



Introductions



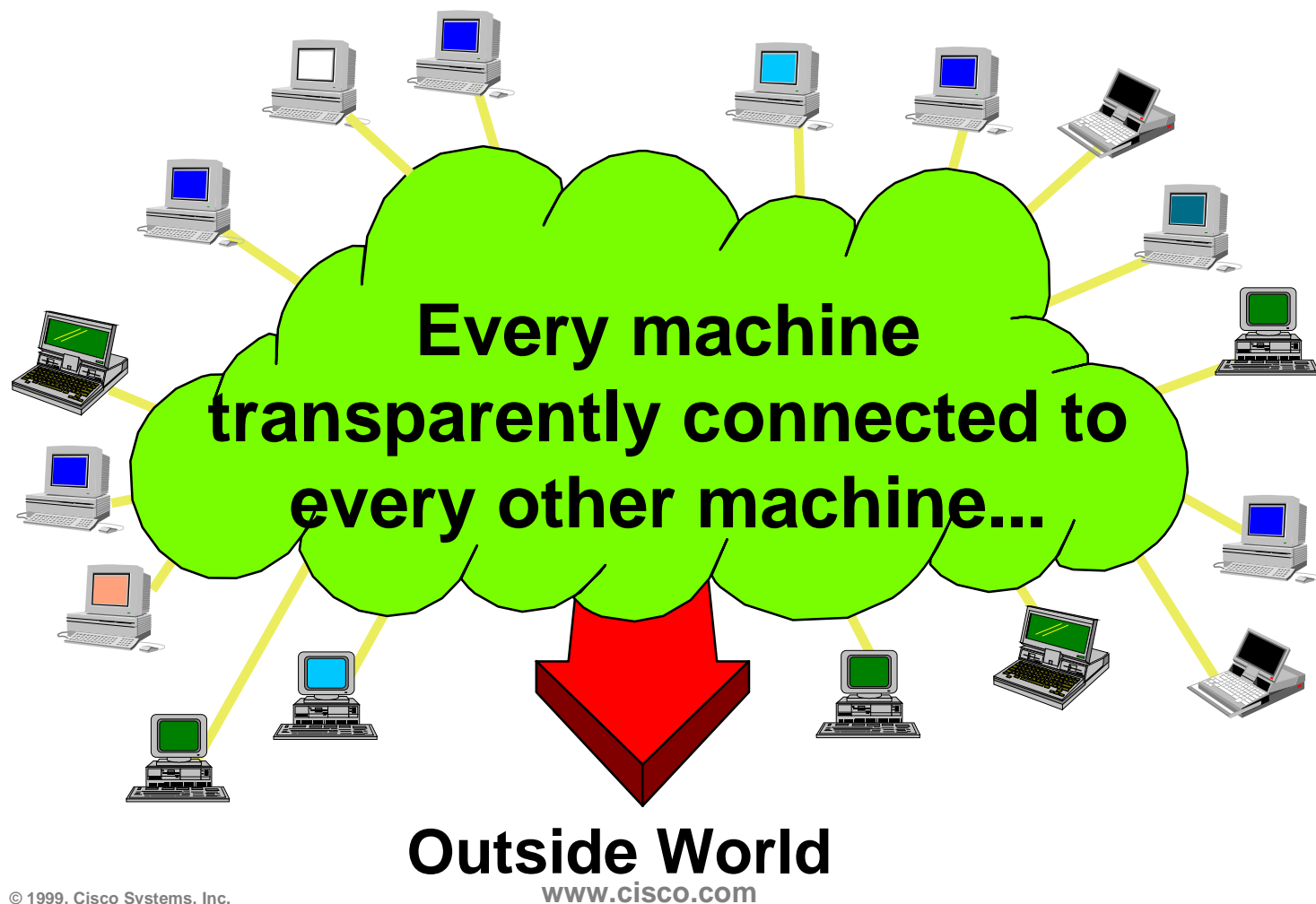
Who is Barry?

- **Barry Raveendran Greene**
 - ✓ **bgreene@cisco.com**
- **CTO Corporate Consulting**
- **Working on networks for the past 21 years.**
 - ✓ **Airfield and Military systems for the first decade (including the ARPANET and MilNet)**
 - ✓ **Internet Specific for the past decade**

Current Specializations

- **Building, Running, Scaling, and Making Money in the ISP Business**
- **Content Networking**
- **ISP Routing**
- **ISP Security**
- **Internet eXchange Points (IXPs)**
- **Trans-Oceanic Backbones**

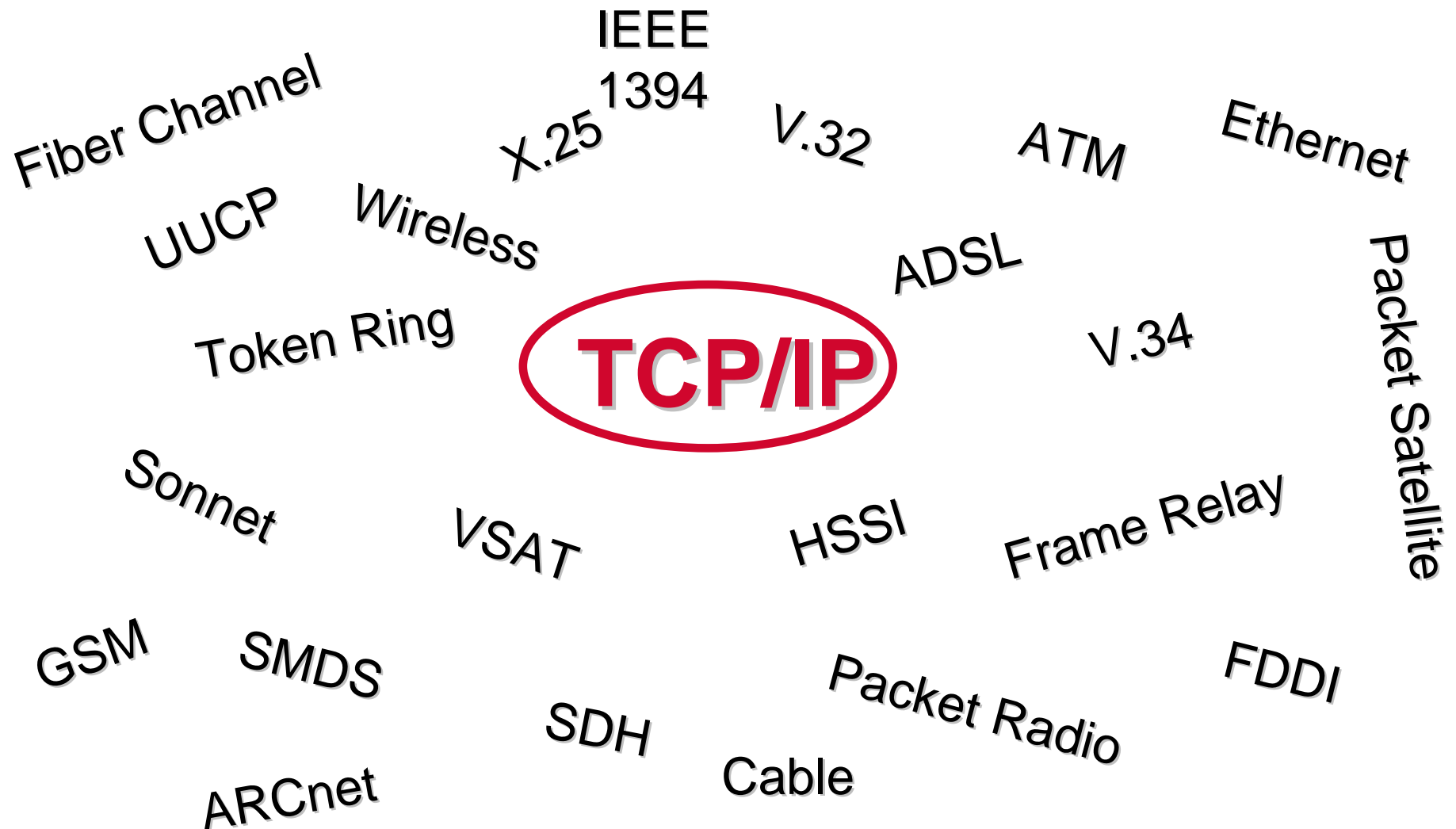
Internet - A Different Perspective



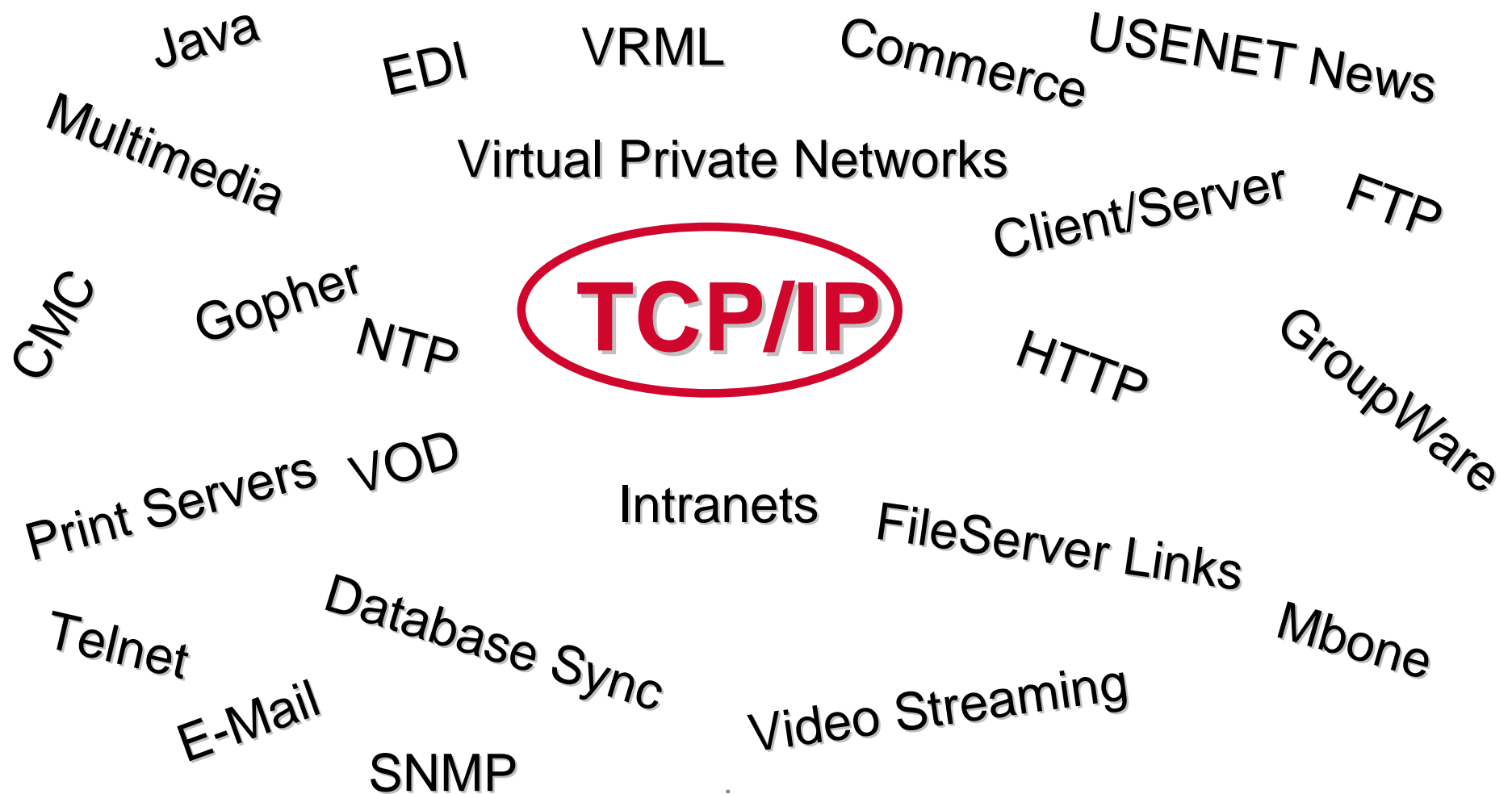
Internet - A Different Perspective

**Every Person
Virtually Connected to
Each Other....**

Why is the Internet A Success?



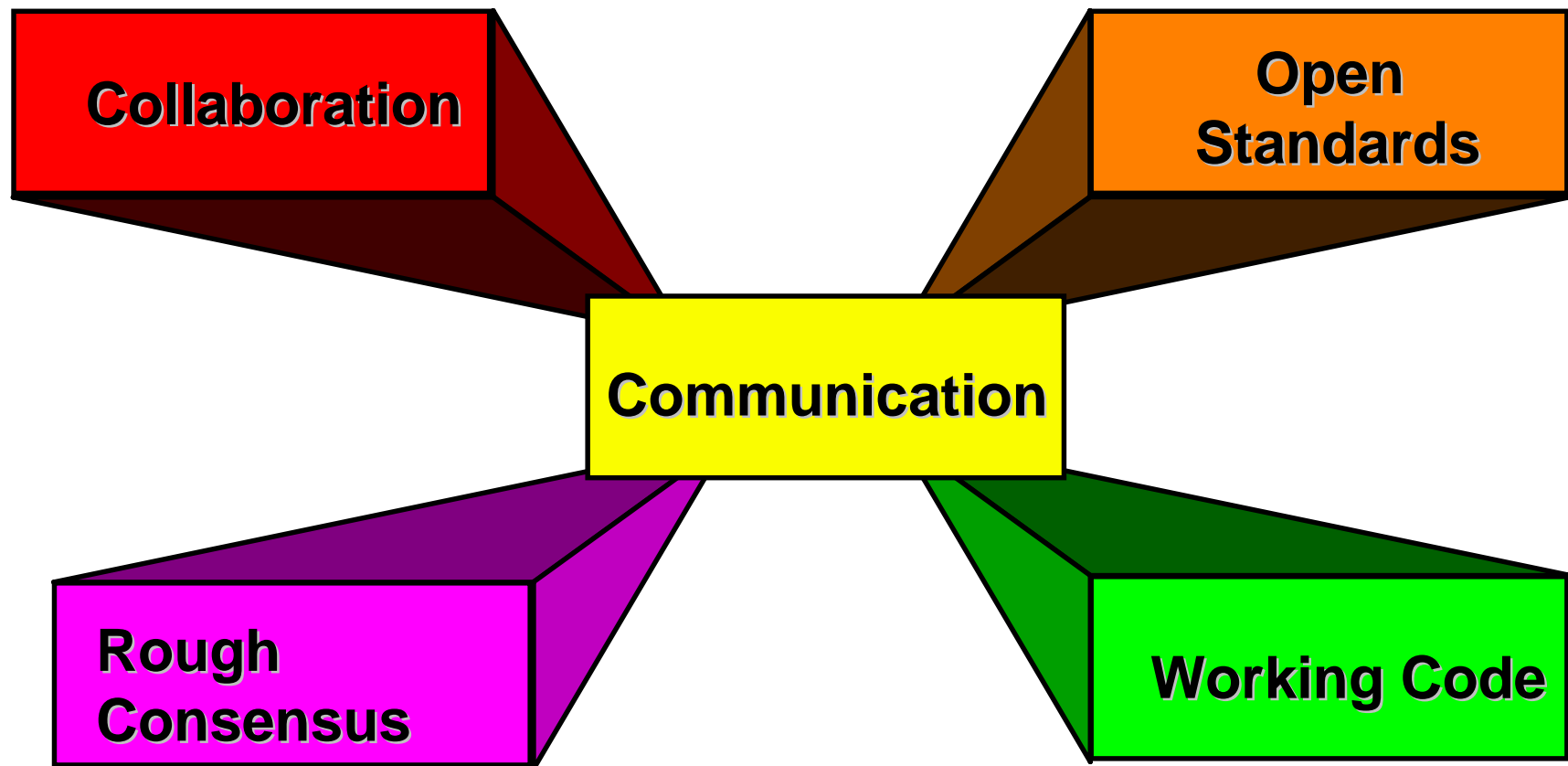
Why is the Internet A Success?



Why is the Internet A Success?

The Core Values of the Internet

From the perspective of the IETF



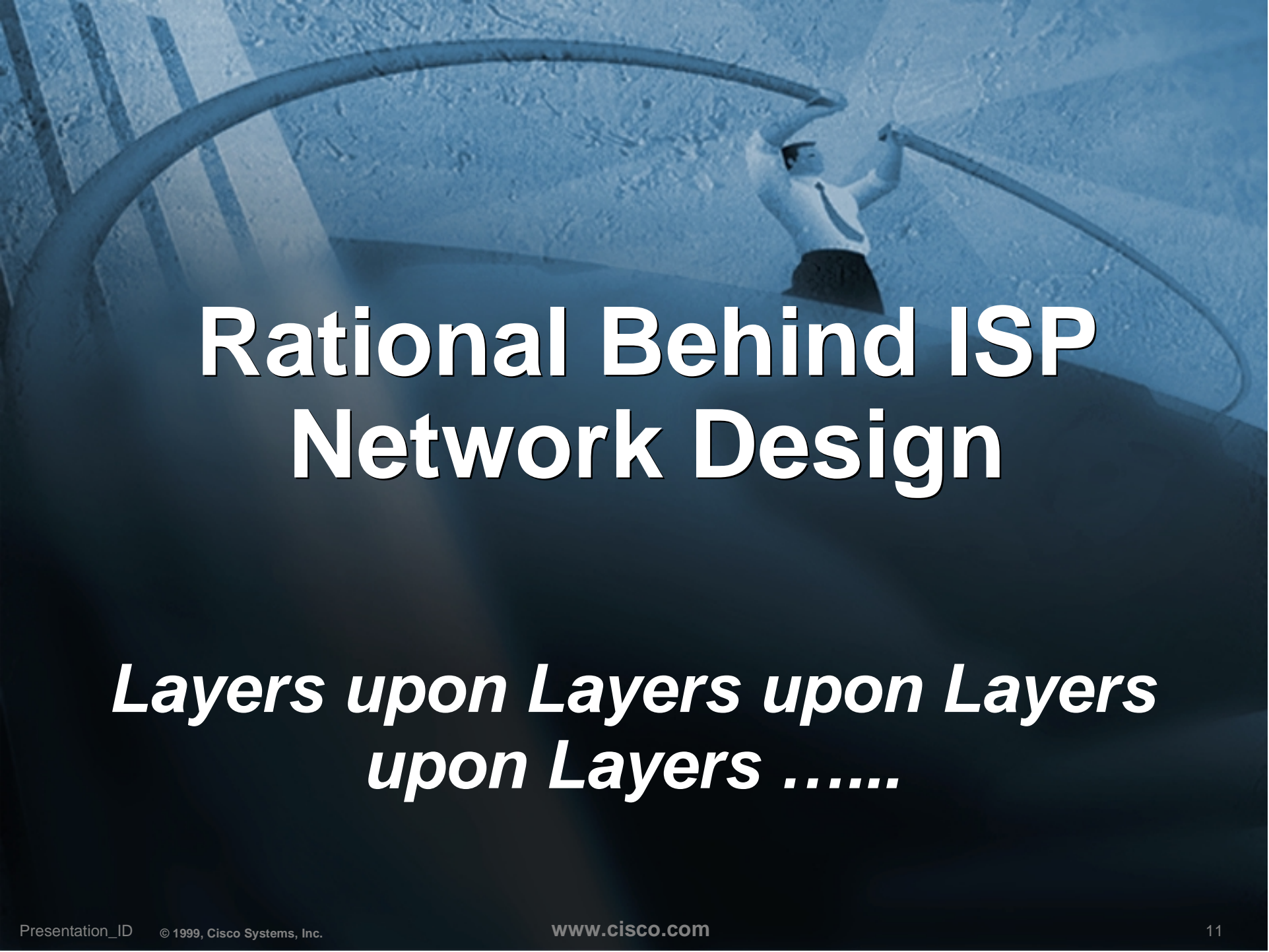


Introductions to ISP Design Fundamentals



Agenda

- **Rational Behind ISP Network Design**
- **Point of Presence Topologies**
- **Adding Services to the Architecture**
- **Impact of Services on the Network**



Rational Behind ISP Network Design

*Layers upon Layers upon Layers
upon Layers*

The Free On-line Dictionary of Computing

Architecture: Design; the way components fit together; it may also be used for any complex system, e.g. “software architecture”, “network architecture”

Network Design and Architecture...

- ... can be critical
- ... can contribute to the success of the network
- ... can contribute to the failure of the network

Ferguson's Law of Engineering

“

**No amount of magic
knobs will save a
sloppily
designed network**

”

**Paul Ferguson—Consulting Engineer,
Cisco Systems**

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**
 - ✓ **One world revolves around private organizational networks and another concerns the global Internet**
- **Growth in the Internet is faster than any other technology introduced to the public-at-large**

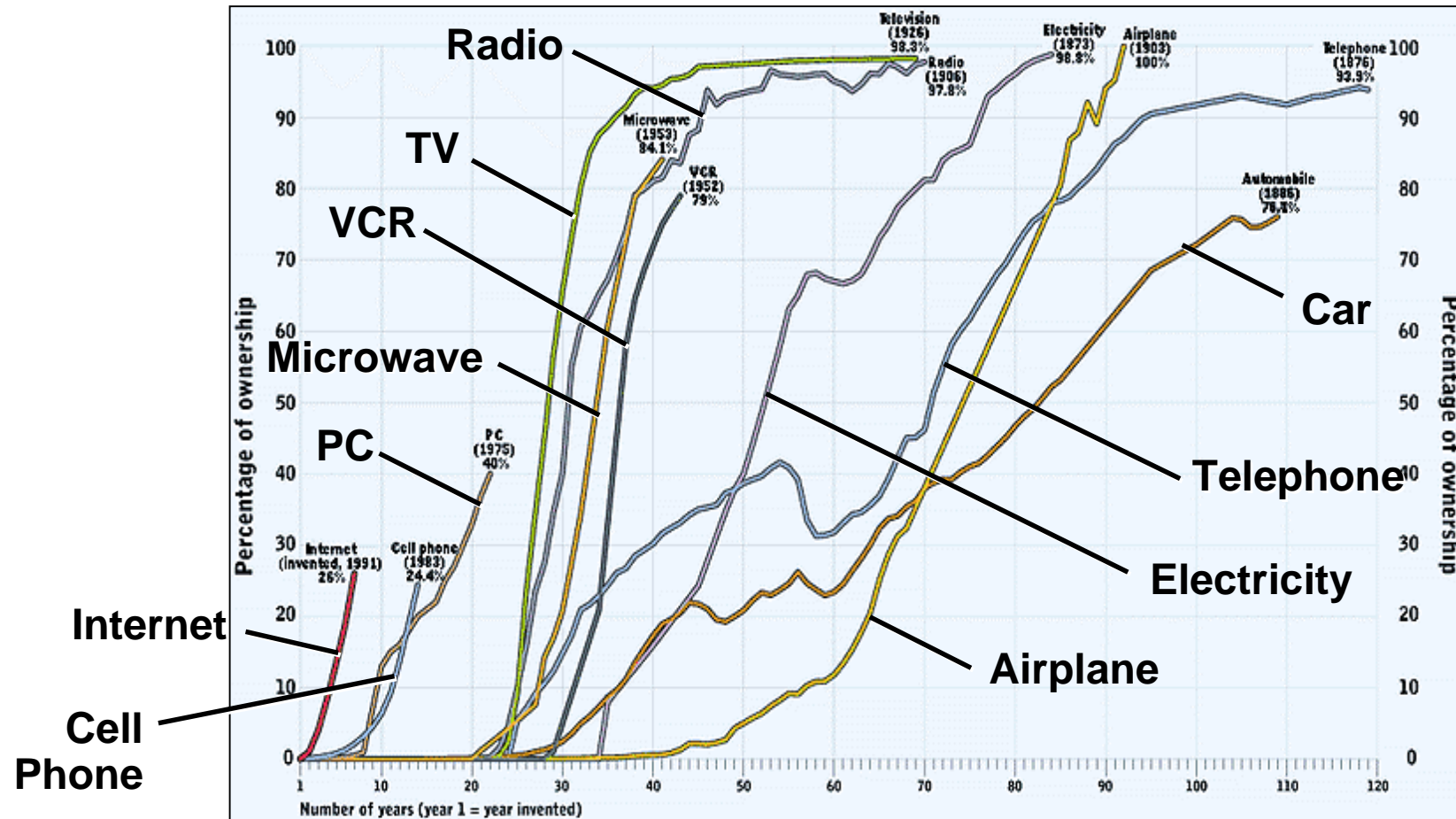
Scaling is the #1 Problem on the Internet

“

***If you're not scared yet,
you don't understand the
problem!***

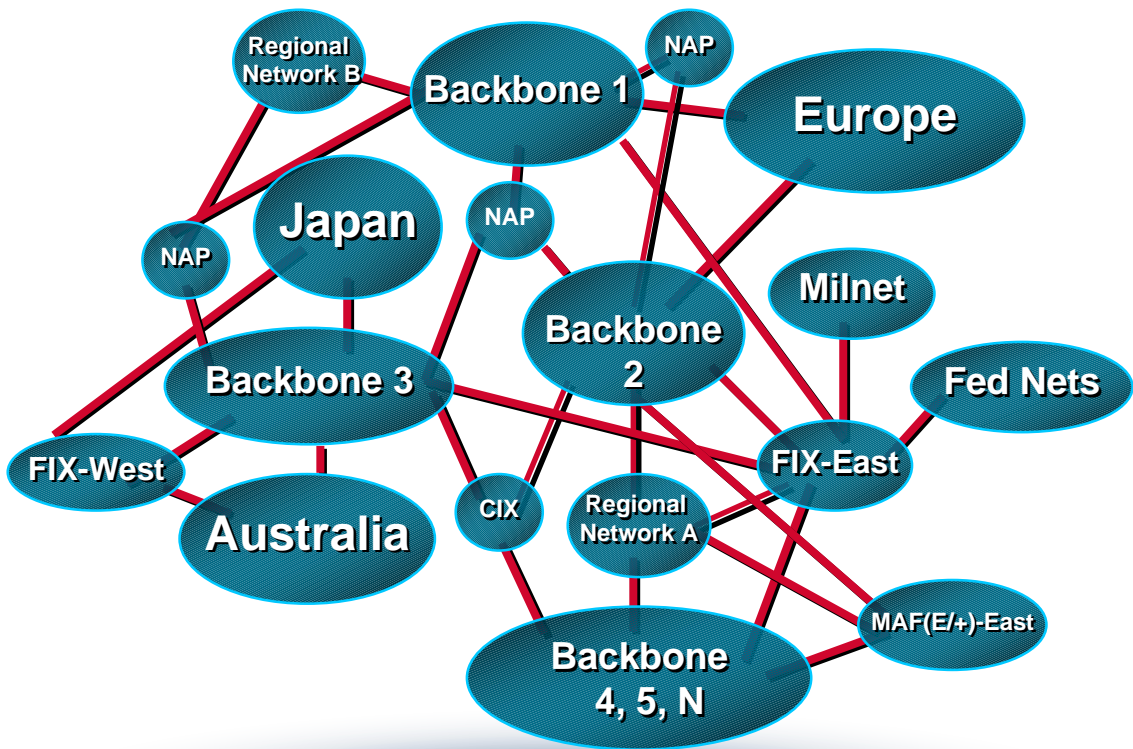
”

Technology Adoption



Basic Scaling Concepts

- **Hierarchy**
- **Discipline**
- **Information reduction**
- **Consistency**
- **Planning**





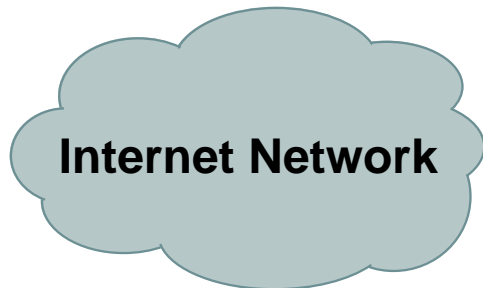
Old World vs New World

More Issues to Consider

Key Design Principles

- **Internet/L3 Networks**

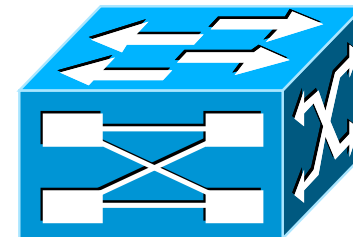
- ✓ Build the redundancy into the **system**.



- **Telco Voice and L2 Networks**

- ✓ Put all the redundancy into a **box**.

VS



Key Design Principles

- **Internet Years**

- ✓ **Very Rapid Change**
- ✓ **1 Year = 3 Months**

VS

- **Telco Years**

- ✓ **Slow Consistent Change**
- ✓ **1 Year = 3 Years**

Key Design Principles

- **Internet Growth**

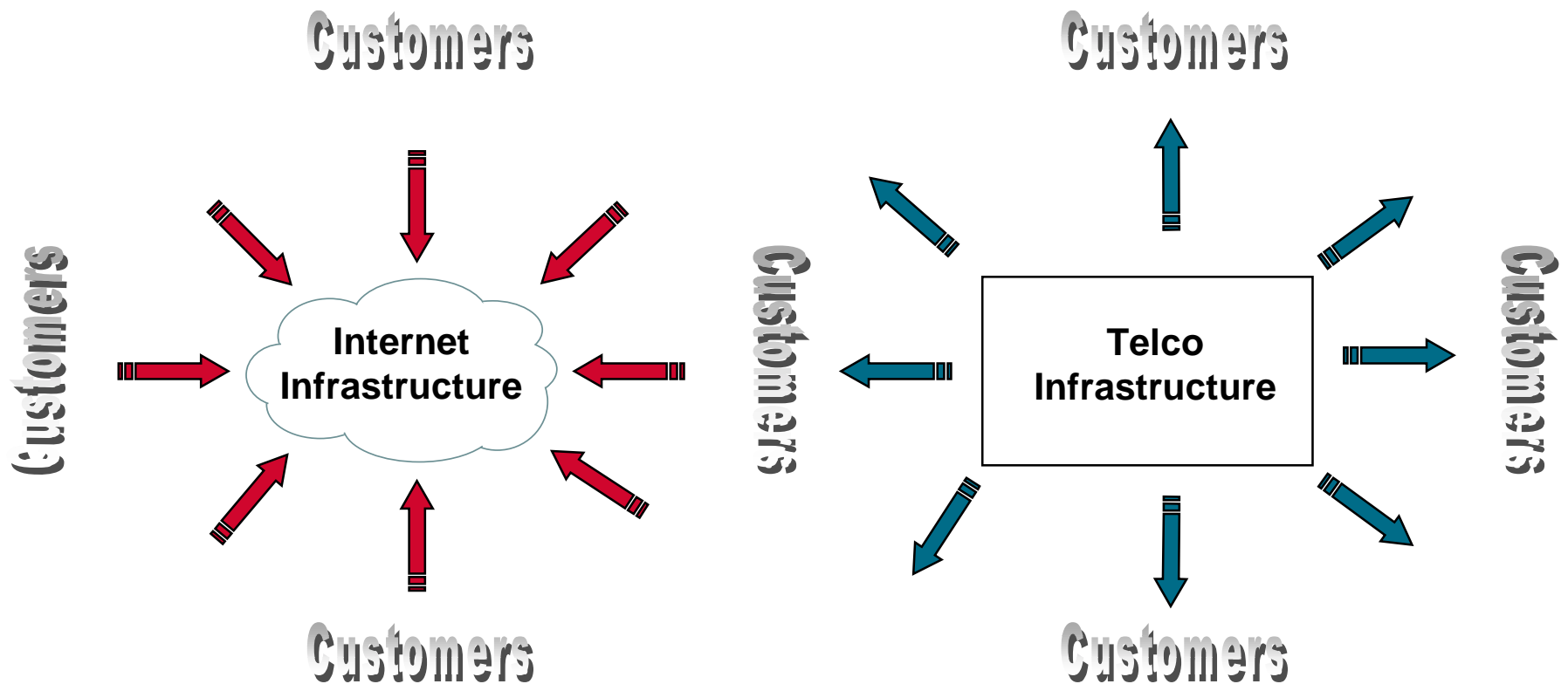
- ✓ **+ 100% per year
for the past 10
years**

VS

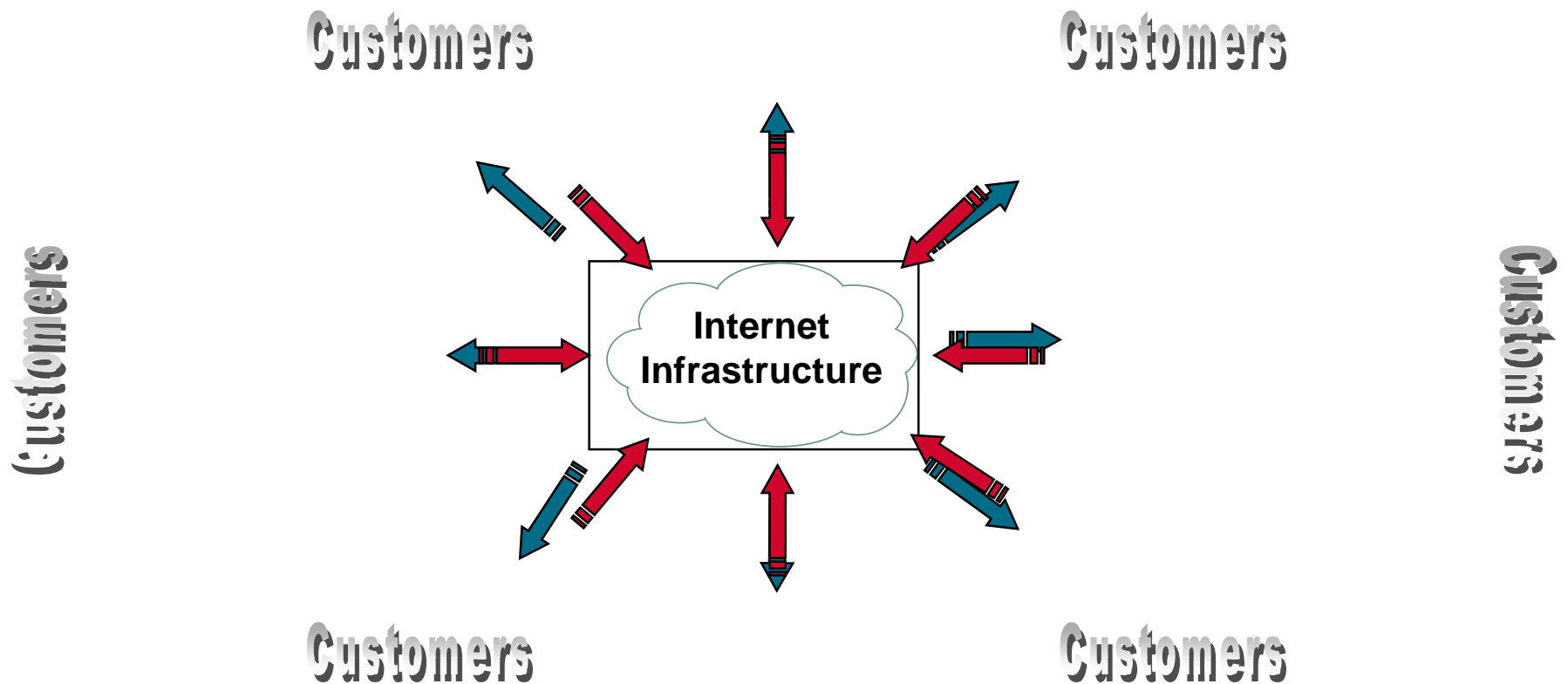
- **Telco Growth**

- ✓ **Telephony is 5%
to 25% year for
the past 10
years**

Key Design Principles



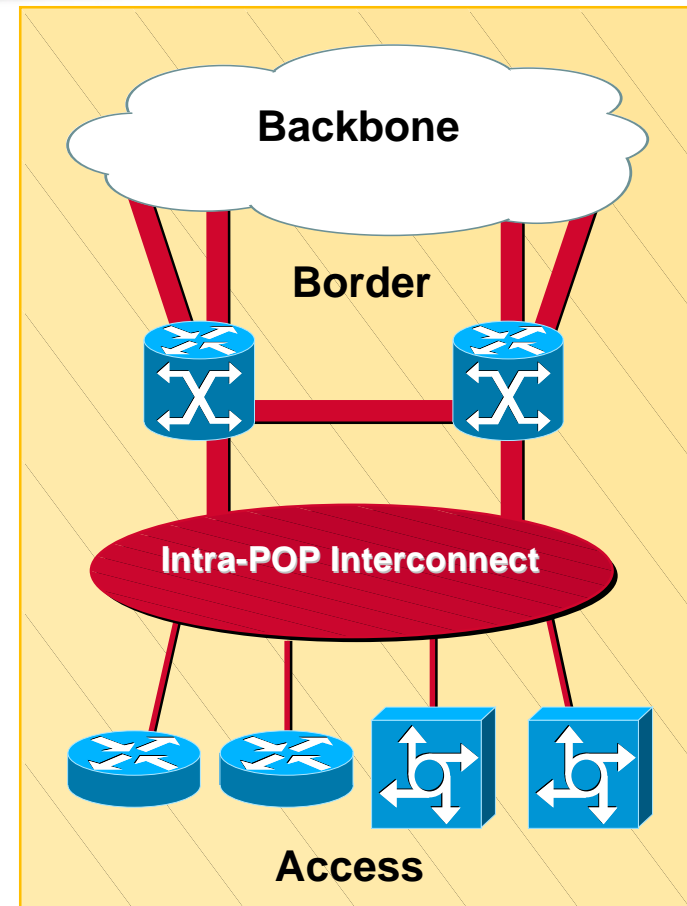
Key Design Principles



Key Design Principles

- **Triple Layered POP Redundancy**

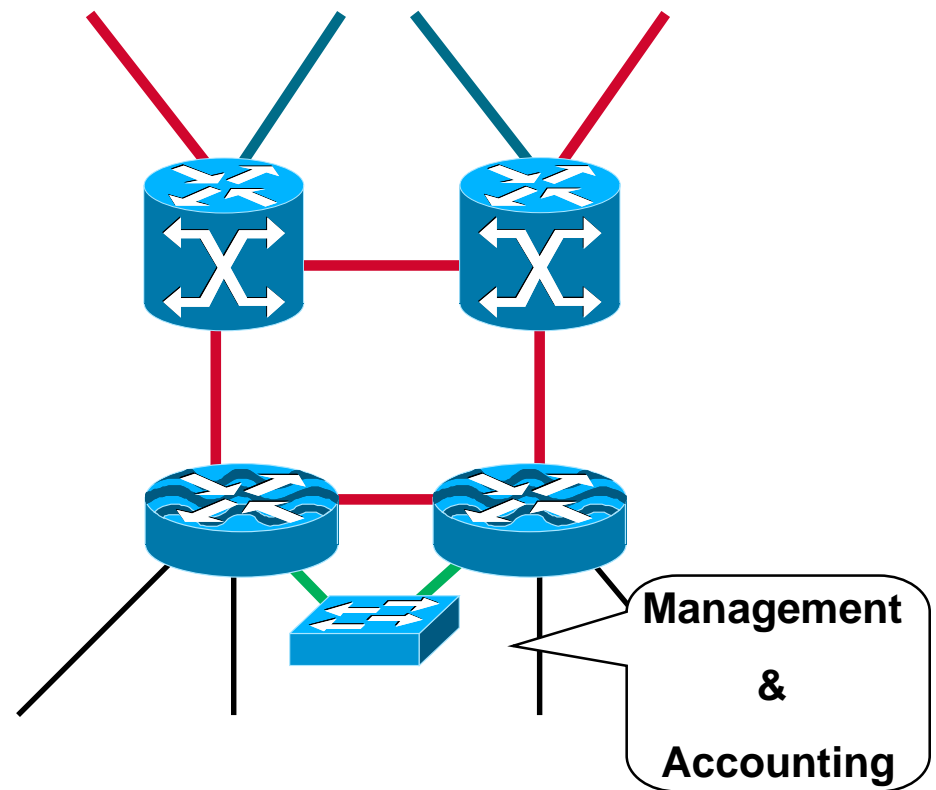
- ✓ Two connection to the backbone from any boarder router.
- ✓ Two boarder routers, load balanced w/ one able to take the full load.
- ✓ Two POP interconnect devices and/or a physical failover medium (FDDI or POS)



Key Design Principles

- **Interconnection for Management, Security, and Accounting services**

- ✓ Netflow Devices - FlowCollector
- ✓ Syslog collector for all network devices
- ✓ SNMP collector (PC Based UNIX)
- ✓ Security Auditing Tools (NetSonar)



Key Design Principles

“

Do not throw out the baby with the bath water. +100 years of build networks is experience that cannot be ignored.

”

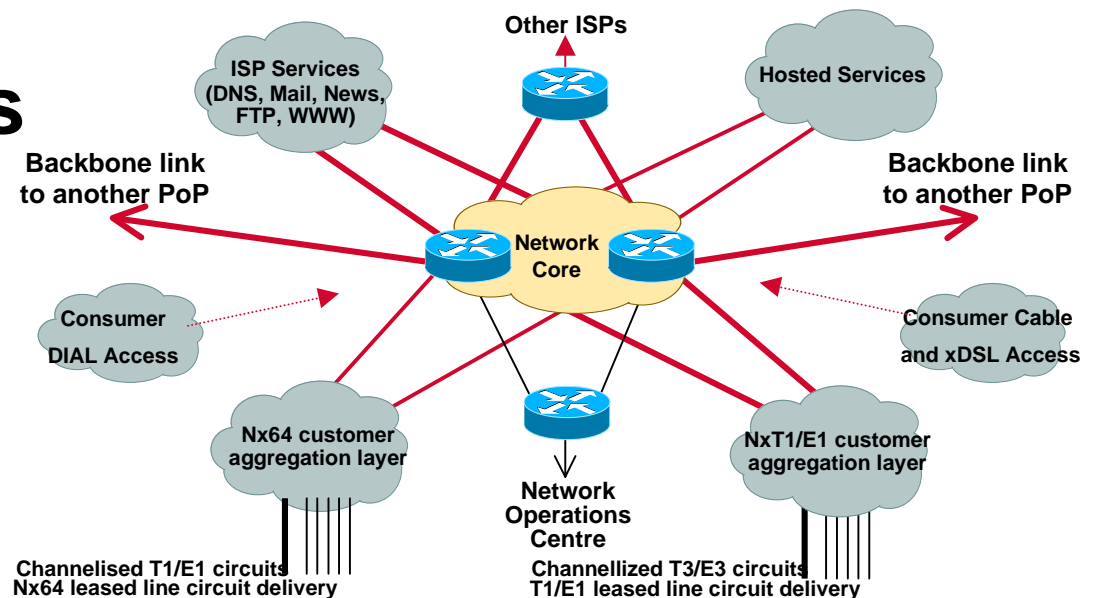
Core Influences to ISP Design

- **Modular Design**
- **Functional Design**
- **Tiered/Hierarchical Design**
- **Multiple Levels of Redundancy**
- **Routing Protocol Hierarchy**
- **Build for IP Forwarding First - then add services**

Modular Design

Organize the Network
into separate and
repeatable modules

- ✓ Backbone
- ✓ POP
- ✓ Hosting Services
- ✓ ISP Services
- ✓ Support/NOC

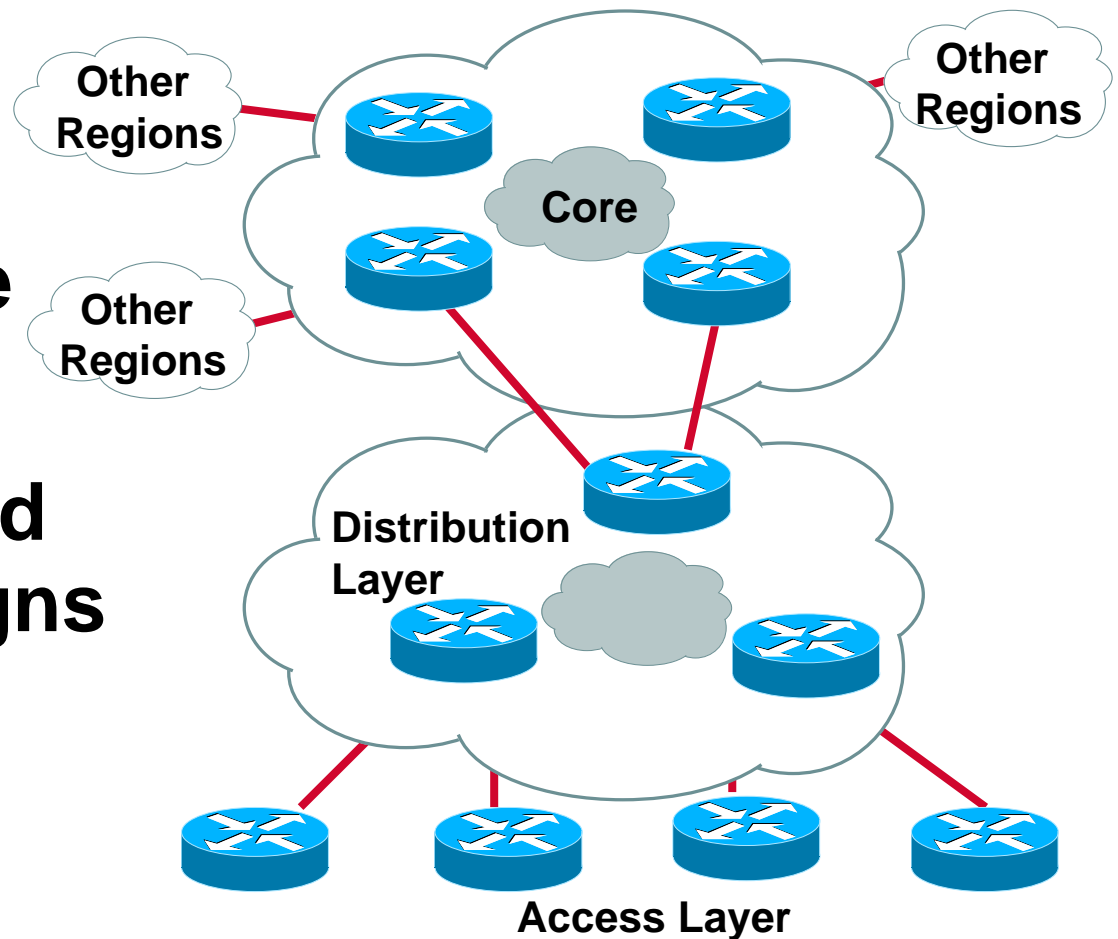


Functional Design

- **One *Box* cannot do everything!** (no matter how hard people have tried in the past)
- **Each router/switch in a network has a well-defined set of functions.**
- **The various *boxes* interact with each other.**
- **ISP Networks are a systems approach to design.**

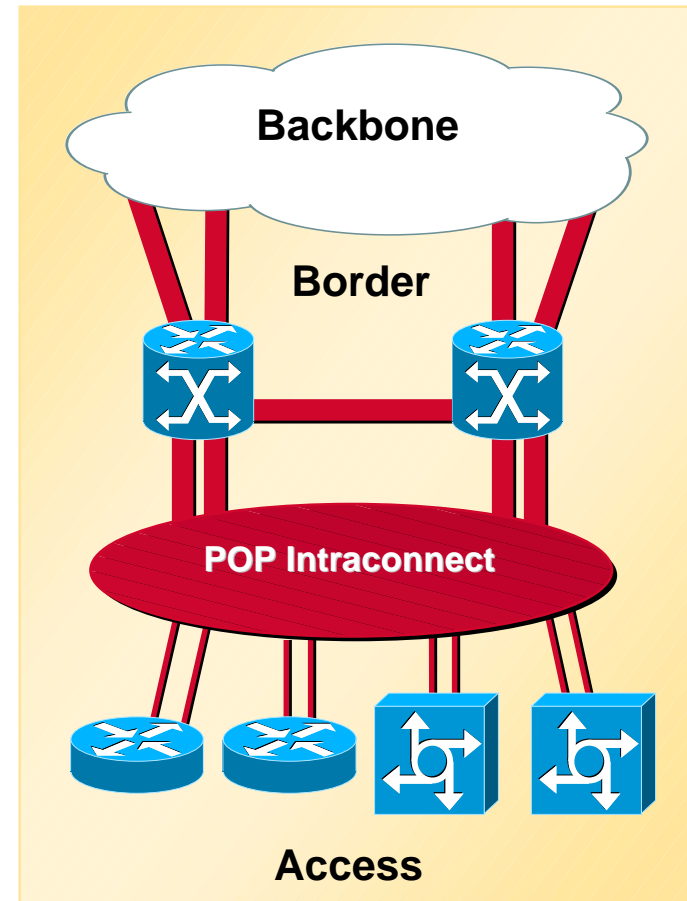
Tiered/Hierarchical Network Design

- **Flat - Meshed Topologies have not scaled.**
- **Hierarchy is used in network designs to scale the network.**



Multiple Levels of Redundancy

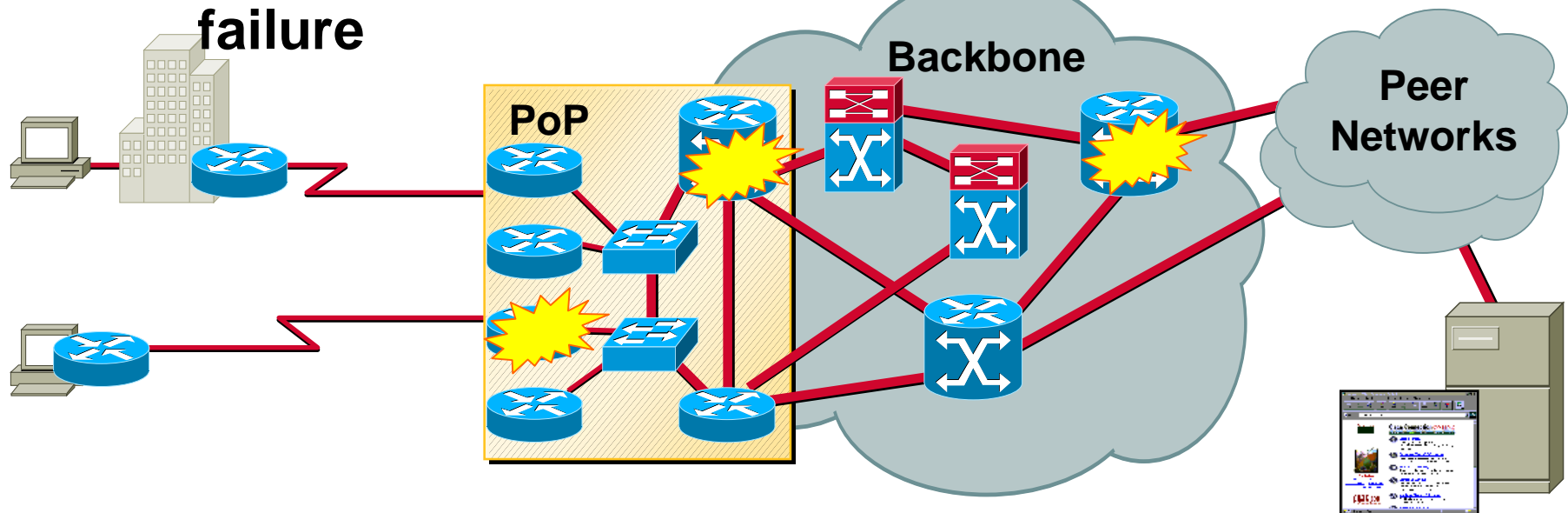
- **Triple Layered POP Redundancy**
 - ✓ Lower-level failures are better
 - ✓ Lower-level failures may trigger higher-level failures
 - ✓ L2: Two of everything at
 - ✓ L3: IGP and BGP provide redundancy and load balancing
 - ✓ L4: TCP re-transmissions recovers during the fail-over



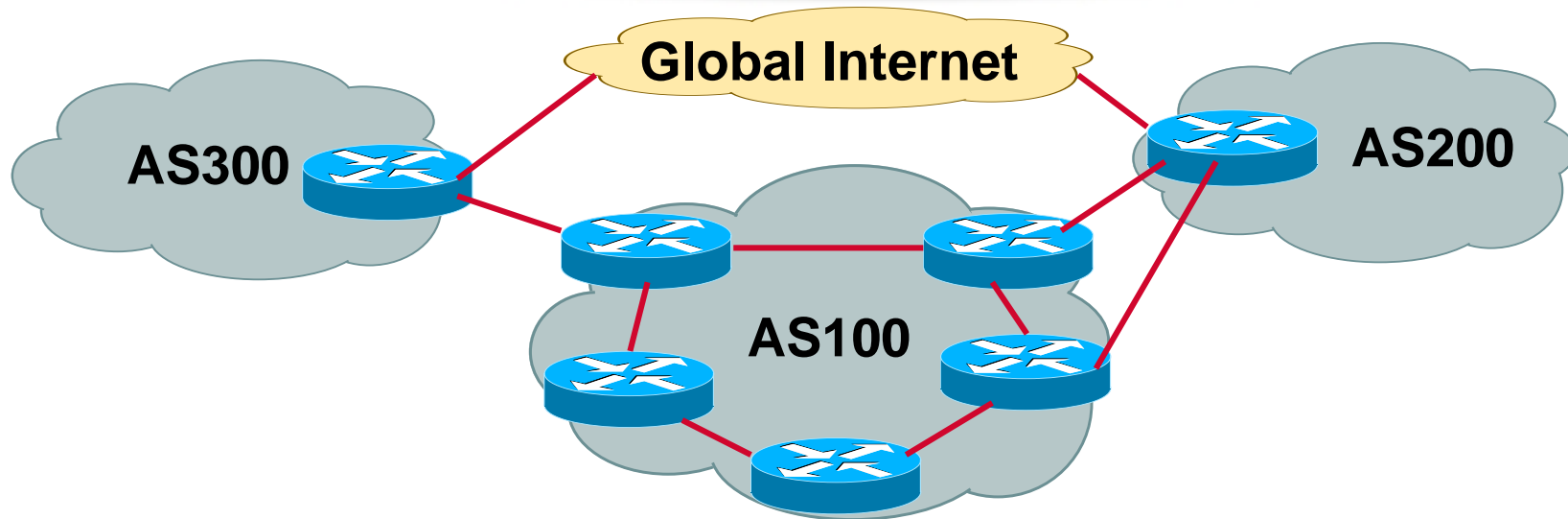
Multiple Levels of Redundancy

- **Objectives -**

- ✓ As little user visibility of a fault as possible
- ✓ Minimize the impact of any fault in any part of the network.
- ✓ Network needs to handle L2, L3, L4, and Router failure

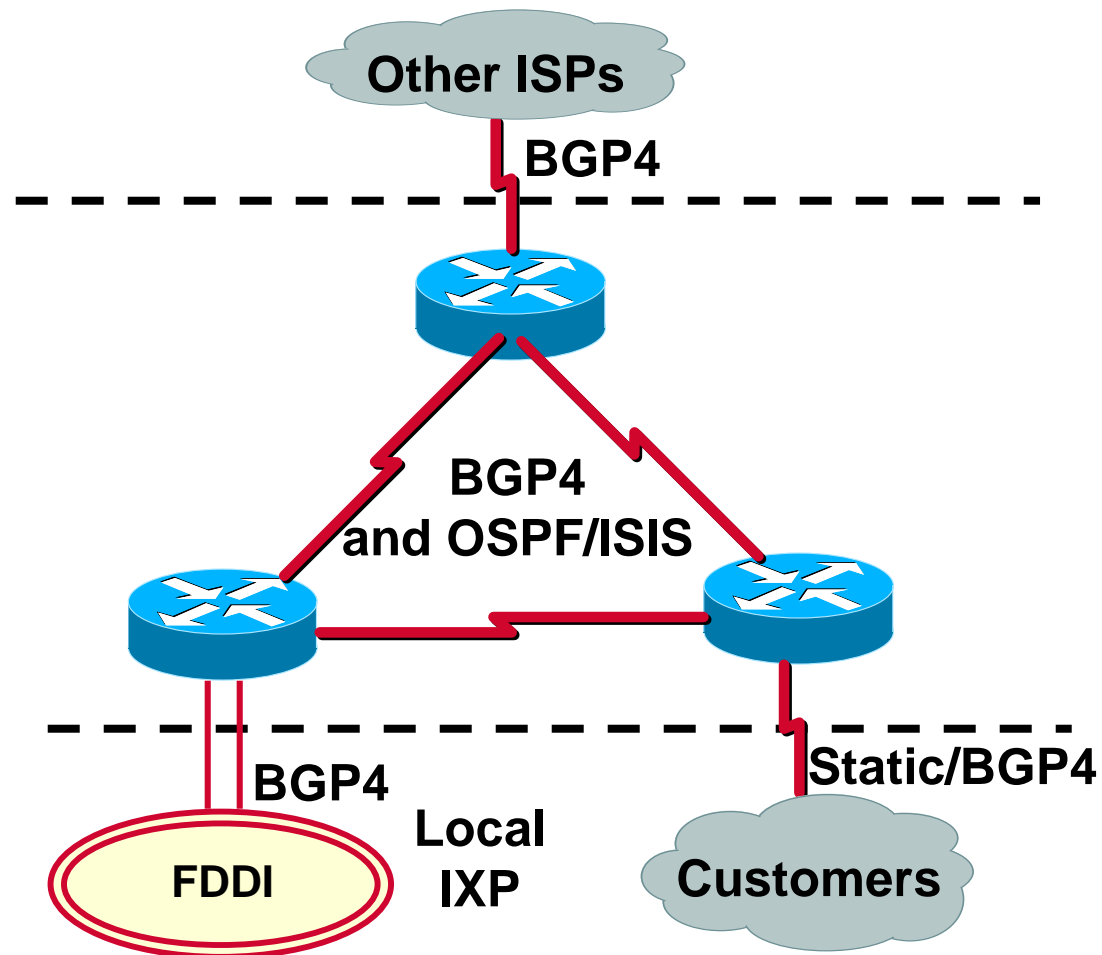


Enterprise Multihoming is Practically a Norm



- **Many situations possible**
 - ✓ multiple sessions to same ISP
 - ✓ secondary for backup only
 - ✓ load-share between primary and secondary
 - ✓ selectively use different ISPs

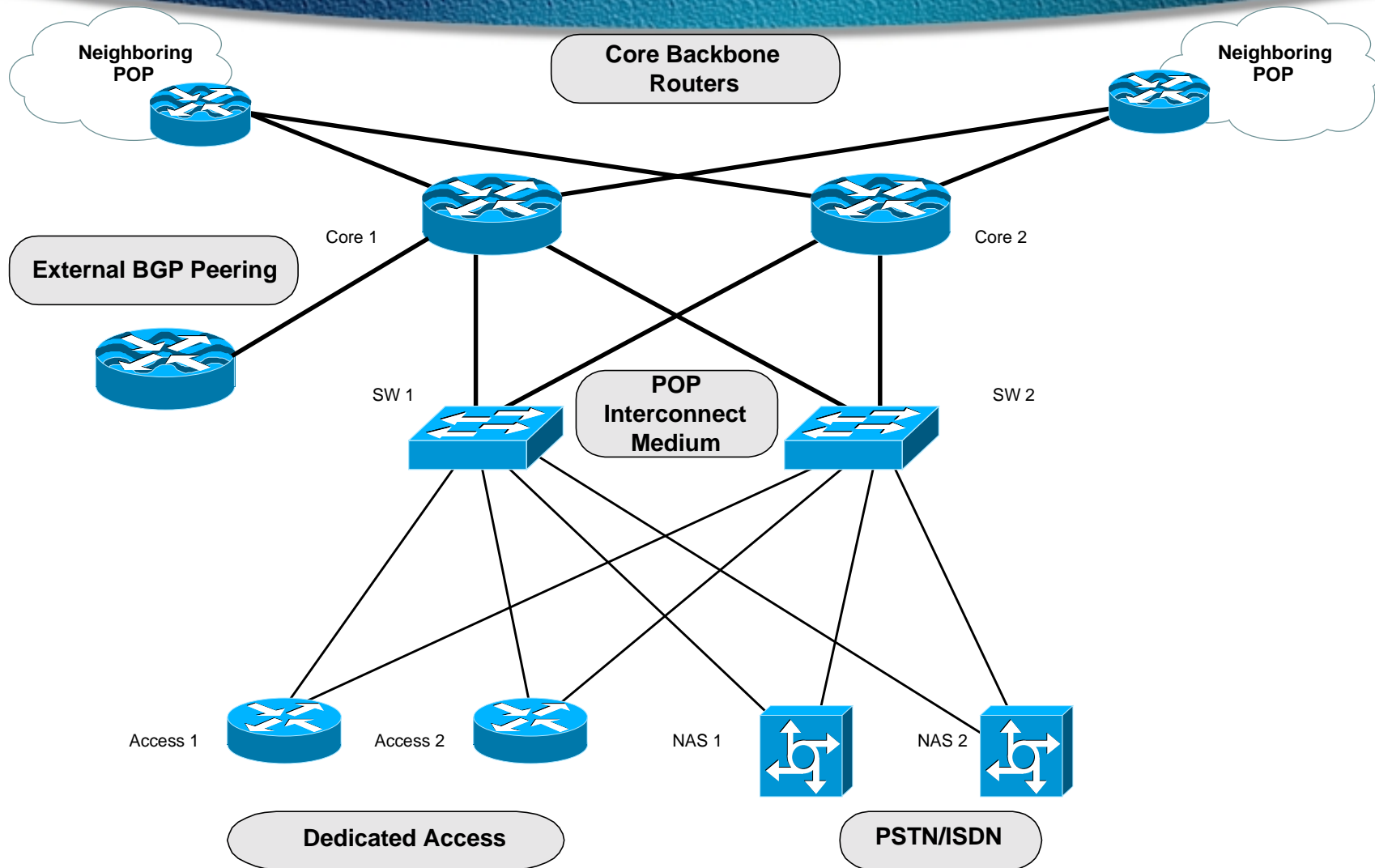
Hierarchy of Routing Protocols





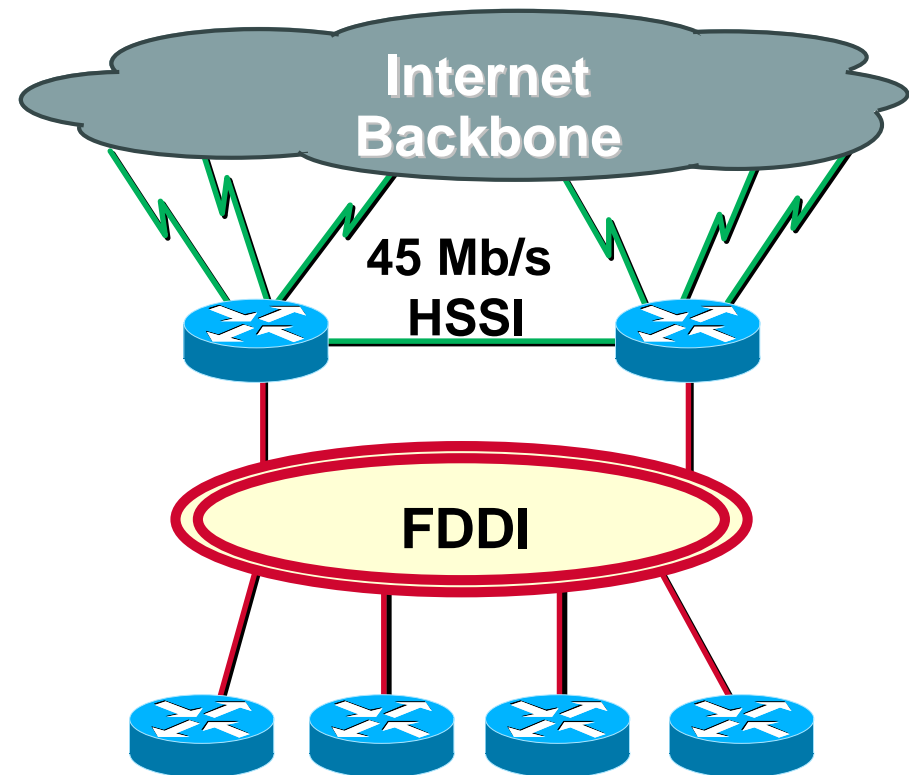
Point of Presence Topologies

PoP Design



Early Internet POP Architecture - NSP

- ✓ Backbone trunks at 45 Mb/s
- ✓ Shared media interconnect within POP:
FDDI, Ethernet, Switched Ethernet
- ✓ Conventional T3 backbone Internet router



Internet POP Architecture - '96/'98

- ✓ **Backbone trunks at 155 Mb/s**

Packet over SONET OC3

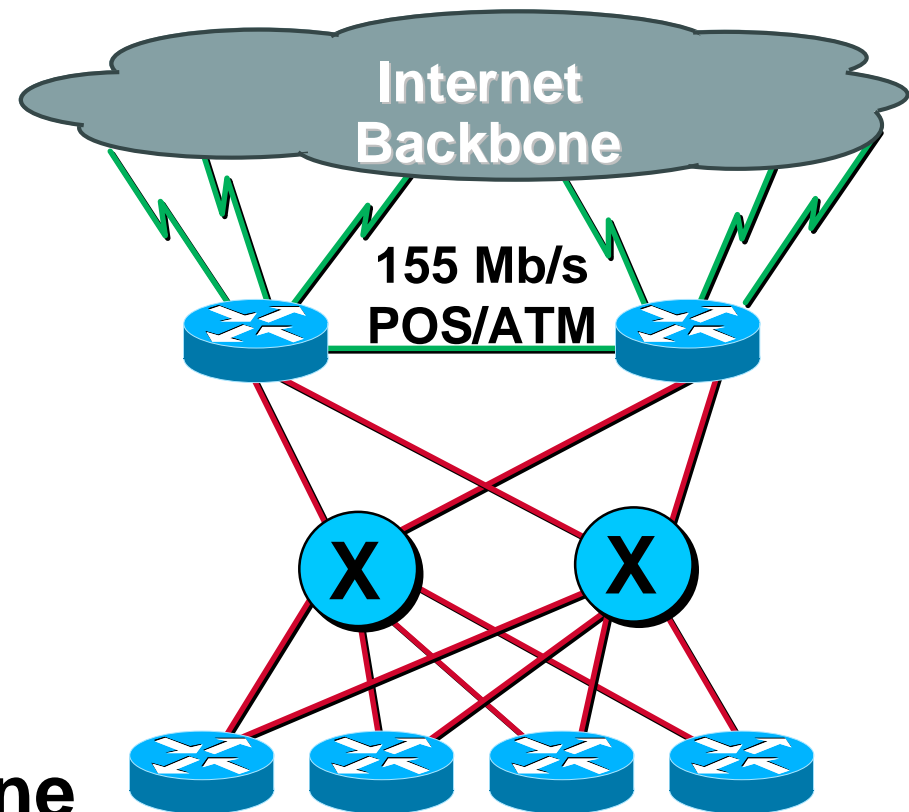
ATM OC3

- ✓ **Switched interconnect within POP:**

Switched FDDI/Fast Ethernet

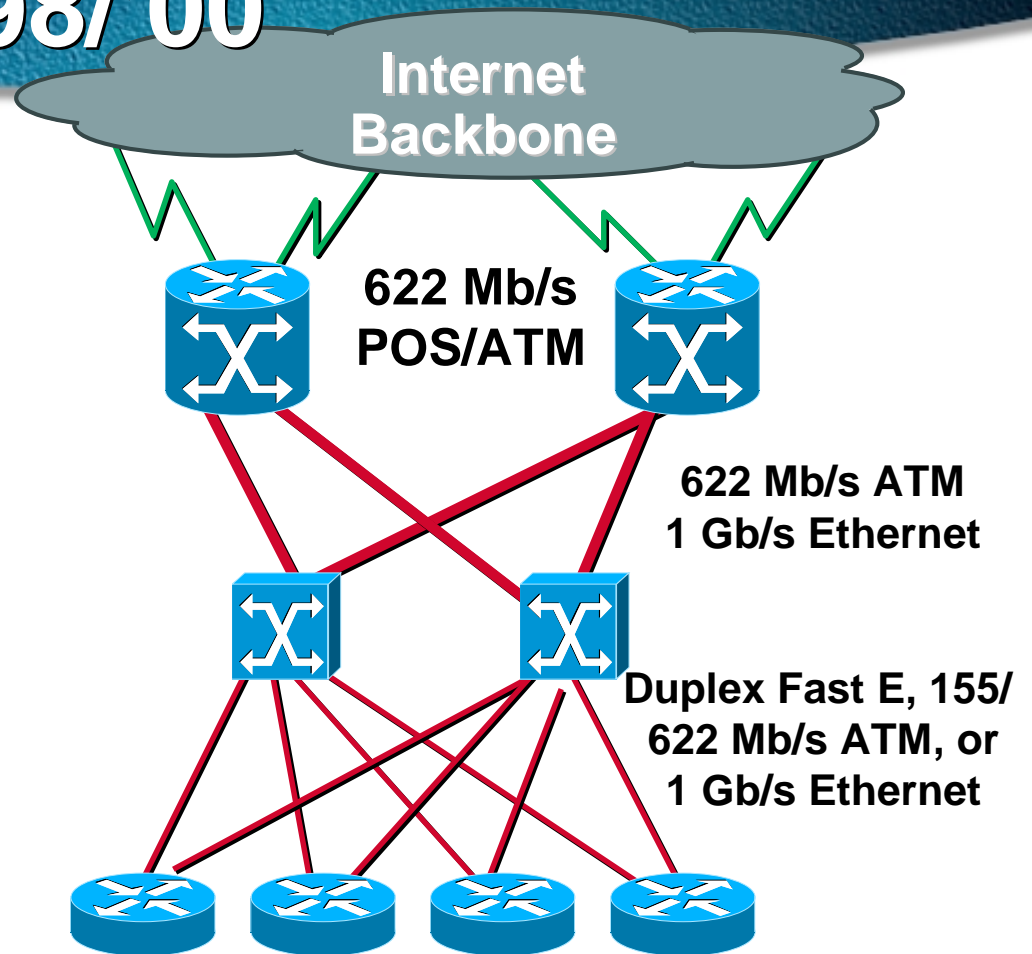
ATM OC3

- ✓ **Advanced OC3 backbone Internet router**



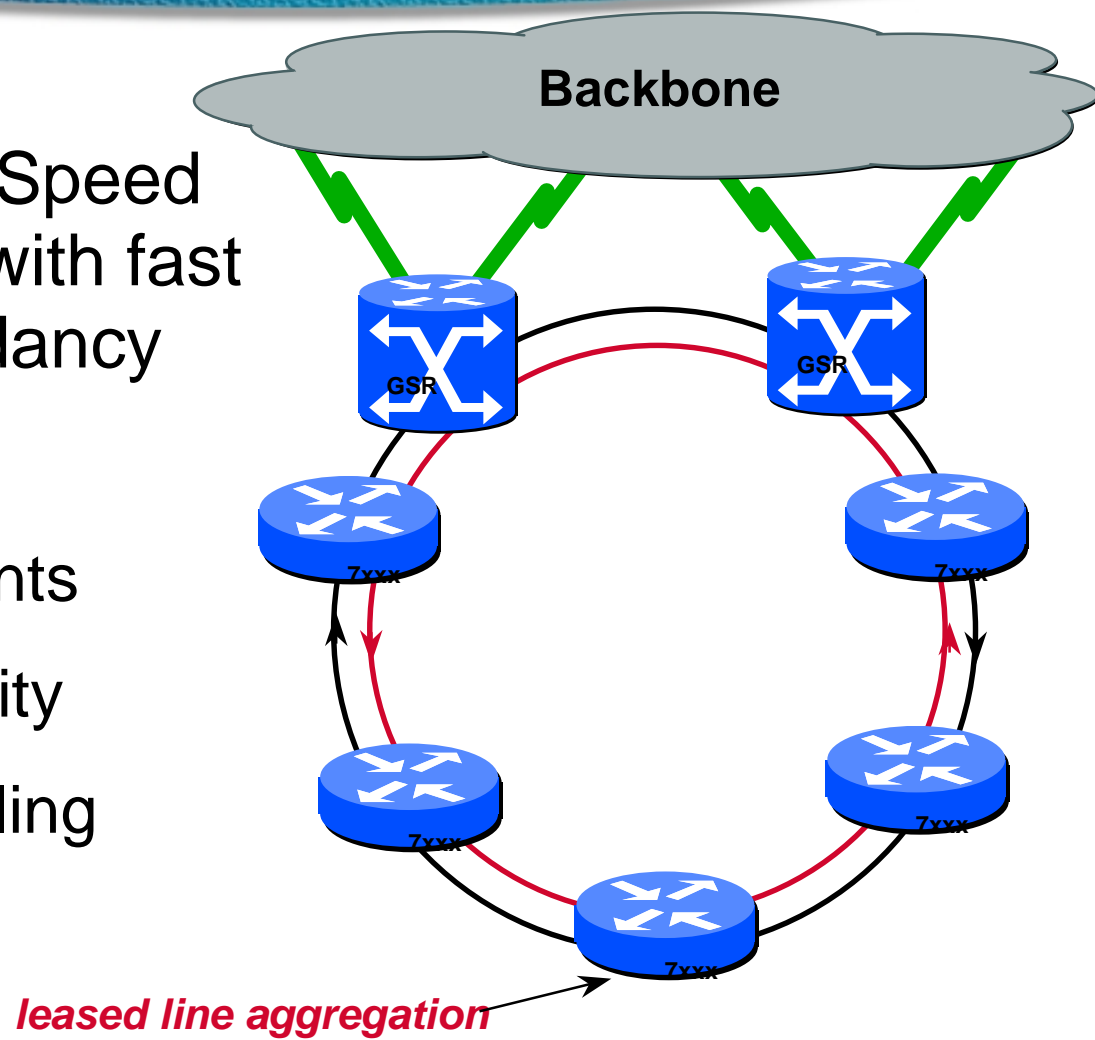
Internet POP Architecture - '98/'00

- ✓ **Backbone trunks at 622 Mb/s**
Packet over SONET OC12
ATM OC12
- ✓ **Switched interconnect within POP:**
ATM at OC3 AND OC12
Ethernet Channel
Gigabit Ethernet (early '98)
POSIP (late '98)
- ✓ **Gigabit OC12 backbone Internet router**



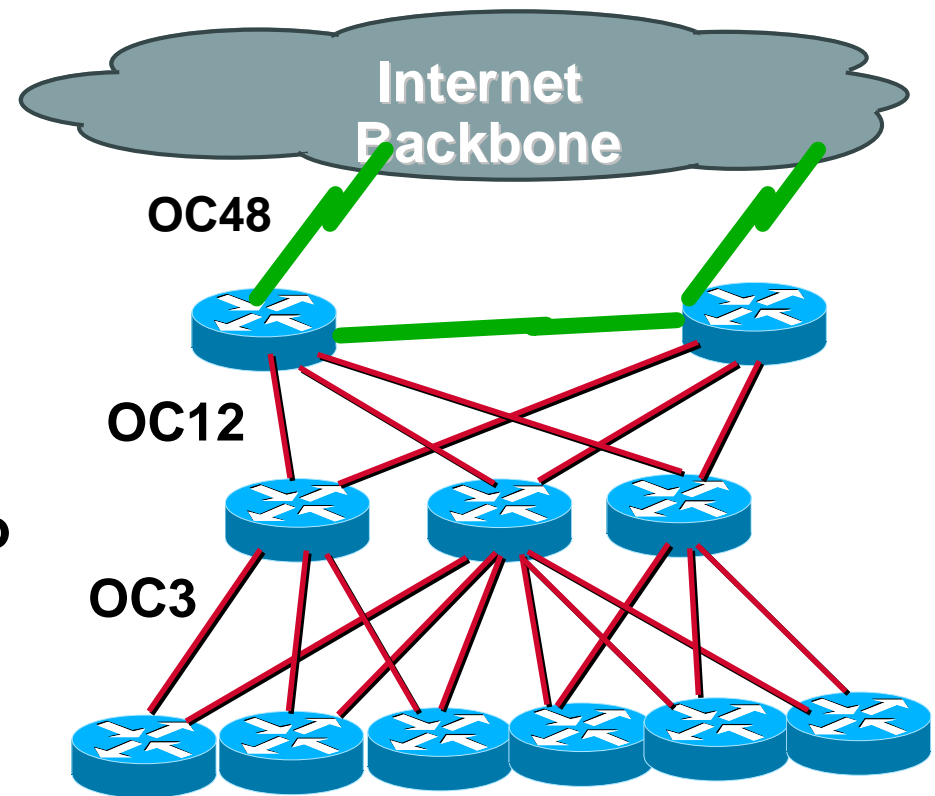
Internet POP Architecture - '99/'01

- **SRP Rings** - High Speed of SDH combined with fast failover and redundancy
 - ✓ High bandwidth
 - ✓ Reduced port counts
 - ✓ Reduced complexity
 - ✓ Proactive self healing

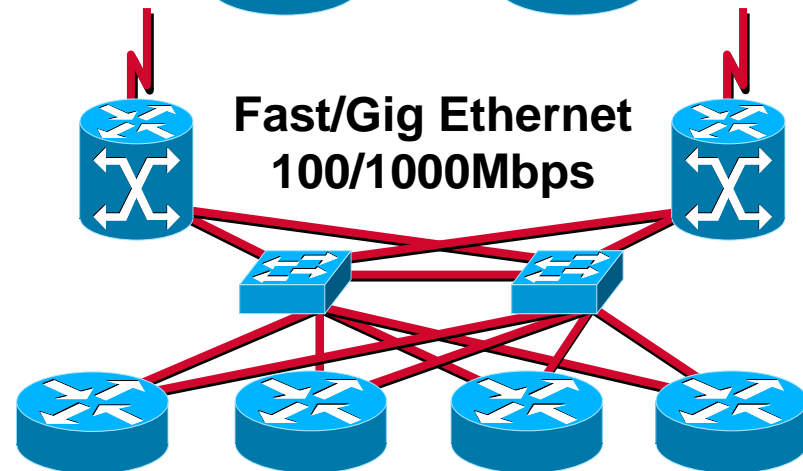
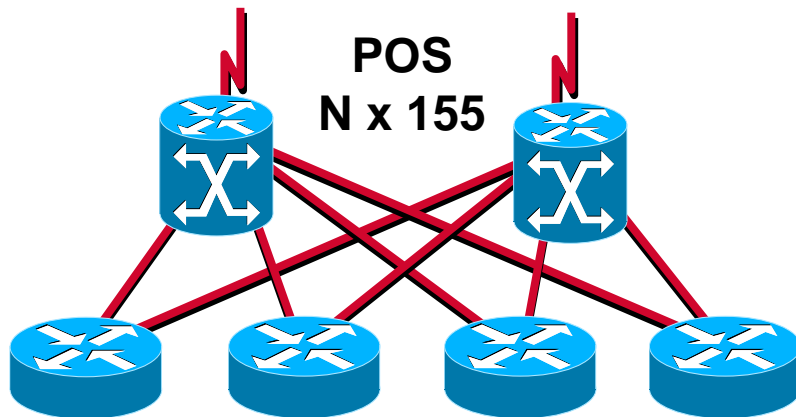
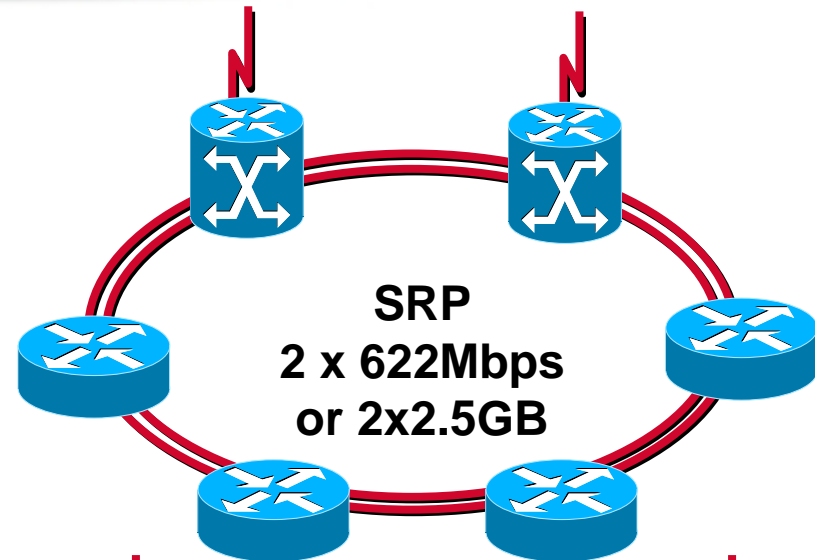
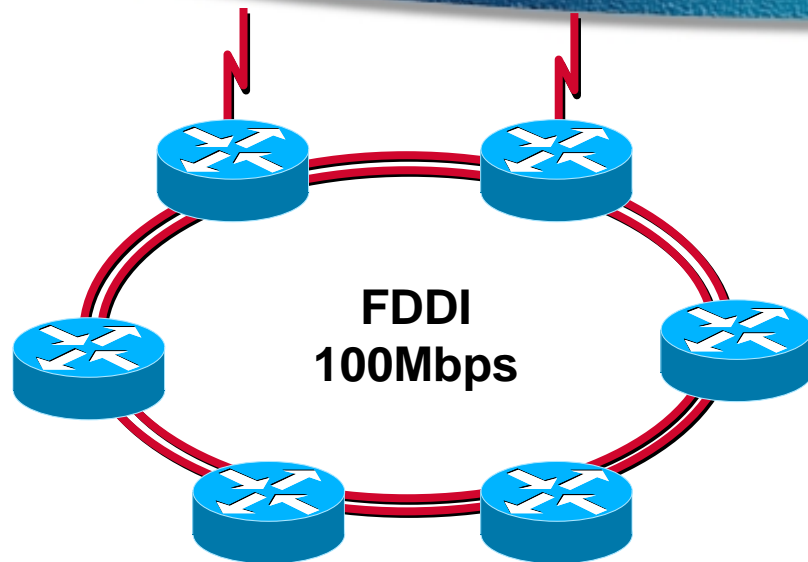


Large POPs - add a 3rd layer

- ✓ Problem: port density!
- ✓ Solution: buy more routers!
- ✓ Customer routers connect to aggregation routers
 - Packet over SONET OC3
 - ATM OC3
- ✓ Aggregation routers connect to backbone routers
- ✓ Scales nicely
- ✓ X CRs to Y ARs to Z BRs
 - ✓ ...where $X > Y > Z$
 - ✓ Be careful not to oversubscribe!



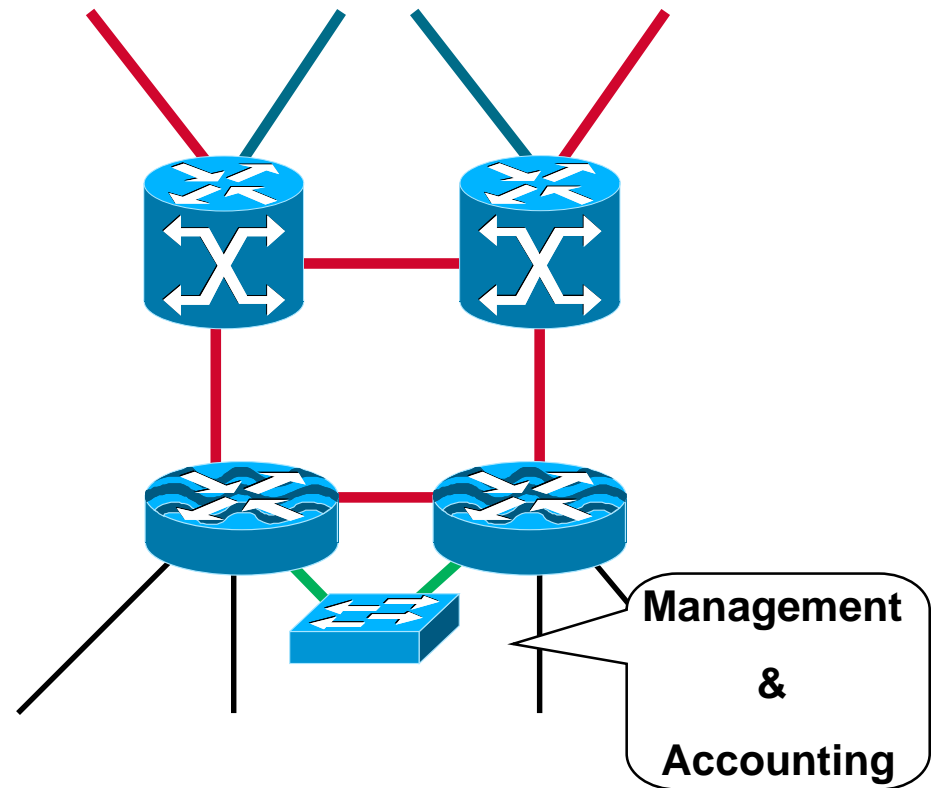
POP Interconnect Summary



Key Design Principles

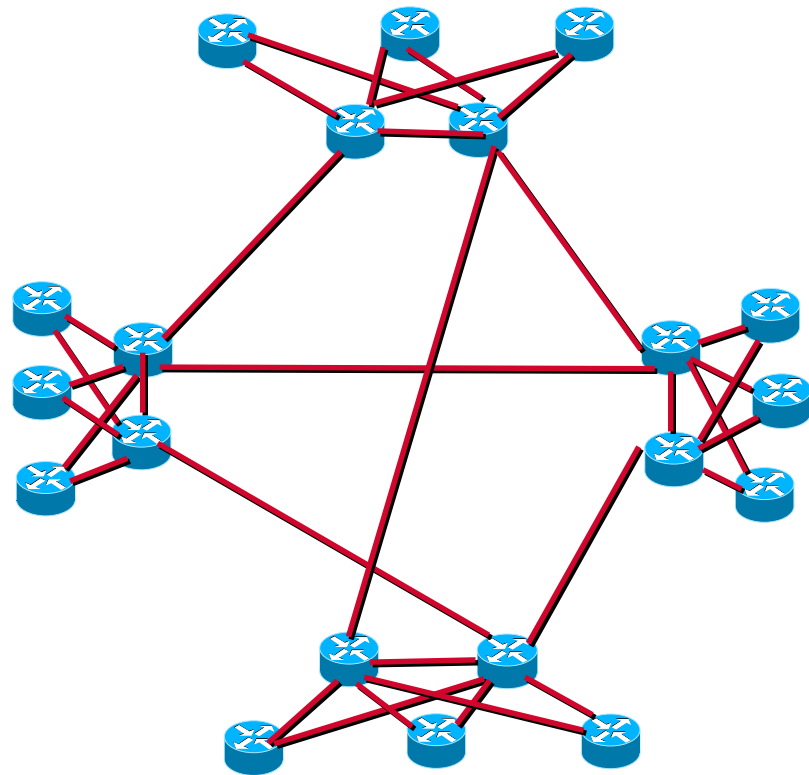
- **Interconnection for Management, Security, and Accounting services**

- ✓ Netflow Devices - FlowCollector
- ✓ Syslog collector for all network devices
- ✓ SNMP collector (PC Based UNIX)
- ✓ Security Auditing Tools (NetSonar)



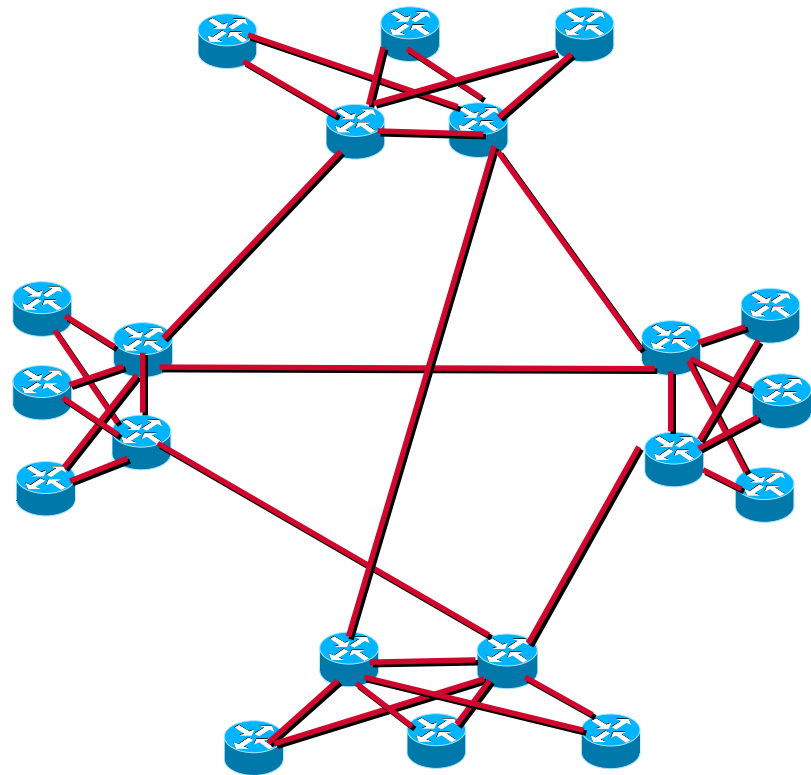
ISP routing Architectures - IP

- IGP = EIGRP, IS-IS, or OSPF
 - ✓ *almost always* IS-IS or OSPF
 - ✓ IS-IS, single level (usually L2)
 - ✓ OSPF, either single area or BB/POP areas
- BGP = all routers in full mesh
 - ✓ mesh accomplished with route reflectors, confederations, actual full mesh
- All routers have all routes, so services could go anywhere



ISP routing Architectures - IP+MPLS

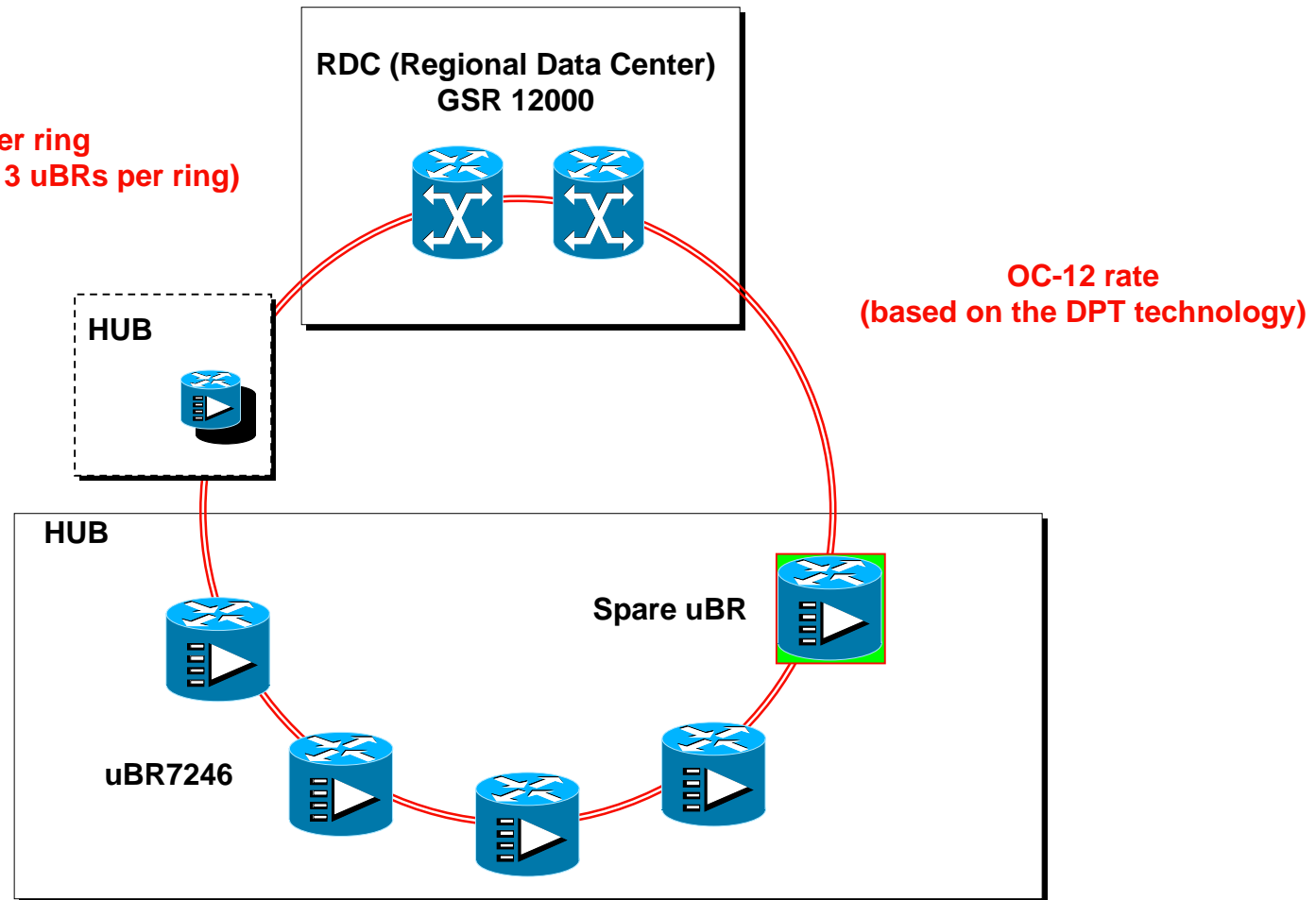
- IGP = EIGRP, IS-IS, or OSPF
 - ✓ *must* be IS-IS or OSPF to use MPLS TE
- BGP = only edge routers need full routes
 - ✓ full-mesh of edge routers using aforementioned mechanisms
 - ✓ packets are forwarded via LDP labels, not IP destination address
- Where to put your services?
 - ✓ cannot hang a cache service off of a router that doesn't have full routes!



Cable Internet Access

Hub Connectivity

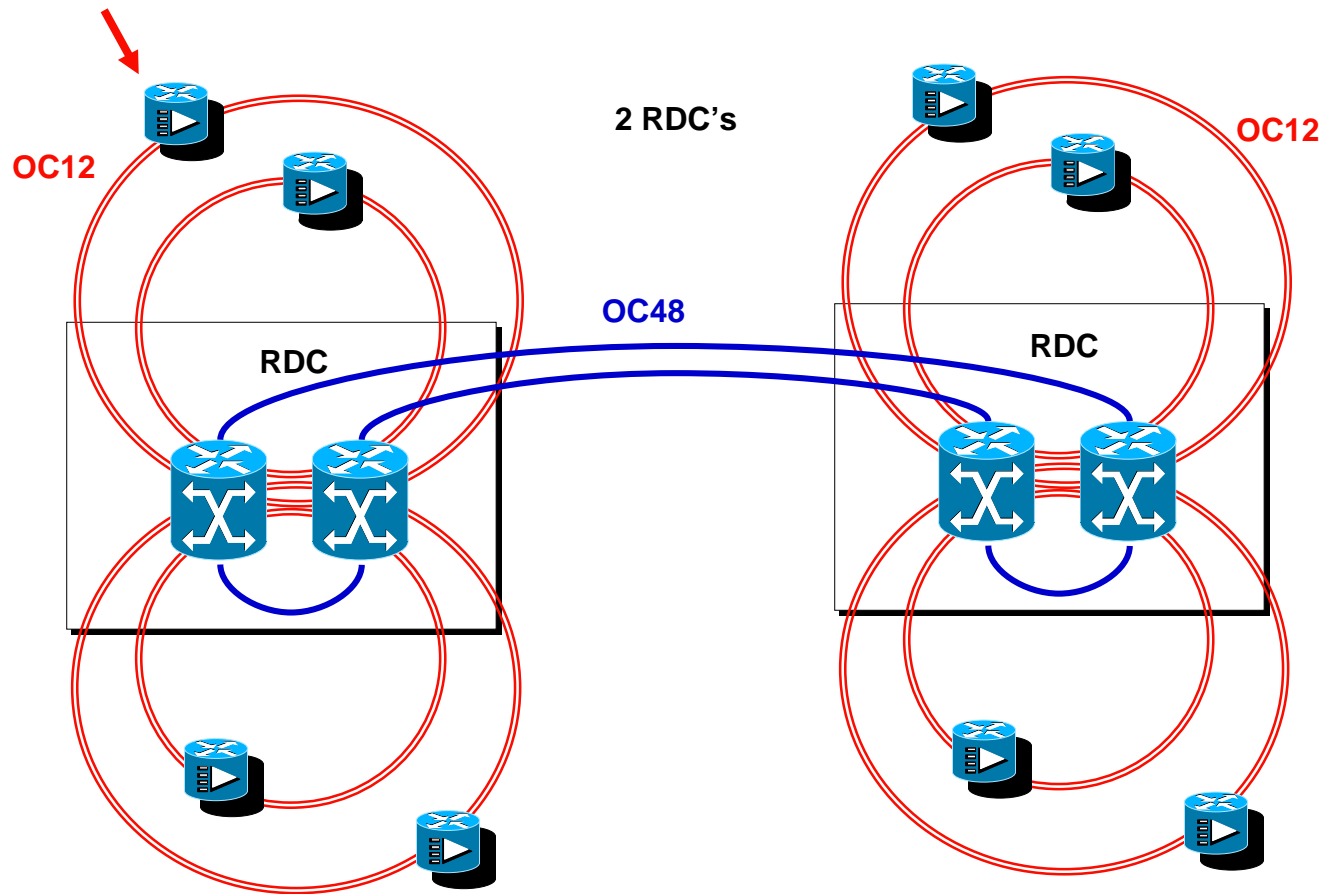
- 2 1/2 HUBS per ring
(either 12 or 13 uBRs per ring)



Cable Internet Access

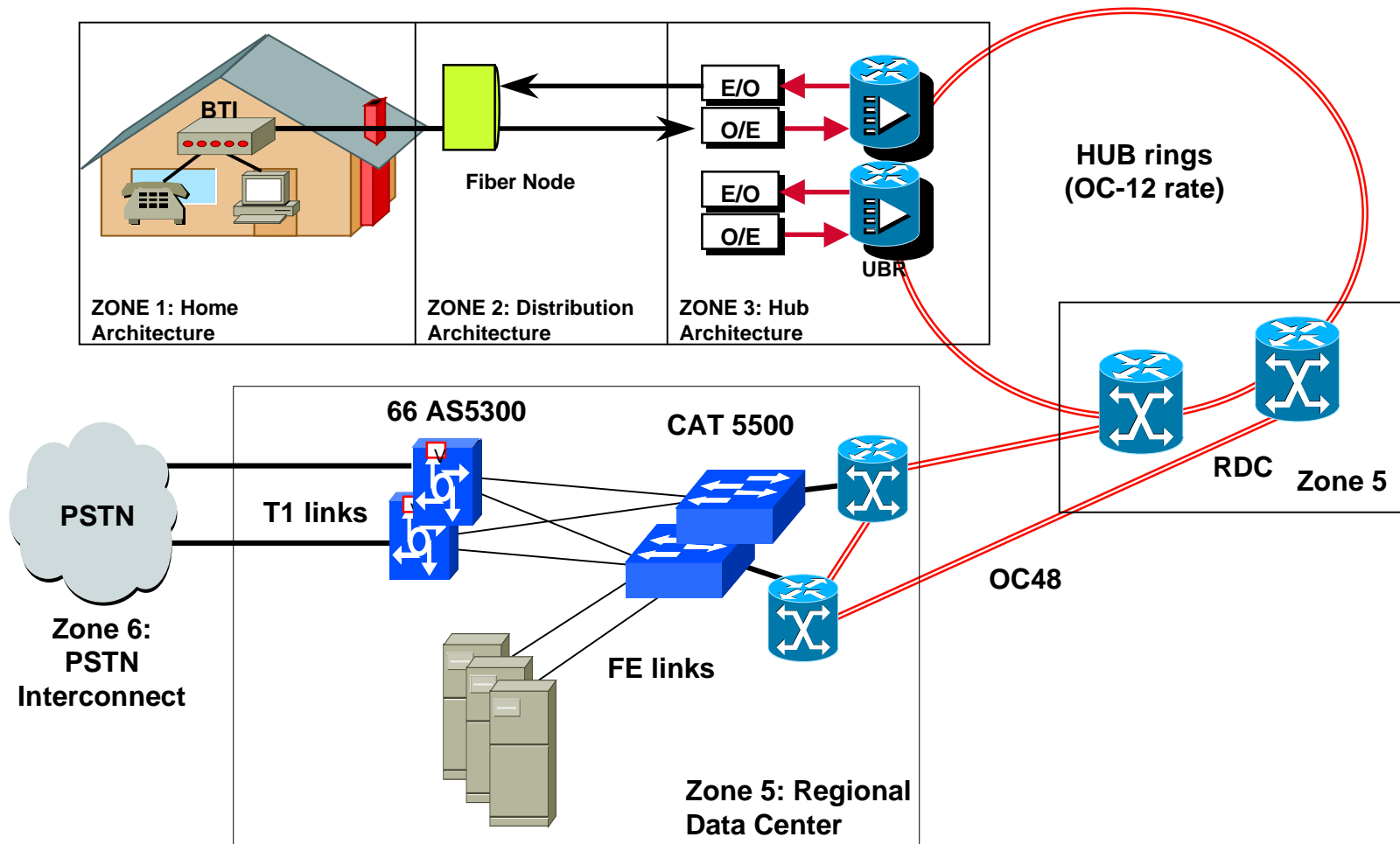
Hub Connectivity

12 or 13 uBR's per Ring (total of 50 uBR's per RDC)



Cable Internet Access

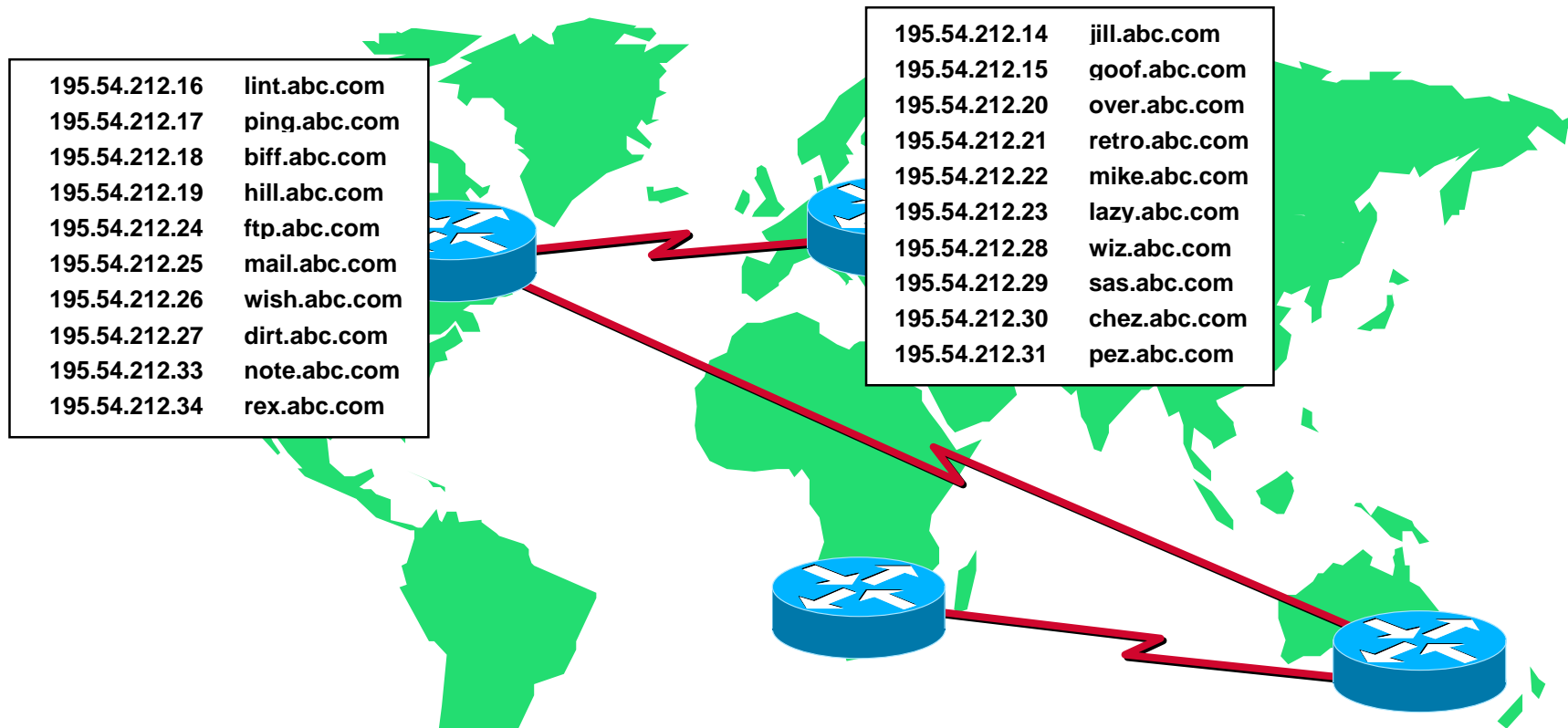
Sample Topology: RoadRunner





Addressing the Infrastructure

Addresses Not Scaling

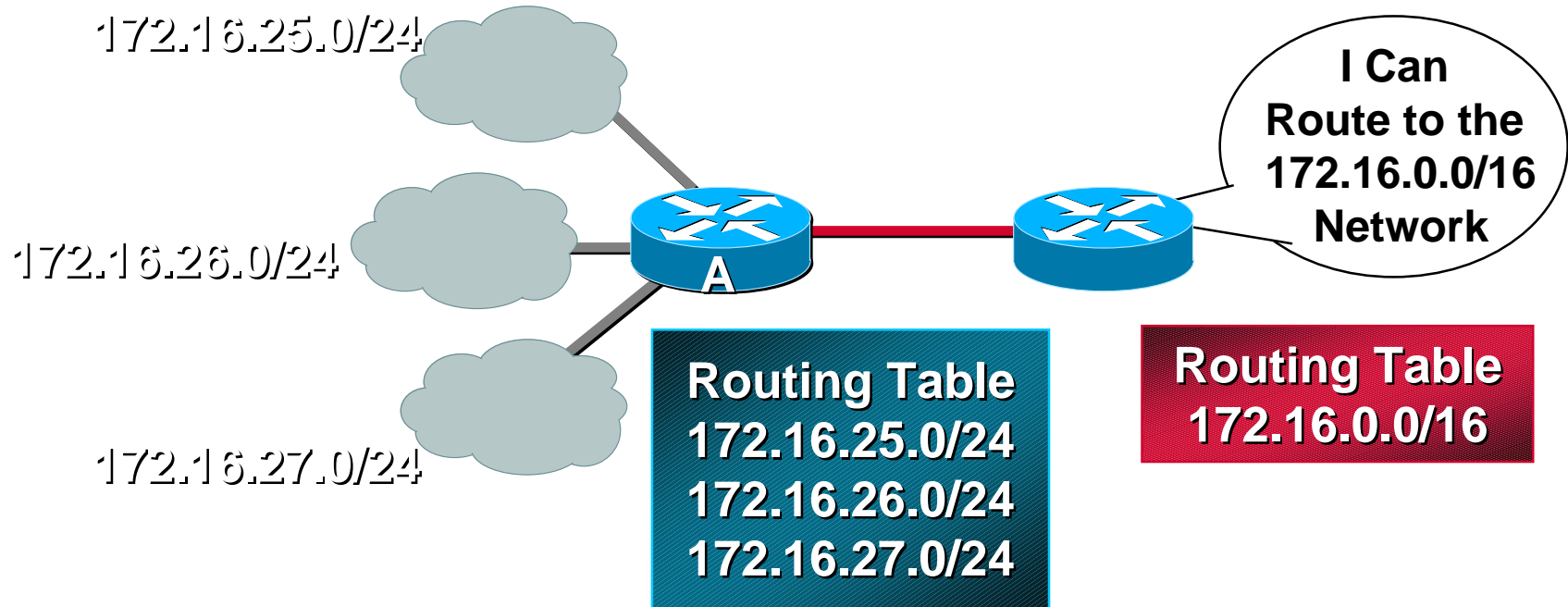


- Too many routes
- Wasted address space

Aggregating Addresses

- **Prefixes in backbone**
 - ✓ **73,000 Entries in Internet on 3/23/00**
 - ✓ **<http://www.employees.org/~tbates/>**
- **Scaling method for routing**

What Is Route Summarization?



- **Routing protocols can summarize addresses of several networks into one address**

Route Aggregation

Old Method:

202.14.64.0
202.14.65.0
202.14.66.0
⋮
202.14.96.0

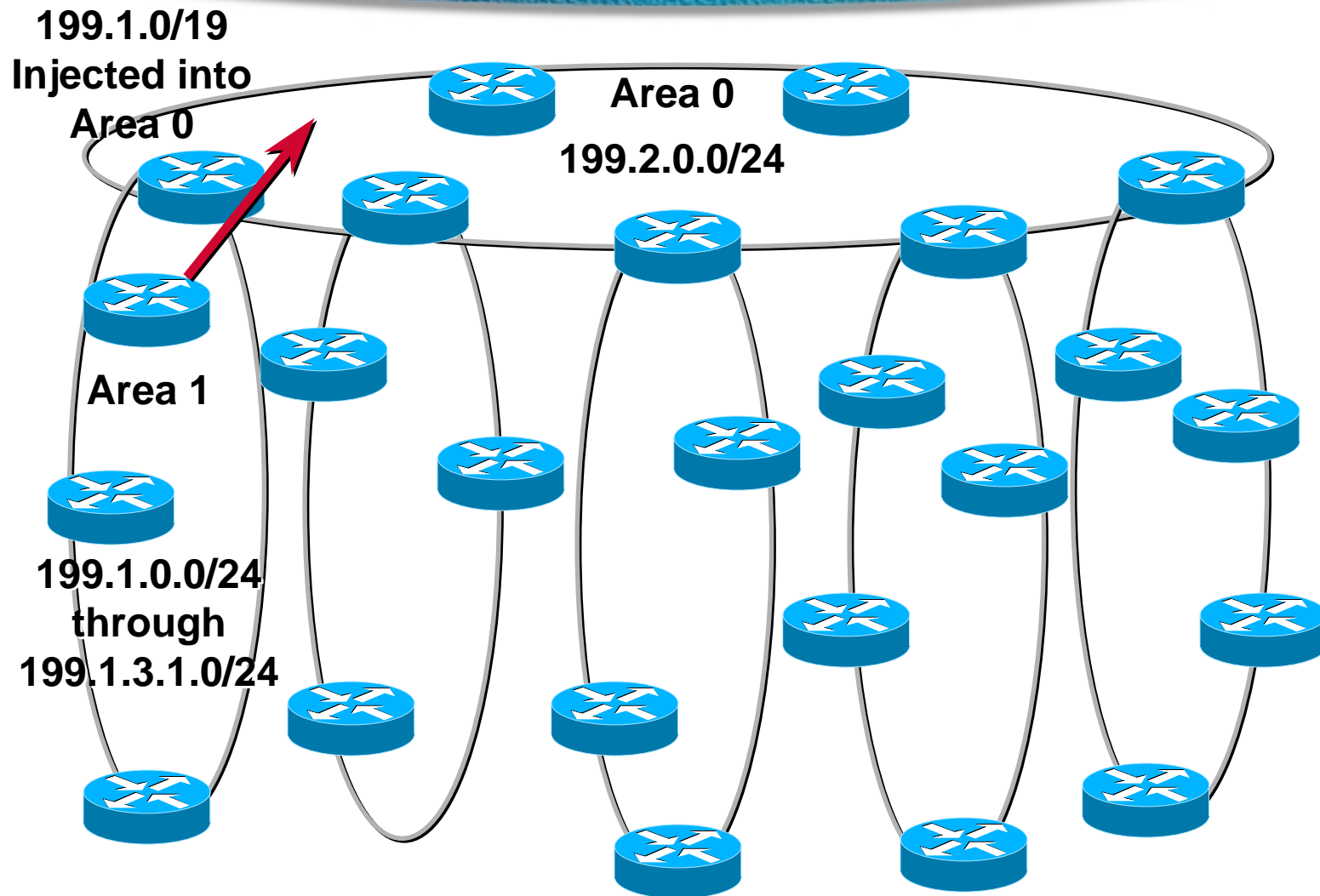
→ 32
Routes

New Method:

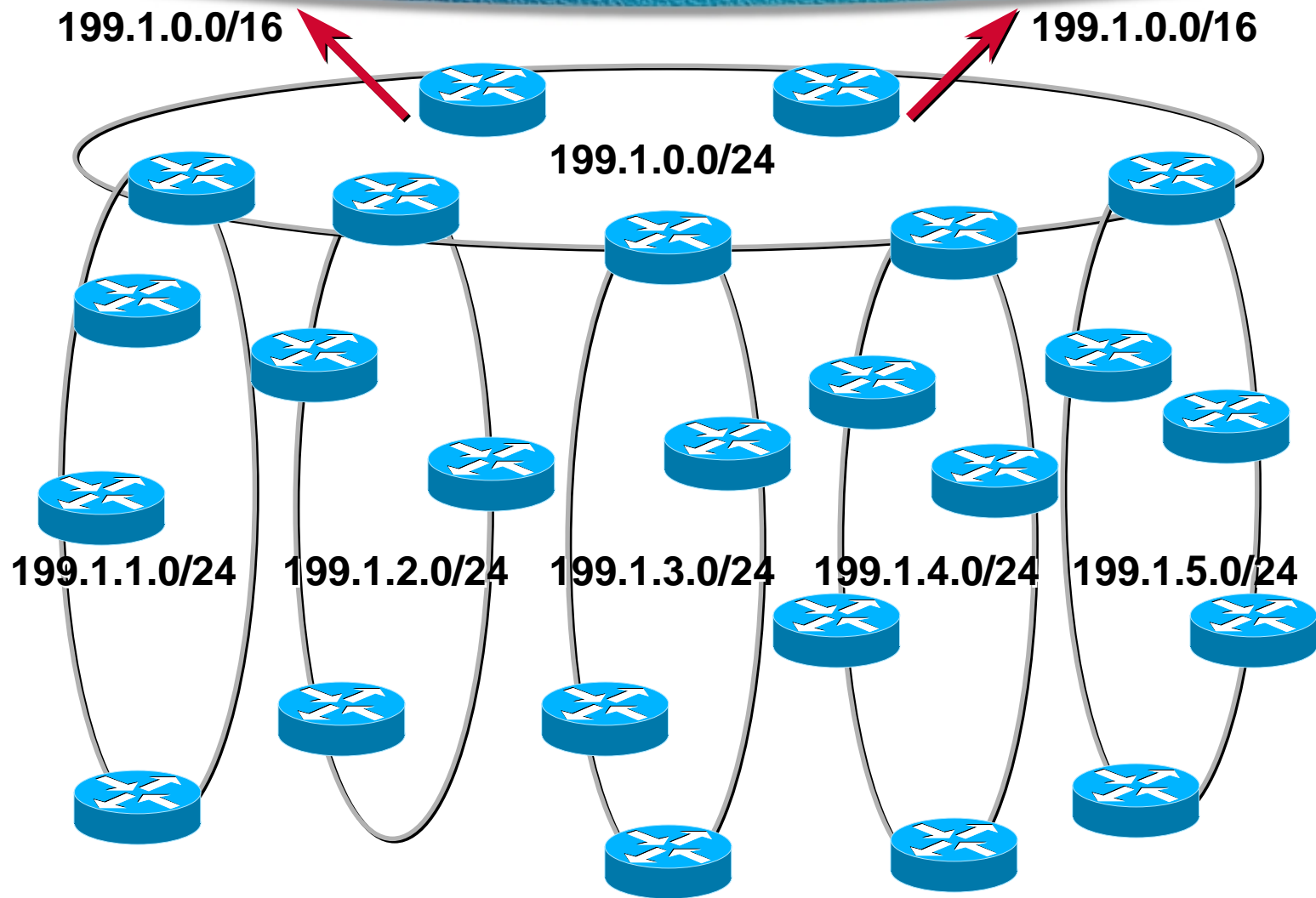
202.14.64.0/19 → 1 Route

- 131.108.0.0 /16 versus 255.255.0.0
- Summarizable blocks of subnets

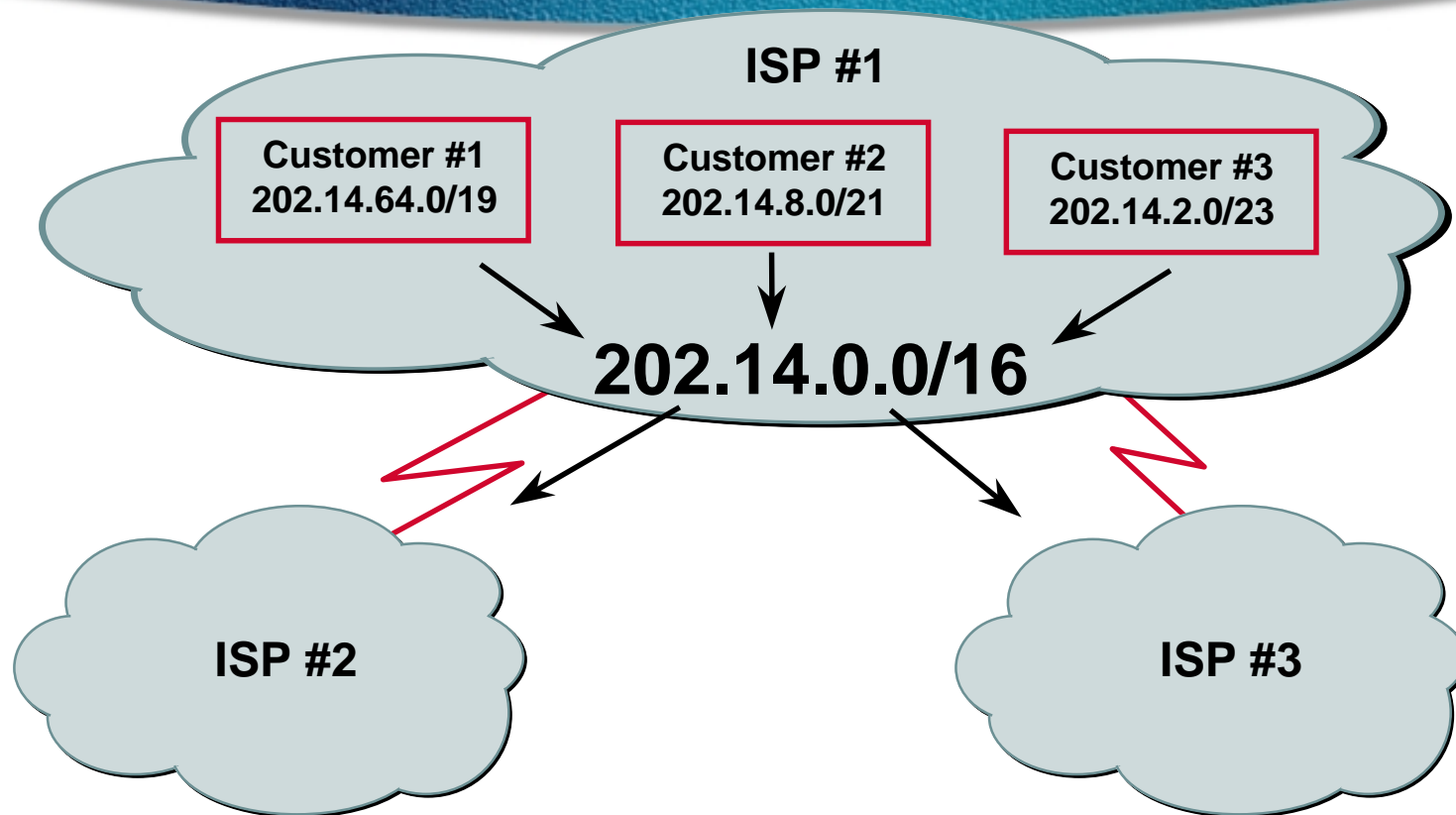
Intra-Domain Route Summarization



Inter-Domain Route Summarization



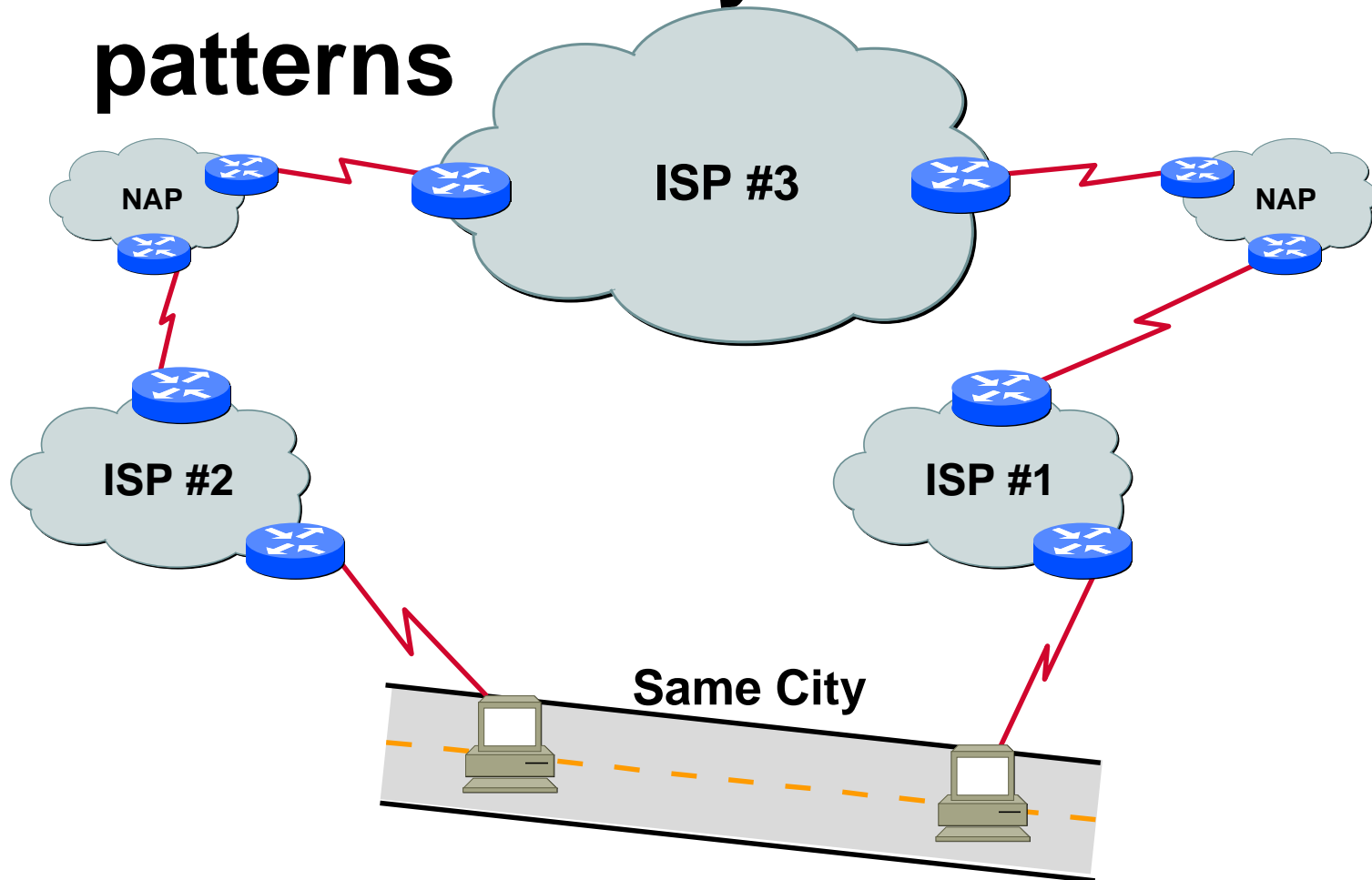
Addressing and ISPs



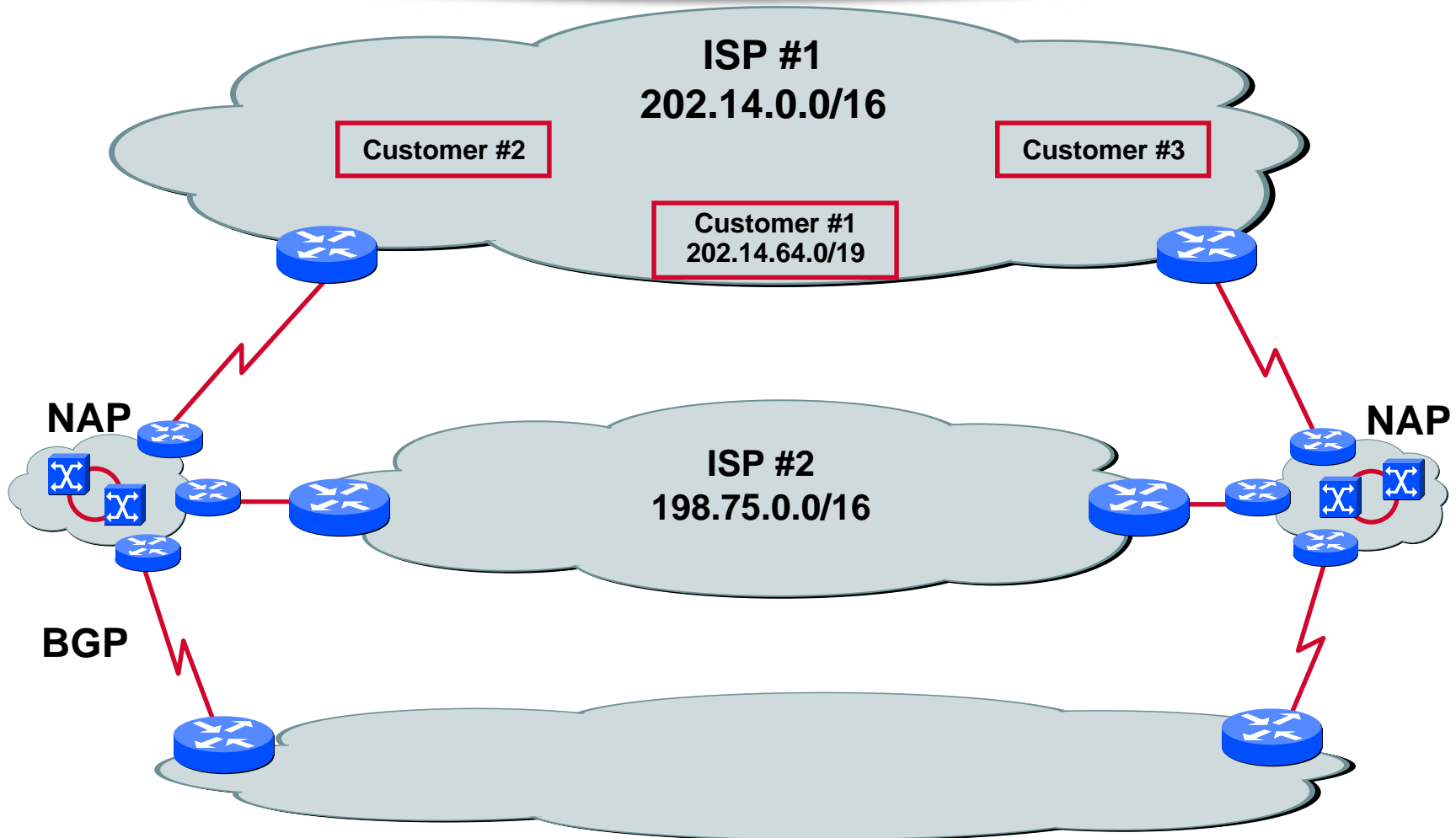
- **Smaller routes aggregated within ISP #1**

Border Gateways

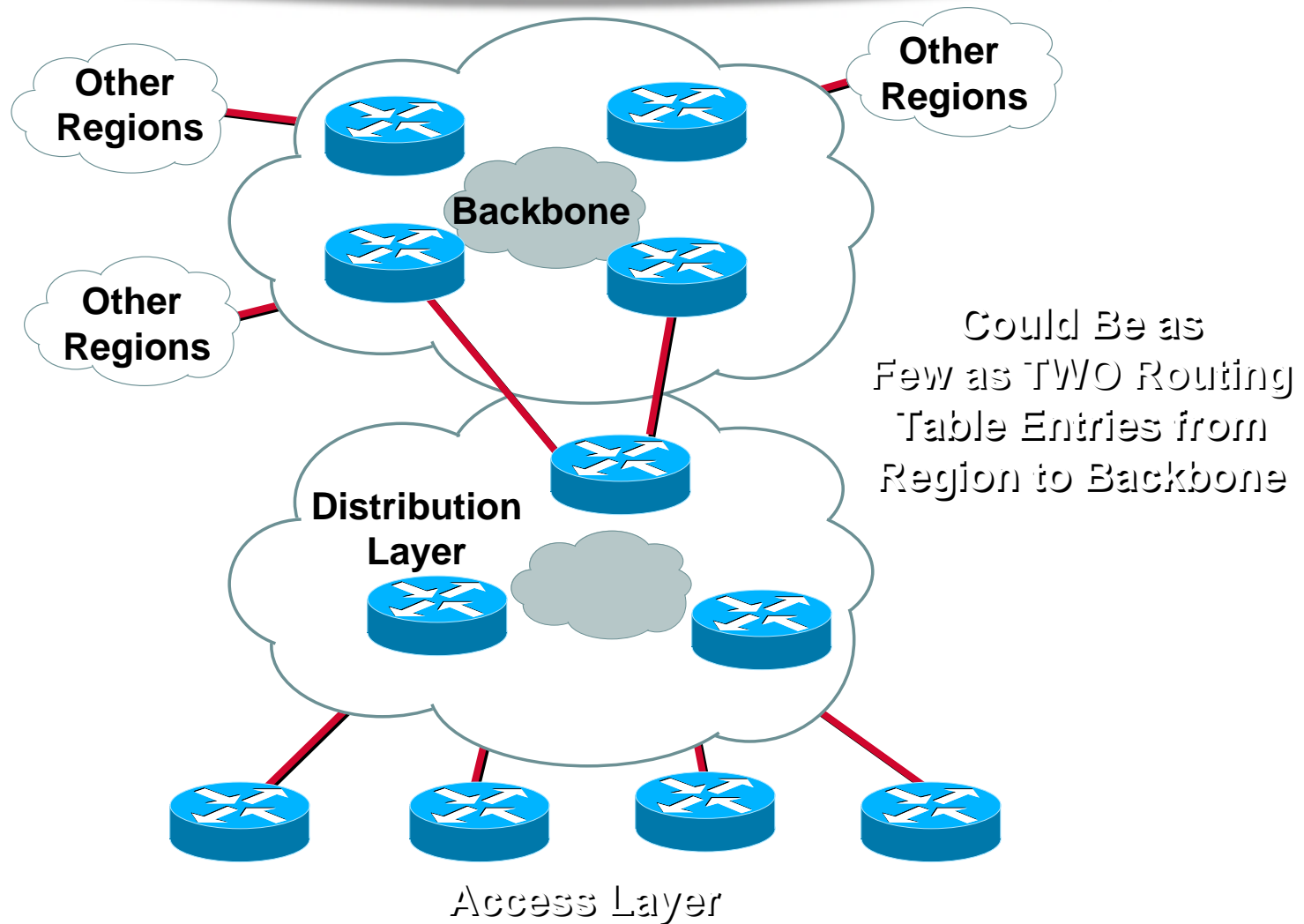
- **How it affects your traffic patterns**



Scaling the Internet: Addressing



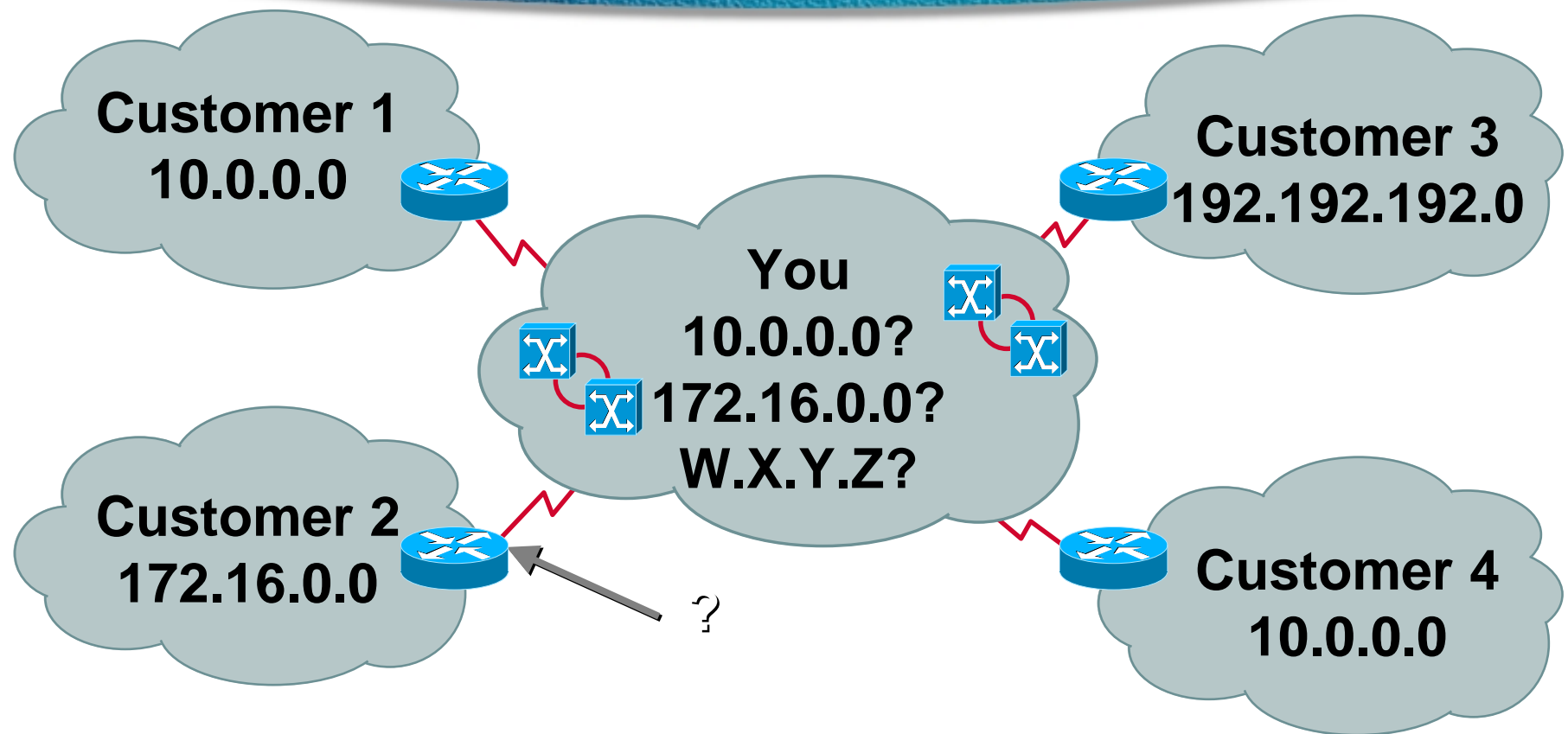
Minimum Routing Table Entries



Separate Infrastructure and Customer Addressing

- **Eases administration of policies**
- **Security and management**
- **Routing management**
- **Don't share network 10!**

Customer vs. Infrastructure



See Session #1306 *"Expanding Connectivity With NAT"*

Classless Routing Protocols

- **Supports Variable Length Subnet Masks (VLSM's) and non-contiguous subnetting**
- **OSPF, integrated IS-IS, EIGRP, RIPv2 and BGP**
- **Classful routing protocols are effectively deprecated**

Default Config for All Routers

- **All routers should have the following configuration commands for full CIDR routing:**
 - ✓ **ip classless**
 - ✓ **ip subnet-zero**
 - ✓ **router bgp XXXX**
 - ✓ **no auto-summary**



Adding Services to the Architecture Cause and Effect

Services?

How many *Services*?

Most network services are applied at the edge!

Edge (one-time) services

- Voice over IP
- MPLS VPNs
- CDNs
- VPDNs
- Managed services
- Dial—DSL—cable

Per-hop services

- MPLS packet forwarding
- DiffServ, other QoS
- Multicast Services

Ask the Right Questions

- **What is the value of the service?**
 - ✓ **Technical merit**
 - ✓ **Cost savings**
 - ✓ **Marketecture**
- **What is the cost of the service?**
 - ✓ **Equipment?**
 - ✓ **Training people to support it?**
 - ✓ **Network buildouts/topology changes?**



Impact of Services on the Network

Who Knows?

- **What will be the impact on existing traffic loads/patterns?**
- **Can the network deliver the performance that your customers/applications desire? delay? jitter (delay variation)?**
- **Make sure to add capacity as you add services - bandwidth is a must.**

Deployment of New Services

- **Is more of a business decision**
- **The technical aspect is to ensure continued network performance—scalability and stability**
- **Try to keep services within your AS**
 - ✓ **end2end control**
 - ✓ **less likelihood of failure/flaps**

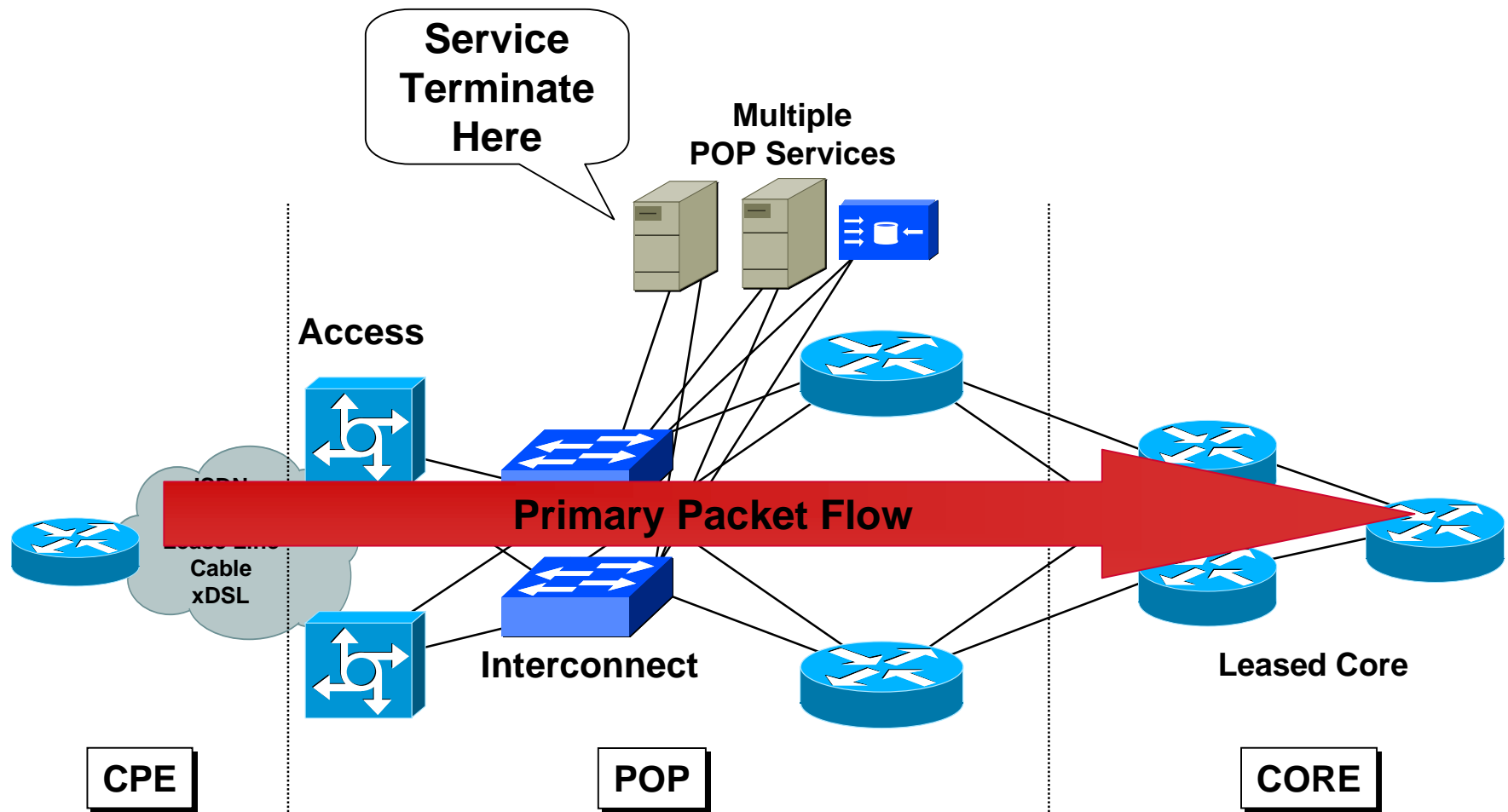
Deploying New Services

- **Don't feed the hype fire**
- **Look *before* you leap!**
- **Don't deploy new technologies and services just for the sake of it; have valid business and technical reasons**

Deploying New Services

- Usually a Service requires a TCP/UDP termination (i.e. TCP's three way handshake)
- Termination should happen out side of the *primary flow path*
- Otherwise, the network is then designed around the single service.

Deploying New Services

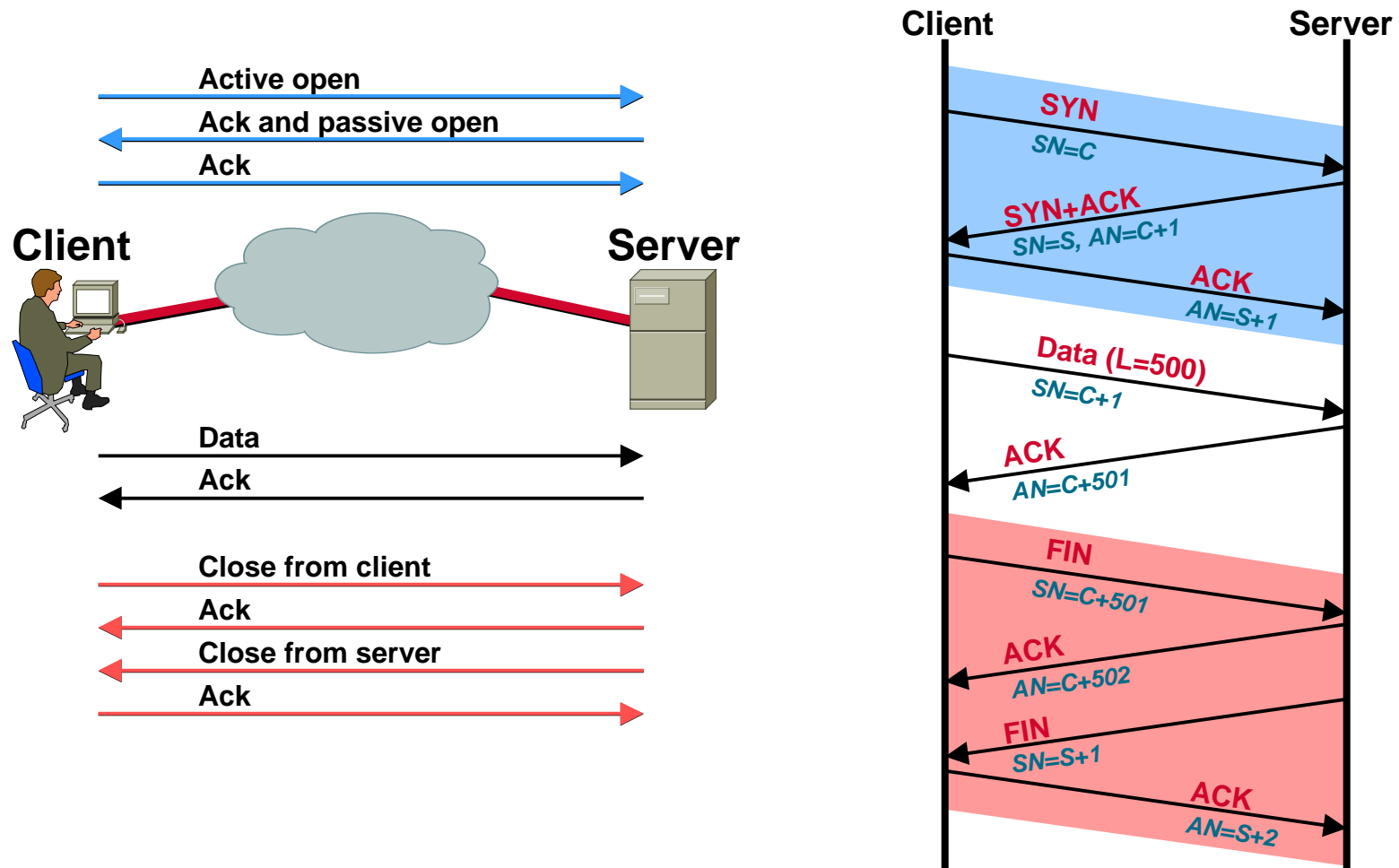




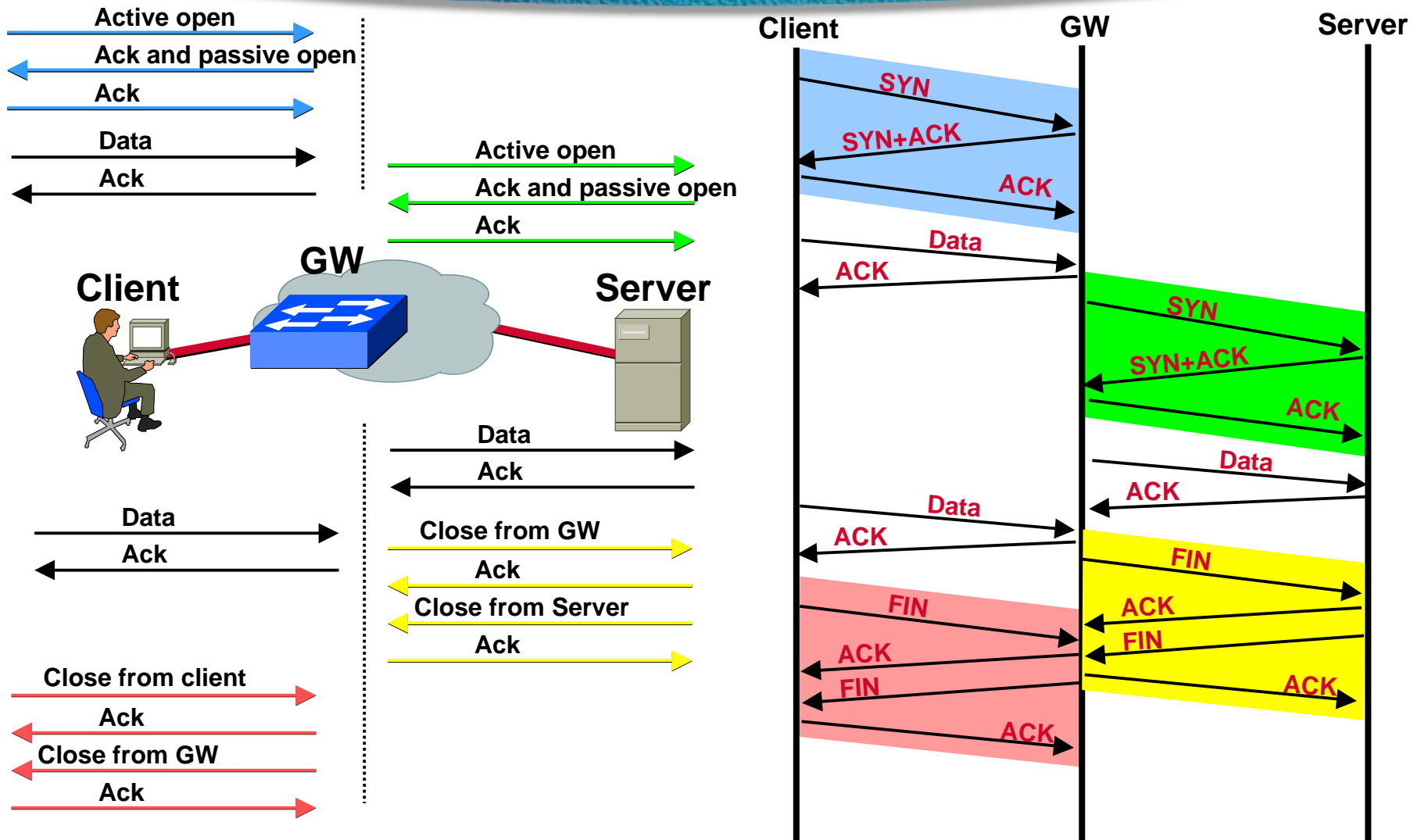
TCP/UDP Termination in The POP

***Factors Effecting any product
design***

Connections Establishment and Termination

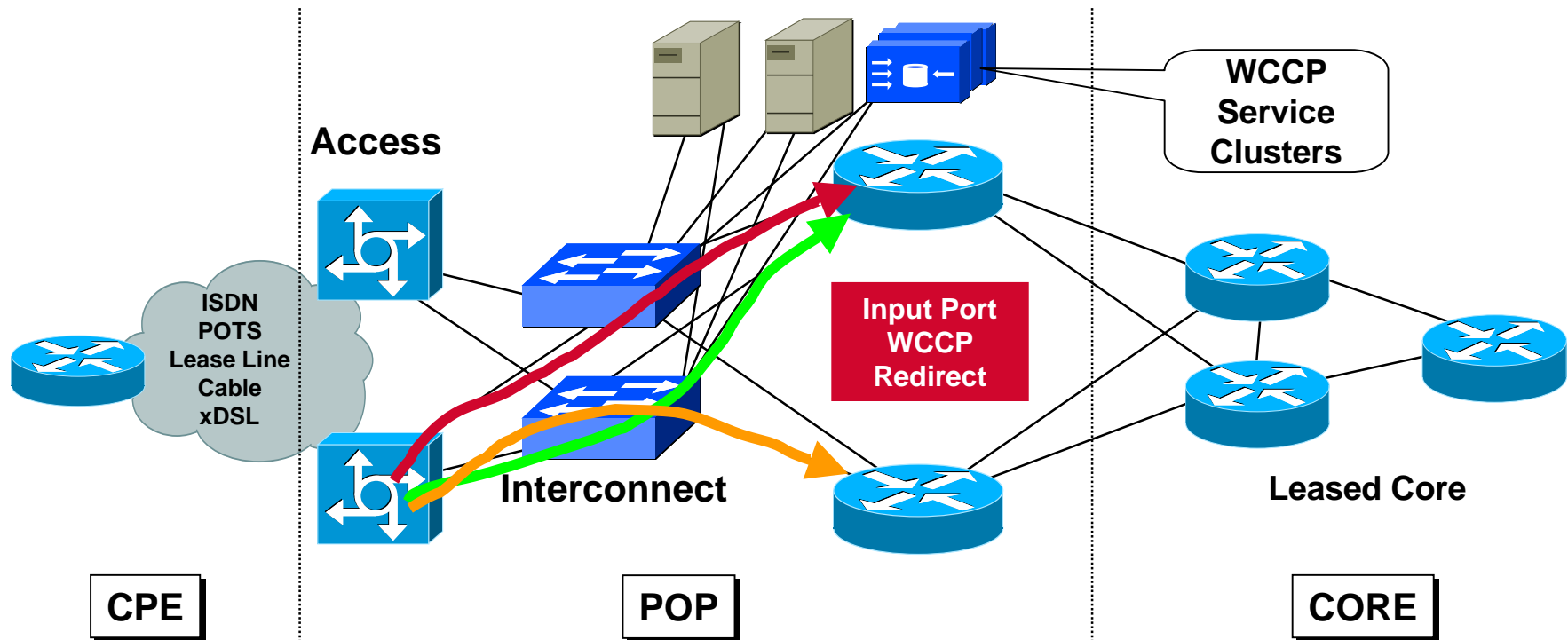


Termination with GW or L7 Switch



Equal Path Load Balancing through the POP

- Work with the multi-level L2/L3 redundancy of the ISP POP. Equal paths in the IGP + CEF leads packet asymmetry.

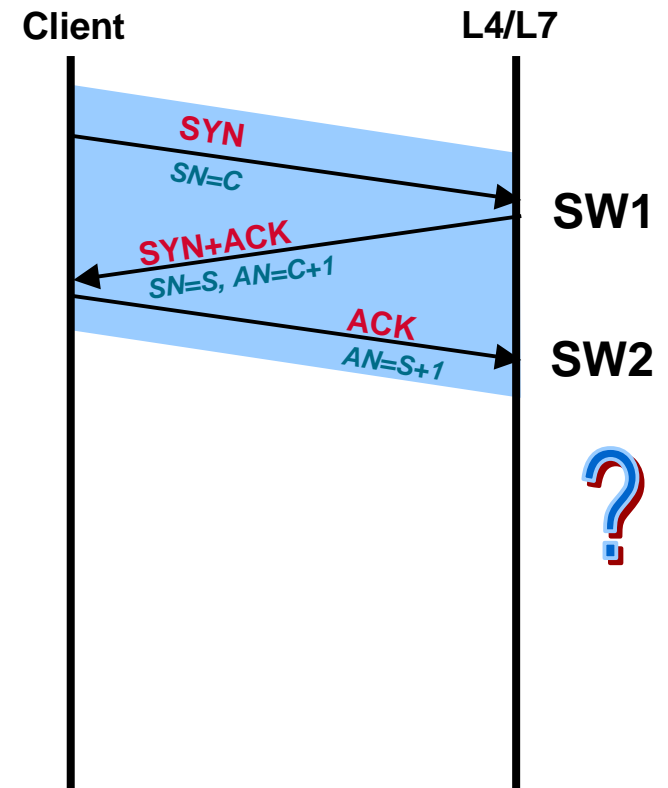
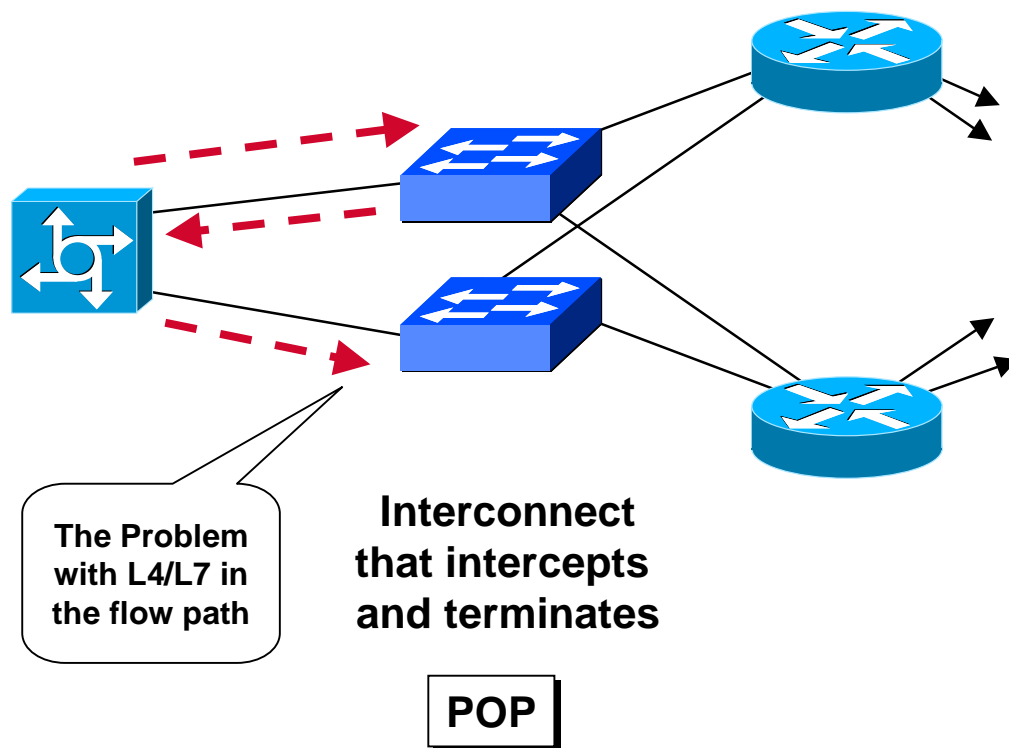


The Major Problem

- **Cannot assume that the NAS boxes will comply to a L4/L7 switches requirements of forced symmetry**
 - ✓ **Per packet round robin - default for most NAS gear. Most Cisco NAS boxes left in this mode (no CEF).**
 - ✓ **Per destination - If full routes - OK. Equal cost defaults in the POP, then back to per packet.**
 - ✓ **Hash of source/destination. Only mode that will work for L4/L7 switches.**

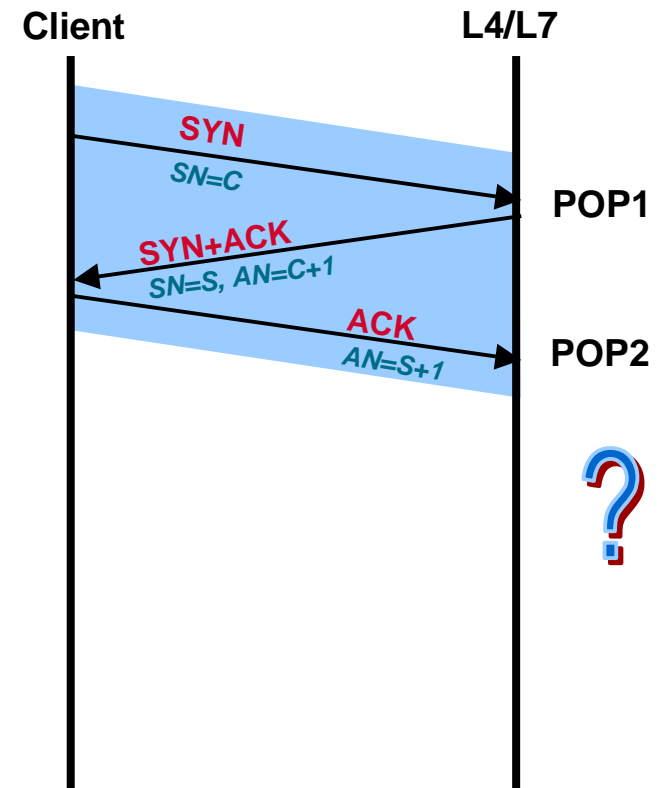
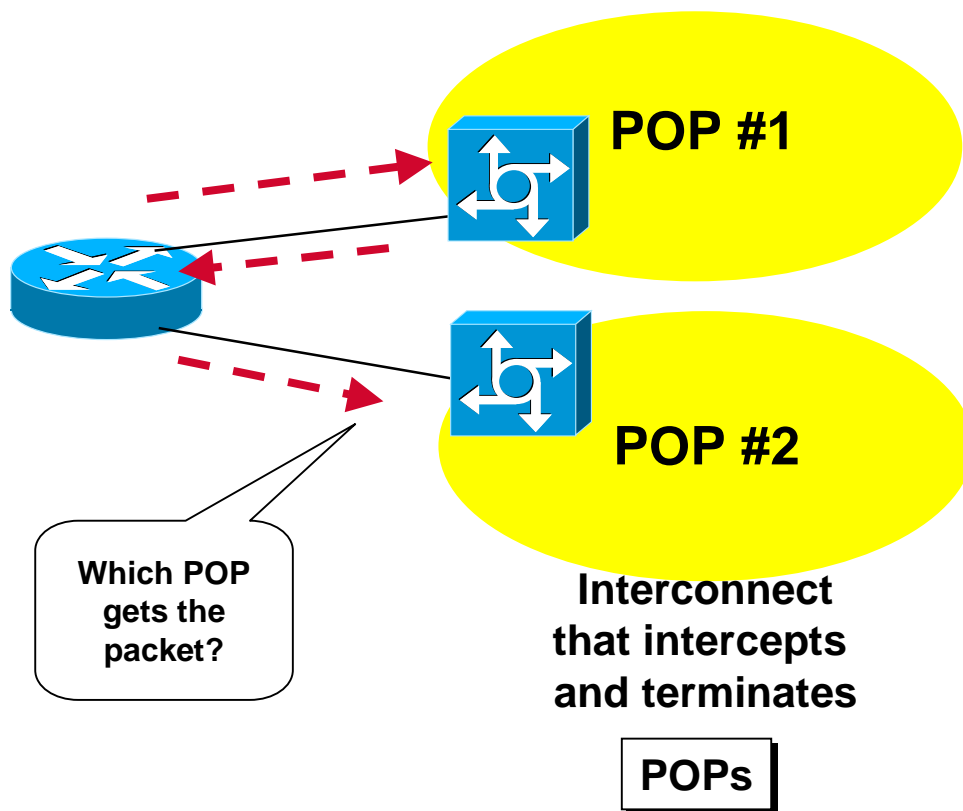
How the Packets Flow through the POP

Worse Case Scenario - Per Packet Load Balancing with four equal cost paths to default routes on the POP GW



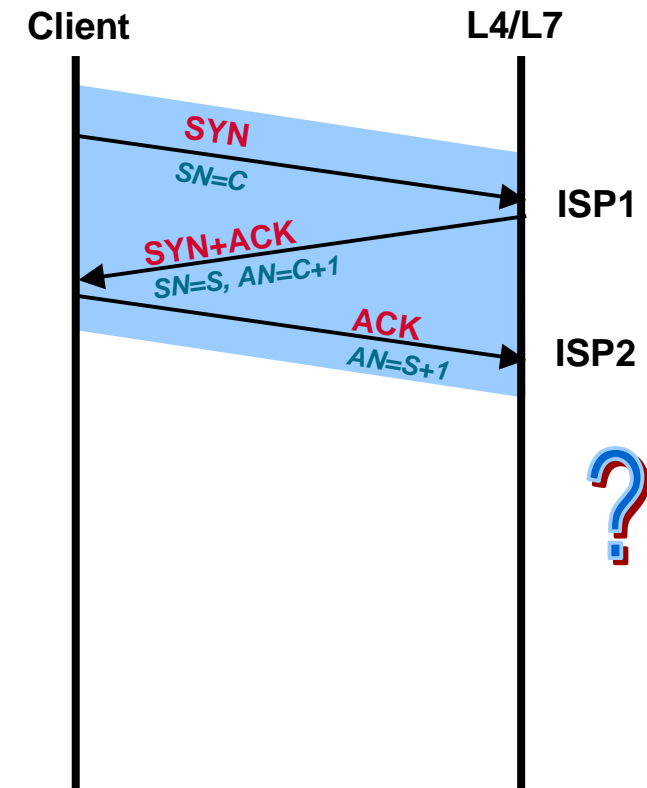
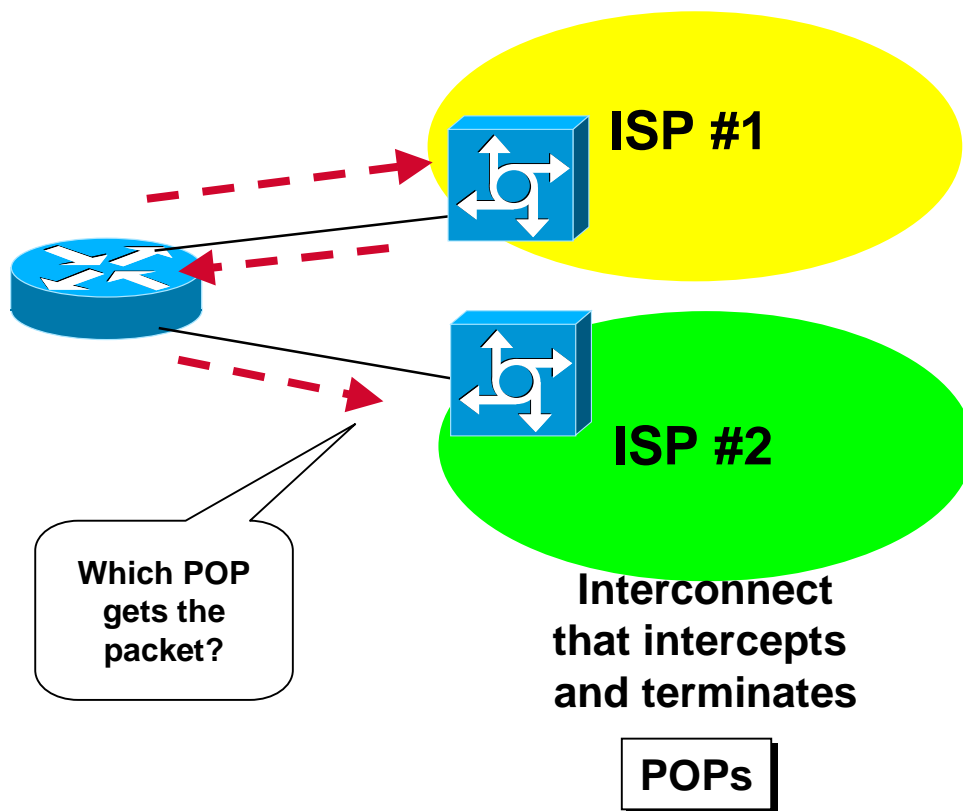
How the Packets Flow through the POP

An Even Worse Case - Per Packet Load Balancing from an multihomed enterprise with two equal cost paths to the same ISP



How the Packets Flow through the POP

Absolute Worse Case - Per Packet Load Balancing from an multihomed enterprise with two equal cost paths to the two or more ISPs



Make No Assumptions!

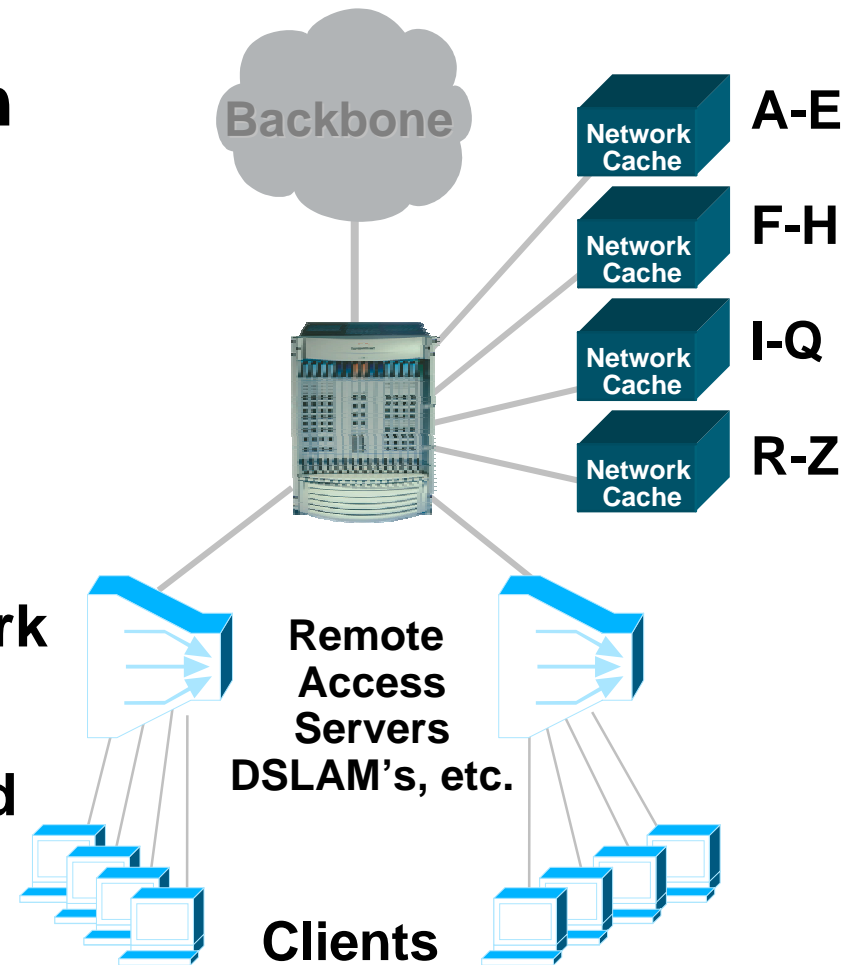
- **Cascading switching paths**
 - ✓ Optimum -> fast switching -> process switching
 - ✓ CEF replaces optimum switching
 - ✓ CEF -> fast switching -> process switching
 - ✓ DCEF -> CEF -> fast switching -> process switching
- **All of this effects the equal cost load balancing technique.**
- **Certain features drop out of CEF into fast or process switching**

What to do?

- **Cannot trust the equal cost load balancing to keep the traffic flow symmetrical.**
- **That leaves two solutions:**
 - ✓ **Force symmetry with one POP interconnect device**
 - ✓ **Separate the identification and redirection functions of transparent interception from the termination function.**

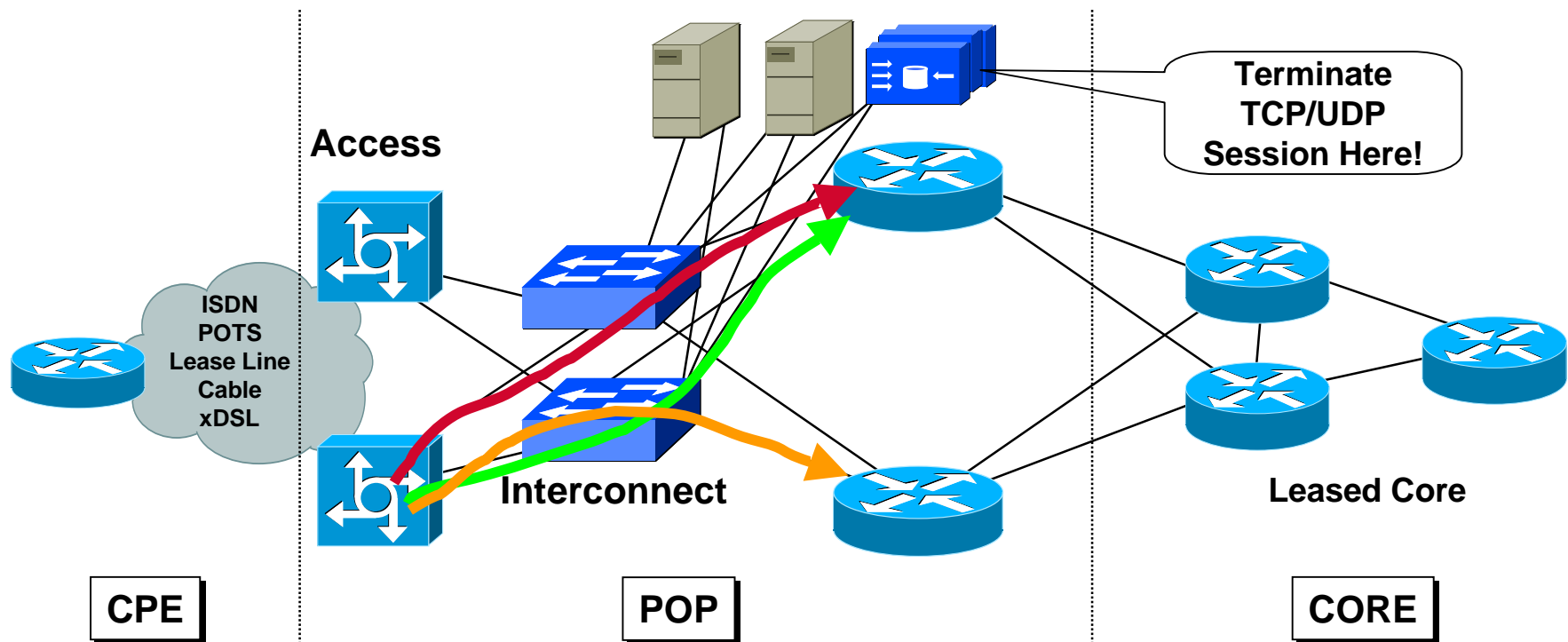
Force Symmetry in the POP

- Use only one L4/L7 switch in the POP
 - ✓ No POP interconnect redundancy
 - ✓ Only Ethernet Topologies - no SRP/DRP or ATM solutions
 - ✓ No five “9s” end-to-end network reliability
 - ✓ Will not work with multi-homed customers



Separate the Transparent Interception Functions

Do a L3/L4 redirect first - then terminate! This works with multi-level L2/L3 redundancy and packet asymmetry in the ISP POP - but not multi-homed customers.



Web Cache Communication Protocol (WCCP)

- **Content Routing Technology first introduced in 1997**
- **Provides mechanism to redirect traffic flows [originally caches] in real-time**
- **Has in-built load-balancing mechanism, scaling, fault tolerance, and service-assurance (failsafe) mechanisms**

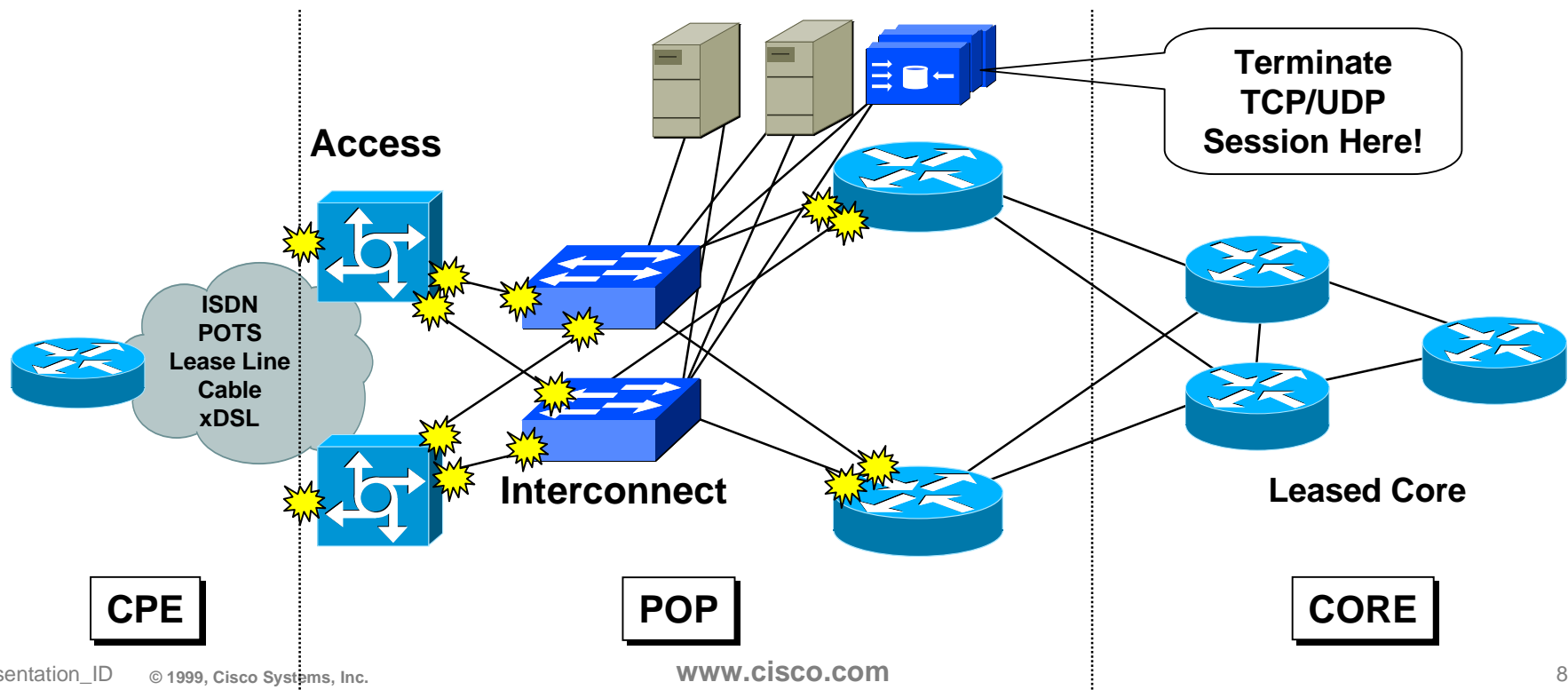
Web Cache Communication Protocol (WCCP)

- **WCCPv1/WCCPv2 implemented by several vendors:**
 - ✓ **Inktomi, NetApps, CacheFlow, Novell, Infolibra - original licensees**
 - ✓ **Squid has a version with WCCPv1 w/ WCCPv2 coming (when Lincoln has time)**



Separate the Transparent Interception Functions

- Provide the ISP with Flexibility on the point of redirection. Do not force an architecture on the customer.





Mindfulness

*Things that will make or break
your business.*

Managing the Network

- **Documentation (for real)**
- **Plan ahead**
- **Know your limits and the limits of your network**
- **Traffic Engineering 101**
- **Service Contract - PICA**

Scaling Operations

- **Few operators allowed to configure backbone infrastructure**
- **Define clear processes/automate customer provisioning**
- **Documentation, simplicity, and repetition**

Empowering People

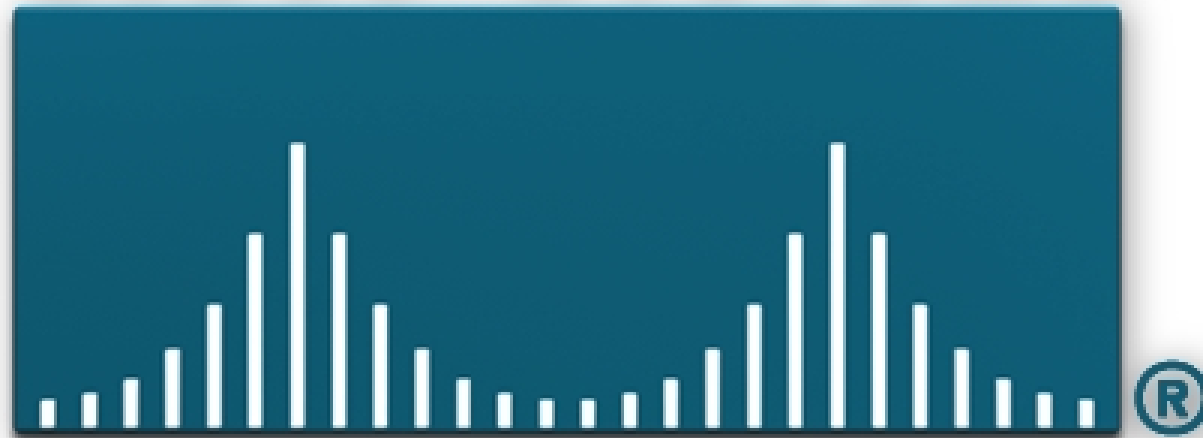
- **People** —not bandwidth, content, or applications—are **THE** most critical factor
- **Raise skills**
- **Provide Tools**



Labs and Testing

- **Cisco IOS™ is the OS for your Network**
- **Test new releases with existing applications, like a new server OS release**
- **Test new applications on a test network before deployment, like a test server**
- **An ounce of preparation is worth...**

CISCO SYSTEMS



EMPOWERING THE INTERNET GENERATIONSM



Addendum - Equal Cost Forwarding

How IP Forwarding works with equal cost paths

Equal Cost Forwarding

- Router discovers multiple paths to a destination via a routing protocol.
- The forward table is updated with multiple entries to that destination:

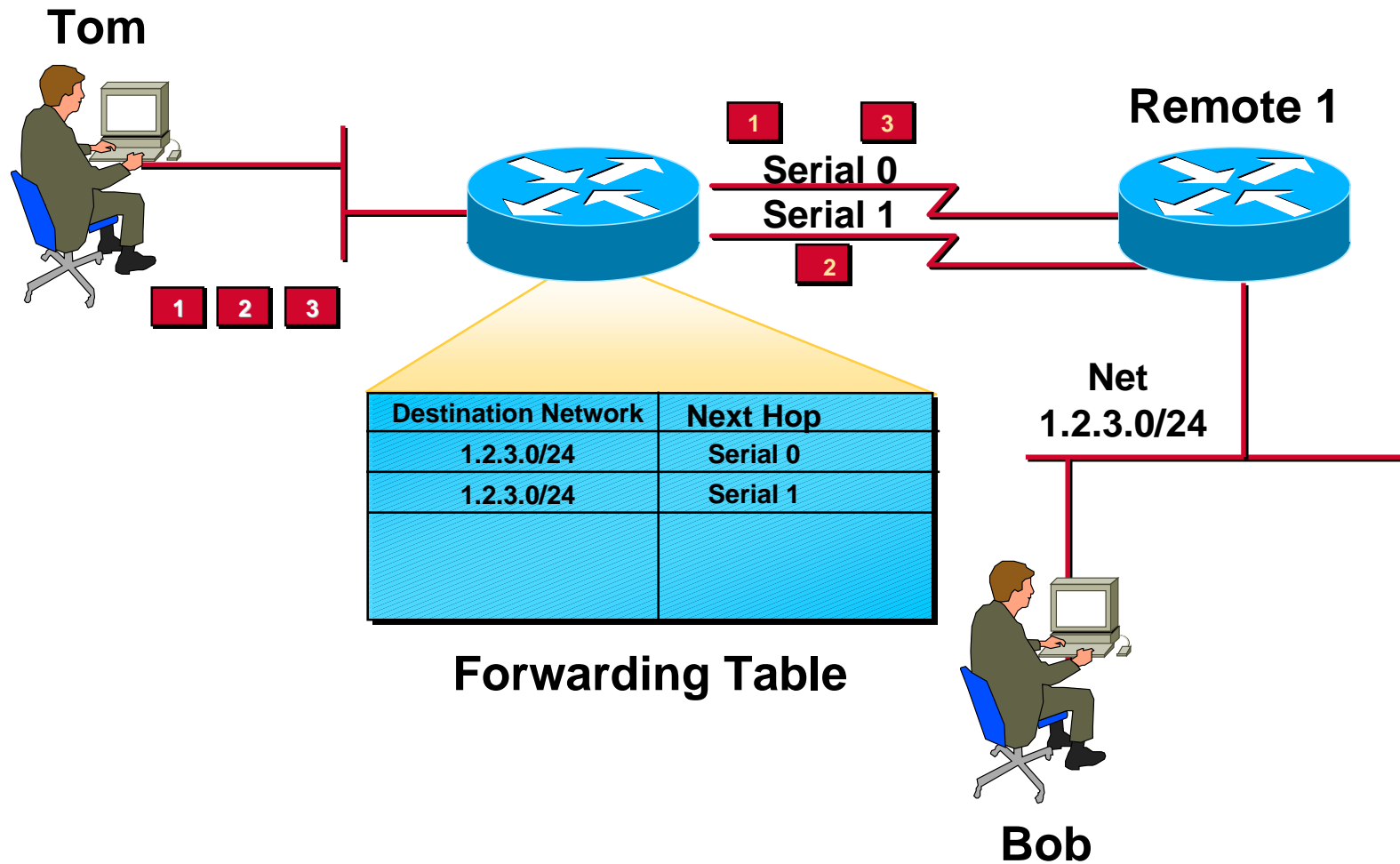
```
router> show ip route  
[...]  
I      192.168.25.0/24 [115/10] via 192.168.24.6  
                                [115/10] via 192.168.24.10  
                                [115/10] via 192.168.24.14  
[...]
```

- If metric are equal, the router will forward the packets using one of several load balancing techniques

Types of Equal Cost Forwarding

- **Per packet round robin**
- **Per destination**
- **Hash of source/destination**

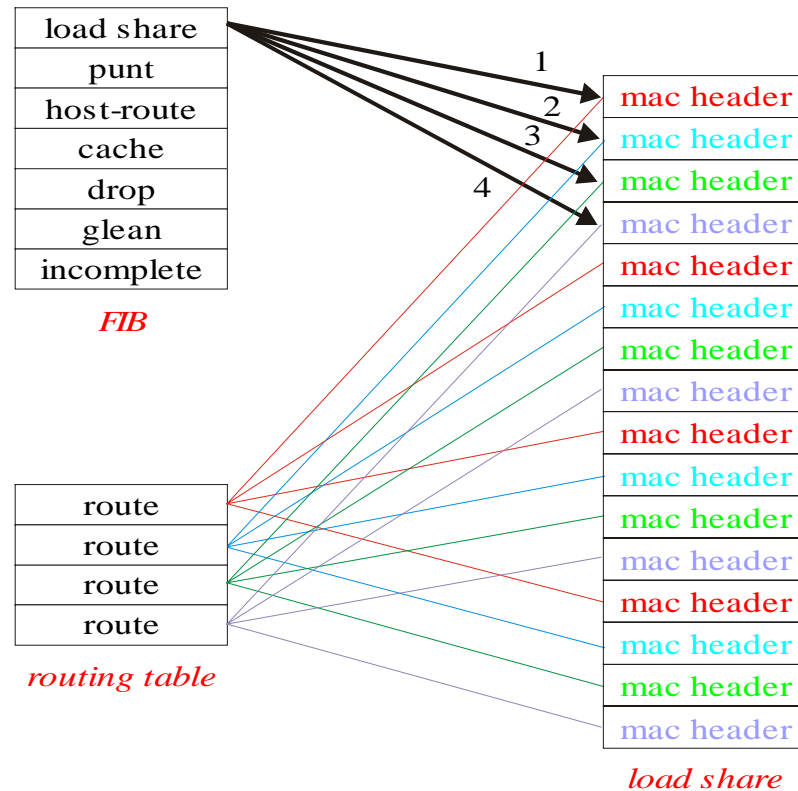
Per Packet Round Robin



Per Packet Load Balancing

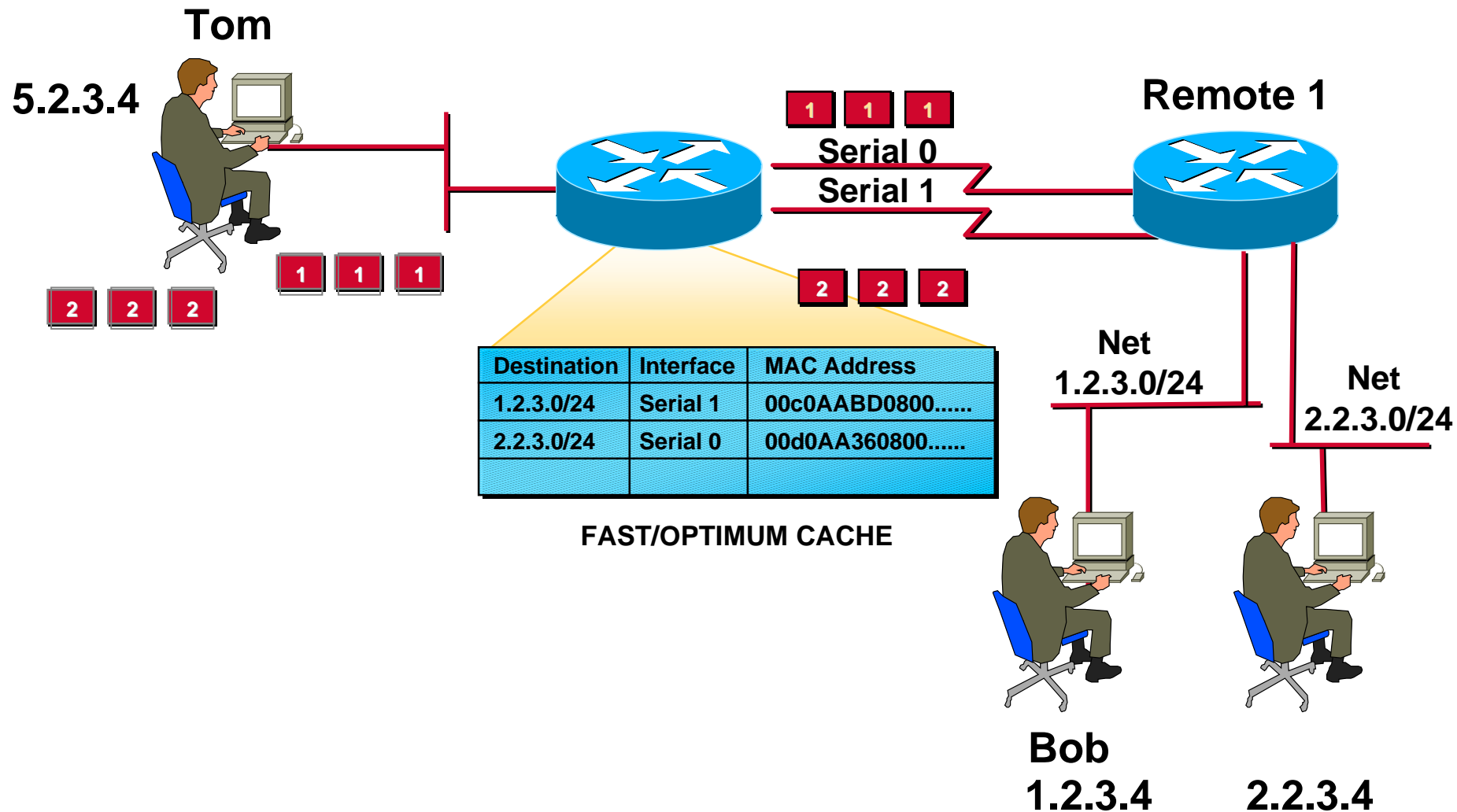
- **Supported on all routers.**
- **Processor intensive - CPU hit.**
- **Prone to out of order packets in some situations.**
- **CEF has per-packet load balancing - but with out the CPU hit on the processor.**

CEF's Per Packet Load Balancing



Round robin between 16 buckets.

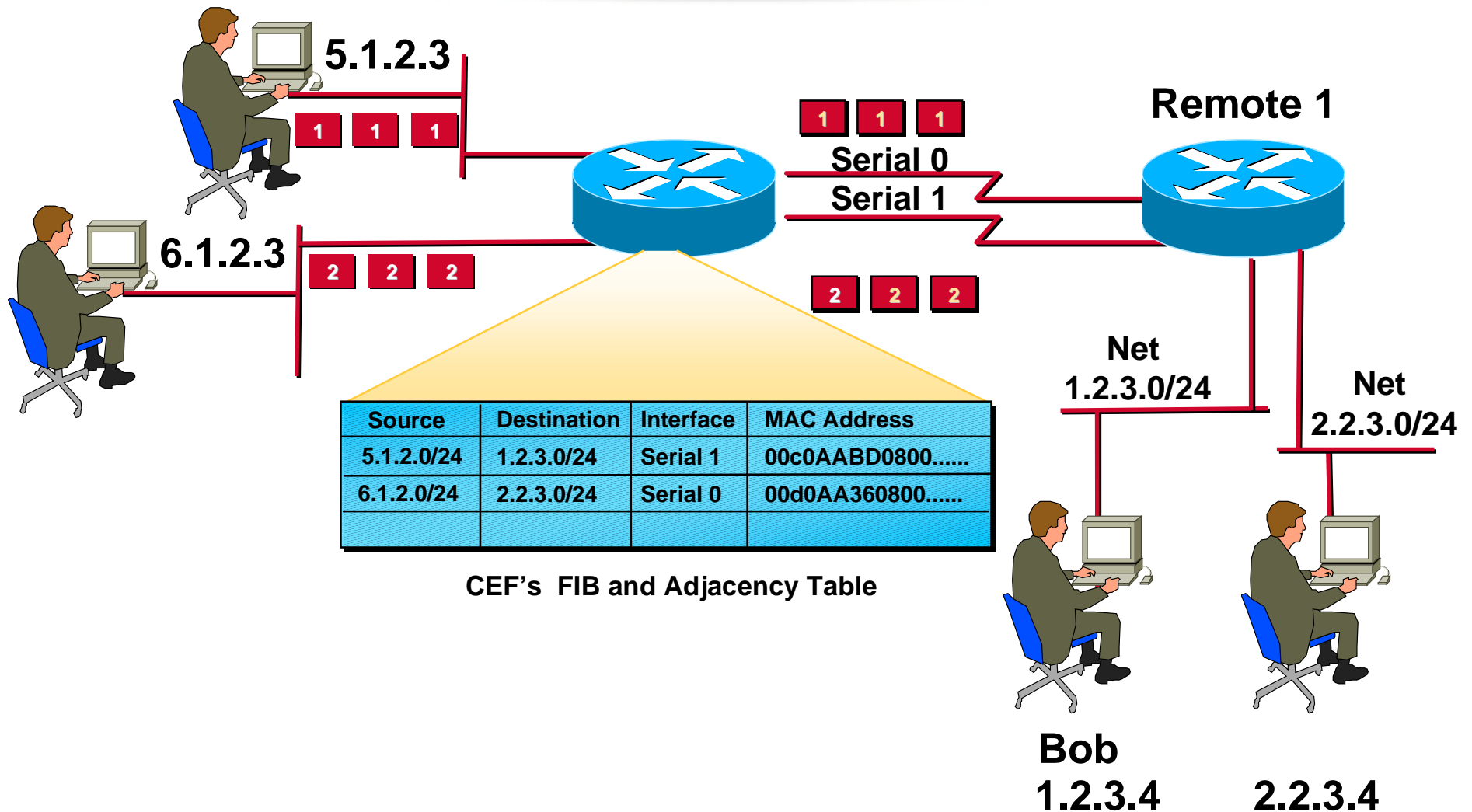
Per-Destination



Per Destination Load Balancing

- Based on the *cached* destination
- Supported on cached based switching - fast switching, optimum switching, and optimum/flow switching
- Eliminated the out of order packet issues, but creates the issue of traffic polarization.

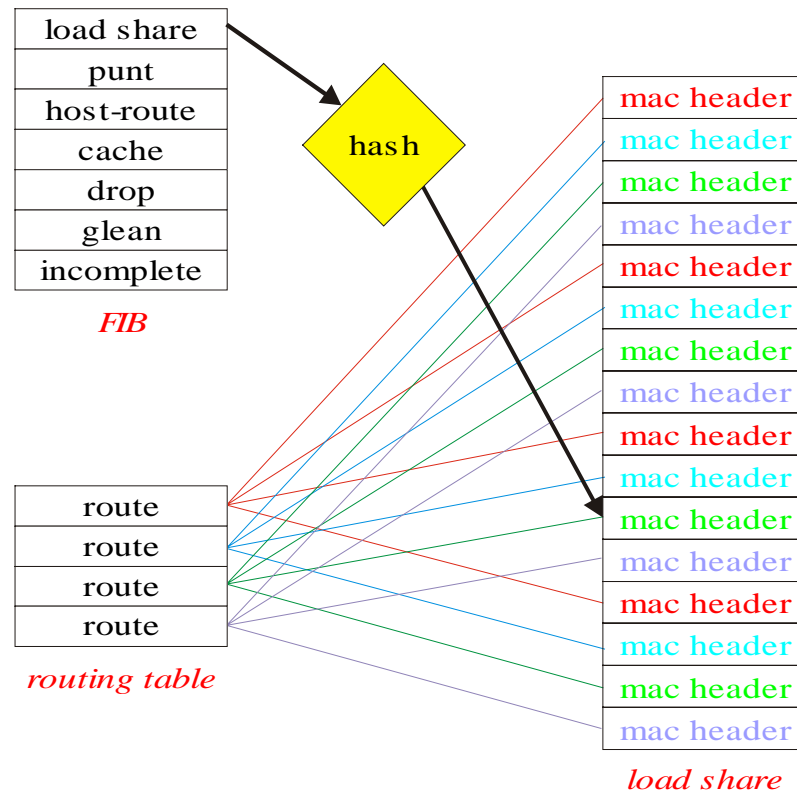
Hashed Source/Destination



Hashed Source/Destination

- **Based on a hash of the source and destination addresses.**
- **This insures that each source/destination pair will be sent out the same interface.**
- **In some cases, it will have traffic polarization issues (fix coming for CEF).**

CEF'S Hashed Source/Destination



By default, then, traffic is load balanced based on both the source and destination address; each packet sourced from a given host and destined to a given network will always choose the same next hop.

How many Paths?

IGP in the POP with POP GWs Originating Default to the POP Devices w/ flat ethernet.

**Four paths with each
over each link**

**8 equal paths to
default**

