the rest

© 2009 Cisco Systems, Inc. All rights reserved.

Things to think about...

- Would using a route server benefit you?
  - Helpful when BGP knowledge is limited (but is NOT an excuse not to learn BGP)
  - Avoids having to maintain a large number of eBGP peers
  - But can you afford to lose policy control? (An ISP not in control of their routing policy is what?)

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong…

The different ways IXP operators harm their IXP…

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Concept

- Some Service Providers attempt to cash in on the reputation of IXPs
- Market Internet transit services as “Internet Exchange Point”
  - “We are exchanging packets with other ISPs, so we are an Internet Exchange Point!”
  - So-called Layer-3 Exchanges — really Internet Transit Providers
  - Router used rather than a Switch
  - Most famous example: SingTelIX

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Competition

- Too many exchange points in one locale
  - Competing exchanges defeats the purpose
- Becomes expensive for ISPs to connect to all of them

- An IXP:
  - is **NOT** a competition
  - is **NOT** a profit making business

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Rules and Restrictions

- IXPs try to compete with their membership
  - Offering services that ISPs would/do offer their customers
- IXPs run as a closed privileged club e.g.:
  - Restrictive membership criteria (closed shop)
- IXPs providing access to end users rather than just Service Providers
- IXPs interfering with ISP business decisions e.g. Mandatory Multi-Lateral Peering

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Technical Design Errors

- Interconnected IXPs
  - IXP in one location believes it should connect directly to the IXP in another location
  - Who pays for the interconnect?
  - How is traffic metered?
  - Competes with the ISPs who already provide transit between the two locations (who then refuse to join IX, harming the viability of the IX)
  - Metro interconnections work ok (e.g. LINX)

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Technical Design Errors

- ISPs bridge the IXP LAN back to their offices
  - “We are poor, we can’t afford a router”
  - Financial benefits of connecting to an IXP far outweigh the cost of a router
  - In reality it allows the ISP to connect any devices to the IXP LAN — with disastrous consequences for the security, integrity and reliability of the IXP

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Routing Design Errors

- Route Server implemented from Day One
  - ISPs have no incentive to learn BGP
  - Therefore have no incentive to understand peering relationships, peering policies, &c
  - Entirely dependent on operator of RS for troubleshooting, configuration, reliability
    - RS can’t be run by committee!
- Route Server is to help scale peering at LARGE IXPs

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Routing Design Errors

- iBGP Route Reflector used to distribute prefixes between IXP participants
- Claimed Advantage (1):

  Participants don't need to know about or run BGP
- Actually a Disadvantage

  IXP Operator has to know BGP

  ISP not knowing BGP is big commercial disadvantage

  ISPs who would like to have a growing successful business need to be able to multi-home, peer with other ISPs, etc — these activities require BGP

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Routing Design Errors (cont)

- Route Reflector Claimed Advantage (2):

  Allows an IXP to be started very quickly

- Fact:

  IXP is only an Ethernet switch — setting up an iBGP mesh with participants is no quicker than setting up an eBGP mesh

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Routing Design Errors (cont)

- Route Reflector Claimed Advantage (3):
  - IXP operator has full control over IXP activities
- Actually a Disadvantage
  - ISP participants surrender control of:
    - Their border router; it is located in IXP's AS
    - Their routing and peering policy
  - IXP operator is single point of failure
    - If they aren't available 24x7, then neither is the IXP
    - BGP configuration errors by IXP operator have real impacts on ISP operations

© 2009 Cisco Systems, Inc. All rights reserved.

What can go wrong? Routing Design Errors (cont)

- Route Reflector Disadvantage (4):
  - Migration from Route Reflector to “correct” routing configuration is highly non-trivial
  - ISP router is in IXP’s ASN
    - Need to move ISP router from IXP’s ASN to the ISP’s ASN
    - Need to reconfigure BGP on ISP router, add to ISP’s IGP and iBGP mesh, and set up eBGP with IXP participants and/or the IXP Route Server

© 2009 Cisco Systems, Inc. All rights reserved.

More Information

© 2009 Cisco Systems, Inc. All rights reserved.

Exchange Point Policies & Politics

- AUPs
  - Acceptable Use Policy
  - Minimal rules for connection
- Fees?
  - Some IXPs charge no fee
  - Other IXPs charge cost recovery
  - A few IXPs are commercial
- Nobody is obliged to peer
  - Agreements left to ISPs, not mandated by IXP

© 2009 Cisco Systems, Inc. All rights reserved.

Exchange Point etiquette

- Don’t point default route at another IXP participant
- Be aware of third-party next-hop
- Only announce your aggregate routes
  - Read RIPE-399 first (www.ripe.net/docs/ripe-399.html)
- Filter! Filter! Filter!
  - And do reverse path check

© 2009 Cisco Systems, Inc. All rights reserved.

Exchange Point Examples

- LINX in London, UK
- TorIX in Toronto, Canada
- AMS-IX in Amsterdam, Netherlands
- SIX in Seattle, Washington, US
- PA-IX in Palo Alto, California, US
- JPNAP in Tokyo, Japan
- DE-CIX in Frankfurt, Germany
- …
- All use Ethernet Switches

© 2009 Cisco Systems, Inc. All rights reserved.

Features of IXPs (1)

- Redundancy & Reliability
  - Multiple switches, UPS
- Support
  - NOC to provide 24x7 support for problems at the exchange
- DNS, Route Collector, Content & NTP servers
  - ccTLD & root servers
  - Content redistribution systems such as Akamai
  - Route Collector – Routing Table view

© 2009 Cisco Systems, Inc. All rights reserved.

Features of IXPs (2)

- Location
  - neutral co-location facilities
- Address space
  - Peering LAN
- AS Number
  - If using Route Collector/Server
- Route servers (optional, for larger IXPs)
- Statistics
  - Traffic data – for membership

© 2009 Cisco Systems, Inc. All rights reserved.

More info about IXPs

- http://www.ep.net/ep-main.html

  Excellent resource for ip address allocation for exchanges, locations of XPs in the world, AUPs and other policies

- http://www.pch.net/documents

  Another excellent resource of IXP locations, papers, IXP statistics, etc

- http://www.telegeography.com/ee/ix/index.php

  A collection of IXPs and interconnect points for ISPs

© 2009 Cisco Systems, Inc. All rights reserved.

Summary

- L2 IXP – most commonly deployed
  - The core is an ethernet switch
  - ATM and other old technologies are obsolete
- L3 IXP – nowadays is a marketing concept used by wholesale ISPs
  - Does not offer the same flexibility as L2
  - Not recommended unless there are overriding regulatory or political reasons to do so
  - **Avoid!**

© 2009 Cisco Systems, Inc. All rights reserved.

Internet Exchange Point Design

ISP/IXP Workshops

© 2009 Cisco Systems, Inc. All rights reserved.