Peering, Transit and IXP Design

Philip Smith

MENOG 13

Kuwait

15th – 24th September 2013

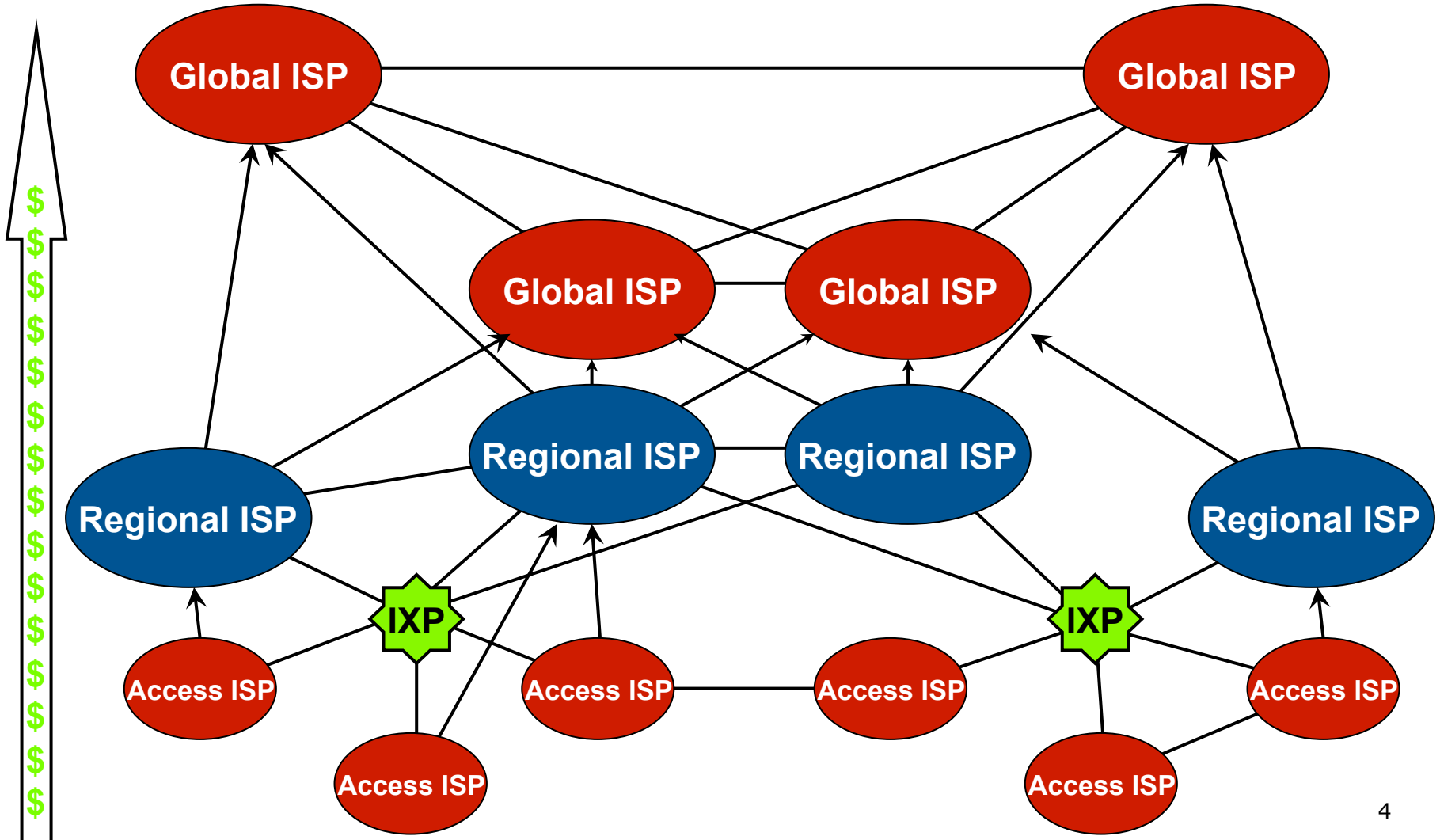
Presentation Slides

- Will be available on
  - http://thyme.apnic.net/ftp/seminars/MENOG13-IXP-NetworkDesign.pdf
  - And on the MENOG13 website
- Feel free to ask questions any time

The Internet

- Internet is made up of ISPs of all shapes and sizes
  - Some have local coverage (access providers)
  - Others can provide regional or per country coverage
  - And others are global in scale
- These ISPs interconnect their businesses
  - They don't interconnect with every other ISP (over 43000 distinct autonomous networks) – won't scale
  - They interconnect according to practical and business needs
- Some ISPs provide transit to others
  - They interconnect other ISP networks

Categorising ISPs

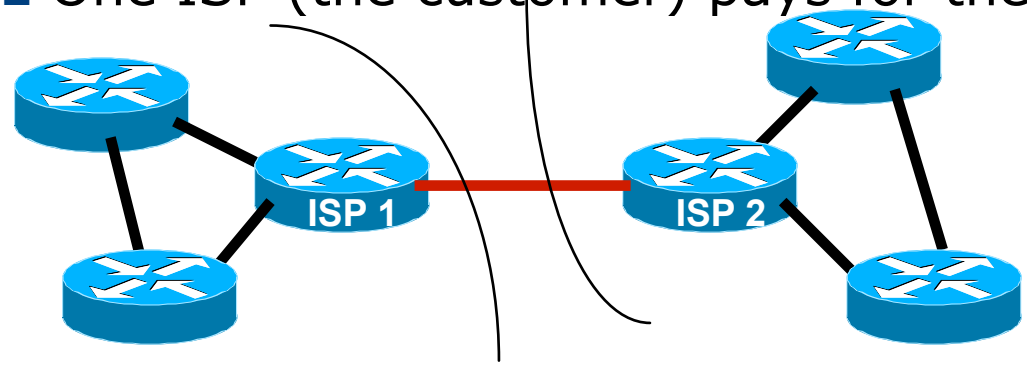


Peering and Transit

- Transit
  - Carrying traffic across a network
  - Usually for a fee
  - Example: Access provider connects to a regional provider
- Peering
  - Exchanging routing information and traffic
  - Usually for no fee
  - Sometimes called settlement free peering
  - Example: Regional provider connects to another regional provider

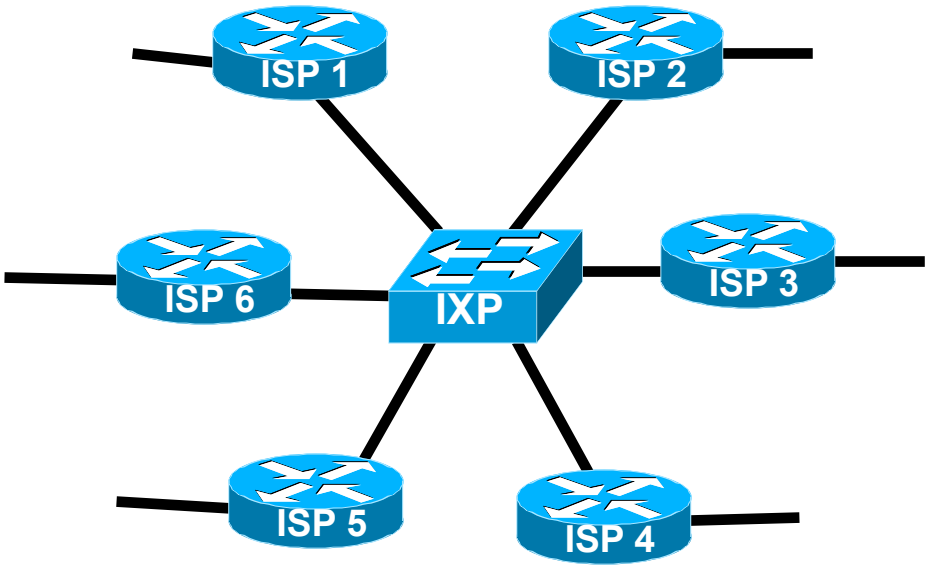
Private Interconnect

- Two ISPs connect their networks over a private link
  - Can be peering arrangement
    - No charge for traffic
    - Share cost of the link
  - Can be transit arrangement
    - One ISP charges the other for traffic
    - One ISP (the customer) pays for the link



Public Interconnect

- Several ISPs meeting in a common neutral location and interconnect their networks
  - Usually is a peering arrangement between their networks



ISP Goals

- Minimise the cost of operating the business
- Transit
  - ISP has to pay for circuit (international or domestic)
  - ISP has to pay for data (usually per Mbps)
  - Repeat for each transit provider
  - Significant cost of being a service provider
- Peering
  - ISP shares circuit cost with peer (private) or runs circuit to public peering point (one off cost)
  - No need to pay for data
  - Reduces transit data volume, therefore reducing cost

Transit – How it works

- Small access provider provides Internet access for a city's population
  - Mixture of dial up, wireless and fixed broadband
  - Possibly some business customers
  - Possibly also some Internet cafes
- How do their customers get access to the rest of the Internet?
- ISP buys access from one, two or more larger ISPs who already have visibility of the rest of the Internet
  - This is transit – they pay for the physical connection to the upstream and for the traffic volume on the link

Peering – How it works

- If two ISPs are of equivalent sizes, they have:
  - Equivalent network infrastructure coverage
  - Equivalent customer size
  - Similar content volumes to be shared with the Internet
  - Potentially similar traffic flows to each other's networks
- This makes them good peering partners
- If they don't peer
  - They both have to pay an upstream provider for access to each other's network/customers/content
  - Upstream benefits from this arrangement, the two ISPs both have to fund the transit costs

The IXP's role

- Private peering makes sense when there are very few equivalent players
  - Connecting to one other ISP costs X
  - Connecting to two other ISPs costs 2 times X
  - Connecting to three other ISPs costs 3 times X
  - Etc… (where X is half the circuit cost plus a port cost)
- The more private peers, the greater the cost
- IXP is a more scalable solution to this problem

The IXP's role

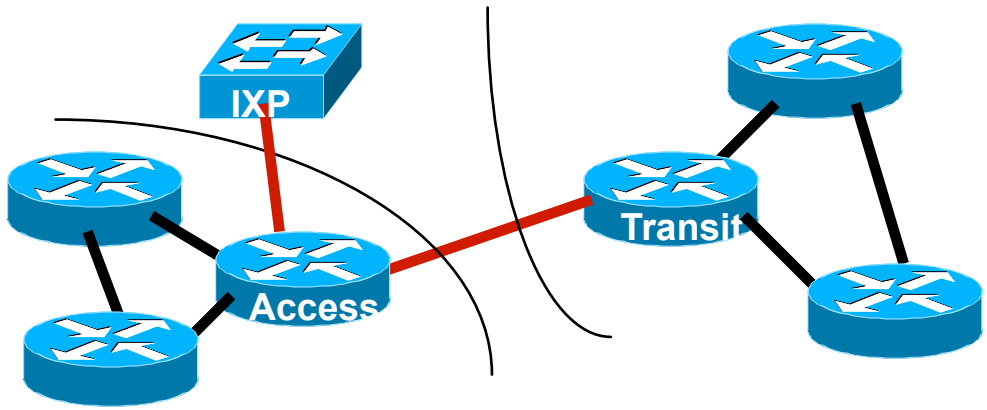
- Connecting to an IXP
  - ISP costs: one router port, one circuit, and one router to locate at the IXP
- Some IXPs charge annual "maintenance fees"
  - The maintenance fee has potential to significantly influence the cost balance for an ISP
- Generally connecting to an IXP and peering there becomes cost effective when there are at least three other peers
  - The real $ amount varies from region to region, IXP to IXP

Who peers at an IXP?

- Access Providers
  - Don't have to pay their regional provider transit fees for local traffic
  - Keeps latency for local traffic low
  - 'Unlimited' bandwidth through the IXP (compared with costly and limited bandwidth through transit provider)
- Regional Providers
  - Don't have to pay their global provider transit for local and regional traffic
  - Keeps latency for local and regional traffic low
  - 'Unlimited' bandwidth through the IXP (compared with costly and limited bandwidth through global provider)

The IXP's role

- Global Providers can be located close to IXPs
  - Attracted by the potential transit business available
- Advantageous for access & regional providers
  - They can peer with other similar providers at the IXP
  - And in the same facility pay for transit to their regional or global provider
  - (Not across the IXP fabric, but a separate connection)



Connectivity Decisions

- Transit
  - Almost every ISP needs transit to reach rest of Internet
  - One provider = no redundancy
  - Two providers: ideal for traffic engineering as well as redundancy
  - Three providers = better redundancy, traffic engineering gets harder
  - More then three = diminishing returns, rapidly escalating costs and complexity
- Peering
  - Means low (or zero) cost access to another network
  - Private or Public Peering (or both)

Transit Goals

1. Minimise number of transit providers
   - But maintain redundancy
   - 2 is ideal, 4 or more is bad
2. Aggregate capacity to transit providers
   - More aggregated capacity means better value
     - Lower cost per Mbps
   - 4x 45Mbps circuits to 4 different ISPs will almost always cost more than 2x 155Mbps circuits to 2 different ISPs
     - Yet bandwidth of latter (310Mbps) is greater than that of former (180Mbps) and is much easier to operate

Peering or Transit?

- How to choose?
- Or do both?
- It comes down to cost of going to an IXP
  - Free peering
  - Paying for transit from an ISP co-located in same facility, or perhaps close by
- Or not going to an IXP and paying for the cost of transit directly to an upstream provider
  - There is no right or wrong answer, someone has to do the arithmetic

Private or Public Peering

- Private peering
  - Scaling issue, with costs, number of providers, and infrastructure provisioning
- Public peering
  - Makes sense the more potential peers there are (more is usually greater than "two")
- Which public peering point?
  - Local Internet Exchange Point: great for local traffic and local peers
  - Regional Internet Exchange Point: great for meeting peers outside the locality, might be cheaper than paying transit to reach the same consumer base

Local Internet Exchange Point

- Defined as a public peering point serving the local Internet industry
- Local means where it becomes cheaper to interconnect with other ISPs at a common location than it is to pay transit to another ISP to reach the same consumer base
  - Local can mean different things in different regions!

Regional Internet Exchange Point

- These are also “local” Internet Exchange Points
- But also attract regional ISPs and ISPs from outside the locality
  - Regional ISPs peer with each other
  - And show up at several of these Regional IXPs
- Local ISPs peer with ISPs from outside the locality
  - They don’t compete in each other’s markets
  - Local ISPs don’t have to pay transit costs
  - ISPs from outside the locality don’t have to pay transit costs
  - Quite often ISPs of disparate sizes and influences will happily peer – to defray transit costs

Which IXP?

- How many routes are available?
  - What is traffic to & from these destinations, and by how much will it reduce cost of transit?
- What is the cost of co-lo space?
  - If prohibitive or space not available, pointless choosing this IXP
- What is the cost of running a circuit to the location?
  - If prohibitive or competitive with transit costs, pointless choosing this IXP
- What is the cost of remote hands/assistance?
  - If no remote hands, doing maintenance is challenging and potentially costly with a serious outage

Internet Exchange Point

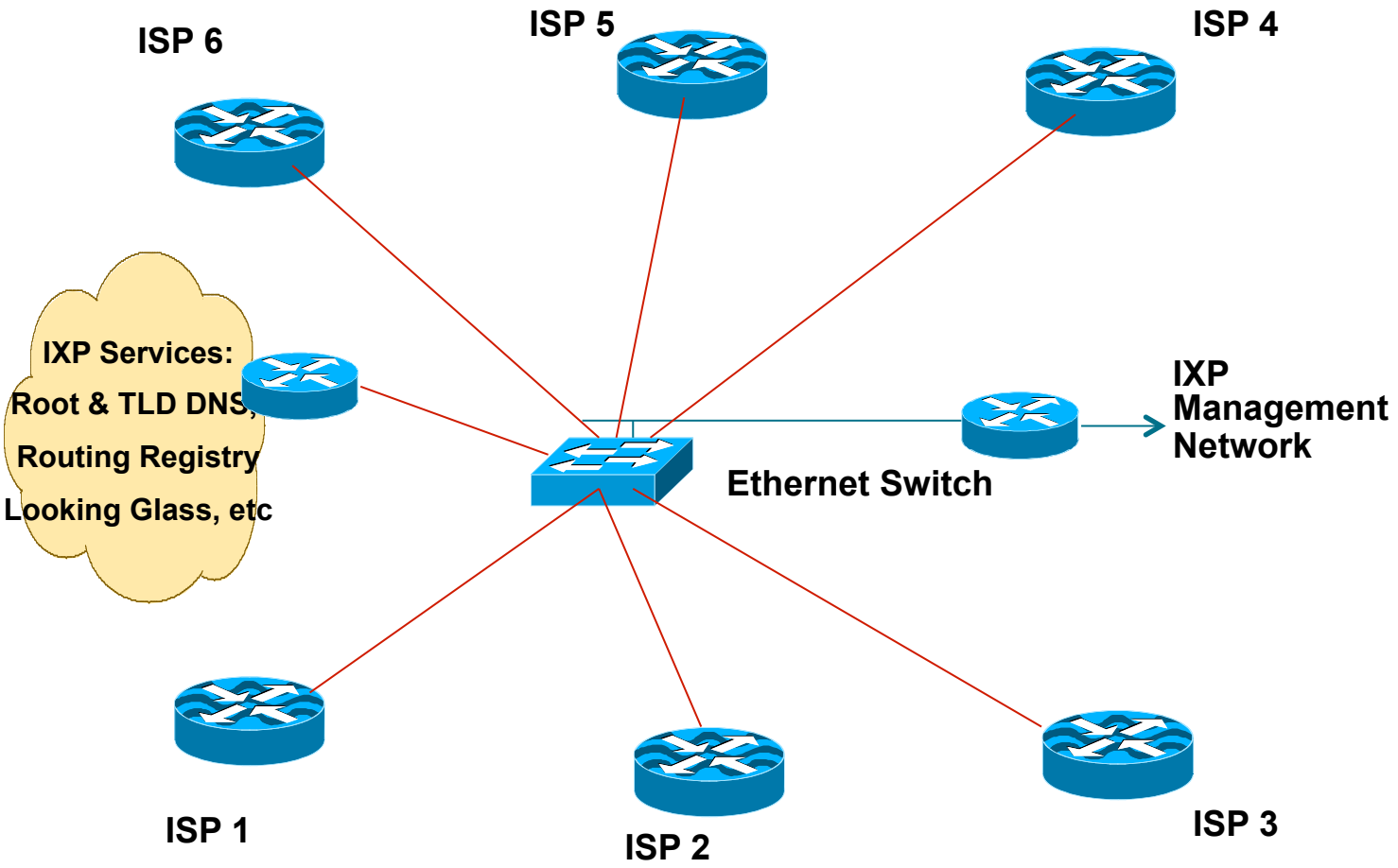
- Solution
  - Every ISP participates in the IXP
  - Cost is minimal – one local circuit covers all domestic traffic
  - International circuits are used for just international traffic – and backing up domestic links in case the IXP fails
- Result:
  - Local traffic stays local
  - QoS considerations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

Exchange Point Design

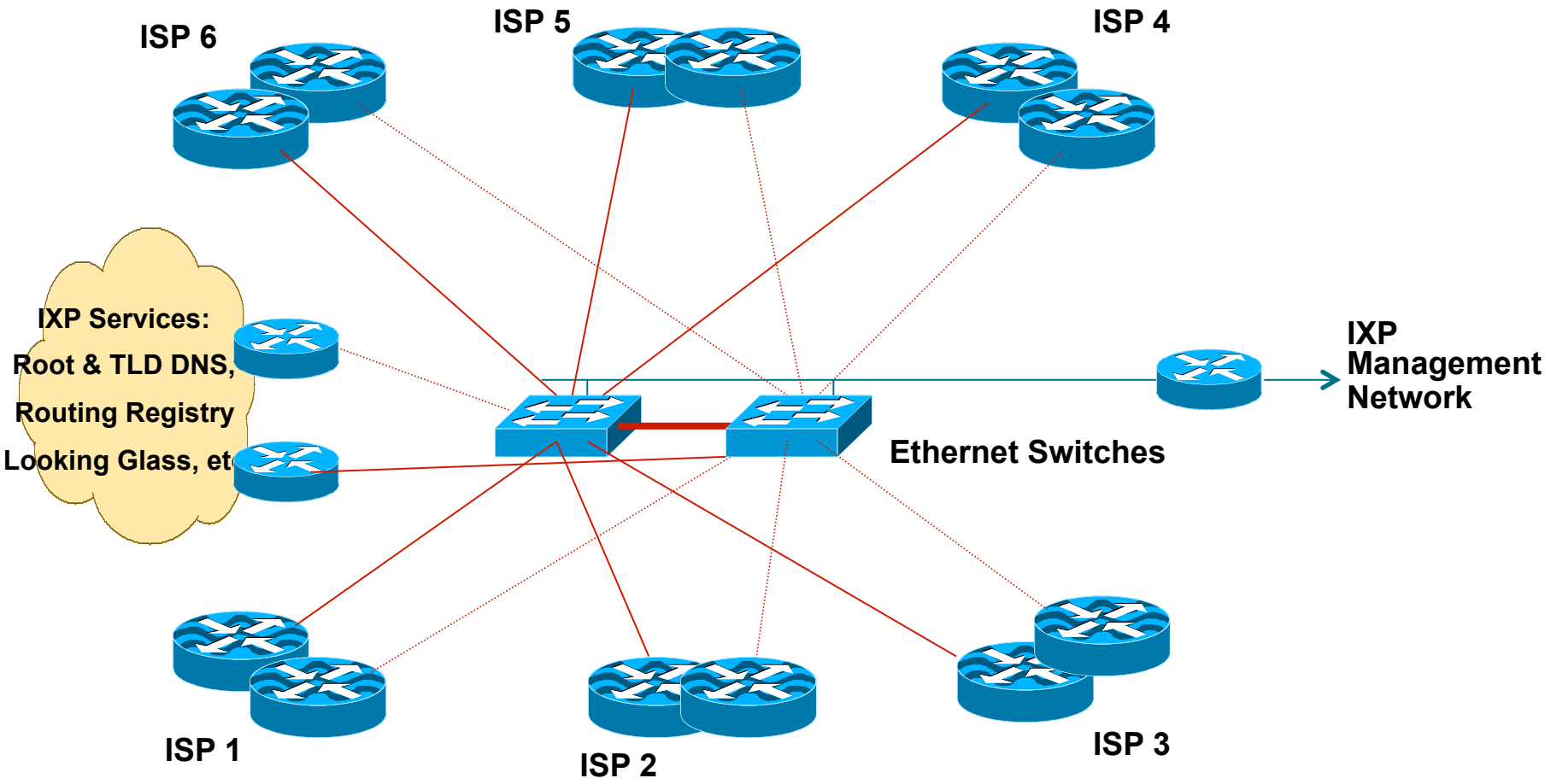
IXP Design

- Very simple concept:
  - Ethernet switch is the interconnection media
    - IXP is one LAN
  - Each ISP brings a router, connects it to the ethernet switch provided at the IXP
  - Each ISP peers with other participants at the IXP using BGP
- Scaling this simple concept is the challenge for the larger IXPs

Layer 2 Exchange



Layer 2 Exchange



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 Root & TLD, NTP servers
  - Routing Registry and Looking Glass

Layer 2 Exchange

- Requires neutral IXP management
  - Usually funded equally by IXP participants
  - 24x7 cover, support, value add services
- Secure and neutral location
- Configuration
  - IPv4 /24 and IPv6 /64 for IXP LAN
  - ISPs require AS, basic IXP does not

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

“Layer 3 IXP”

- Layer 3 IXP is marketing concept used by Transit ISPs
- Real Internet Exchange Points are only Layer 2

IXP Design Considerations

Exchange Point Design

- The IXP Core is an Ethernet switch
  - It must be a managed switch
- Has superseded all other types of network devices for an IXP
  - From the cheapest and smallest managed 12 or 24 port 10/100 switch
  - To the largest switches now handling high densities of 10GE and 100GE interfaces

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

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 1Gbps, 10Gbps, 40Gbps or 100Gbps connections

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

Peering (more)

- Mandatory Multi-Lateral Peering (MMLP)
  - Each participant is forced to peer with every other participant as part of their IXP membership
  - **Has no history of success** — the practice is strongly discouraged
- Multi-Lateral Peering (MLP)
  - Each participant peers with every other participant (usually via a Route Server)
- 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

Routing

- ISP border routers at the IXP must 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

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

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

Charging

- IXPs should be run at minimal cost to participants
- Examples:
  - Datacentre hosts IX for free
    - Because ISP participants then use data centre for co-lo services, and the datacentre benefits long term
  - IX operates cost recovery
    - Each member pays a flat fee towards the cost of the switch, hosting, power & management
  - Different pricing for different ports
    - One slot may handle 24 10GE ports
    - Or one slot may handle 96 1GE ports
    - 96 port 1GE card is tenth price of 24 port 10GE card
    - Relative port cost is passed on to participants

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/Cacti to produce IX throughput graphs for member (or public) information

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

Services to Offer

- Route Collector
  - Route collector shows the reachability information available at the exchange
- Looking Glass
  - One way of making the Route Collector routes available for global view (e.g. www.traceroute.org)
  - Public or members only access
  - Useful for members to check BGP filters
  - Useful for everyone to check route availability at the IX

Services to Offer

- Route Server
  - A Route Collector that also sends the prefixes it has collected to its peers
  - Like a Route Collector, usually a router or Unix based system running BGP
  - Does not forward packets
  - Useful for scaling eBGP sessions for larger IXPs
  - Participation needs to be optional
    - And will be used by ISPs who have open peering policies

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

What can go wrong?

- High annual fees
  - Should be cost recovery
- Charging for traffic between participants
  - Competes with commercial transit services
- Competing IXPs
  - Too expensive for ISPs to connect to all
- Too many rules & restrictions
  - Want all network operators to participate
- Mandatory Multi-Lateral Peering
  - Has no history of success
- Interconnected IXPs
  - Who pays for the interconnection?
- Etc…

Conclusion

- IXPs are technically very simple to set up
- Little more than:
  - An ethernet switch
  - Neutral secure reliable location
  - Consortium of members to operate it
- Political aspects can be more challenging:
  - Competition between ISP members
  - “ownership” or influence by outside parties