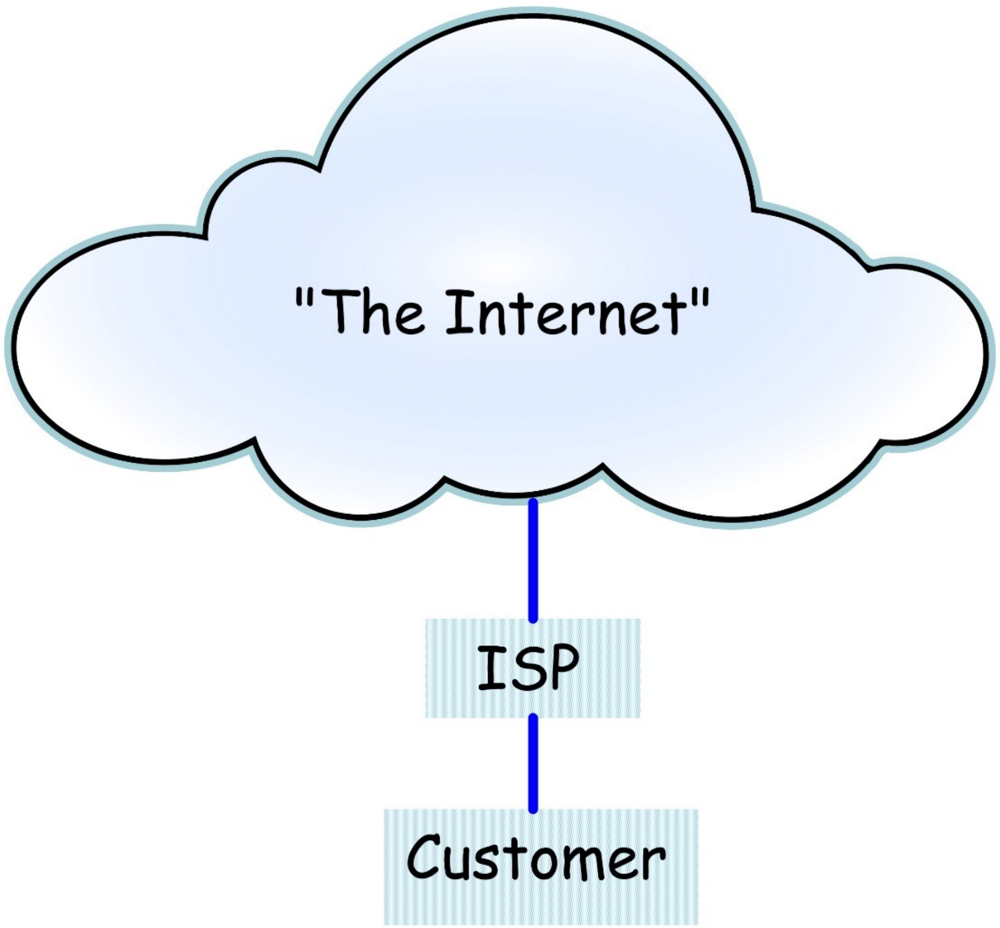
Internet Peering

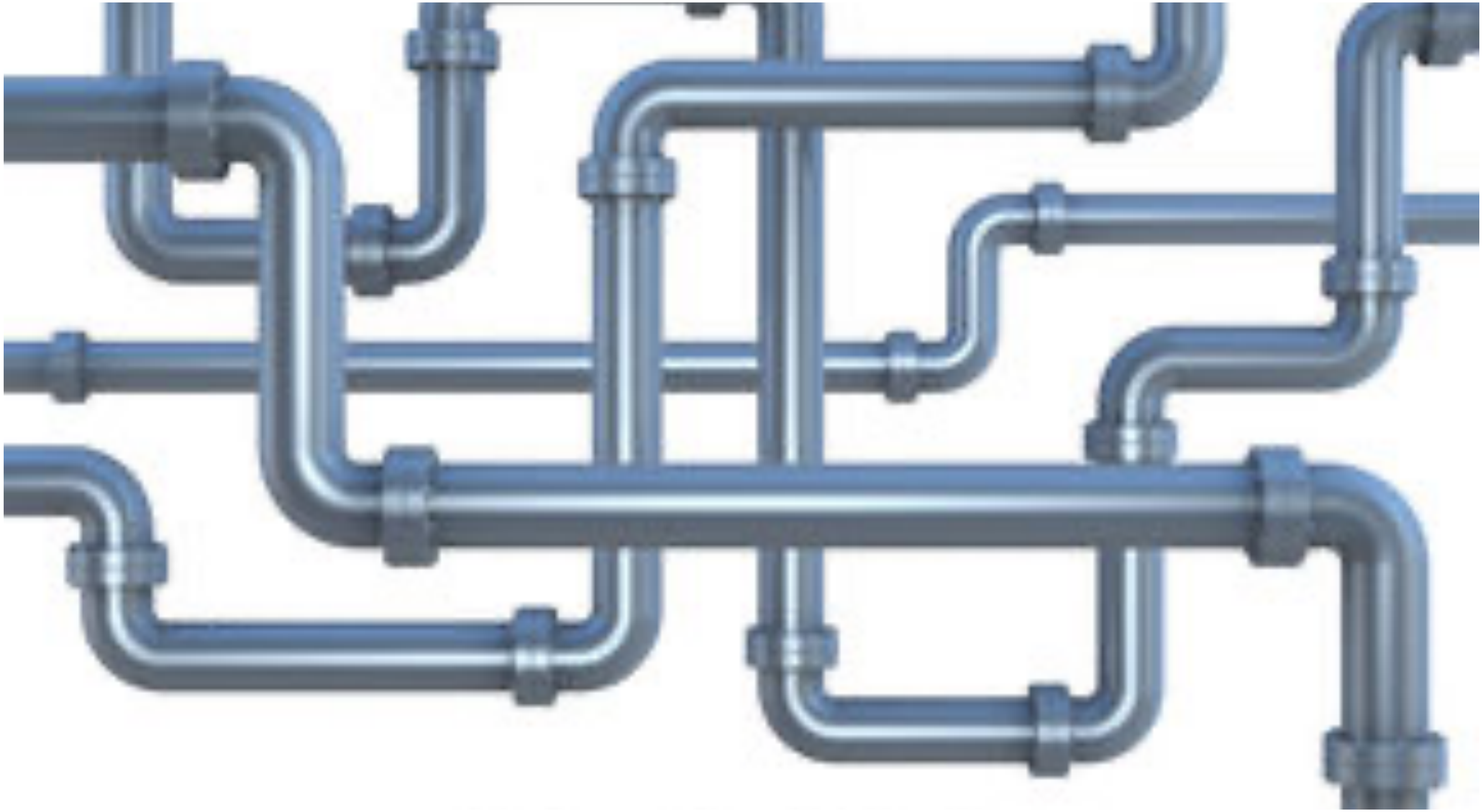
Why, How, Where, …

APNIC

Customer's Expectation



But it's really just…

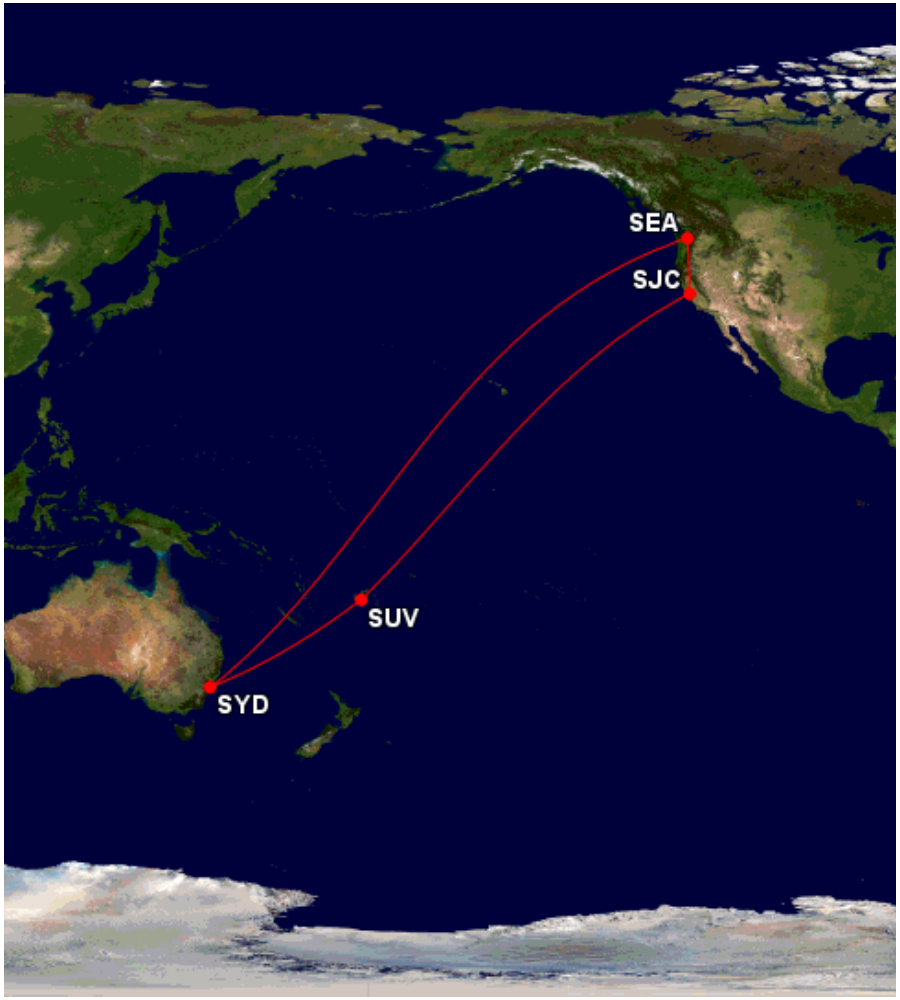


Until this happens

Or this

```
 2  core1-vl400.vcc.kidanet.com.fj (113.20.64.49)  327.221 ms  204.824 ms  12.070 ms
 3  202.170.41.85 (202.170.41.85)  1.564 ms  2.537 ms  1.383 ms
 4  202.170.33.2 (202.170.33.2)  2.826 ms  2.738 ms  1.563 ms
 5  gi0-2-1-4.rcr21.b001848-1.sjc01.atlas.cogentco.com (38.122.92.249)  123.644 ms  123.736 ms  123.017 ms
 6  be2063.ccr21.sjc01.atlas.cogentco.com (154.54.1.161)  124.323 ms
    be2095.ccr22.sjc01.atlas.cogentco.com (154.54.3.137)  124.578 ms
    be2063.ccr21.sjc01.atlas.cogentco.com (154.54.1.161)  123.759 ms
 7  be3144.ccr41.sjc03.atlas.cogentco.com (154.54.5.102)  124.418 ms  124.695 ms
    be3142.ccr41.sjc03.atlas.cogentco.com (154.54.1.194)  123.785 ms
 8  zayo.sjc03.atlas.cogentco.com (154.54.10.194)  126.692 ms  125.425 ms  124.132 ms
 9  ae16.cr2.sjc2.us.zip.zayo.com (64.125.31.14)  126.694 ms  123.851 ms  124.828 ms
10  ae27.cs2.sjc2.us.eth.zayo.com (64.125.30.232)  142.824 ms  142.947 ms  142.736 ms
11  ae3.cs2.sea1.us.eth.zayo.com (64.125.29.41)  142.369 ms  142.763 ms  142.015 ms
12  ae28.mpr1.sea1.us.zip.zayo.com (64.125.29.105)  142.880 ms  144.592 ms  142.519 ms
13  64.125.193.130.i223.above.net (64.125.193.130)  162.471 ms  163.139 ms  162.358 ms
14  xe-1-0-1.pe2.brwy.nsw.aarnet.net.au (202.158.194.120)  163.443 ms  162.016 ms  163.059 ms
15  ae9.bb1.a.syd.aarnet.net.au (113.197.15.57)  162.210 ms  163.574 ms  162.243 ms
16  ge-1-1-0.bb1.a.suv.aarnet.net.au (202.158.194.226)  198.100 ms  197.932 ms
```

What’s wrong with this picture?



- Fintel customer in Suva
- Accessing content at the University of the South Pacific in Suva
- Packet travels > 25,000km
- Physical distance < 10km
- Adding latency
- Possibly jitter too
- Using expensive submarine capacity

IP Transit

- Provide access to "The Internet"
- Requires a circuit to an "upstream" ISP
  - Could be local (domestic) or international
  - Submarine circuits are fixed capacity, not tied to usage
- Also requires service from the "upstream" ISP
  - Billing is based on usage, typically 95th percentile
  - Or based on the speed of the connection (rate-limited or not)
- Repeat to get the level of redundancy required
  - Two circuits to the same "upstream" ISP
  - Circuits to two, or more, "upstream" ISPs

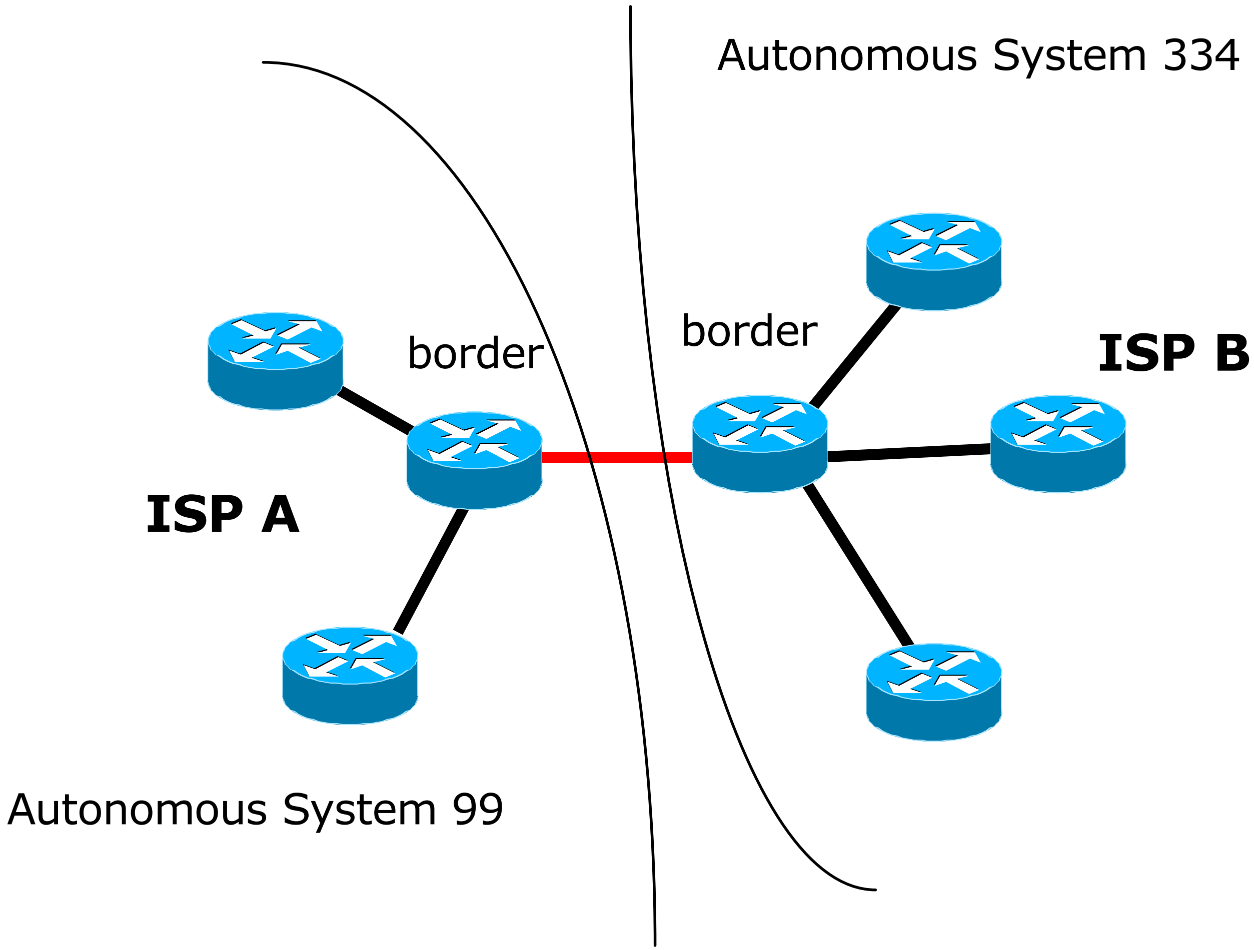
Interconnection (aka Peering)

- Connection to a "peer" network
  - Exchange of traffic to customers of each peer
- Requires a circuit to the peer (or to an Internet Exchange)
  - Fixed cost based on capacity of the link
  - May also require a cross connect in a data centre
  - Could be fixed cost or more likely monthly recurring fee
- Traffic is settlement free mostly
- Cost is the same if zero bytes exchanged or link saturated
  - Don't saturate the link, customers will be grumpy ☺

We compete, why interconnect?

- International Connections…
  - If satellite, RTT is around 550ms per hop
  - Compared to local traffic < 10ms round trip
- International bandwidth
  - Costs significantly more than domestic bandwidth
  - Don’t congest it with local traffic
    - Wastes money
  - Harms overall performance (end-user experience)
- Lose-lose if not interconnect locally

Private Interconnect



Interconnection (aka Peering)

- Local (loop) connections
- Not in a customer/transit relationship
- Sharing customer & infrastructure routes only
  - Routes that generate revenue for you
- Share costs
  - Two circuits, pay for one each

Results of Peering

- Both save money
- Local traffic stays local
- Better performance, better QoS, …
- Expensive international bandwidth available for actual international traffic
- Everyone is happy (except submarine cable and satellite owners)
- It is win-win

Scaling peering

- What happens when new ISPs enter the equation?
  - Just repeat the process?

- Private peering means that each ISP has to buy circuits to every other peer (perhaps 2 for redundancy)
  - For (n) peers in total, each peer needs (n-1) half circuits
    - Eg 10 peers in total => 9 half circuits for each

Why an Internet eXchange Point (IXP)

- Private peering relies on just the two parties making best use of the circuit
  - by building dedicated circuits to each peer
    - n peers in total => n(n-1)/2 circuits in total
- With an IXP:
  - Every participant has to buy just one whole circuit
    - From their premises to the IXP fabric
  - Improve latency performance between peers where traffic volume wouldn't justify a dedicated circuit
  - Maximizes the opportunity to fill the circuit
    - Peak traffic may not be the same across all peers

Internet eXchange Point (IXP)

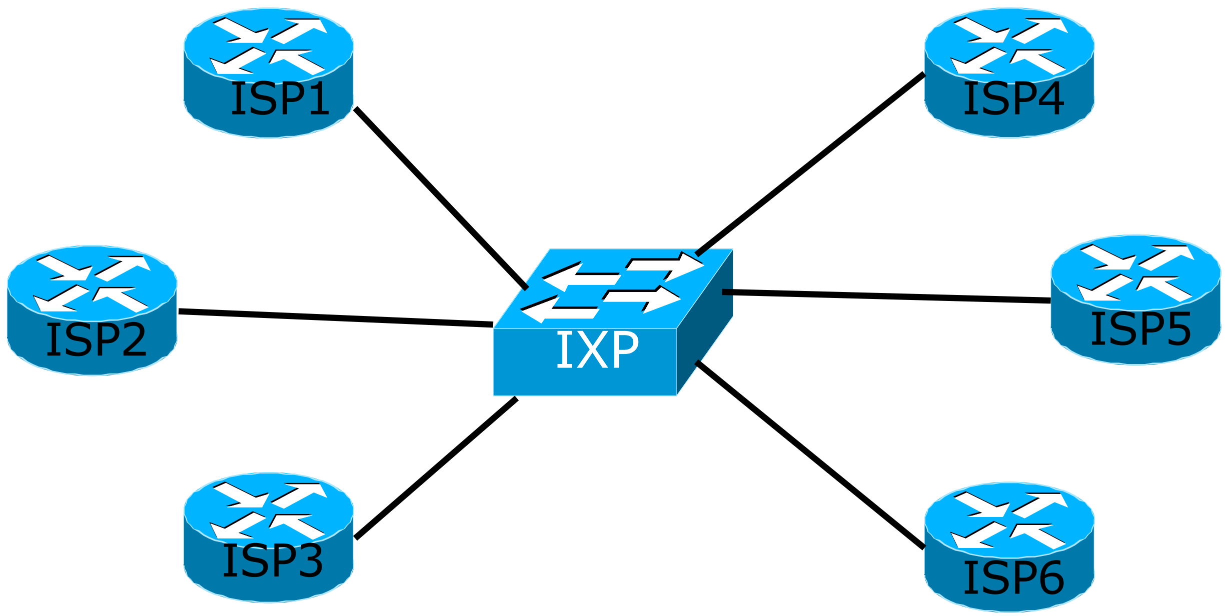
- Need a location or facility that ISPs can access and can connect to each other over a common shared media
  - Eg: Ethernet switch
- Should be a NEUTRAL venue
- Needs to have multiple telco circuit providers and/or allow any licenced provider to install services
- Needs controlled environment & access

Internet eXchange Points

- Variety of shapes and sizes
  - Commercial
  - Community
  - Tbps to Mbps
  - Single location or Metropolitan Area scoped
  - Purely a traffic exchange
  - Value added services

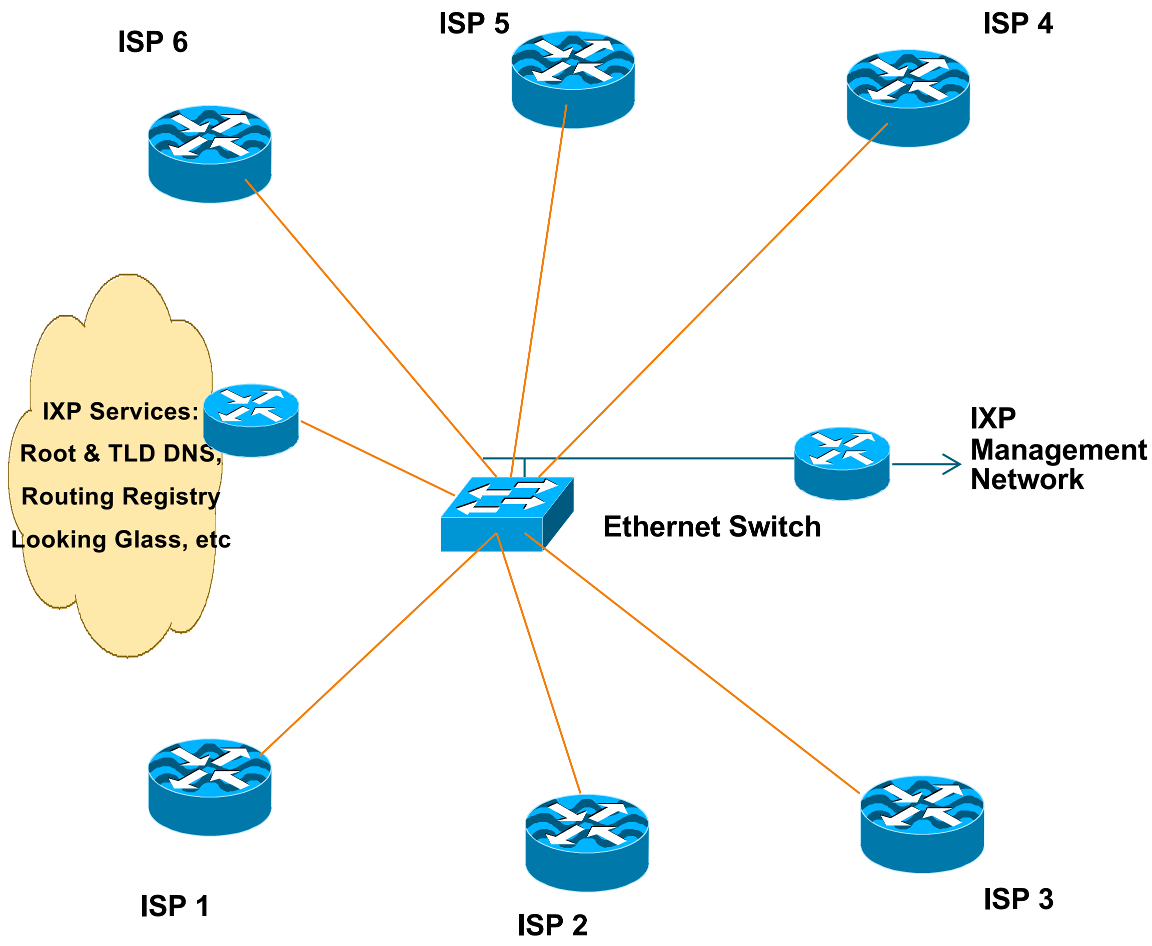
- Layer 2 exchange point
  - Ethernet Switches (100Gbps/10Gbps/1Gbps/100Mbps)

Internet eXchange Point

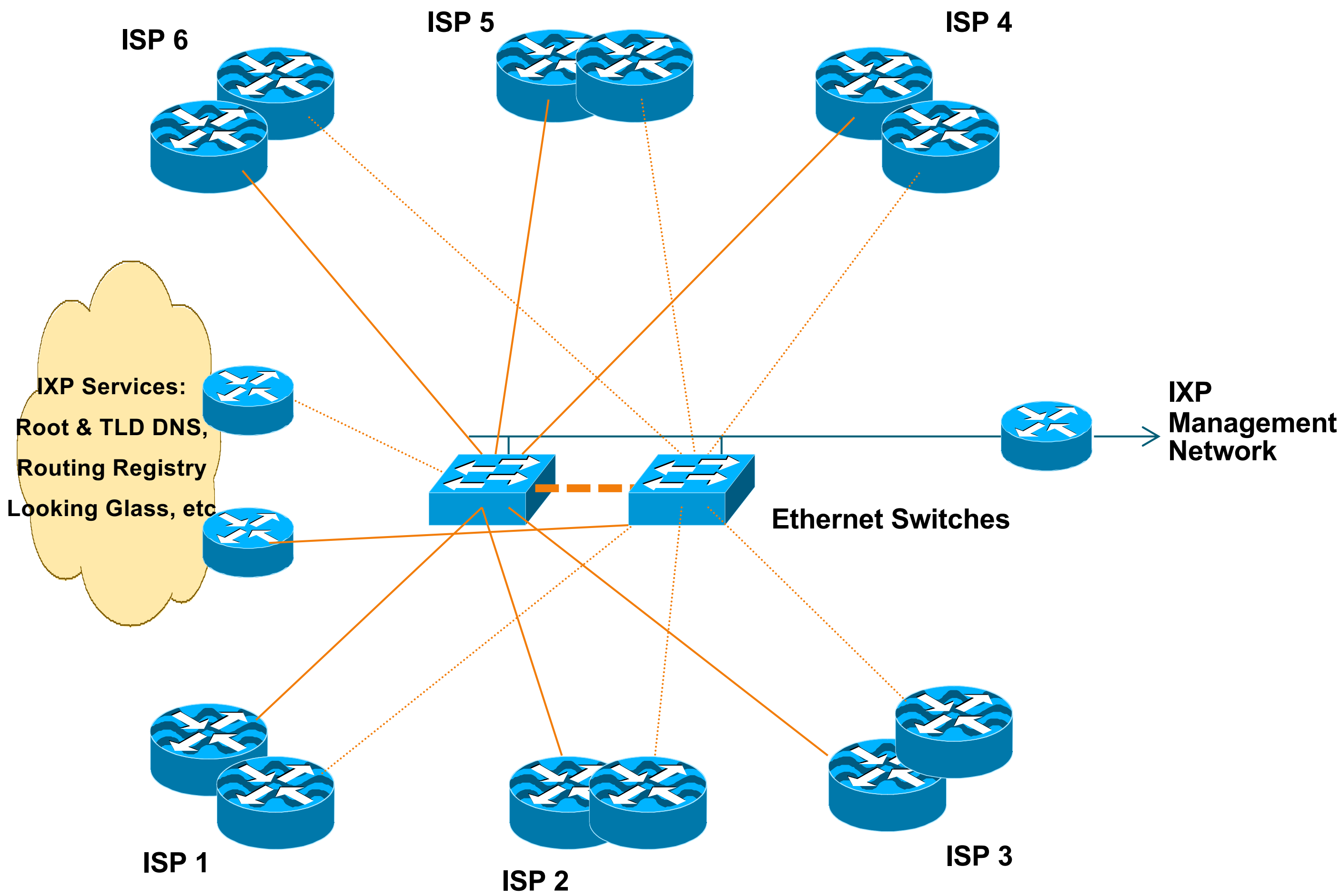


- Border routers in different Autonomous Systems

Layer 2 Exchange



Layer 2 Exchange



Layer 2 Exchange

- Two switches for redundancy
- ISPs use dual routers for redundancy or load-sharing
- Offer services for the "common good"
  - Internet portals and search engines
  - DNS Root & TLDs, 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
  - Private address space if non-transit and no value add services
  - Otherwise public IPv4 (/24) and IPv6 (/48, /56, /64)
  - ISPs require ASN, basic IXP does not
    - Route Servers need ASN

Layer 2 Exchange

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

Defining some terms

Types of Peering

- Private Peering
- Bi-lateral Peering
- Multi-lateral Peering

Private Peering

- Dedicated circuit between two peers
  - Can use a cross connect within a data centre
  - Or via dark fibre, telco circuit, microwave, …
- Used where traffic levels high between two peers
- Expensive, cost shared between only two parties
  - Often in pairs; each peer pays for one
- But ultimate in control

Bi-lateral Peering

- Uses an Ethernet switch at an Internet Exchange
- Single cross connect to the switch
  - Peer can be remote (e.g. using Metro-Ethernet)
- Dedicated BGP peering between two peers
- Relies on the IXP to manage the switch
- Bandwidth shared by multiple peering relationships
- But direct relationship between the two peers
  - More control (granularity)
  - If bad things happen can turn down BGP on one peer

Multi-lateral Peering (MLPA)

- Uses an Ethernet switch at an Internet Exchange
- Single cross connect to the switch
- Single BGP peering session to a “route server”
- Easiest to setup, only one session
  - Automatically peer with everyone else
- Reliant on IXP for both switch and route server
- Relationship is with the IXP
- Lesser control (granularity)
  - If a peer has a problem less options to workaround

Types of Peering Policy

- Open
- Selective
- Restrictive

Open Peering

- "Have a pulse peering"
- Will peer with anyone
  - Typically bi-lateral or multi-lateral at an existing facility
  - Negligible additional cost so why not?
- Typically content providers have open peering policy

Selective Peering

- Conditional peering
  - Ex: at an IXP, will ONLY peer bilaterally and NOT with the RS
- Some negotiation may be necessary
- May have some rules that peers must fulfil
  - volumes, ratios, number of multiple connects
- May only peer outside of primary market

Restrictive Peering

- Rules!
- Has a (written) policy that defines if they will peer
  – Often with rules, which are set so that they don't peer
- Often involves a minimum level of traffic
  – Could require a test peering to check conformance
- Also can include a "ratio" in/out traffic levels

Controlling costs

Cost tied to circuit size (not byte count)

- Peering is typically settlement free
  - No charge for the traffic exchanged

- Cost to peer
  - Router interface
  - Circuit to the peering fabric
  - Charges imposed by the IXP
  - All fixed, either capital expenditure or monthly recurring fee

Choosing a IXP

- Some markets have more than one
- Even if there is only one IXP it might appear in multiple locations
  - E.g. LINX is built on two rings through multiple data centres across London
- Best location might be dictated by availability of IPLC, transit, or other factors

Which IXP?

- How many routes are available?
  - How many other operators/providers are at the IX?
  - What is the traffic to and from these destinations, and how much will it reduce the transit cost?
- What is the cost of co-lo space?
  - Availability of power, type of cabinet, …
- What is the cost of a circuit to the location?
  - If similar to transit costs are you getting a benefit?
- What is the cost of remote-hands?
  - For maintenance purposes to avoid serious outages

Remote locations

- If building to a remote location
- Make sure remote hands work at times when it's important to you
  - Their 9-5 is not normally your office hours
- Check the skill set of the remote hands
  - Maybe engage a local consultant to help

Worked Example

Single International Transit versus Local IXP + Regional IXP + Transit

APNIC

Worked Example

- ISP A is local access provider
  - Some business customers (around 200 fixed links)
  - Some co-located content provision (datacentre with 100 servers)
  - Some consumers on broadband (5000 DSL/Cable/Wireless)
- They have a single transit provider
  - Connect with a 16Mbps international leased link to their transit's PoP
  - Transit link is highly congested

Worked Example (2)

- There are two other ISPs serving the same locality
  - There is no interconnection between any of the three ISPs
  - Local traffic (between all 3 ISPs) is traversing International connections
- Course of action for our ISP:
  - Work to establish local IXP
  - Establish presence at overseas co-location
- **First Step**
  - Assess local versus international traffic ratio
  - Use NetFlow on border router connecting to transit provider

Worked Example (3)

- Local/Non-local traffic ratio
  - Local = traffic going to other two ISPs
  - Non-local = traffic going elsewhere
- Example: balance is 30:70
  - Of 16Mbps, that means 5Mbps could stay in country and not congest International circuit
  - 16Mbps transit costs $50 per Mbps per month
    - local traffic charges = $250 per month, or $3K per year for local traffic
  - Circuit costs $100K per year => $30K is spent on local traffic
- Total is $33K per year for local traffic

Worked Example (4)

- IXP cost:
  - Simple 8 port 10/100 managed switch plus co-lo space over 3 years could be around US$30K total => $3K per year per ISP
  - One router to handle 5Mbps would be around $9K, good for 3 years => $3K per year
  - One local 10Mbps circuit from ISP location to IXP location would be around $5K per year, no traffic charges
  - Per ISP total: $11K
  - Somewhat cheaper than $33K
  - Business case for local peering is straightforward - $22K saving per annum

Worked Example (5)

- After IXP establishment
  - 5Mbps removed from International link
  - Leaving 5Mbps for more International traffic – and that fills the link within weeks of the local traffic being removed
- Next step is to assess transit charges and optimise costs
  - ISPs visits several major regional IXPs
  - Assess routes available
  - Compares routes available with traffic generated by those routes from its NetFlow data
  - Discovers that 30% of traffic would transfer to one IXP via peering

Example: South Asian ISP @ LINX

- Date: May 2013

- Data:
  - Route Server plus bilateral peering offers 70K prefixes
  - IXP traffic averages 247Mbps/45Mbps
  - Transit traffic averages 44Mbps/4Mbps

- Analysis:
  - 85% of inbound traffic comes from 70K prefixes available by peering
  - 15% of inbound traffic comes from remaining 380K prefixes from transit provider

Example: South Asian ISP @ HKIX

- Date: May 2013

- Data:
  - Route Server plus bilateral peering offers 67K prefixes
  - IXP traffic is 159Mbps/20Mbps
  - Transit traffic is 108Mbps/50Mbps

- Analysis:
  - 60% of inbound traffic comes from 67K prefixes available by peering
  - 40% of inbound traffic comes from remaining 383K prefixes from transit provider

Example: South Asian ISP

- Summary:
  - Traffic by Peering: 406Mbps/65Mbps
  - Traffic by Transit: 152Mbps/54Mbps

  - 73% of incoming traffic is by peering
  - 55% of outbound traffic is by peering

Example: South Asian ISP

- Router at remote co-lo
  - Benefits: can select peers, easy to swap transit providers
  - Costs: co-lo space and remote hands

- Overall advantage:
  - Can control what goes on the expensive connectivity “back to home”

Value propositions

- Peering at a local IXP
  - Reduces latency & transit costs for local traffic
  - Improves Internet quality perception
- Participating at a Regional IXP
  - A means of offsetting transit costs
- Managing connection back to home network
- Improving Internet Quality perception for customers