ISP Network Design

ISP/IXP Workshops

ISP Network Design

- PoP Topologies and Design
- Backbone Design
- ISP Systems Design
- Addressing
- Routing Protocols
- Security
- Out of Band Management
- Operational Considerations

Point of Presence Topologies

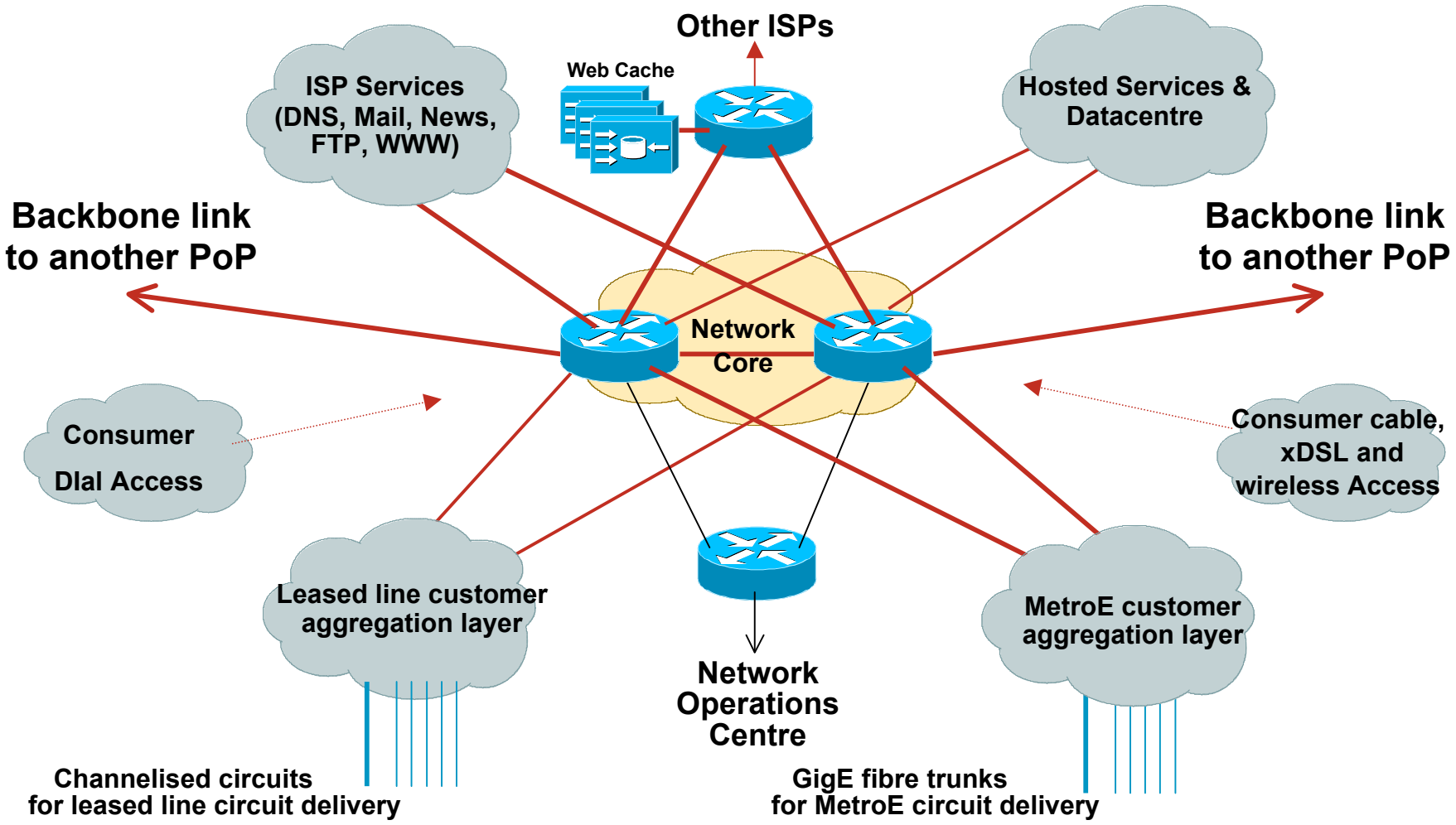
PoP Topologies

- 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 Smaller ISPs

- 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 per module
  - Core routers are the route-reflectors
  - Remaining routers are clients & peer with route-reflectors only

Modular Routing Protocol Design Larger ISPs

- Modular IGP implementation
  - IGP “area” per module (but avoid overloading core routers)
  - Core routers in backbone area (Area 0/L2)
  - Aggregation/summarisation where possible into the core
- Modular iBGP implementation
  - BGP route reflector cluster per module
  - Dedicated route-reflectors adjacent to core routers
  - Clients peer with route-reflectors only

Point of Presence Design

PoP Modules

- Low Speed customer connections
  - PSTN/ISDN dialup
  - Low bandwidth needs
  - Low revenue, large numbers
- Leased line customer connections
  - E1/T1 speed range
  - Delivery over channelised media
  - Medium bandwidth needs
  - Medium revenue, medium numbers

PoP Modules

- Broad Band customer connections
  - xDSL, Cable and Wireless
  - High bandwidth needs
  - Low revenue, large numbers
- MetroE & Highband customer connections
  - Trunk onto GigE or 10GigE of 10Mbps and higher
  - Channelised OC3/12 delivery of E3/T3 and higher
  - High bandwidth needs
  - High revenue, low numbers

PoP Modules

- PoP Core
  - Two dedicated routers
  - High Speed interconnect
  - Backbone Links **ONLY**
  - *Do not touch them!*
- Border Network
  - Dedicated border router to other ISPs
  - The ISP's "front" door
  - Transparent web caching?
  - **Two** in backbone is minimum guarantee for redundancy

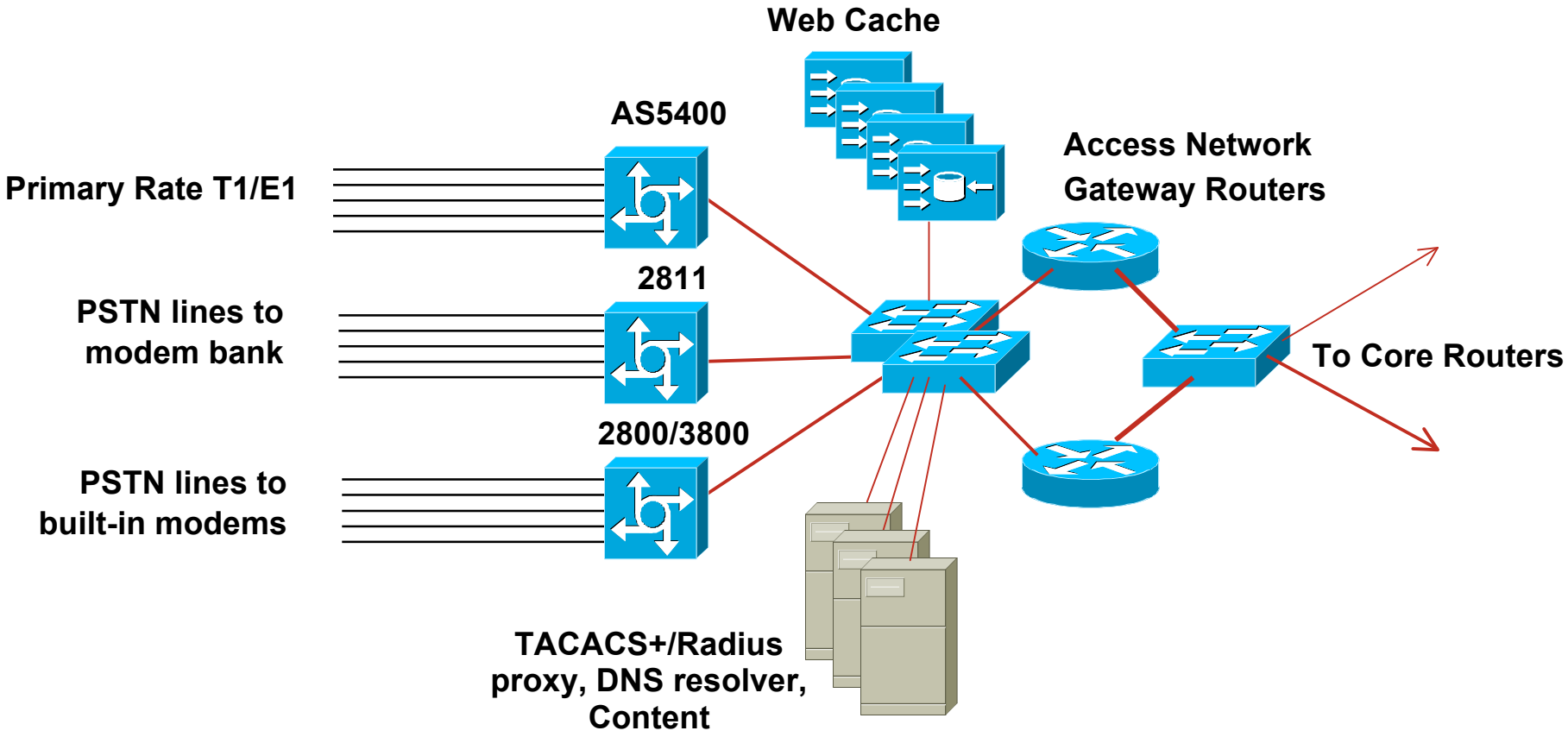
PoP Modules

- ISP Services
  - DNS (cache, secondary)
  - News (still relevant?)
  - Mail (POP3, Relay, Anti-virus/anti-spam)
  - WWW (server, proxy, cache)
- Hosted Services/DataCentres
  - Virtual Web, WWW (server, proxy, cache)
  - Information/Content Services
  - Electronic Commerce

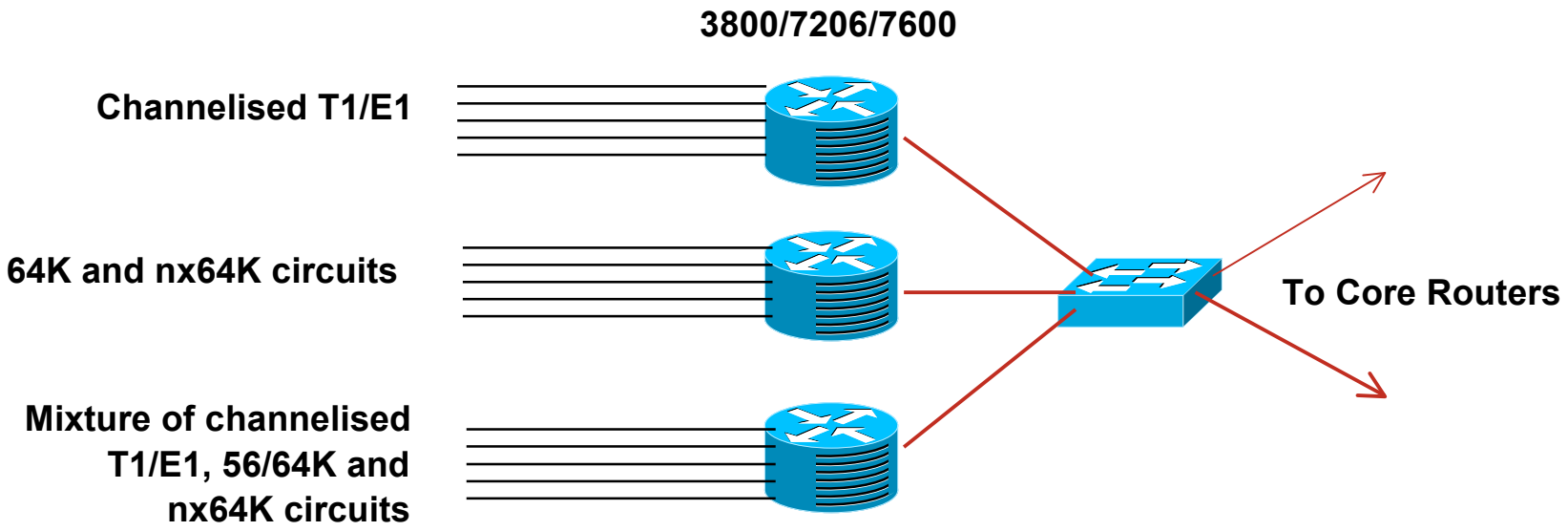
PoP Modules

- Network Operations Centre
  - Consider primary and backup locations
  - Network monitoring
  - Statistics and log gathering
  - Direct but secure access
- Out of Band Management Network
  - The ISP Network “Safety Belt”

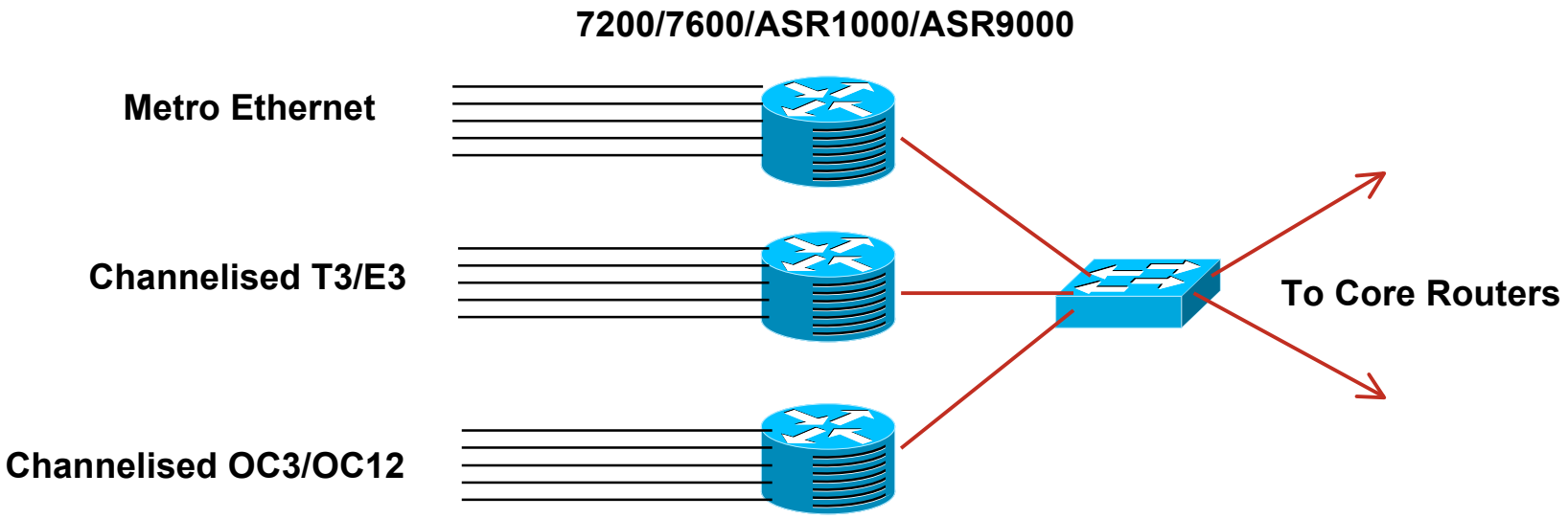
Low Speed Access Module



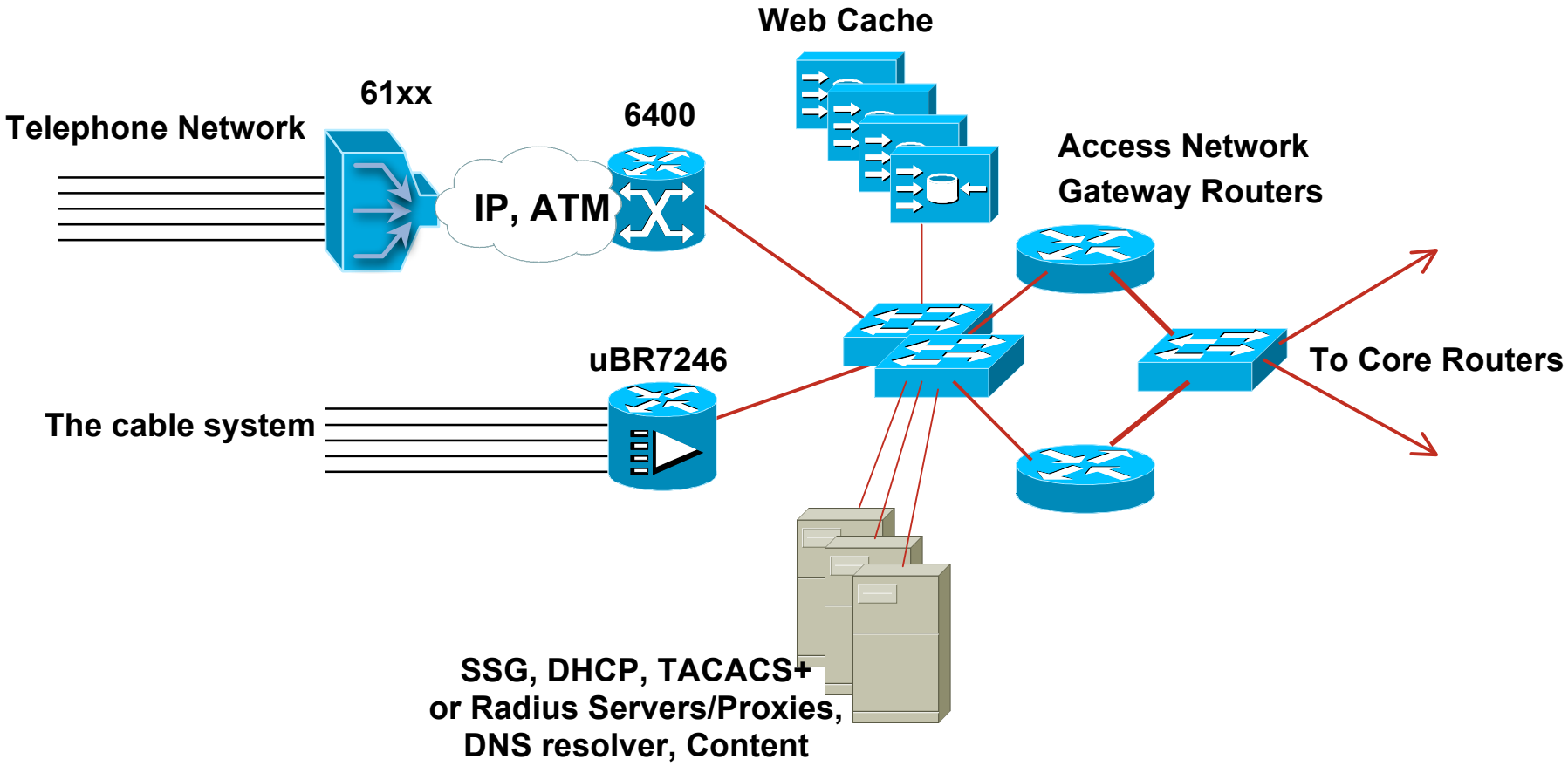
Medium Speed Access Module



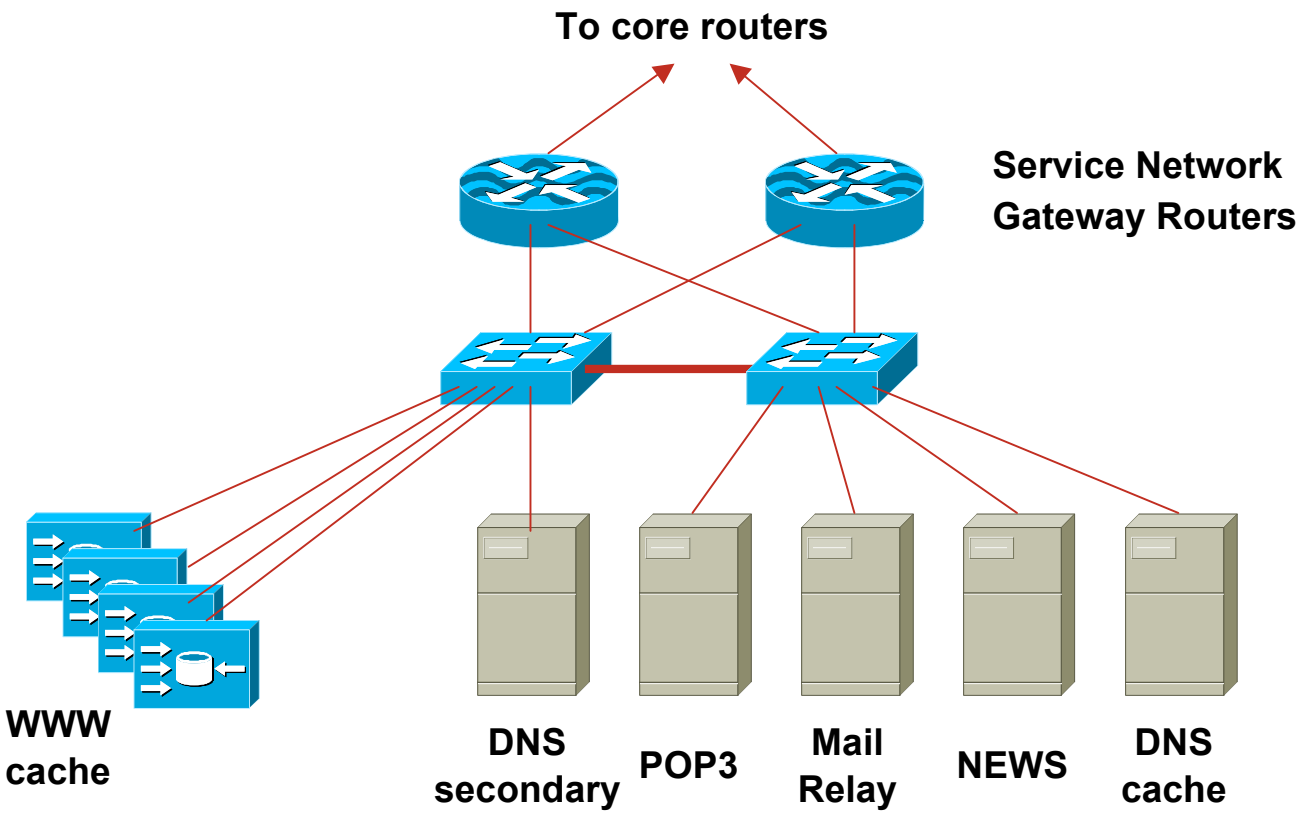
High Speed Access Module



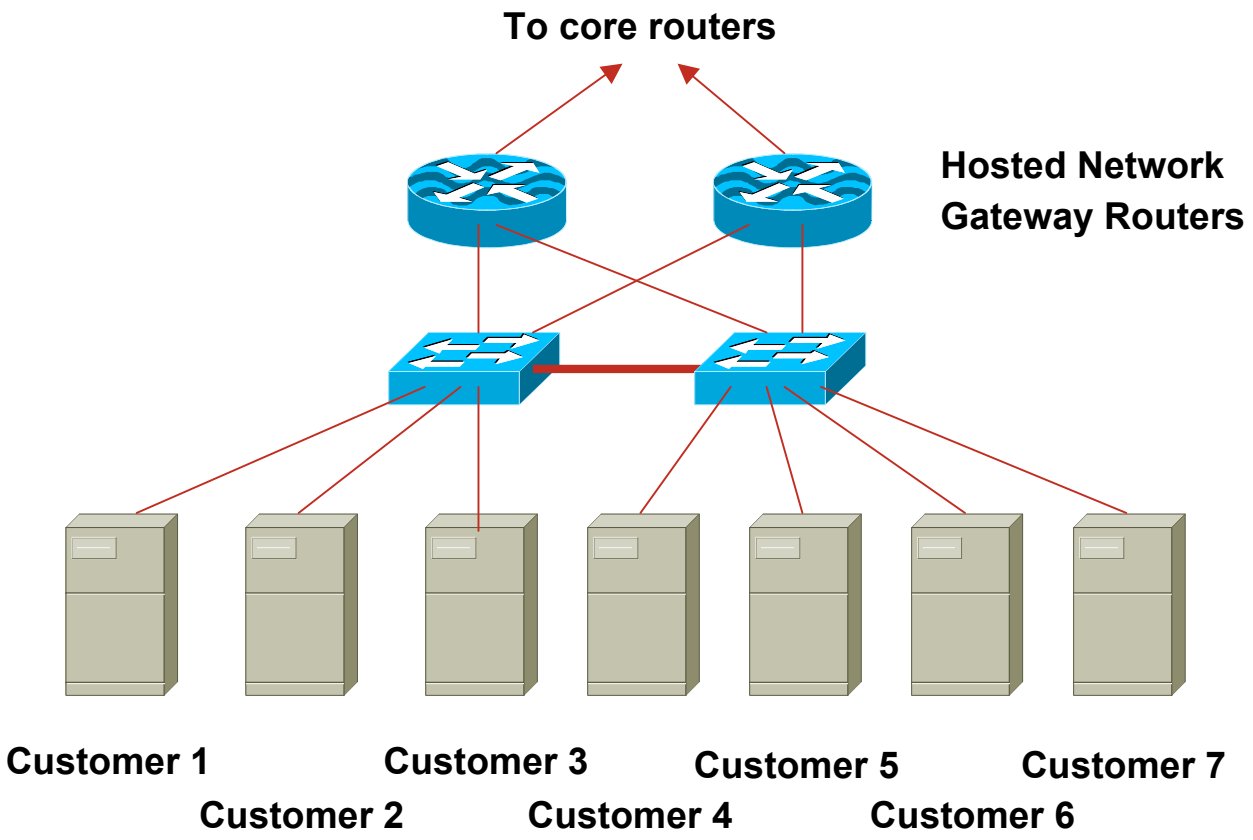
Broad Band Access Module



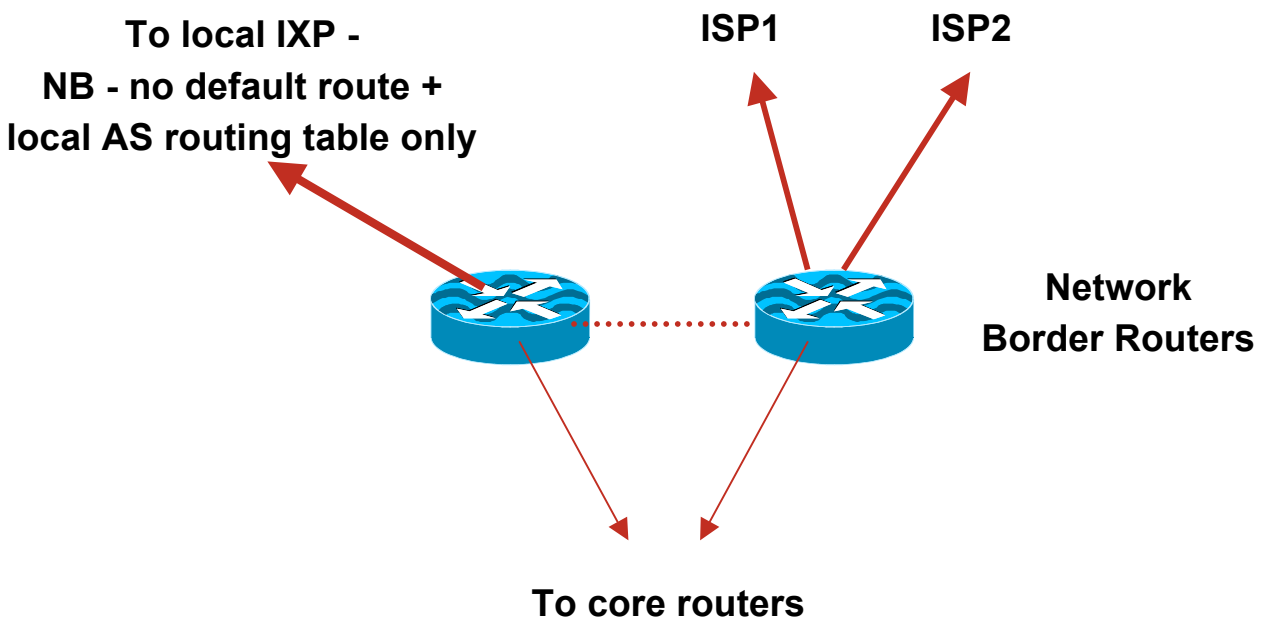
ISP Services Module



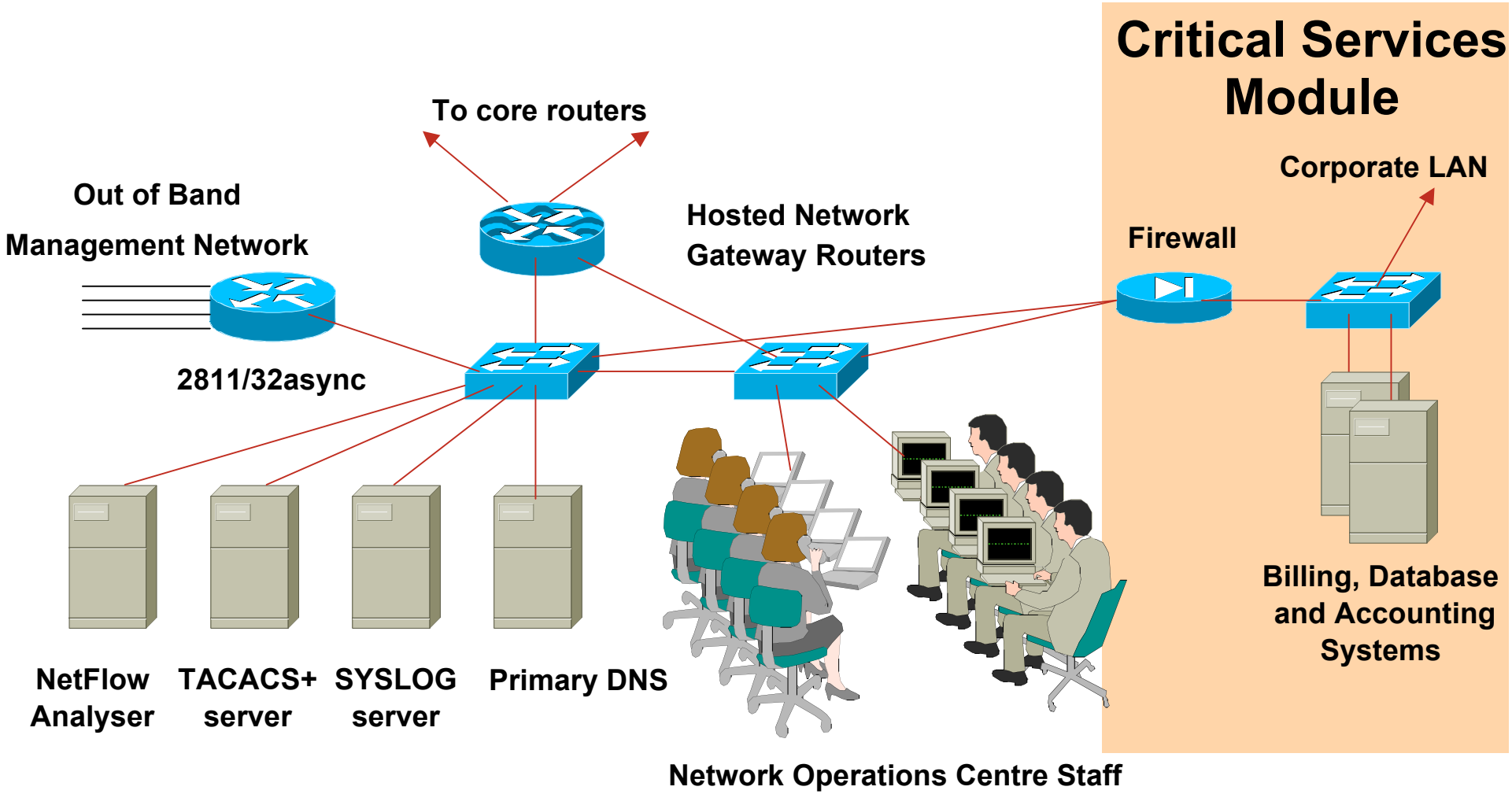
Hosted Services Module



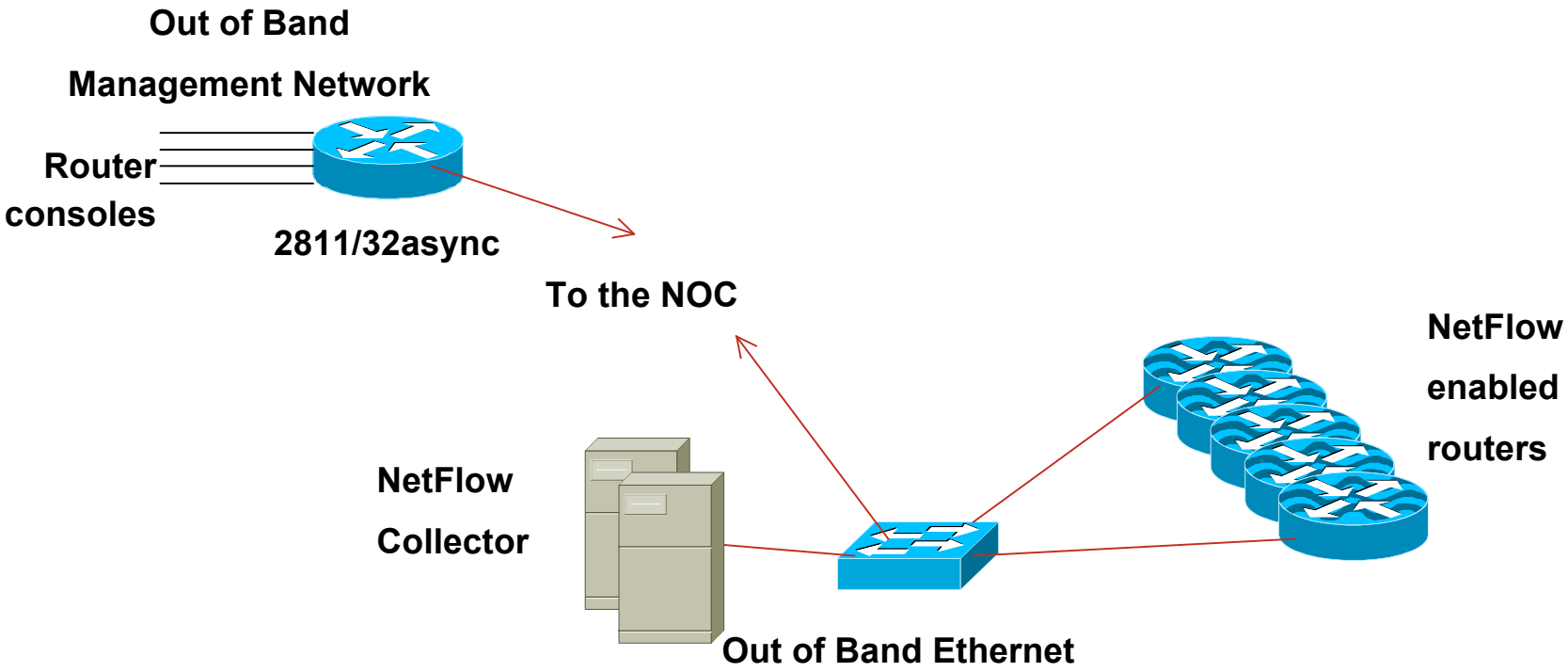
Border Module



NOC Module



Out of Band Network



Backbone Network Design

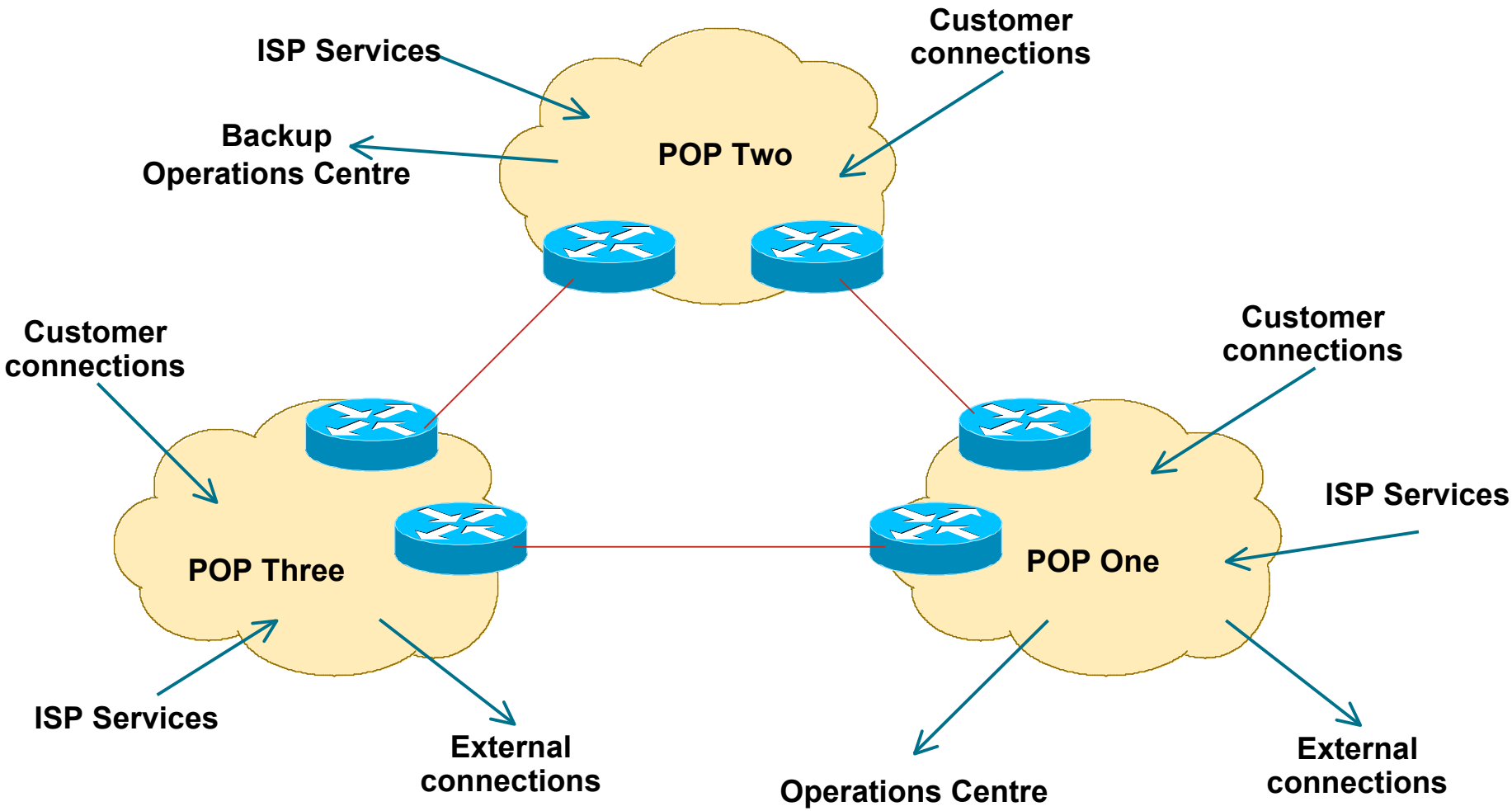
Backbone Design

- Routed Backbone
- Switched Backbone
  - Virtually obsolete
- Point-to-point circuits
  - nx64K, T1/E1, T3/E3, OC3, OC12, GigE, OC48, 10GigE, OC192
- ATM/Frame Relay service from telco
  - T3, OC3, OC12,… delivery
  - Easily upgradeable bandwidth (CIR)
  - Almost vanished in availability now

Distributed Network Design

- PoP design “standardised”
  - operational scalability and simplicity
- ISP essential services distributed around backbone
- NOC and “backup” NOC
- Redundant backbone links

Distributed Network Design



Backbone Links

- ATM/Frame Relay

  Virtually disappeared due to overhead, extra equipment, and shared with other customers of the telco

  MPLS has replaced ATM & FR as the telco favourite

- Leased Line/Circuit

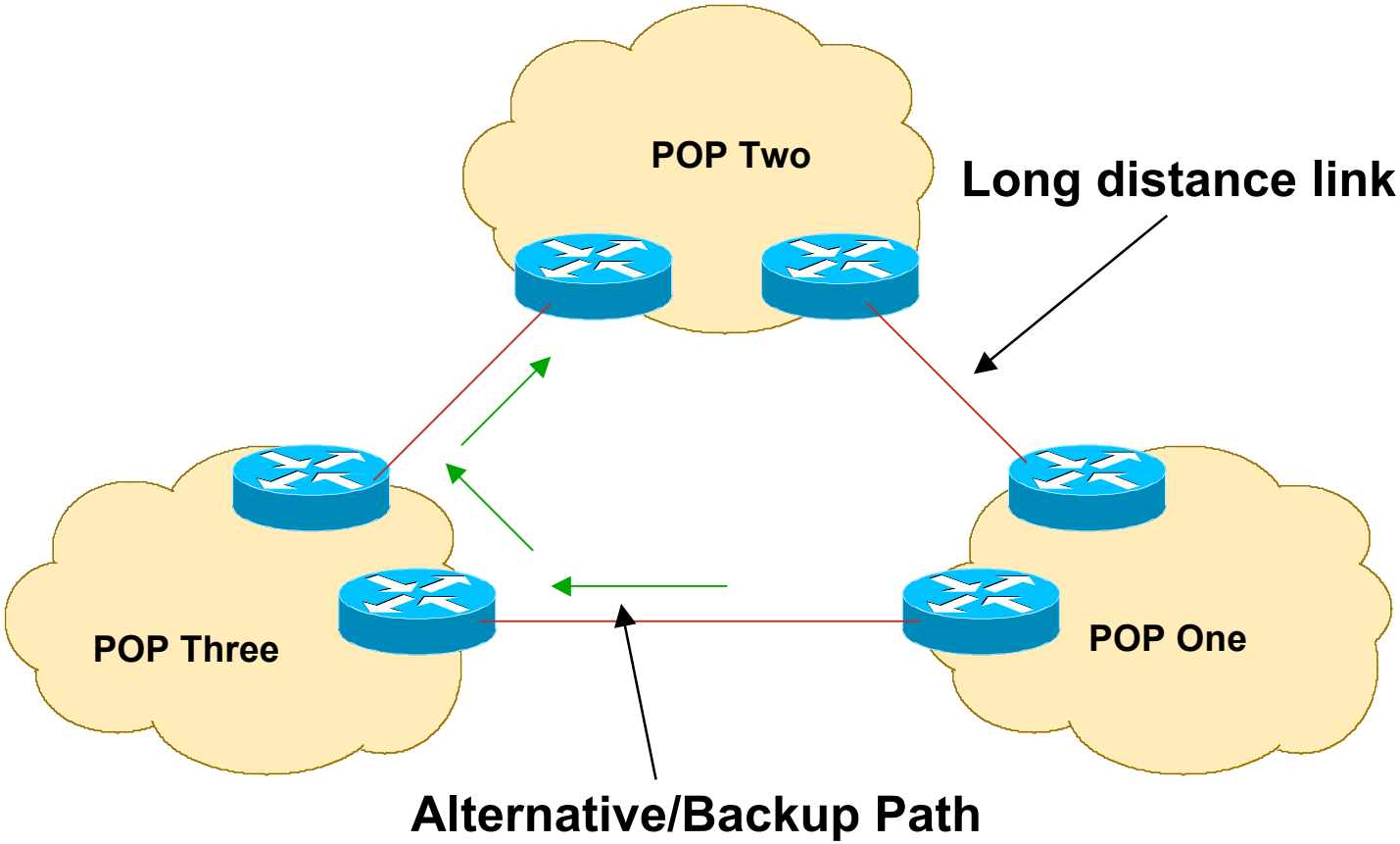
  Most popular with backbone providers

  IP over Optics and Metro Ethernet very common in many parts of the world

Long Distance Backbone Links

- Tend to cost more
- Plan for the future (at least two years ahead) but stay in budget
  - Unplanned “emergency” upgrades can be disruptive without redundancy
- Allow sufficient capacity on alternative paths for failure situations
  - Sufficient can be 20% to 50%
  - Some businesses choose 0% – meaning they have no spare capacity at all!!

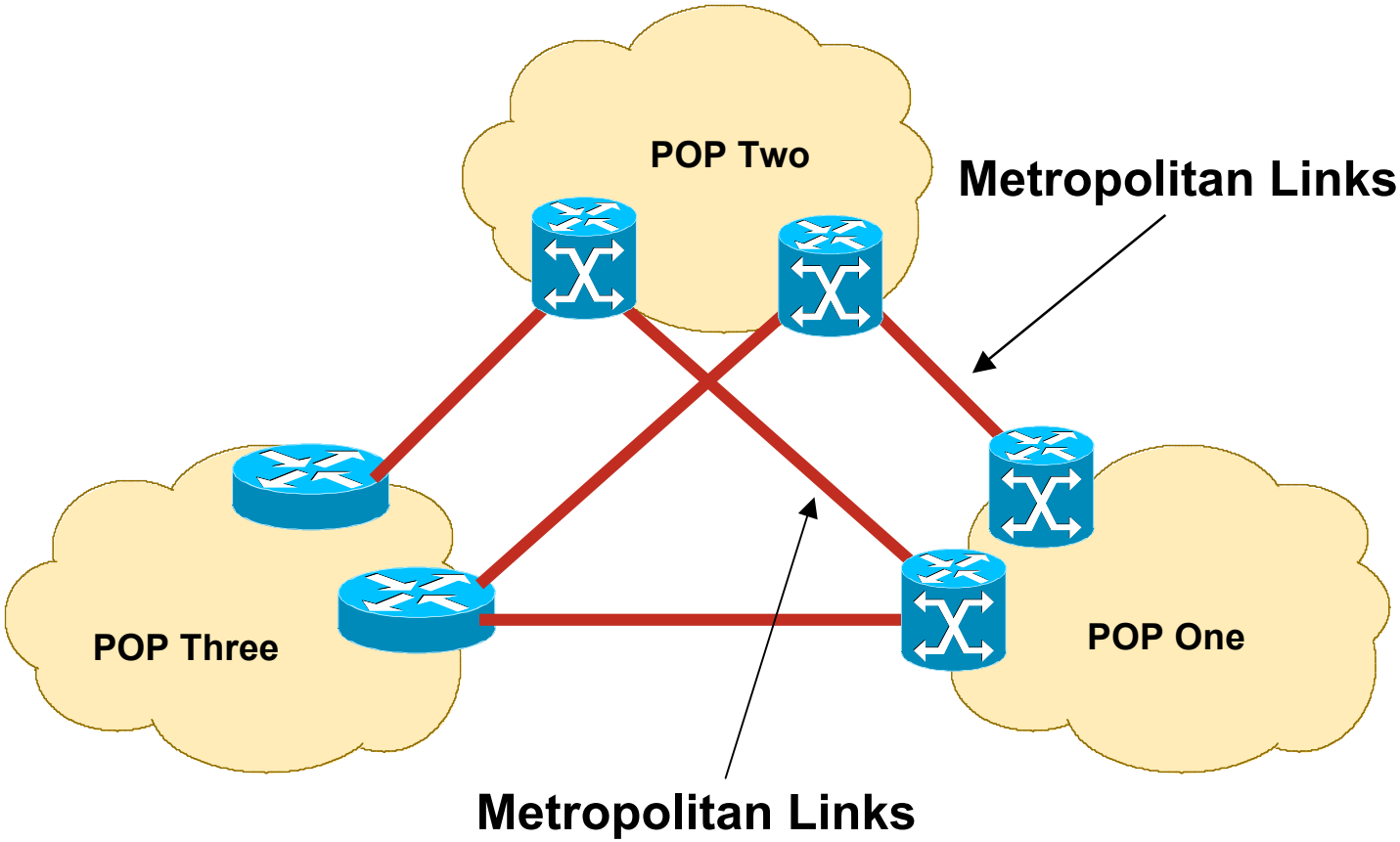
Long Distance Links



Metropolitan Area Backbone Links

- Tend to be cheaper
  - Circuit concentration
  - Choose from multiple suppliers
- Think big
  - More redundancy
  - Less impact of upgrades
  - Less impact of failures

Metropolitan Area Backbone Links



Traditional Point to Point Links

ISP Services

DNS, Mail, News

Design and Placement

ISP Services

- Most ISP services such as DNS, Mail, etc are easily deliverable on low budget hardware platforms
  - Single Rack Unit in height (1RU)
  - Dual processor is “default” now
  - RAM is very cheap (may as well use 2Gbytes or more)
  - Hard drives are very cheap (SCSI more reliable)
  - Unix like operating systems (FreeBSD, Debian, Ubuntu, CentOS) are very common
    - In addition to commercial operating systems such as Solaris, RedHat Enterprise Linux &c
  - Minimal overhead, minimal OS install, plus the service required

ISP Services: DNS

- Domain Name System
  - Provides name and address resolution
  - Servers need to be differentiated, properly located and specified
    - Primary nameserver
    - Secondary nameserver
    - Caching nameserver – resolver

ISP Services: DNS

- Primary nameserver
  - Holds ISP zone files
    - Forward zone (list of name to address mappings) for all ISP's and any customer zones
    - Reverse zone (list of address to name mappings) for all ISP's address space
  - Hardware & OS: easily satisfied by simple specification
  - Located in secure part of net, e.g. NOC LAN
  - Usually run as "hidden master" – secondary nameservers are the official listed nameservers

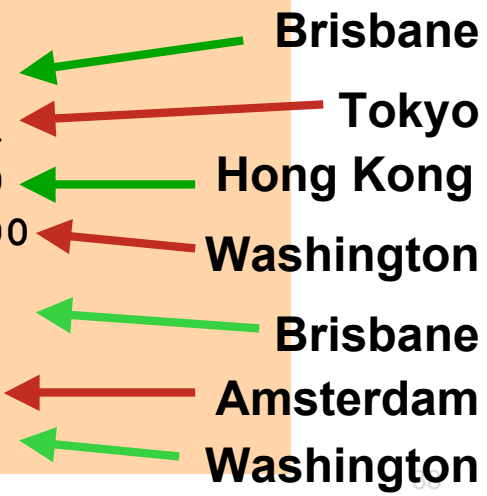
ISP Services: DNS

- Secondary nameserver
  - Holds copies of ISP zone files
  - At least two are required, more is better
  - Hardware & OS: easily satisfied by simple specification
  - Strongly recommended to be geographically separate from each other and the primary DNS
    - At different PoPs
    - On a different continent e.g. via services offered by ISC, PCH and others
    - At another ISP

ISP Services:
Secondary DNS Example

```
$ dig apnic.net ns
;; ANSWER SECTION:
apnic.net.              10800   NS      ns1.apnic.net.
apnic.net.              10800   NS      ns3.apnic.net.
apnic.net.              10800   NS      ns4.apnic.net.
apnic.net.              10800   NS      ns5.apnic.com.
apnic.net.              10800   NS      cumin.apnic.net.
apnic.net.              10800   NS      ns-sec.ripe.net.
apnic.net.              10800   NS      tinnie.arin.net.
apnic.net.              10800   NS      tinnie.apnic.net.

;; ADDITIONAL SECTION:
ns1.apnic.net.          3600    A       202.12.29.25
ns3.apnic.net.          3600    A       202.12.28.131
ns4.apnic.net.          3600    A       202.12.31.140
ns5.apnic.com.          10800   A       203.119.43.200
cumin.apnic.net.        3600    A       202.12.29.59
tinnie.apnic.net.       3600    A       202.12.29.60
ns-sec.ripe.net.        113685  A       193.0.0.196
tinnie.arin.net.        10800   A       199.212.0.53
```



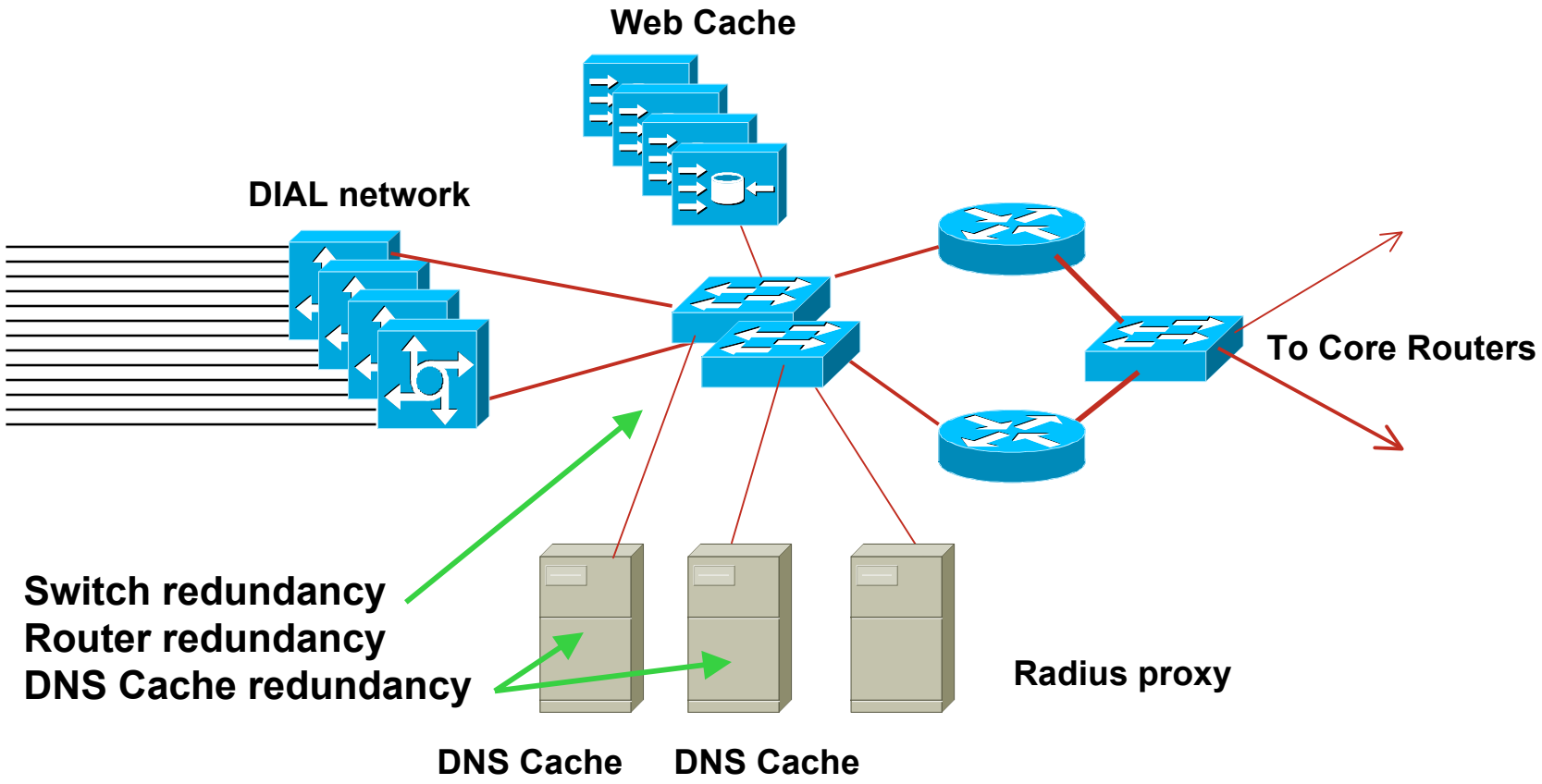
ISP Services: Secondary DNS Example

- apnic.net zone
  - Primary DNS in Brisbane (ns1.apnic.net)
  - Secondary DNS run all over the world by APNIC:
    - Brisbane
    - Hong Kong
    - Tokyo
    - Washington
  - Zone secondaried by
    - RIPE NCC in Amsterdam
    - ARIN in Washington
  - Geographical and service provider redundancy – this is the perfect example!

ISP Services: DNS

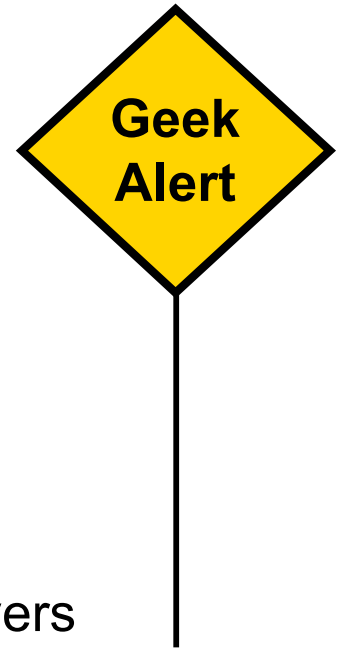
- Caching nameserver
  - This is the resolver – it is the DNS cache
  - Your customers use this as resolver, NEVER your primary or secondary DNS
  - Provides very fast lookups
  - Does NOT secondary any zones
  - One, or preferably two per PoP (redundancy)
  - Hardware & OS: easily satisfied by simple specification

ISP Services:
Caching Nameserver



- DIAL users automatically given the IP addresses of DNS caches when they dial in

ISP Services:
Anycasting the Caching Nameserver



- One trick of the trade
  - assign two unique IP addresses to be used for the two DNS resolver systems
  - use these two IP addresses in every PoP
  - route the two /32s across your backbone
  - even if the two resolver systems in the local PoP are down, the IGP will ensure that the next nearest resolvers will be reachable
  - Known as IP Anycast

ISP Services: DNS

- Efficient and resilient design
  - Primary DNS – keep it secure
  - Secondary DNS – geographical and provider redundancy
    - Don’t ever put them on the same LAN, switched or otherwise
    - Don’t put them in the same PoP
  - Caching DNS – one or two per PoP
    - Reduces DNS traffic across backbone
    - More efficient, spreads the load

ISP Services: DNS

- Software
  - Make sure that the BIND distribution on the Unix system is up to date
    - **The vendor's distribution is rarely current**
  - Pay attention to bug reports, security issues
  - Reboot the DNS cache on a regular (e.g. monthly) basis
    - Clears out the cache
    - Releases any lost RAM
    - Accepted good practice by system administrators

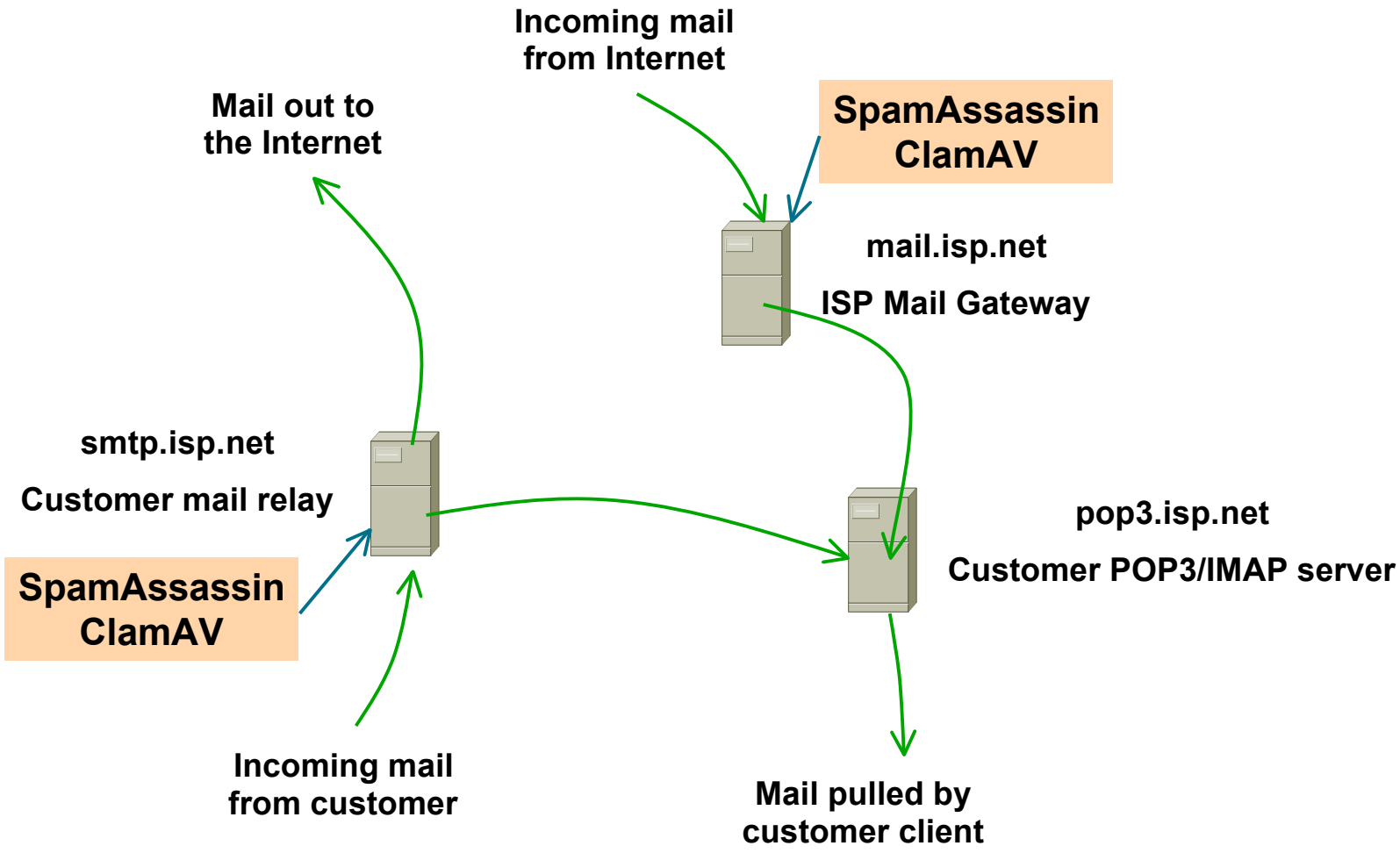
ISP Services: DNS

- Implementation
  - Put all your hosts, point-to-point links and loopbacks into the DNS
    - Under your ISP's domain name
    - Use sensible/meaningful names
  - Put all your hosts, point-to-point links and loopbacks into the REVERSE DNS also
    - Don't forget about in-addr.arpa and ip6.arpa – many ISPs do
    - Some systems demand forward/reverse DNS mapping before allowing access

ISP Services: Mail

- Must have at least two mail hosts (MX records) for all supported domains
  - Geographical separation helps
- Dedicated POP3 server
  - Consumers/mobile users get mail from here
- SMTP gateway dedicated to that function
  - Consumers/mobile users send mail via here
- Mail relay open to CUSTOMERS only!
  - Don’t let outside world use your mail relay
- Block port 25 outbound for all customers
  - Insist that outbound e-mail goes through SMTP relay
  - SMTP relay does virus (ClamAV) and spam (Spamassassin) filtering

ISP Services:
Mail Configuration



ISP Services: Mail Example

- cisco.com mail (MX records)
  - primary MX are 6 systems in San Jose
  - Three backup MXes in RTP, Amsterdam and Sydney
  - backup MX only used if primary unavailable

```
$ dig cisco.com mx

;; ANSWER SECTION:
cisco.com.          86400   MX      10 sj-inbound-a.cisco.com.
cisco.com.          86400   MX      10 sj-inbound-b.cisco.com.
cisco.com.          86400   MX      10 sj-inbound-c.cisco.com.
cisco.com.          86400   MX      10 sj-inbound-d.cisco.com.
cisco.com.          86400   MX      10 sj-inbound-e.cisco.com.
cisco.com.          86400   MX      10 sj-inbound-f.cisco.com.
cisco.com.          86400   MX      15 rtp-mx-01.cisco.com.
cisco.com.          86400   MX      20 ams-inbound-a.cisco.com.
cisco.com.          86400   MX      25 syd-inbound-a.cisco.com.
```

ISP Services: Mail

- Software

  Make sure that the MAIL and POP3 distributions on the Unix system are up to date

  The vendor distributions are rarely current

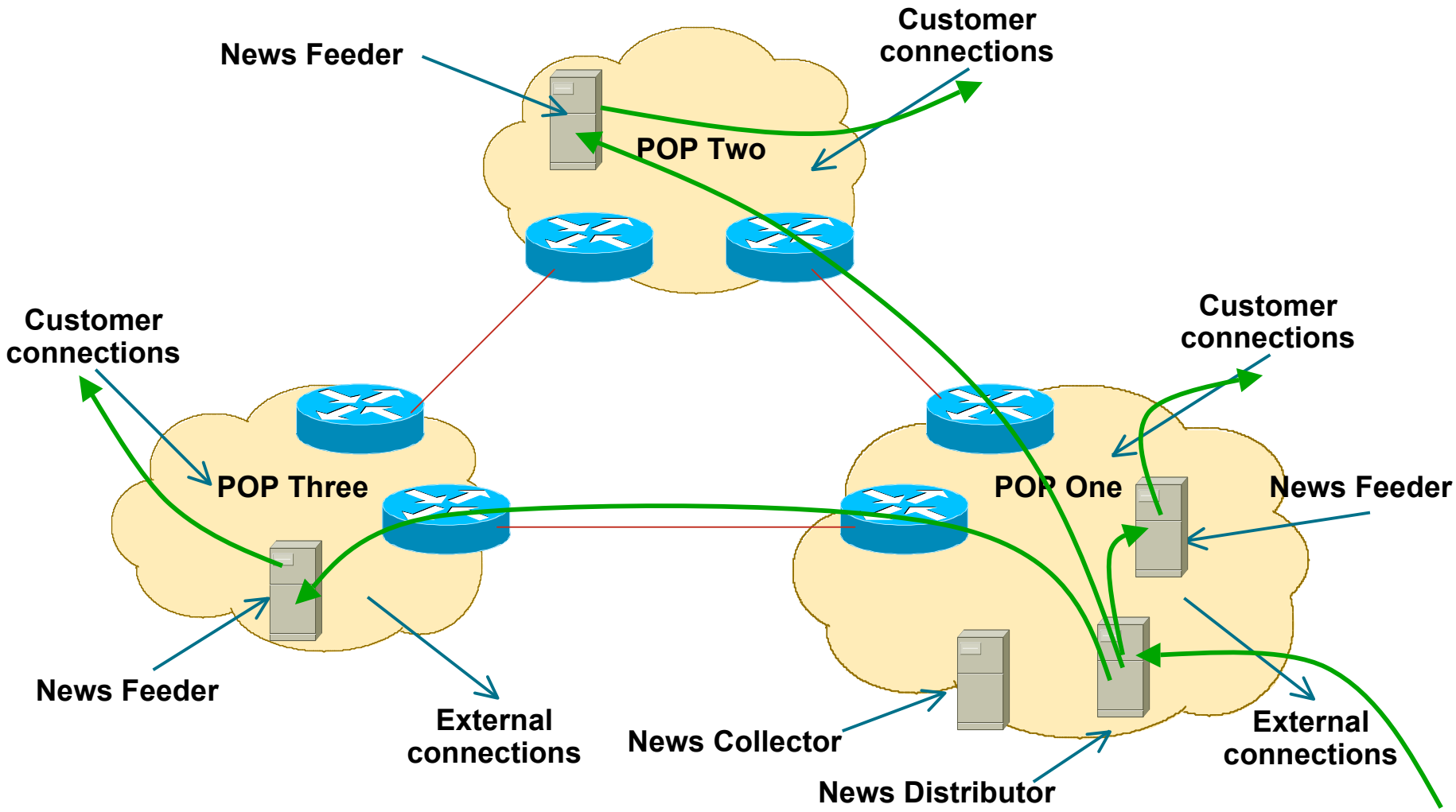
  Pay attention to bug reports, security issues, unsolicited junk mail complaints

IMPORTANT: Do NOT allow non-customers to use your mail system as a relay

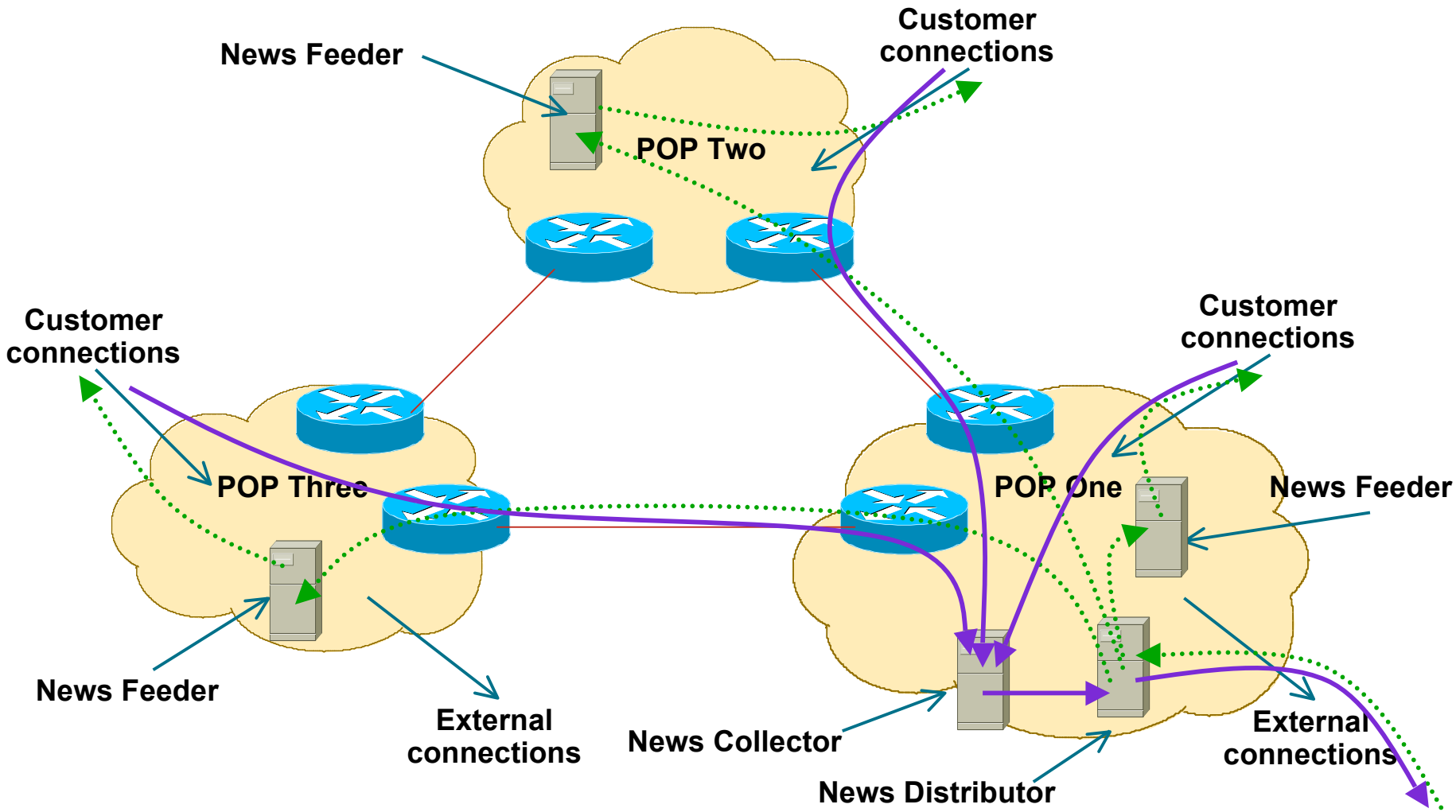
ISP Services: News

- News servers provide a Usenet news feed to customers
- Distributed design required
  - Incoming newsfeed to one large server
  - Distributed to feed servers in each PoP
  - Feed servers provide news feed to customers
  - Outgoing news goes to another server
  - Separate reading news system
  - Separate posting news system

ISP Services:
News System Placement



ISP Services:
News System Placement



ISP Services: News

- Software
  - Make sure that the Internet News distribution on the Unix system is up to date
    - **The vendor distributions are rarely current**
  - Pay attention to bug reports, security issues, unsolicited junk posting complaints

IMPORTANT: Do NOT allow non-customers to use your news system for posting messages

Addressing

Where to get IP addresses and AS numbers

- Your upstream ISP
- Africa
  - AfriNIC – http://www.afrinic.net
- Asia and the Pacific
  - APNIC – http://www.apnic.net
- North America
  - ARIN – http://www.arin.net
- Latin America and the Caribbean
  - LACNIC – http://www.lacnic.net
- Europe and Middle East
  - RIPE NCC – http://www.ripe.net/info/ncc

Internet Registry Regions

Getting IP address space

- Take part of upstream ISP's PA space

  or

- Become a member of your Regional Internet Registry and get your own allocation

  Require a plan for a year ahead

  General policies are outlined in RFC2050, more specific details are on the individual RIR website

- There is still plenty of IPv4 address space

  Registries require high quality documentation

  When applying for IPv4 addresses, get an IPv6 allocation too!

What about RFC1918 addressing?

- RFC1918 defines IP addresses reserved for private Internets
  - Not to be used on Internet backbones
  - **http://www.ietf.org/rfc/rfc1918.txt**
- Commonly used within end-user networks
  - NAT used to translate from private internal to public external addressing
  - Allows the end-user network to migrate ISPs without a major internal renumbering exercise
- Most ISPs filter RFC1918 addressing at their network edge
  - **http://www.cymru.com/Documents/bogon-list.html**

What about RFC1918 addressing?

- List of well known problems with this approach for an SP backbone:
  - Breaks Path MTU Discovery
  - Potential conflicts with usage of private addressing inside customer networks
  - Security through obscurity does not provide security
  - Troubleshooting outside the local network becomes very hard
    - Router interface addresses are only locally visible
    - Internet becomes invisible from the router
  - Troubleshooting of connectivity issues on an Internet scale becomes impossible
    - Traceroutes and pings provide no information
    - No distinction between “network invisible” and “network broken”
  - Increases operational complexity of the network infrastructure and routing configuration

Private versus Globally Routable IP Addressing

- Infrastructure Security: not improved by using private addressing

  Still can be attacked from inside, or from customers, or by reflection techniques from the outside

- Troubleshooting: made an order of magnitude harder

  No Internet view from routers

  Other ISPs cannot distinguish between **down** and **broken**

- Performance: PMTUD breakage

- Summary:

  ALWAYS use globally routable IP addressing for ISP Infrastructure

Addressing Plans – ISP Infrastructure

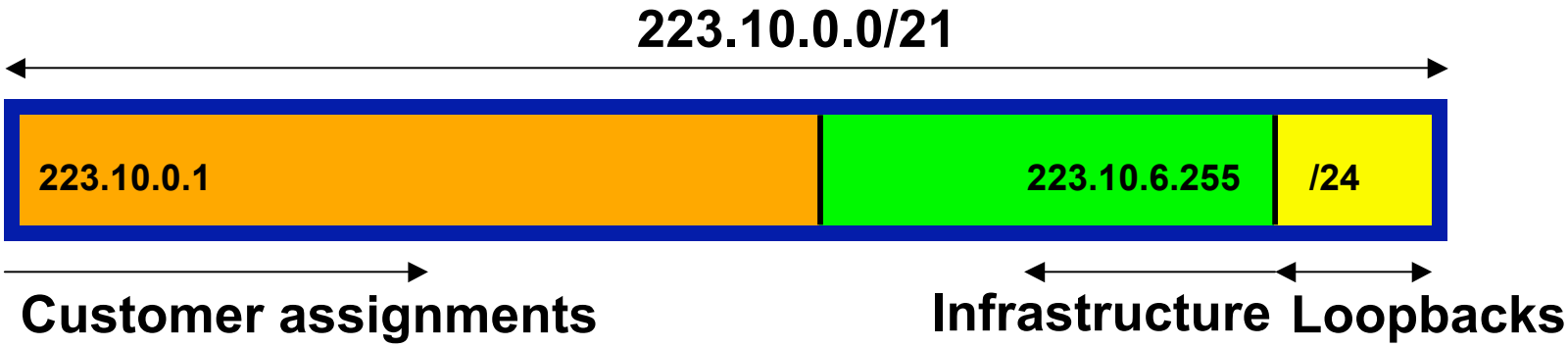
- Address block for router loop-back interfaces
- Address block for infrastructure
  - Per PoP or whole backbone
  - Summarise between sites if it makes sense
  - Allocate according to genuine requirements, not historic classful boundaries
- Similar allocation policies should be used for IPv6 as well
  - ISPs just get a substantially larger block (relatively) so assignments within the backbone are easier to make

Addressing Plans – Customer

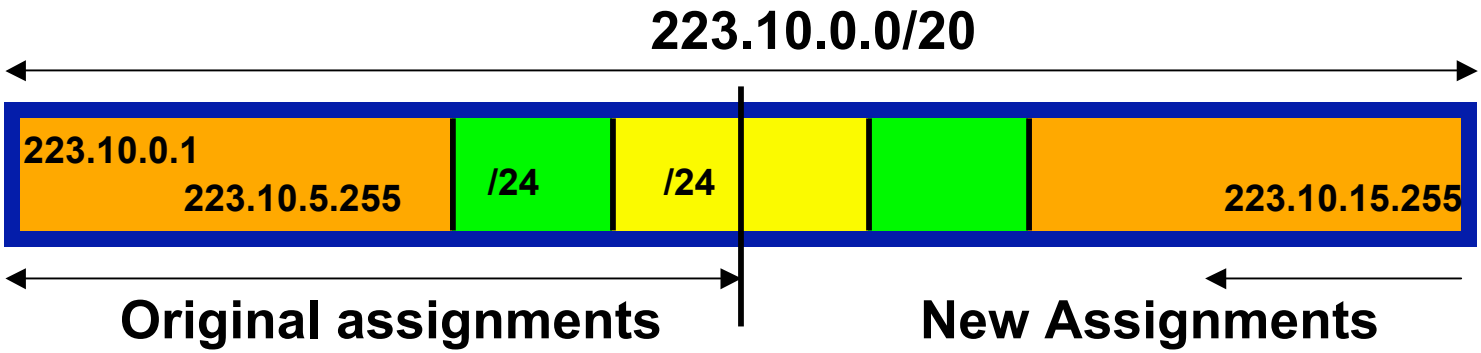
- Customers are assigned address space according to need
- Should not be reserved or assigned on a per PoP basis
  - ISP iBGP carries customer nets
  - Aggregation not required and usually not desirable

Addressing Plans – ISP Infrastructure

- Phase One



- Phase Two



Addressing Plans Planning

- Registries will usually allocate the next block to be contiguous with the first allocation
  - Minimum allocation could be /21
  - Very likely that subsequent allocation will make this up to a /20
  - So plan accordingly

Addressing Plans (contd)

- Document infrastructure allocation
  - Eases operation, debugging and management
- Document customer allocation
  - Contained in iBGP
  - Eases operation, debugging and management
  - Submit network object to RIR Database

Routing Protocols

Routing Protocols

- IGP – Interior Gateway Protocol
  - carries infrastructure addresses, point-to-point links
  - examples are OSPF, ISIS, EIGRP...
- EGP – Exterior Gateway Protocol
  - carries customer prefixes and Internet routes
  - current EGP is BGP version 4
- No connection between IGP and EGP

Why Do We Need an IGP?

- ISP backbone scaling
  - Hierarchy
  - Modular infrastructure construction
  - Limiting scope of failure
  - Healing of infrastructure faults using dynamic routing with fast convergence

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

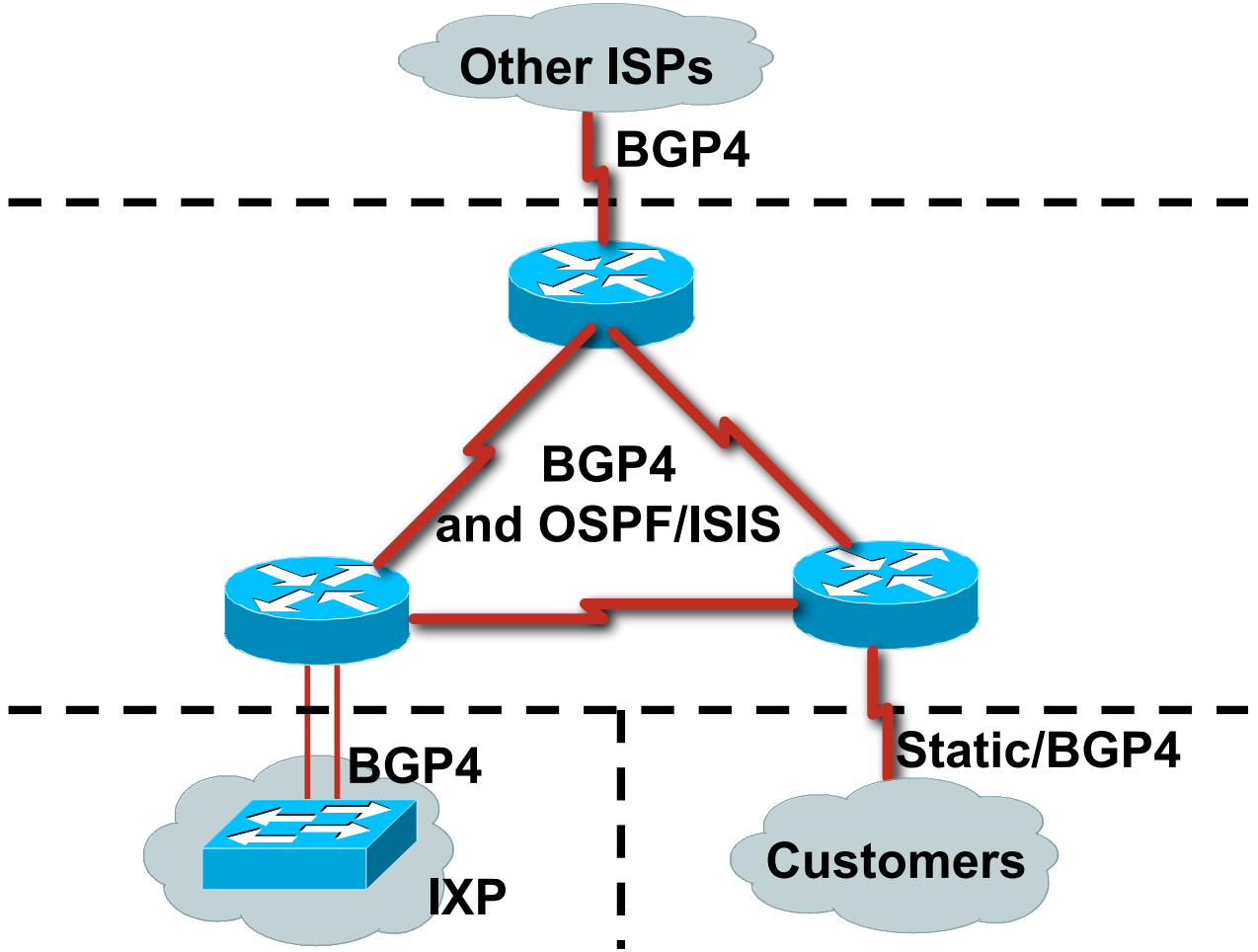
Interior versus Exterior Routing Protocols

- Interior
  - Automatic neighbour discovery
  - Generally trust your IGP routers
  - Prefixes 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



Routing Protocols: Choosing an IGP

- Review the “Introduction to Link State Protocols” presentation
  - i.e. – OSPF and ISIS have very similar properties
- ISP usually chooses between OSPF and ISIS
  - Choose which is appropriate for your operators’ experience
  - In IOS, both OSPF and ISIS have sufficient “nerd knobs” to tweak the IGP’s behaviour
  - OSPF runs on IP
  - ISIS runs on infrastructure, alongside IP

Routing Protocols: IGP Recommendations

- Keep the IGP routing table as small as possible

  If you can count the routers and the point to point links in the backbone, that total is the number of IGP entries you should see

- IGP details:

  Should only have router loopbacks, backbone WAN point-to-point link addresses, and network addresses of any LANs having an IGP running on them

  Strongly recommended to use inter-router authentication

  Use inter-area summarisation if possible

Routing Protocols: More IGP recommendations

- To fine tune IGP table size more, consider:
  - Using “ip unnumbered” on customer point-to-point links – saves carrying that /30 in IGP
    - (If customer point-to-point /30 is required for monitoring purposes, then put this in iBGP)
  - Use contiguous addresses for backbone WAN links in each area – then summarise into backbone area
  - Don’t summarise router loopback addresses – as iBGP needs those (for next-hop)
  - Use iBGP for carrying anything which does not contribute to the IGP Routing process

Routing Protocols: iBGP Recommendations

- iBGP should carry everything which doesn't contribute to the IGP routing process
  - Internet routing table
  - Customer assigned addresses
  - Customer point-to-point links
  - Dial network pools, passive LANs, etc

Routing Protocols: More iBGP Recommendations

- Scalable iBGP features:
  - Use neighbour authentication
  - Use peer-groups to speed update process and for configuration efficiency
  - Use communities for ease of filtering
  - Use route-reflector hierarchy
    - Route reflector pair per PoP (overlaid clusters)

Security

Security

- ISP Infrastructure security
- ISP Network security
- Security is **not optional**!
- ISPs need to:
  - Protect themselves
  - Help protect their customers from the Internet
  - Protect the Internet from their customers
- The following slides are general recommendations
  - Do more research on security before deploying any network

ISP Infrastructure Security

- Router security
  - Usernames, passwords, vty filters, TACACS+
  - Disable telnet on vtys, only use SSH
  - vty filters should only allow NOC access, no external access
  - See IOS Essentials for the recommended practices for ISPs

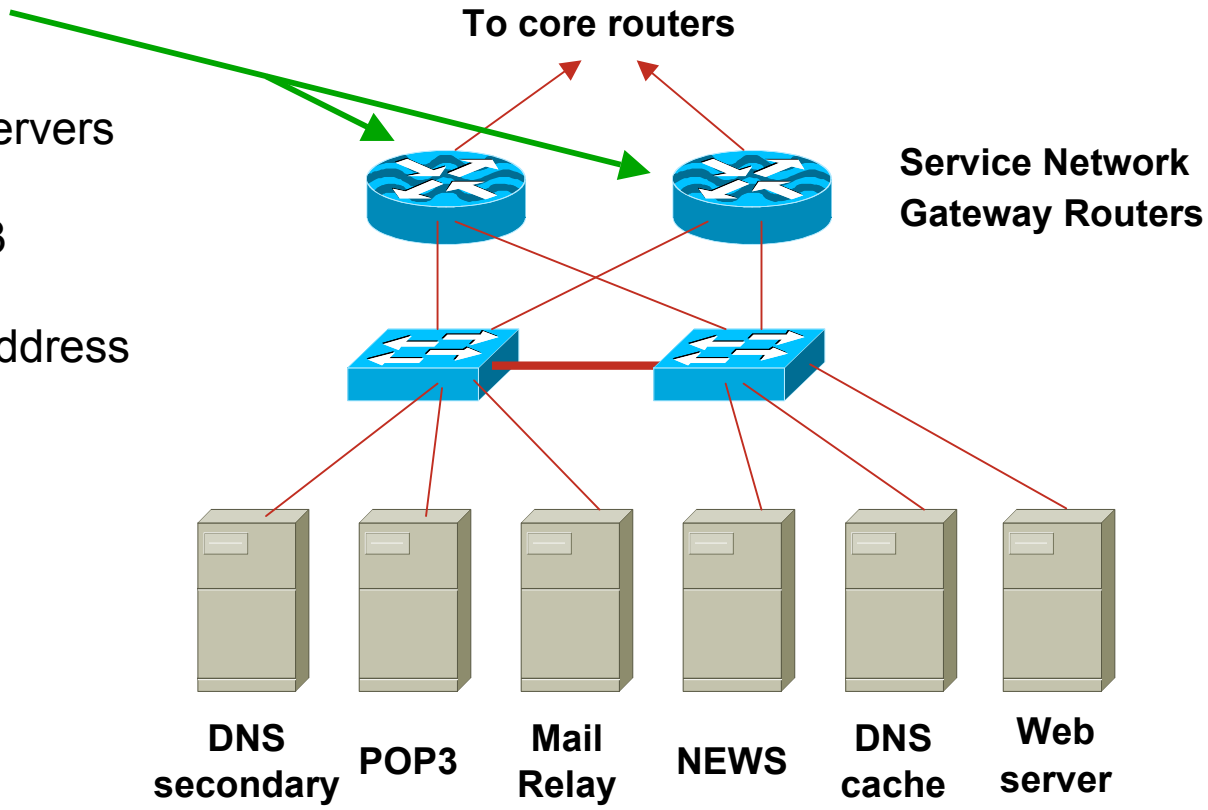
ISP Infrastructure Security

- ISP server security
  - Usernames, passwords, TCP wrappers, IPTABLES
  - Protect **all** servers using routers with strong filters applied
- Hosted services security
  - Protect network from hosted servers using routers with strong filters
  - Protect hosted servers from Internet using routers with strong filters

ISP Infrastructure Security
ISP Server Protection

Access-list examples:

Allow tcp/established to all servers
ICMP
DNS 2ary: udp/53 and tcp/53
POP3: tcp/110
Mail Relay: tcp/25 and ISP address range only
News: tcp/119 and ISP address range only
DNS Cache: udp/53
Web server: tcp/80



Other necessary filters:

All servers: SSH (tcp/22) from NOC LAN only

ISP Infrastructure Security
Hosted Server Protection

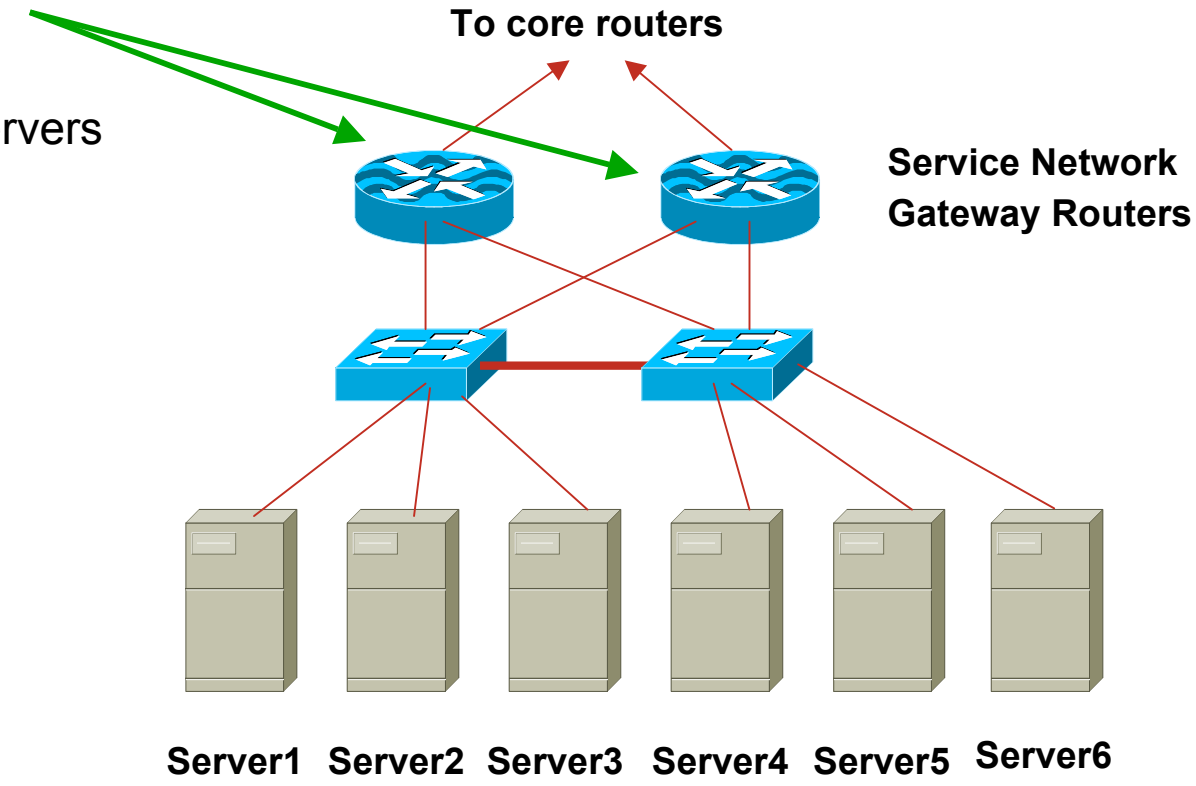
Access-list examples:

Inbound

Allow tcp/established to all servers
ICMP
Web server: tcp/80
SSH for customer access
Any other ports for services
sold to customers

Outbound

ICMP
Allow DNS udp/53 and
tcp/53
Block all access to ISP
address range



ISP Infrastructure Security

- Premises security
  - Locks – electronic/card key preferred
  - Secure access – 24x7 security arrangements
  - Environment control – good aircon
- Staff responsibility
  - Password policy, strangers, temp staff
  - Employee exit procedures
- RFC2196
  - (Site Security Handbook)
- RFC3871
  - (Operational Security Requirements for Large ISP IP Network Infrastructure)

ISP Network Security

- Denial of Service Attacks
  - eg: “smurfing”
  - see http://www.denialinfo.com
- Effective filtering
  - Network borders – see Cisco ISP Essentials
  - Customer connections – unicast RPF on **ALL** of them
  - Network operation centre
  - ISP corporate network – behind firewall

ISP Network Security
Secure external access

- How to provide staff access from outside
  - Set up ssh gateway (Unix system with ssh daemon and nothing else configured)
  - Provide ssh client on all staff laptops
  - ssh available on Unix and Windows
  - ssh is Secure Shell – encrypted link
- How not to provide access from outside
  - telnet, rsh, rlogin – these are all insecure
  - Open host – insecure, can be compromised

Ingress & Egress Route Filtering

Your customers should not be sending *any* IP packets out to the Internet with a source address other then the address you have allocated to them!

Out of Band Management

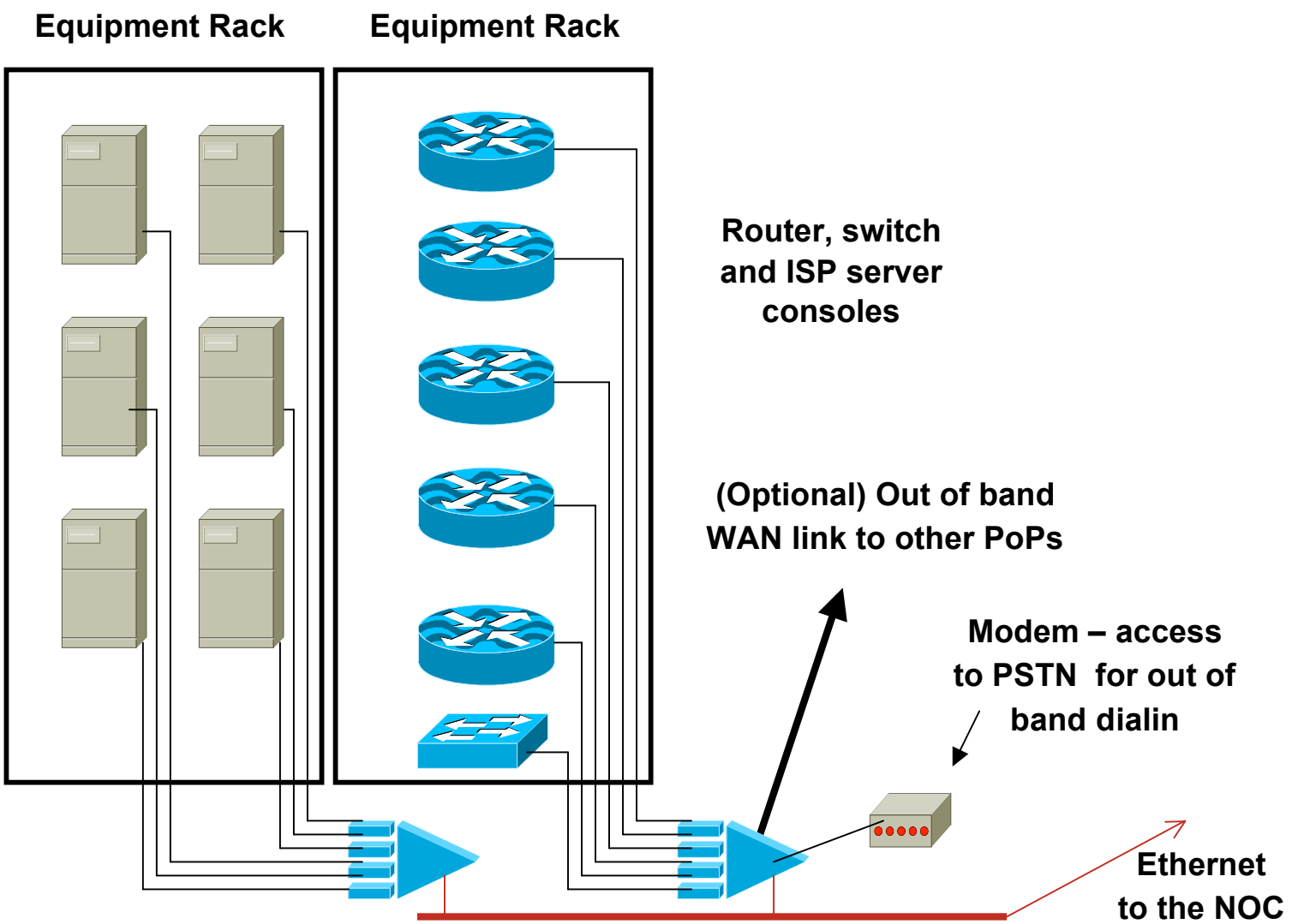
Out of Band Management

- **Not optional!**
- Allows access to network equipment in times of failure
- Ensures quality of service to customers
  - Minimises downtime
  - Minimises repair time
  - Eases diagnostics and debugging

Out of Band Management

- OoB Example – Access server:
  - modem attached to allow NOC dial in
  - console ports of all network equipment connected to serial ports
  - LAN and/or WAN link connects to network core, or via separate management link to NOC
- Full remote control access under all circumstances

Out of Band Network



Out of Band Management

- OoB Example – Statistics gathering:
  - Routers are NetFlow and syslog enabled
  - Management data is congestion/failure sensitive
  - Ensures management data integrity in case of failure
- Full remote information under all circumstances

Test Laboratory

Test Laboratory

- Designed to look like a typical PoP
  - Operated like a typical PoP
- Used to trial new services or new software under realistic conditions
- Allows discovery and fixing of potential problems before they are introduced to the network

Test Laboratory

- Some ISPs dedicate equipment to the lab
- Other ISPs “purchase ahead” so that today’s lab equipment becomes tomorrow’s PoP equipment
- Other ISPs use lab equipment for “hot spares” in the event of hardware failure

Test Laboratory

- Can’t afford a test lab?
  - Set aside one spare router and server to trial new services
  - Never ever try out new hardware, software or services on the live network
- Every major ISP in the US and Europe has a test lab
  - It’s a serious consideration

Operational Considerations

Operational Considerations

Why design the world's best network when you have not thought about what operational good practices should be implemented?

Operational Considerations Maintenance

- Never work on the live network, no matter how trivial the modification may seem
  - Establish maintenance periods which your customers are aware of
    - e.g. Tuesday 4-7am, Thursday 4-7am
- Never do maintenance on a Friday
  - Unless you want to work all weekend cleaning up
- Never do maintenance on a Monday
  - Unless you want to work all weekend preparing

Operational Considerations Support

- Differentiate between customer support and the Network Operations Centre
  - Customer support fixes customer problems
  - NOC deals with and fixes backbone and Internet related problems
- Network Engineering team is last resort
  - They design the next generation network, improve the routing design, implement new services, etc
  - They do not and should not be doing support!

Operational Considerations
NOC Communications

- NOC should know contact details for equivalent NOCs in upstream providers and peers
- Or consider joining the INOC-DBA system
  - Voice over IP phone system using SIP
  - Runs over the Internet
  - www.pch.net/inoc-dba for more information

ISP Network Design

Summary

ISP Design Summary

- KEEP IT SIMPLE & STUPID ! (KISS)
- Simple is elegant is scalable
- Use Redundancy, Security, and Technology to make life easier for yourself
- Above all, ensure quality of service for your customers

ISP Network Design

ISP/IXP Workshops