INTERNET SERVICE PROVIDER(ISP)

Affordable Home Connectivity to
End users

ACCESS THAT KEEPS PACE

Designing Scalable ISP Networks

Acknowledgements

- This material originated from the Cisco ISP/IXP Workshop Programme developed by Philip Smith & Barry Greene
  - I'd like to acknowledge the input from many network operators in the ongoing development of these slides, especially Mark Tinka of SEACOM for his contributions

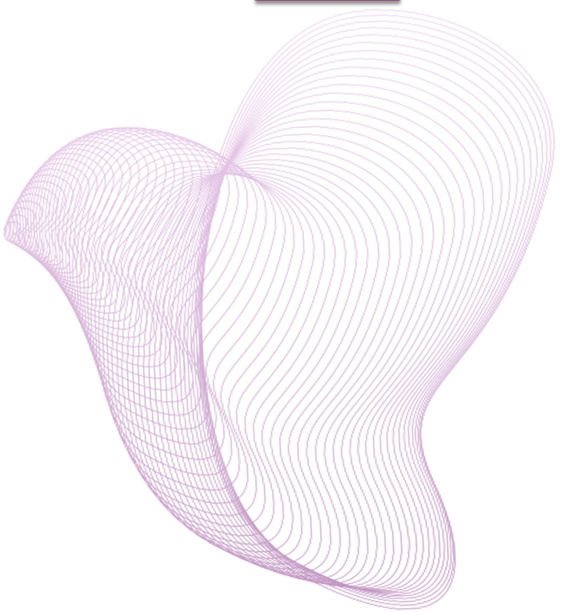
- Bug fixes and improvements are welcomed
  - Please email *workshop (at) bgp4all.com*

Philip Smith

What Is a Well-Designed Network?

One that takes into consideration some main factors :

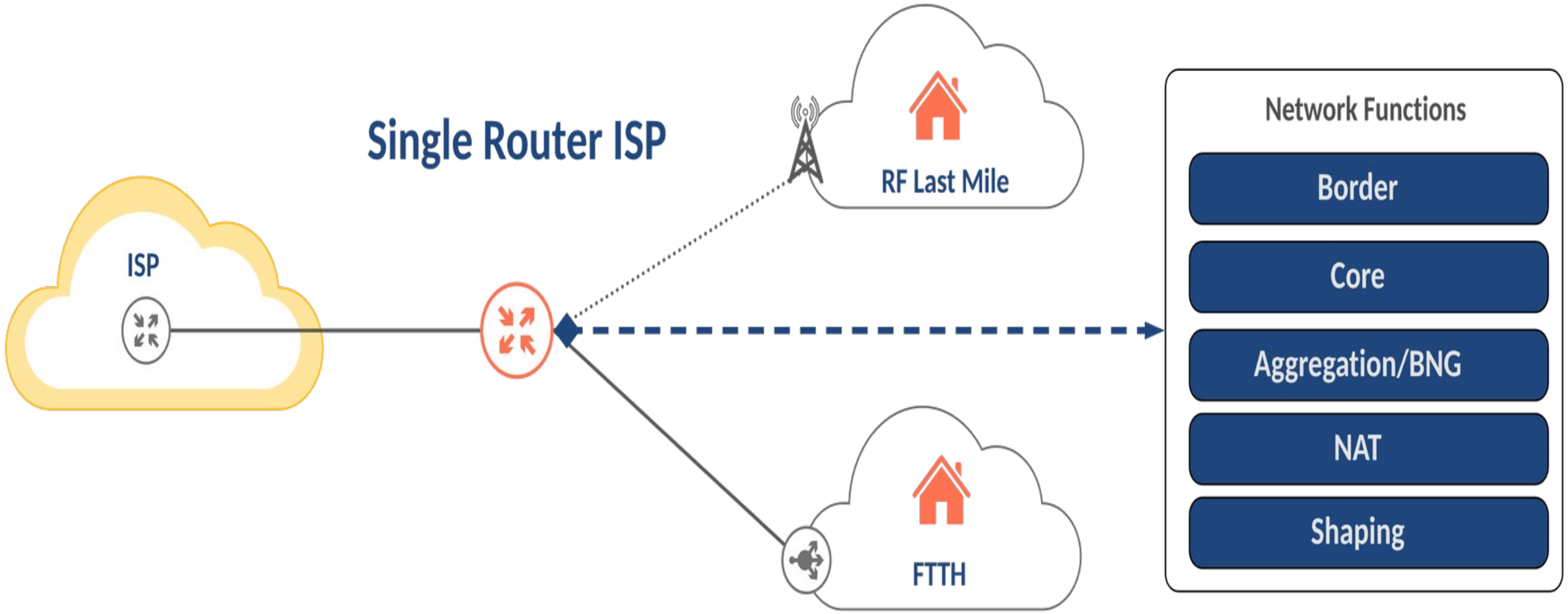
- Topological/protocol hierarchy
- Redundancy
- Addressing aggregation (IGP and BGP)
- Scaling
- Policy implementation (core/edge)
- Management/maintenance/operations
- Cost



One Must Acknowledge that…

- **Two different worlds exist**
  - We must admit two worlds currently exist in the ISP space, Lager ISP (Transit & Corporate), Smaller ISP(End User).
  - Larger ISP are financially capable of a achieving a fully redundant/Scalable Network.
  - Smaller ISP/ Startups grow from a One Router ISP perspective (All in one)

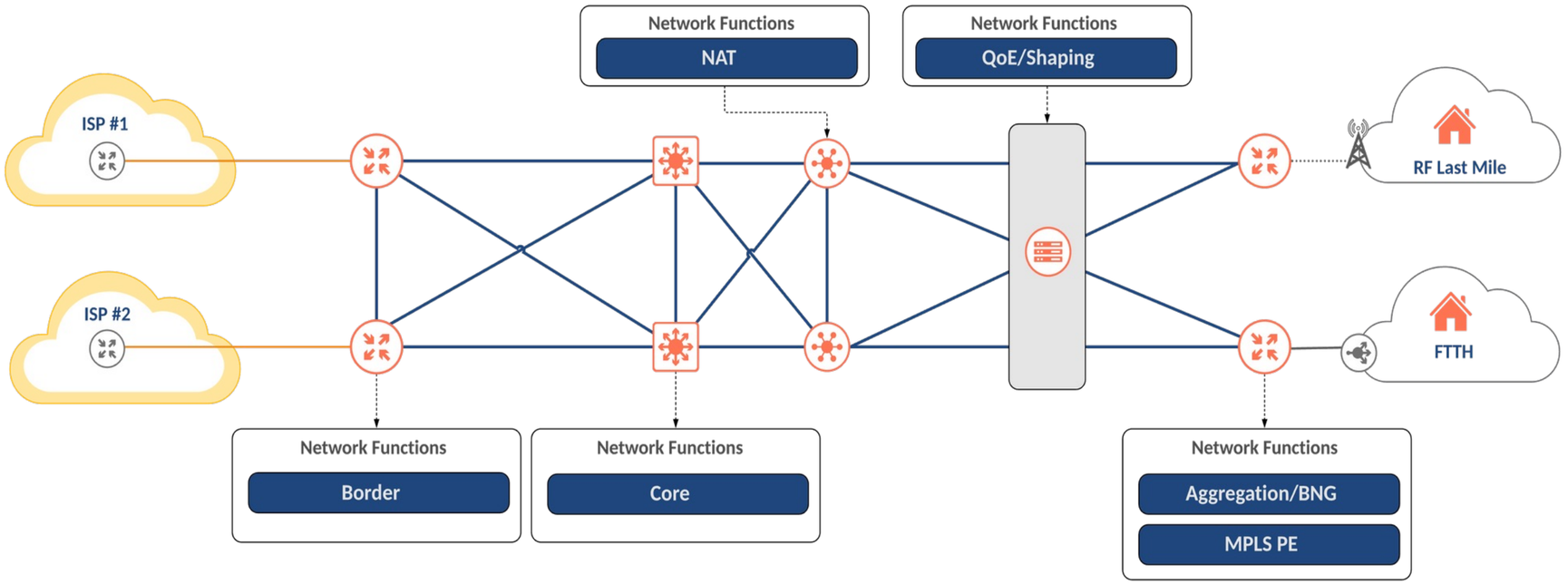
Single Router ISP



Modular Designed ISP – (Seperated Functions)

Overview of an ISP capable of 10,000+ subscribers with separated functions

Leveraging whitebox and commodity vendors for scale - At first glance, this may seem like a lot of gear, but as we go through the individual functions, the flexibility and scalability will become clearer. A decade ago this architecture would be unaffordable with vendors like Cisco and Juniper, but thanks to whitebox and commodity vendors like IP Infusion, Edge Core, MikroTik and UfiSpace, WISPs all over the world have production networks capable of 10s of thousands of subscribers using this design.



Point of presence Layers

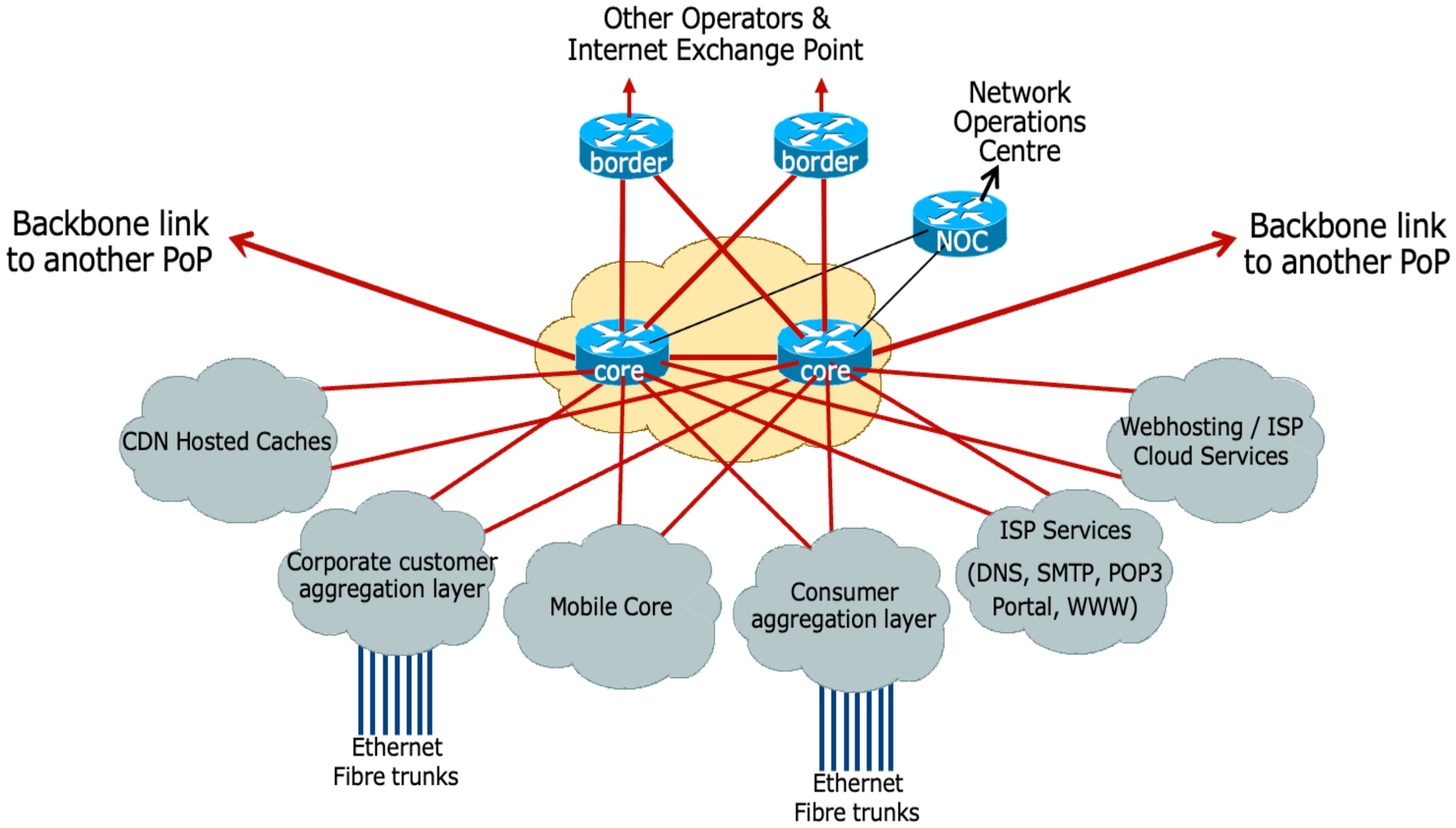
POP Topologies

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

POP Design

- Modular Design
- Aggregation Services separated according to:
  - connection speed
  - customer service
  - contention ratio
  - security considerations

Modular PoP Design



Modular Routing Protocol Design

- Modular IGP implementation
  - IGP “area” per POP
  - Core routers in backbone area (Area 0/L2)
  - Aggregation/summarisation where possible into the core
- Modular iBGP implementation
  - BGP route reflector cluster
  - Core routers are the route-reflectors
  - Remaining routers are clients & peer with route-reflectors only

Point of presence Design

PoP Core

- Two dedicated high performance routers
- Technology
  - High Speed interconnect (10Gbps, 100Gbps, 400Gbps)
  - Backbone Links ONLY; no access services
  - *Do not touch them!*
- Service Profile
  - 24x7, high availability, duplicate/redundant design

PoP Core – details

- Router specification
  - High performance control plane CPU
  - Does not need a large number of interface/line cards
    - Only connecting backbone links and links to the various services
- High speed interfaces
  - Aim as high as possible
  - 10Gbps is the typical standard initial installation now
    - Price differential between 1Gbps and 10Gbps justifies the latter
    - when looking at cost per Gbps
  - Many operators using aggregated 10Gbps links, also 100Gbps

Border Network

- Dedicated border routers to connect to other Network Operators
- Technology
  - High speed connection to core
  - Significant BGP demands, routing policy
  - DDoS front-line mitigation
  - Differentiation in use:
    - Connections to Upstream Providers (Transit links)
    - Connections to Private Peers and Internet Exchange Point
- Service Profile
  - 24x7, high availability, duplicate/redundant design

Border Network – details

- Router specification
  - High performance control plane CPU
  - Only needs a few interfaces
    - Only connecting to external operators and to the network core routers
  - Typically a 1RU or 2RU device
- High speed interfaces
  - 10Gbps standard to the core
  - 10Gbps to Internet Exchange Point
  - Ethernet towards peers (1Gbps upwards)
  - Ethernet towards transit providers (1Gbps upwards)

Border Network – details

- Router options:
  - Router dedicated to private peering and IXP connections
    - Only exchange routes originated by respective peers
      - No default, no full Internet routes
    - Control plane CPU needed for BGP routing table, applying policy, and assisting with DDoS mitigation
  - Router dedicated to transit connectivity
    - Must be separate device from private peering/IXP router
      - Usually carries full BGP table and/or default route
    - Control plane CPU needed for BGP routing table, applying policy, and assisting with DDoS mitigation
- Note: the ratio of peering traffic to transit traffic volume is around 3:1 today

Corporate Customer Aggregation

- Business customer connections
  - High value, high expectations
- Technology
  - Fibre to the premises (FTTx or GPON)
  - Aggregated within the PoP module
  - Usually managed service; customer premise router provided by the operator
- Service Profile
  - Typically demand peak performance during office hours
  - Out of hours backups to the “Cloud”

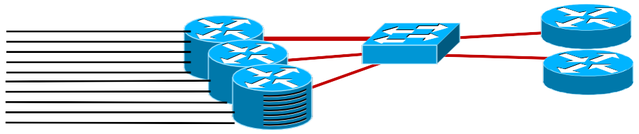
Corporate Customer Aggregation – details

- Router specification
  - Mid-performance control plane CPU
  - High interface densities
- Interface types:
  - 10Gbps uplink to core
  - Multiple 10Gbps trunks
    - Customer connections delivered per VLAN
    - Provided by intermediate ethernet switch or optical equipment

Corporate Customer Aggregation – details

- Router options:
  - Several smaller devices, aggregating multiple 1Gbps trunks to 10Gbps uplinks
    - Typically 1RU routers with 16 physical interfaces
      - 12 interfaces used for customer connections, 4 interfaces for uplinks
    - May need intermediate Distribution Layer (usually ethernet switch) to aggregate to core routers

Direct Fibre Trunk Links



- One larger device, multiple aggregation interfaces, with multiple 10Gbps or single 100Gbps uplink to core
- Typical 8RU or larger with >100 physical interfaces

Consumer Aggregation

- Home users and small business customer connections
  - Low value, high expectations
- Technology:
  - Fibre to the premises (FTTx or GPON)
  - Still find Cable, ADSL and 802.11 wireless used
  - Aggregated within the PoP module
  - Unmanaged service; with customer premise router provided by the customer
- Service Profile
  - Typically demand peak performance during evenings

Consumer Aggregation – details

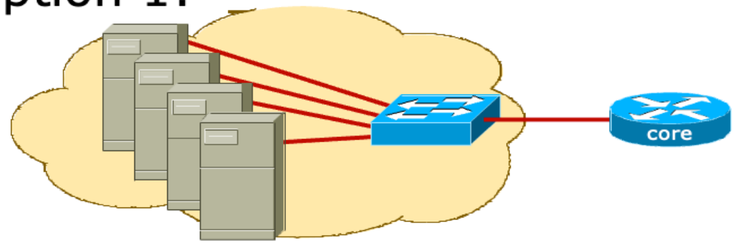
- Router specification
  - Mid-performance control plane CPU
  - High interface densities
- Interface types:
  - 10Gbps uplink to core
  - Multiple 10Gbps trunks
    - Customer connections delivered per VLAN
    - Provided by intermediate ethernet switch or optical equipment

CDN Hosted Services and Caches

- Content provider supplied infrastructure
- Technology:
  - Each CDN provides its own equipment
    - Usually a number of servers & ethernet switch, possibly a router
  - Requires direct and high bandwidth connection to the Core Network
    - Used for cache fill
    - Used to serve end-users
- Service Profile
  - High demand high availability 24x7

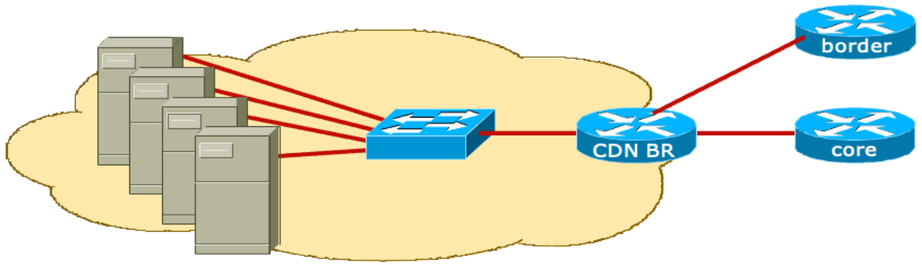
CDN Hosted Services and Caches – details

- Every CDN is different, but follow a similar pattern
  - Option 1:



Connection to Network Operator Core Router

  - Option 2:



Connection to Network Operator Border Router (Transit/Cache-Fill)

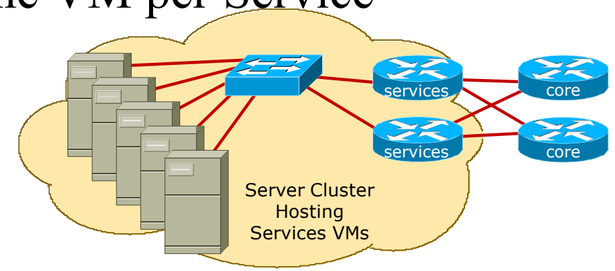
And Core Router (End-User Access)

Network Operator Services

- Infrastructure / Customer services
- Technology:
  - Redundant server cluster behind two routers, hosting virtual machines
  - One virtual machine per service
- Services
  - DNS (2x cache, 2x authoritative)
  - Mail (SMTPS Relay for Customers, POP3S/IMAPS for Customers, SMTP for incoming e-mail)
  - WWW (Operator Website)
- Portal (Customer Self-Service Portal) / Billing Infrastructure

Network Operator Services – details

- Infrastructure is usually multiple 1RU or 2RU servers configured into a cluster
  - Hosting Virtual Machines, one VM per Service
  - Examples:
    - WWW
    - Customer Portal
    - Authoritative DNS
    - DNS Cache (Resolver)
    - SMTP Host (incoming email)
    - SMTPS Relay (outgoing email from customers)
    - POP3S/IMAPS (Secure Mail Host for customers),

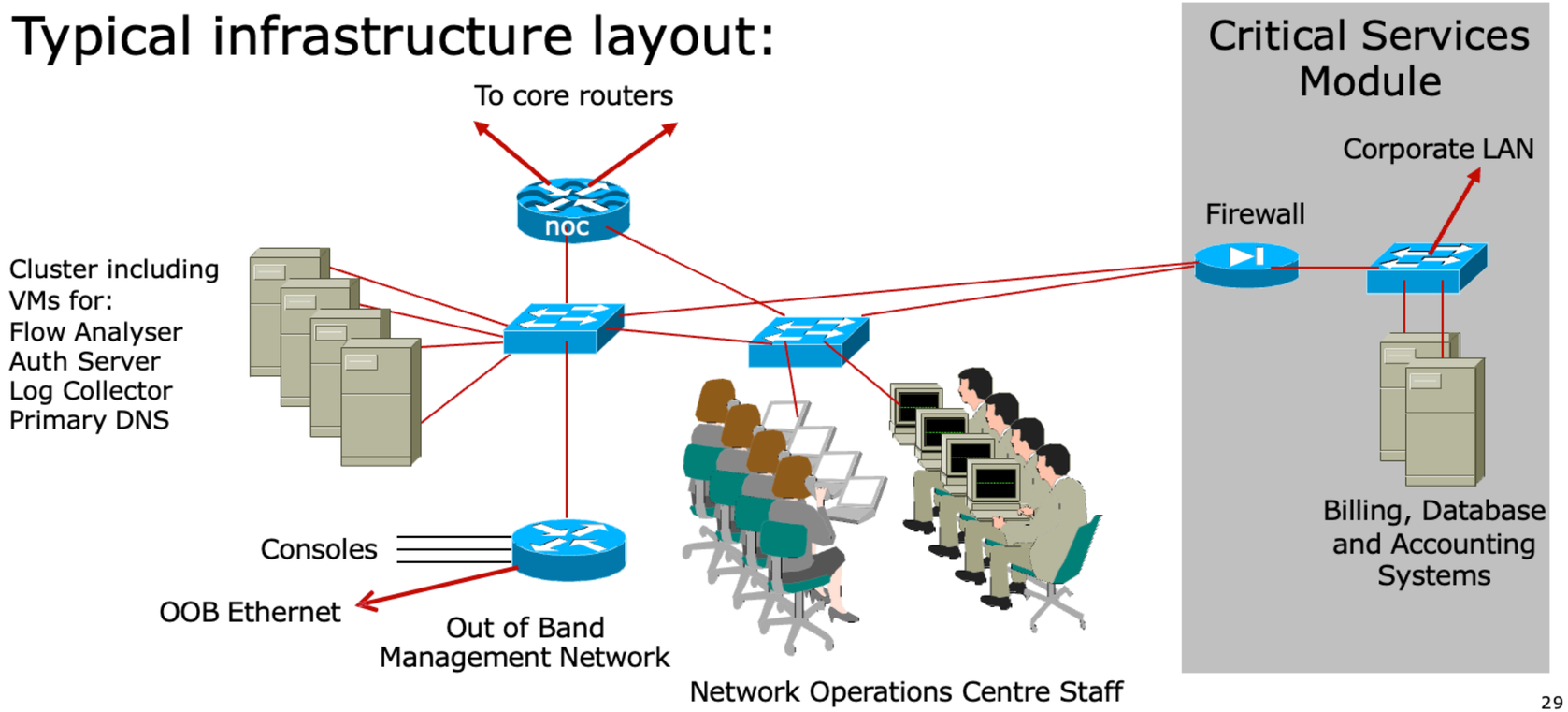


Network Operations Centre

- Management of the network infrastructure
- Technology:
  - Gateway router, providing direct and secure access to the network operator core backbone infrastructure
- Services:
  - Network monitoring
  - Traffic flow monitoring and management
  - Statistics and log gathering
  - RTBH management for DDoS mitigation
  - Out of Band Management Network
    - The Network “Safety Belt”

NOC Module

- Typical infrastructure layout:



Upstream Connectivity and Peering

Transits

- Transit provider is another autonomous system which is used to provide the local network with access to other networks
  - Might be local or regional only
  - But more usually the whole Internet
- Transit providers need to be chosen wisely:
  - Only one
    - no redundancy
  - Too many
    - more difficult to load balance
    - no economy of scale (costs more per Mbps)
    - hard to provide service quality
- **Recommendation: at least two, no more than three**

Common Mistakes

- ISPs sign up with too many transit providers
  - Lots of small circuits (cost more per Mbps than larger ones)
  - Transit rates per Mbps reduce with increasing transit bandwidth purchased
  - Hard to implement reliable traffic engineering that doesn't need daily fine tuning depending on customer activities
- No diversity
  - Chosen transit providers all reached over same satellite or same submarine cable
  - Chosen transit providers have poor onward transit and peering

Peers

- A peer is another autonomous system with which the local network has agreed to exchange locally sourced routes and traffic
- Private peer
  - Private link between two providers for the purpose of interconnecting
- Public peer
  - Internet Exchange Point, where providers meet and freely decide who they will interconnect with
- **Recommendation: peer as much as possible!**

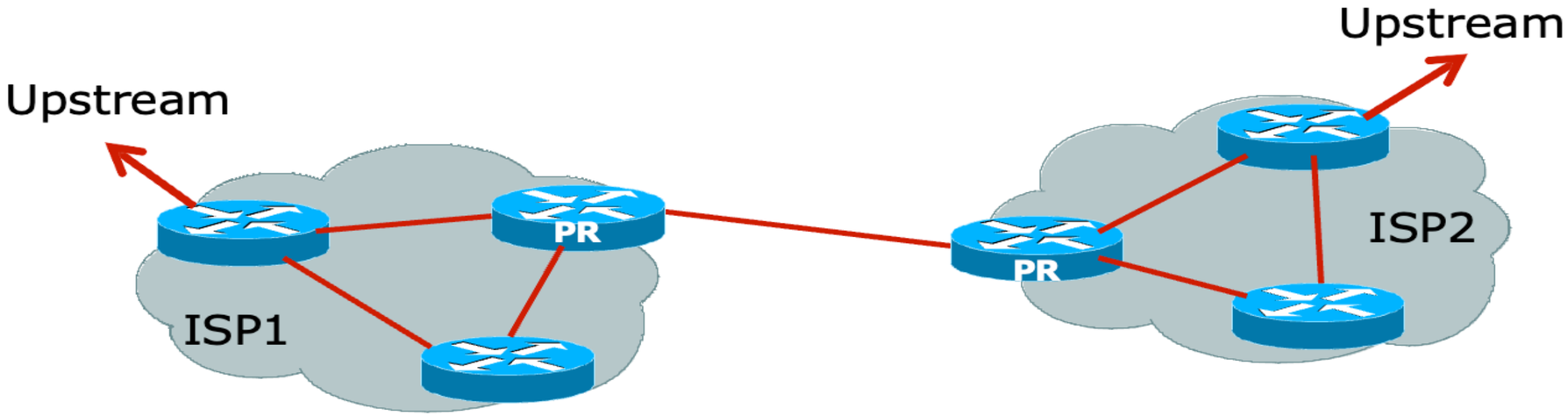
Common Mistakes

- business for a no-cost public peering point
- Mistaking a transit provider’ s “Exchange”
- Not working hard to get as much peering as possible
  - Physically near a peering point (IXP) but not present at it
  - (Transit sometimes is cheaper than peering!!)
- Ignoring/avoiding competitors because they are competition
  - Even though potentially valuable peering partner to give customers a better experience

Private Interconnection

- Two service providers agree to interconnect their networks
  - They exchange prefixes they originate into the routing system (usually their aggregated address blocks)
  - They share the cost of the infrastructure to interconnect
    - Typically each paying half the cost of the link (be it circuit, satellite, microwave, fibre,…)
    - Connected to their respective peering routers
  - Peering routers only carry domestic prefixes

Private Interconnection

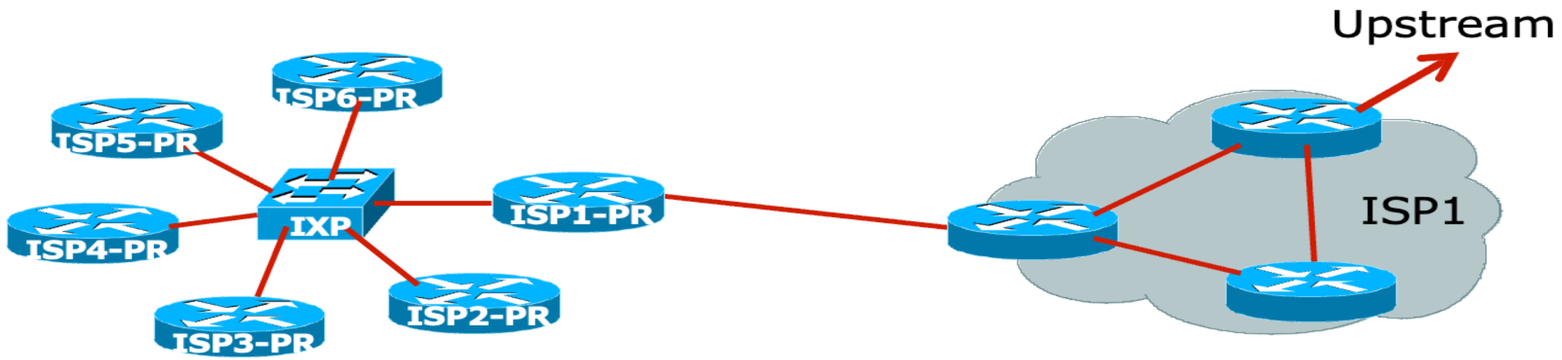


- PR = peering router
  - Runs iBGP (internal) and eBGP (with peer)
  - No default route
  - No "full BGP table"
  - Domestic prefixes only
- Peering router used for all private interconnects

Public Interconnection

- Service provider participates in an Internet Exchange Point
  - It exchanges prefixes it originates into the routing system with the participants of the IXP
  - It chooses who to peer with at the IXP
    - Bi-lateral peering (like private interconnect)
    - Multi-lateral peering (via IXP's route server)
  - It provides the router at the IXP and provides the connectivity from their PoP to the IXP
  - The IXP router carries only domestic prefixes

Public Interconnection



- ISP1-PR = peering router of our ISP
  - Runs iBGP (internal) and eBGP (with IXP peers)
  - No default route
  - No "full BGP table"
  - Domestic prefixes only
- Physically located at the IXP

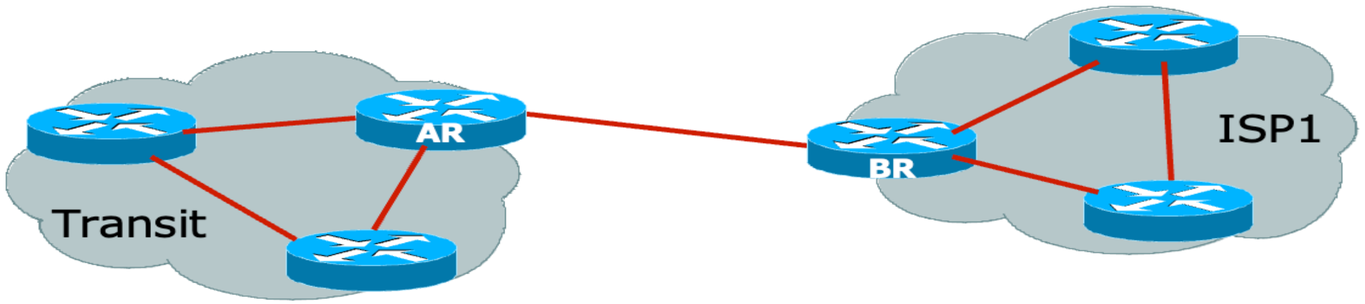
Public Interconnection

- The ISP's router IXP peering router needs careful configuration:
  - It is remote from the domestic backbone
  - Should not originate any domestic prefixes
  - (As well as no default route, no full BGP table)
  - Filtering of BGP announcements from IXP peers (in and out)
- Provision of a second link to the IXP:
  - (for redundancy or extra capacity)
  - Usually means installing a second router
    - Connected to a second switch (if the IXP has two more more switches)
    - Interconnected with the original router (and part of iBGP mesh)

Upstream/Transit Connection

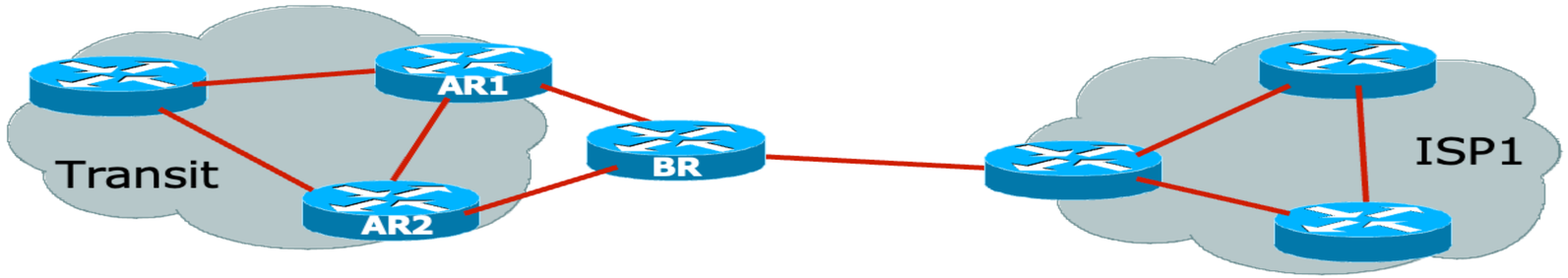
- Two scenarios:
  - Transit provider is in the locality
    - Which means bandwidth is cheap, plentiful, easy to provision, and easily upgraded
  - Transit provider is a long distance away
    - Over undersea cable, satellite, long-haul cross country fibre, etc
- Each scenario has different considerations which need to be accounted for

Local Transit Provider



- BR = ISP's Border Router
  - Runs iBGP (internal) and eBGP (with transit)
  - Either receives default route or the full BGP table from upstream
  - BGP policies are implemented here (depending on connectivity)
  - Packet filtering is implemented here (as required)

Distant Transit Provider



- BR = ISP's Border Router
  - Co-located in a co-lo centre (typical) or in the upstream provider's premises
  - Runs iBGP with rest of ISP1 backbone
  - Runs eBGP with transit provider router(s)
  - Implements BGP policies, packet filtering, etc
  - Does not originate any domestic prefixes

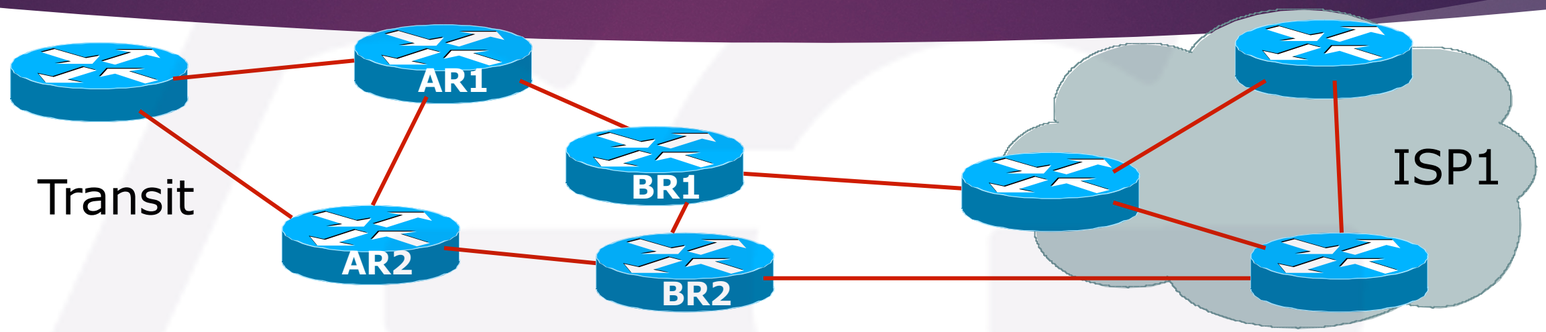
Distant Transit Provider

- Positioning a router close to the Transit Provider's infrastructure is strongly encouraged:
  - Long haul circuits are expensive, so the router allows the ISP to implement appropriate filtering first
  - Moves the buffering problem away from the Transit provider
  - Remote co-lo allows the ISP to choose another transit provider and migrate connections with minimum downtime

Distant Transit Provider

- Other points to consider:
  - Does require remote hands support
  - (Remote hands would plug or unplug cables, power cycle equipment, replace equipment, etc as instructed)
  - Appropriate support contract from equipment vendor(s)
  - Sensible to consider two routers and two long-haul links for redundancy

Distant Transit Provider



- Upgrade scenario:
  - Provision two routers
  - Two independent circuits
  - Consider second transit provider and/or turning up at an IXP

Summary

- Design considerations for:
  - Private interconnects
    - Simple private peering
  - Public interconnects
    - Router co-lo at an IXP
  - Local transit provider
    - Simple upstream interconnect
  - Long distance transit provider
    - Router remote co-lo at datacentre or Transit premises