Arquitectura de Red Resistente

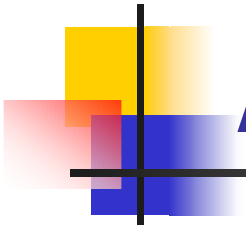
Guayaquil, Ecuador
2004

Mike De Leo
mdeleo@cisco.com

“El portero tiró el enchufe…”

- ¿Por qué se le permitió estar cerca del equipo?
- ¿Por qué se notó el problema después?
- ¿Por qué se tardo 6 semanas en determinar el problema?
- ¿Por qué no habia electricidad redundante?
- ¿Por qué no habia redundancia en la red?

Diseño de Red y Arquitectura…



- …es de importancia crítica
- …contribuye directamente al éxito de la red
- …contribuye directamente al fracaso de la red

Ley de Ingeniería de Ferguson

“Ninguna cantidad de trucos salvan a una red un mal diseño”

Paul Ferguson—Consulting Engineer,
Cisco Systems

¿Qué es una red bien diseñada?

- Una que toma en cuenta estos factores importantes:
  - Infraesctructura física
  - Jerarquía topológica/protocolo
  - Redundancia
  - Agregación de direcciones (IGP y BGP)
  - Escalabilidad
  - Implantación de políticas (dorsal/borde)
  - Administración/mantenimiento/operaciones
  - Costo

Un taburete de tres-patas

- Diseñando la red pensando en resistencia
- Usando tecnología para identificar y eliminar puntos únicos de falla
- Tener un proceso en operación para reducir el error humano
- Todos estos elementos son necesarios e interactúan entre sí
  - Si le falta una pata al taburete, no se queda de pie

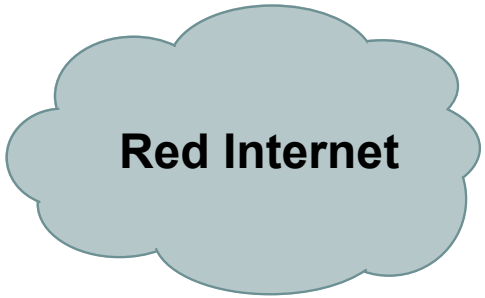
Diseño

Tecnología

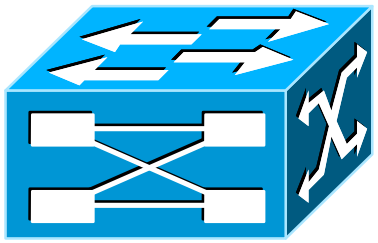
Proceso

Nuevo Mundo vs. Viejo Mundo

- Internet/redes de capa 3
  - Construidos con redundancia en el sistema
- Voz Telefónica y redes de Capa 2
  - Toda la redundancia está puesta en una caja

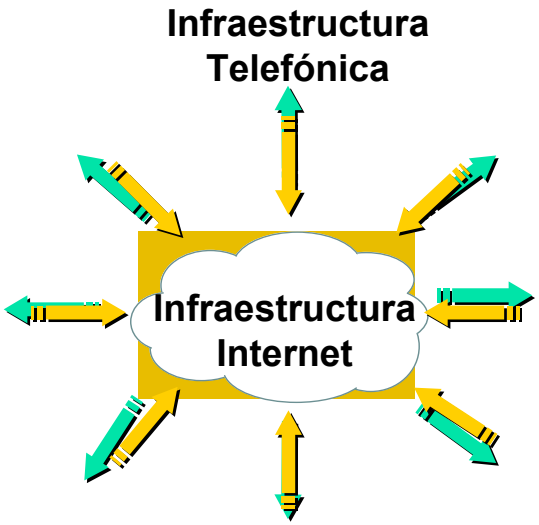


vs.



Nuevo Mundo vs. Viejo Mundo

- A pesar del cambio en la dinámica de Cliente ↔ Proveedor, los conceptos básicos de construir una red no han cambiado
- Los expertos ISP pueden aprender de expertos Telefónicos que tienen más de 100 años de experiencia
- Los expertos telefónicos pueden aprender de los expertos ISP de la dura experiencia de escalar +100% por año



¿Cómo llegamos allí?

"En la era del Internet, la confiabilidad es algo que se debe construir, y no algo que se puede comprar. Esto es trabajo duro, y requiere inteligencia, habilidad y presupuesto. La confiabilidad no es parte del paquete básico."

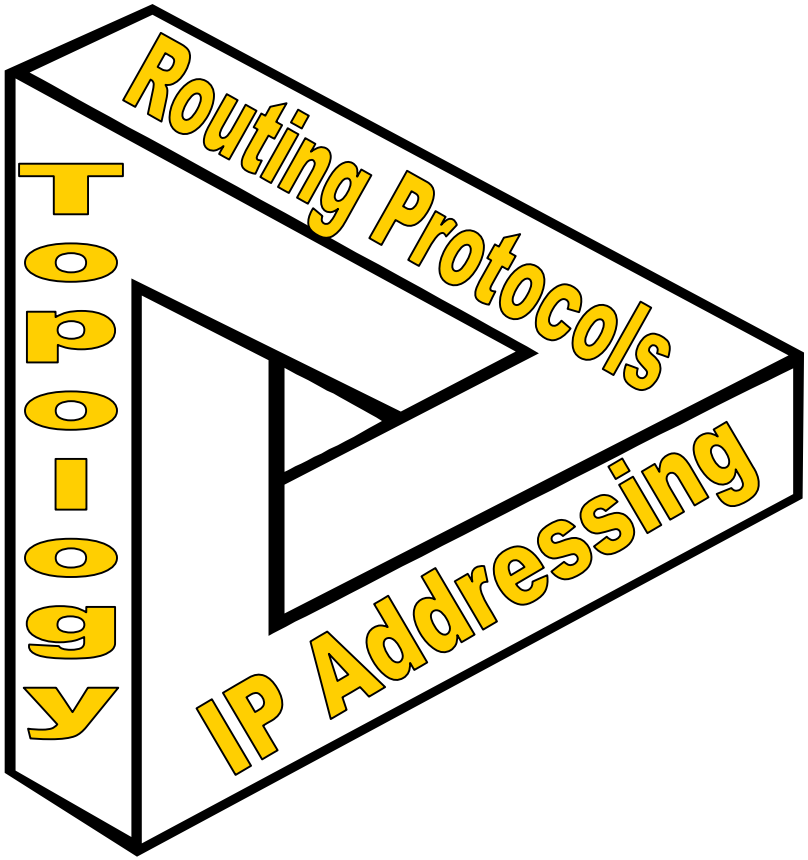
Joel Snyder – Network World Test Alliance 1/10/2000
"Reliability: Something you build, not buy"

Diseño de Red Redundante

Conceptos y Técnicas

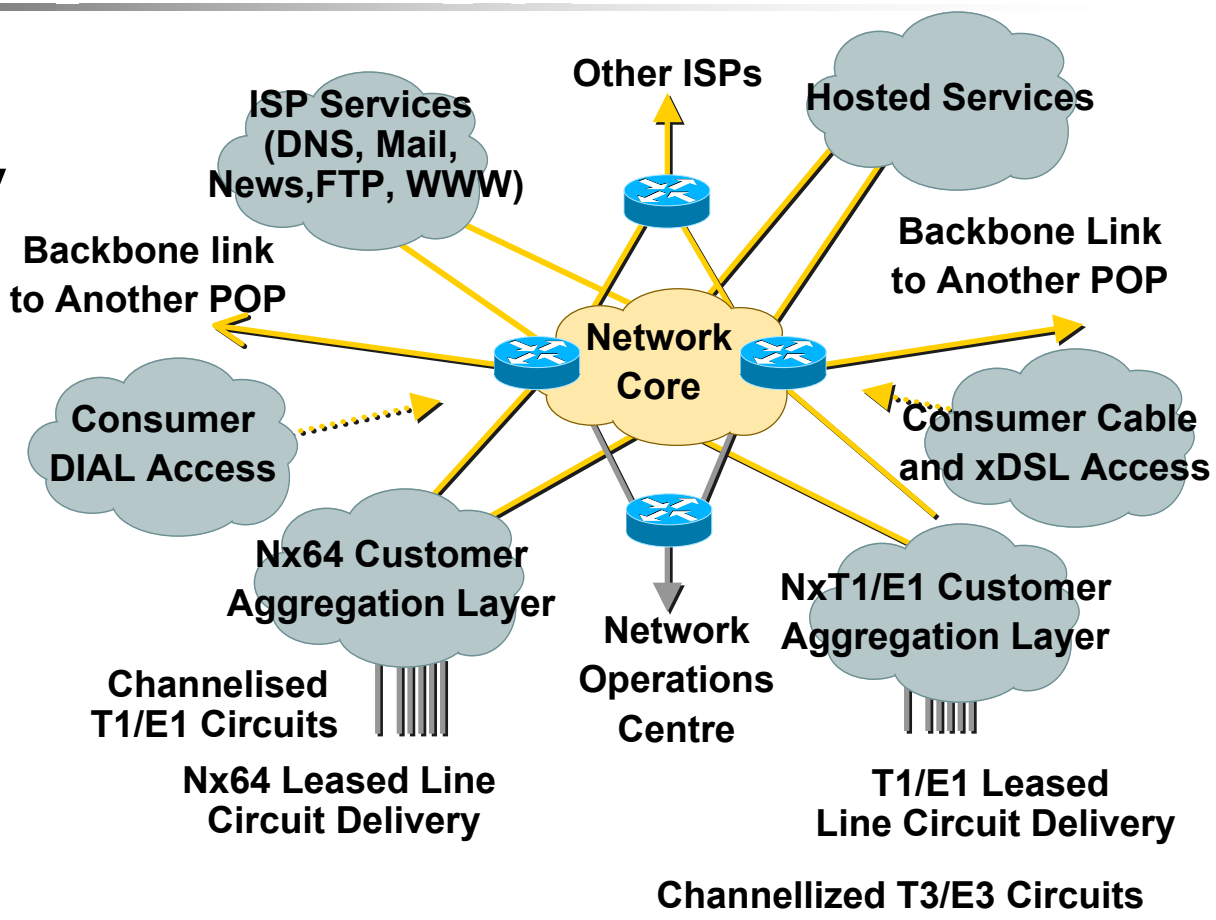
Conceptos Básicos de Escalabilidad para un ISP

- Diseño Modular/Estructurado
- Diseño Funcional
- Disciplina de Diseño con Niveles/Jerárquico



Diseño Modular/Estructurado

- Organice la red en módulos separados y repetibles
  - Dorsal
  - POP (Punto de Presencia)
  - Servicios de Hosting
  - Servicios ISP
  - Soporte/NOC



Diseño Modular/Estructurado

- Modularidad hace fácil escalar la red
  - Diseña unidades pequeñas de la red que se pueden interconectar con otras
  - Cada módulo se puede construir para una función específica en la red
  - Las actualizaciones están construidas alrededor de los módulos, no la red entera

Diseño Funcional

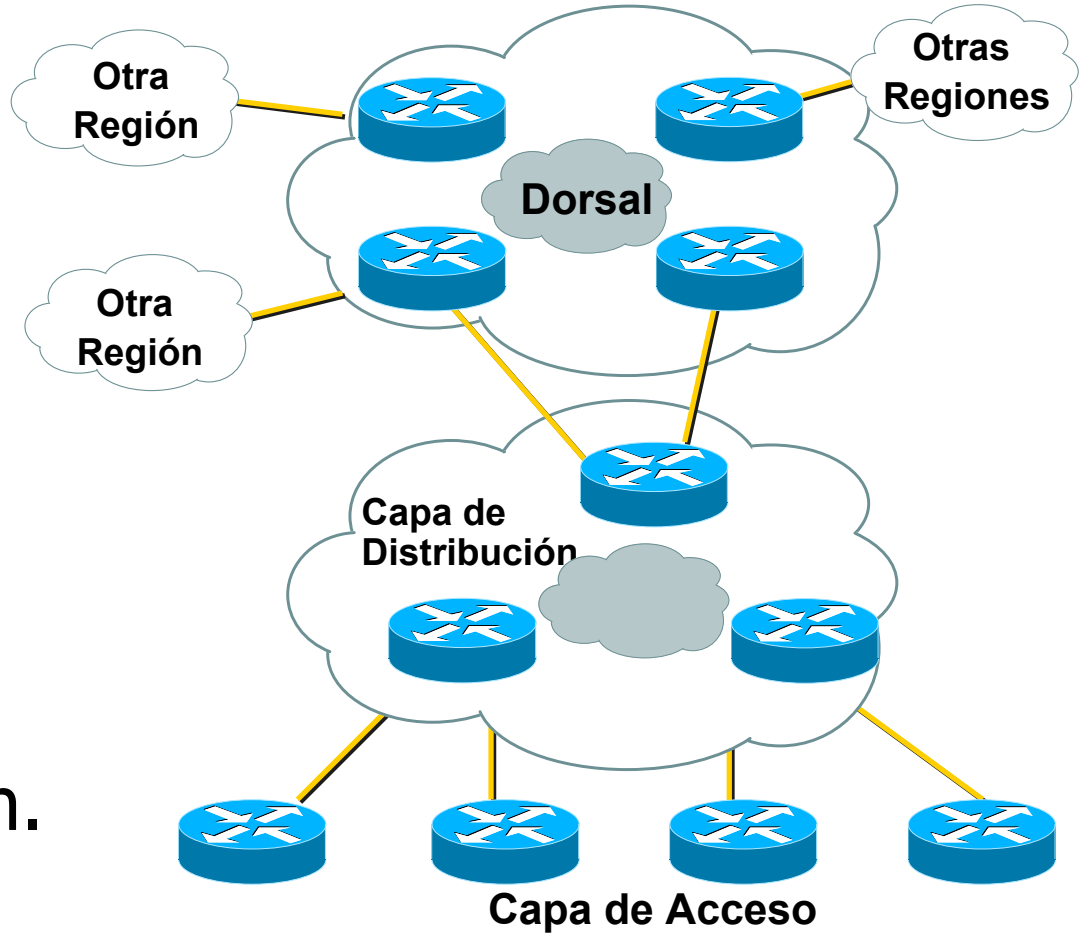
- Una máquina no puede hacer todo

  (no importa que tanto la gente lo ha intentado en el pasado)
- Cada ruteador/switch en una red tiene on conjunto de funciones bien definidos
- Las diversas máquinas interactúan entre ellas
- El equipo se puede seleccionar y colocar funcionalmente en la red basado en sus fortalezas
- Una red ISP es un diseño con perspectiva de sistema

  Funciones que se interconectan e interaccionan forman una solución en red.

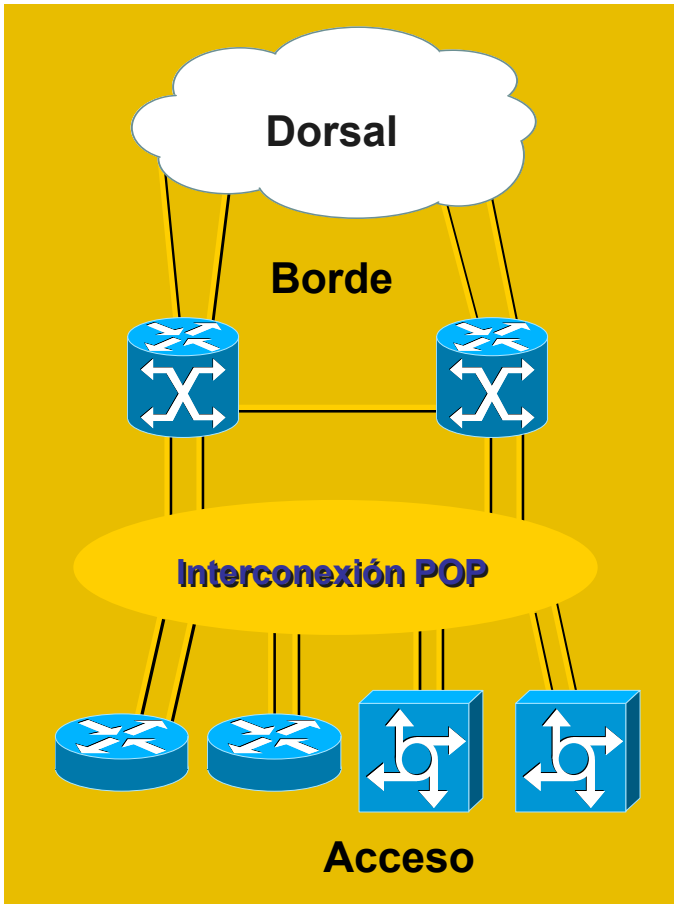
Diseño de Niveles/Jerárquico

- Topologías de malla plana no escalan
- La jerarquía en el diseño se usa para que la red escale
- Buenos conceptos guían, pero se vuelven borrosos cuando se implantan.



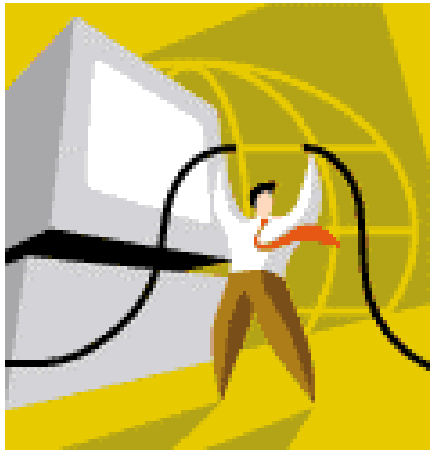
Redundancia en Múltiples Niveles

- Redundancia de 3 capas en el POP
  - Fallas de capas inferiores son mejores
  - Fallas de capas inferiores pueden causar fallas de capas superiores
  - L2: Dos de todo
  - L3: IGP y BGP proveen redundancia y balanceo
  - L4: retrasmisiones TCP recuperan durante una falla



Redundancia de Múltiples Niveles

- Múltiples niveles también significa que uno debe ir profundo – por ejemplo:
  - Cableado de planta externa – circuitos en un solo paquete – fallas por maquinaria
  - Potencia redundante al rack – los circuitos se puede sobrecargar y el técnico puede tropezarse
- Problemas introducidos por mantenimiento (son una de las causas de fallas en los ISPs).



Diseño de Red Redundante

Lo Básico

Lo Básico: Platforma

- Potencia redundante
  - Dos fuentes de electricidad
- Enfriamiento redundante
  - ¿Qué pasa si uno de los ventiladores falla?
- Procesadores de ruteo redundante
  - Una consideración adicional, pero menos importante
  - Un ruteador en pareja es mejor
- Interfaces redundantes
  - Un enlace redundante a un dispositivo en pareja es mejor

Lo Básico: Ambiente

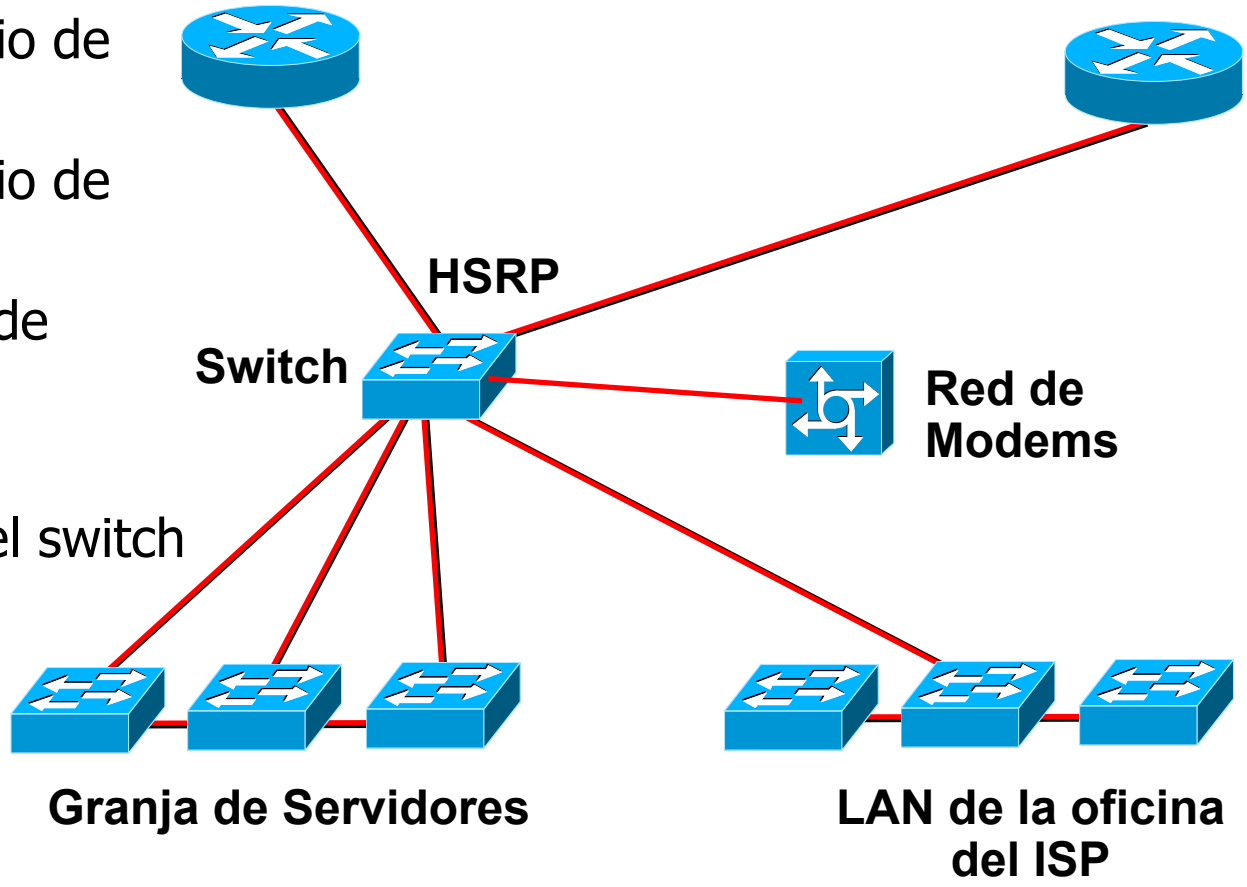
- Electricidad redundante
  - Fuente UPS – protege contra fallas en la red eléctrica
  - Fuente "Sucia" – protege contra fallas en el UPS
- Cableado redundante
  - Rotura de cables dentro de las instalaciones se pueden reparar usando cables "sobrantes"
  - Las instalaciones deben de tener cables externos por dos caminos
- Enfriamiento redundante
  - Instalación de redundancia en aire acondicionado
  - …o cualquier otro sistema de enfriamiento

Diseño Redundante de Red

Dentro del Centro de Datos

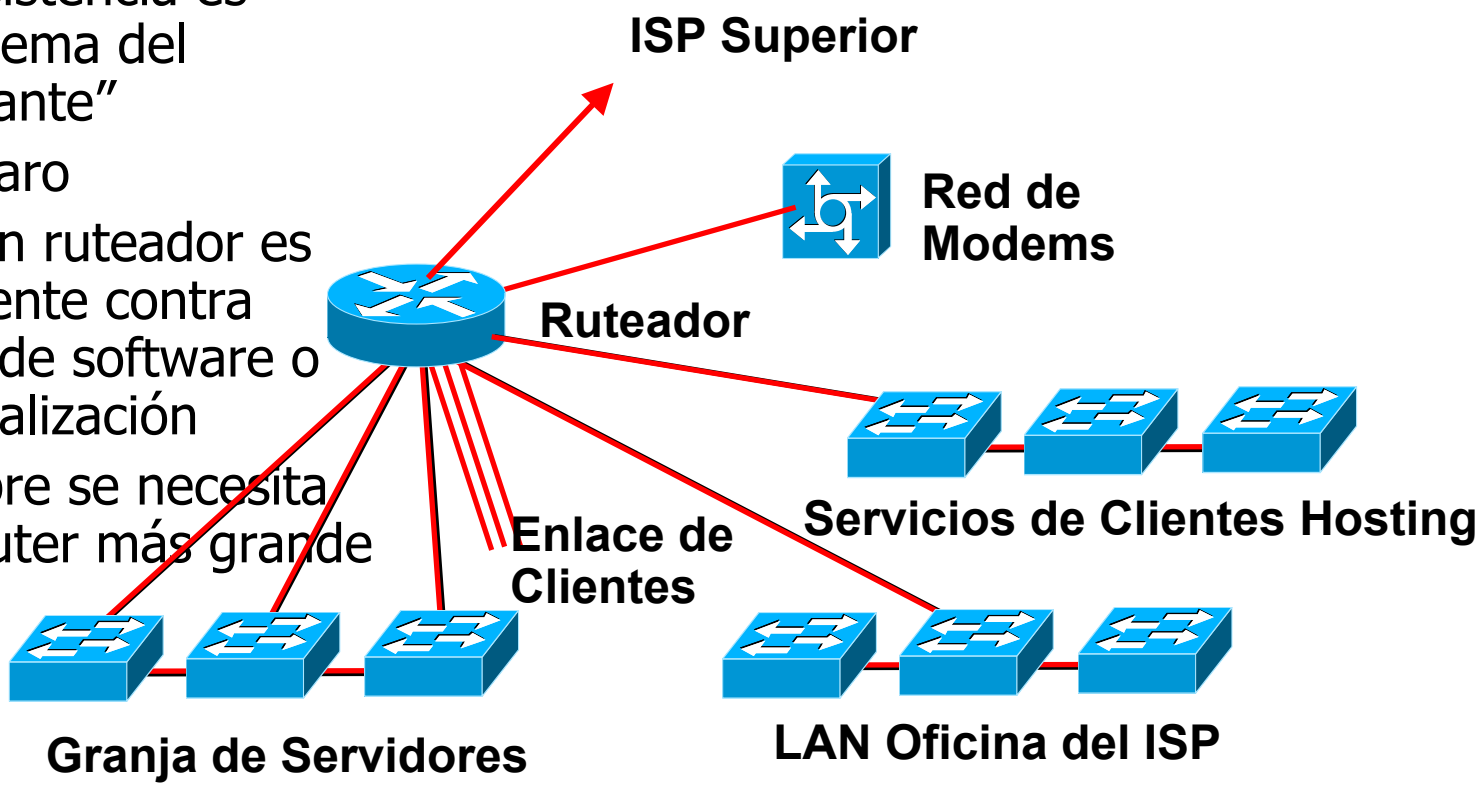
Mala Arquitectura (1)

- Un solo punto de fallo
  - Un solo dominio de colisiones
  - Un solo dominio de seguridad
  - Convergencia de Spanning tree
  - Sin respaldo
  - Desempeño del switch central



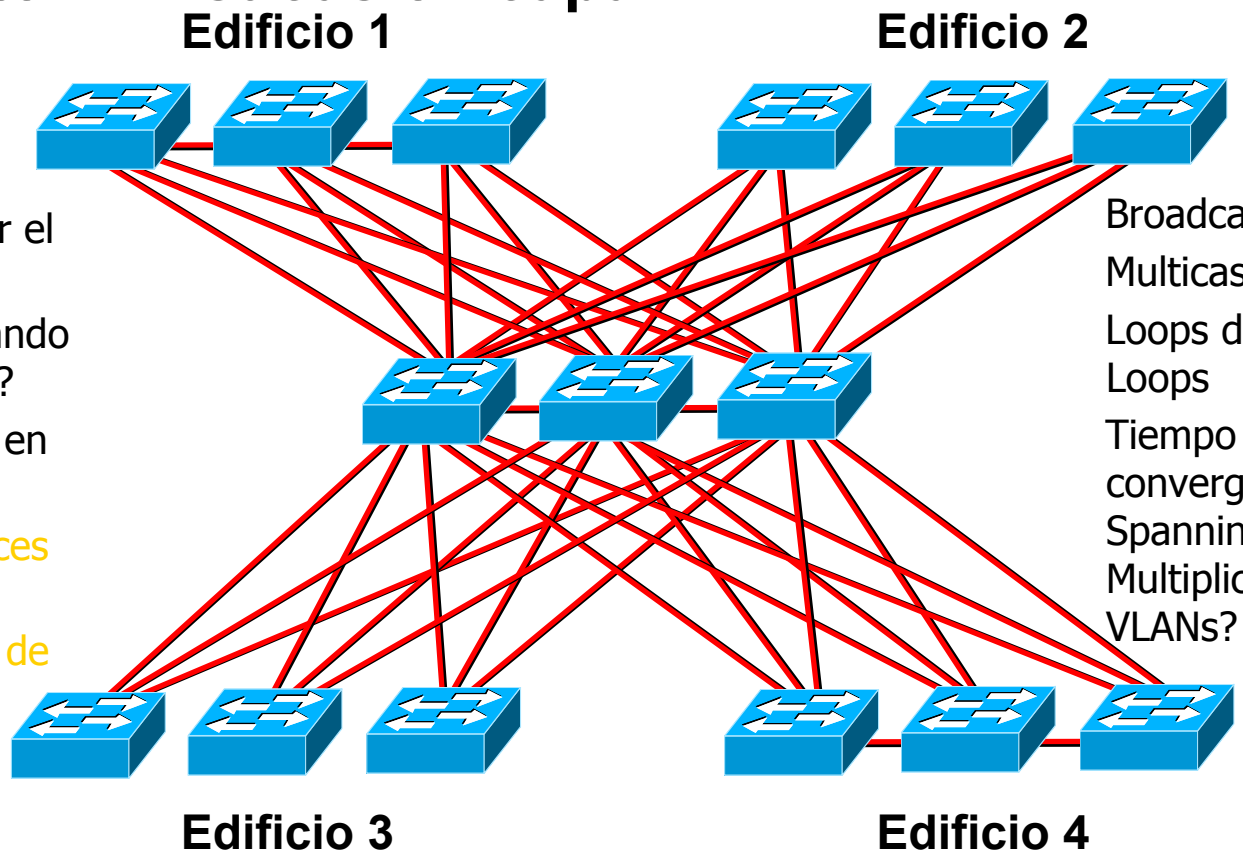
Mala Arquitectura (2)

- Un ruteador central
  - Fácil de construir
  - La resistencia es "problema del fabricante"
  - Más caro
  - Ningún ruteador es resistente contra fallas de software o reinicialización
  - Siempre se necesita un router más grande



¡¡Aun peor!!

- **Evite redes muy enlazadas en malla, no-determinísticas en capa 2**



¿Dónde debe ir el 'root'?

¿Qué pasa cuando algo se rompe?

¿Cuánto tarda en coverger?

¿Cuántos enlaces bloquados?

¡Gran dominio de fallas!

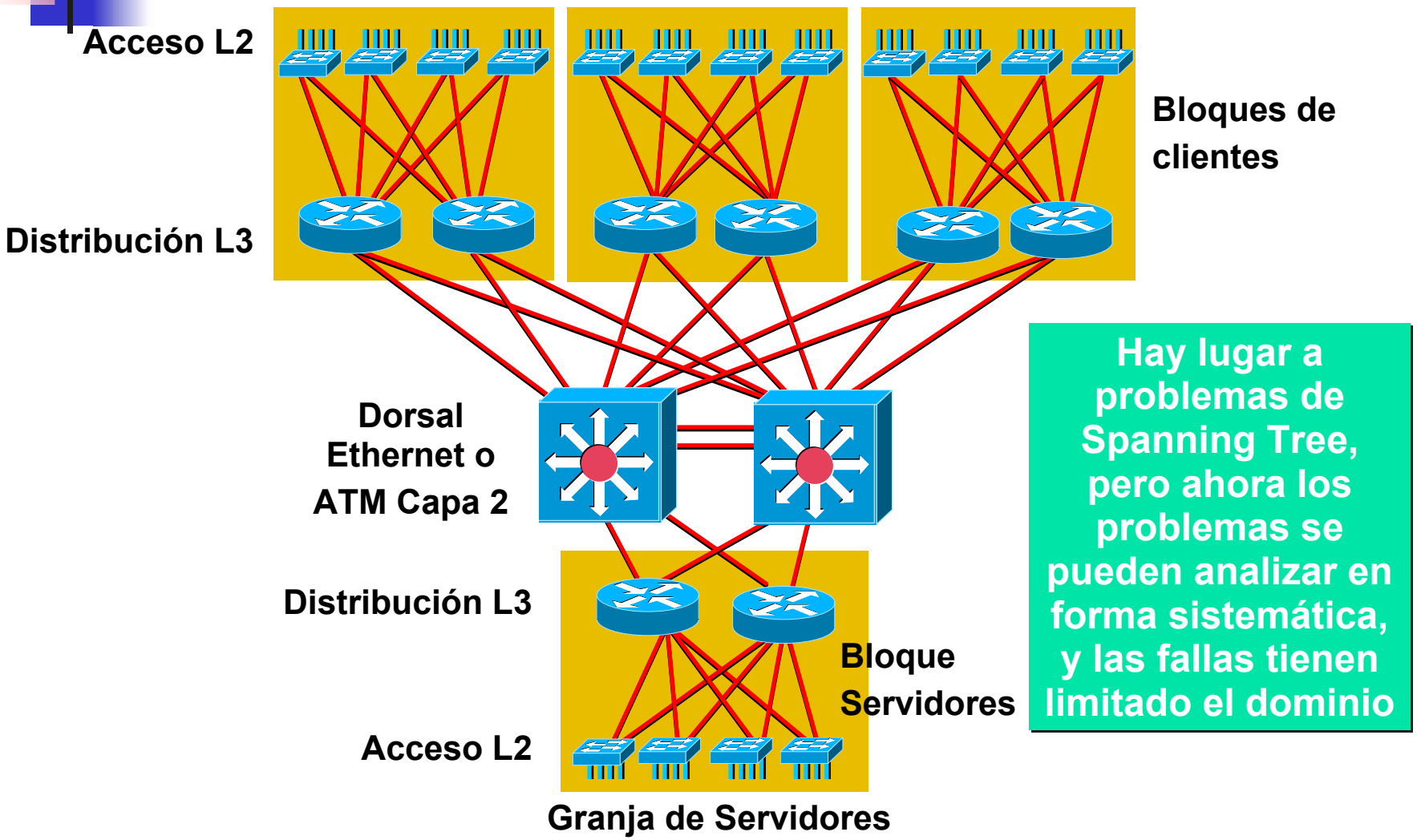
Broadcast Flooding

Multicast Flooding

Loops dentro de Loops

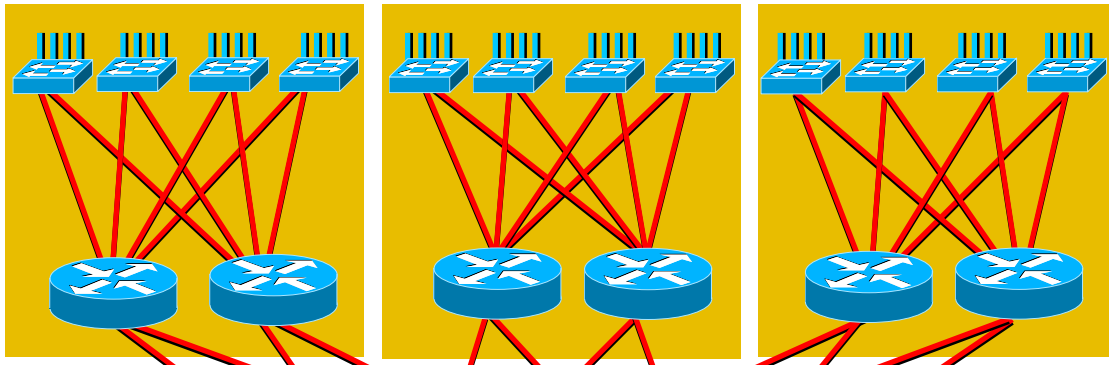
Tiempo de convergencia de Spanning Tree Multiplica por 100 VLANs?

Dorsal Típico (Mejor)



Hay lugar a problemas de Spanning Tree, pero ahora los problemas se pueden analizar en forma sistemática, y las fallas tienen limitado el dominio

La mejor arquitectura



Cliente

Acceso L2

Distribución L3

Múltiples subredes

Muy jerárquico

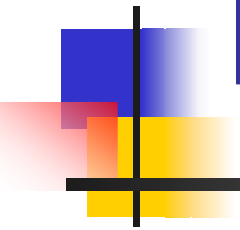
Broadcast y Multicast controlado

Dorsal L3

Distribución L3

Granja de Servidores

Acceso L2

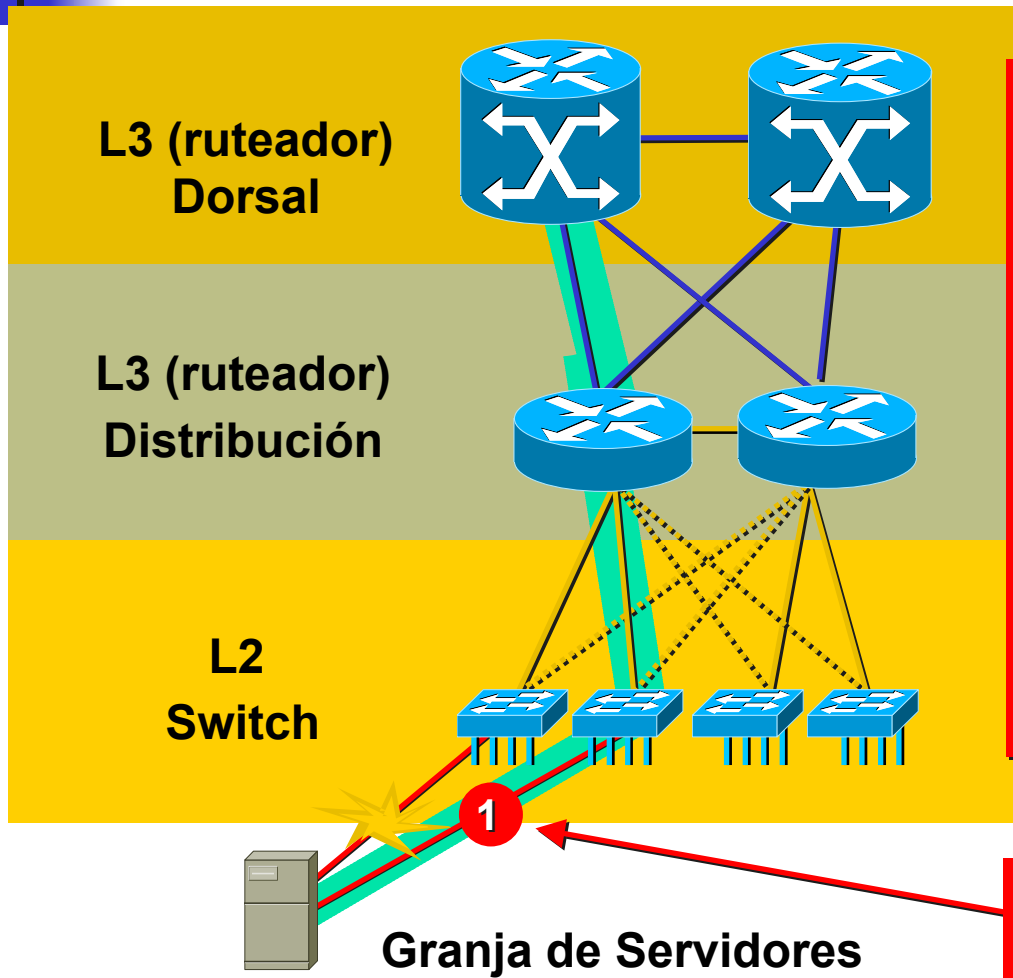


Diseño de Red Redundante

Disponibilidad de Servidores

Tecnología

Servidores Multi-homed



Usanto drivers y NIC adaptivos y tolerante a fallas

NIC tiene una sola dirección IP/MAC (activo un NIC a la vez)

Cuando en enlace de falla es reparado, no regresa para evitar flapping

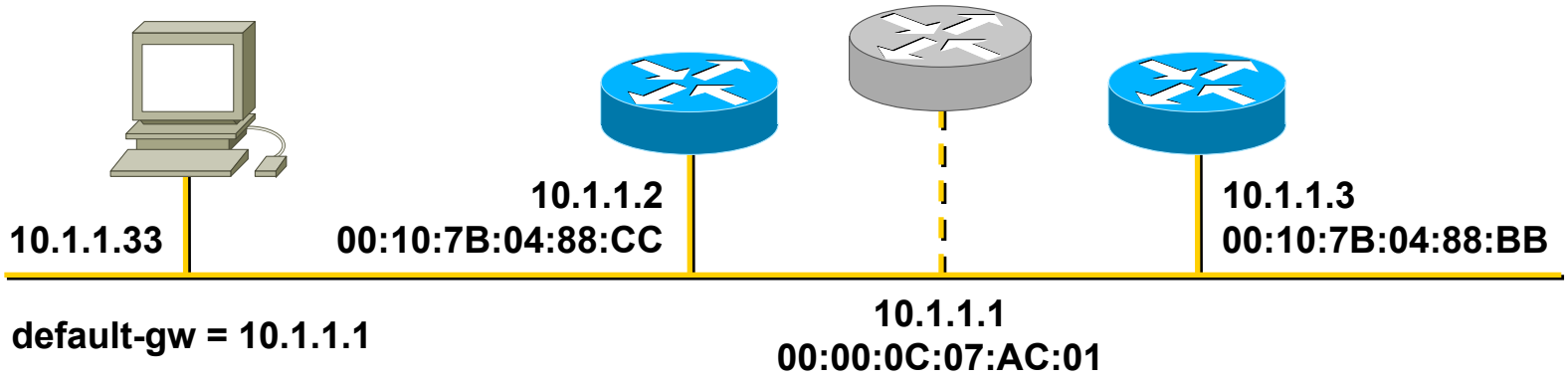
Drivers tolerantes a fallas disponibles de muchos proveedores: Intel, Compaq, HP, Sun

Muchos proveedores tienen dirvers para soportar etherchannel

Servidor Dual-homed — Recuperación del NIC primario (tiempo 1–2 Segundos)

HSRP – Hot Standby Router Protocol

Tecnología

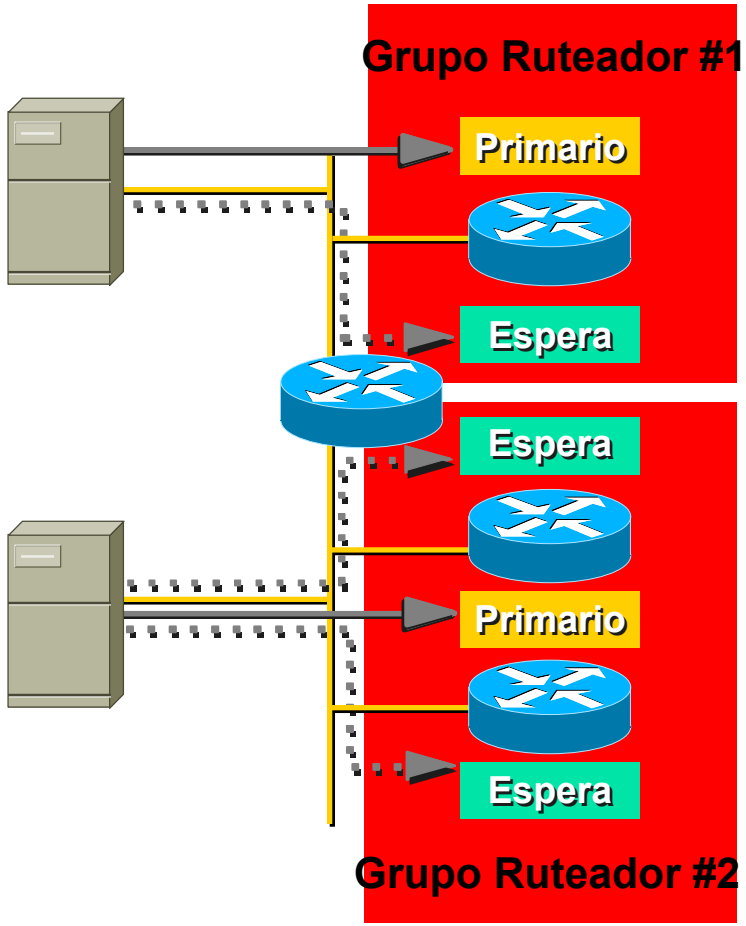


- Cambio transparente de en caso de falla del default router
- Se crea un ruteador “Fantasma”
- Un ruteador es activo, responde a las direcciones fantasmas L2 y L3
- Otros monitorean y toman el lugar de direcciones fantasmas

Tecnología

HSRP – RFC 2281

- El HSR envía multicast hellos cada 3 segs con una prioridad por omisión de 100
- HSR asume el control si tiene la prioridad más alta y tiene preferencia configurado después de retraso (omisión=0) segundos
- HSR resta 10 de la prioridad de la interface si cae

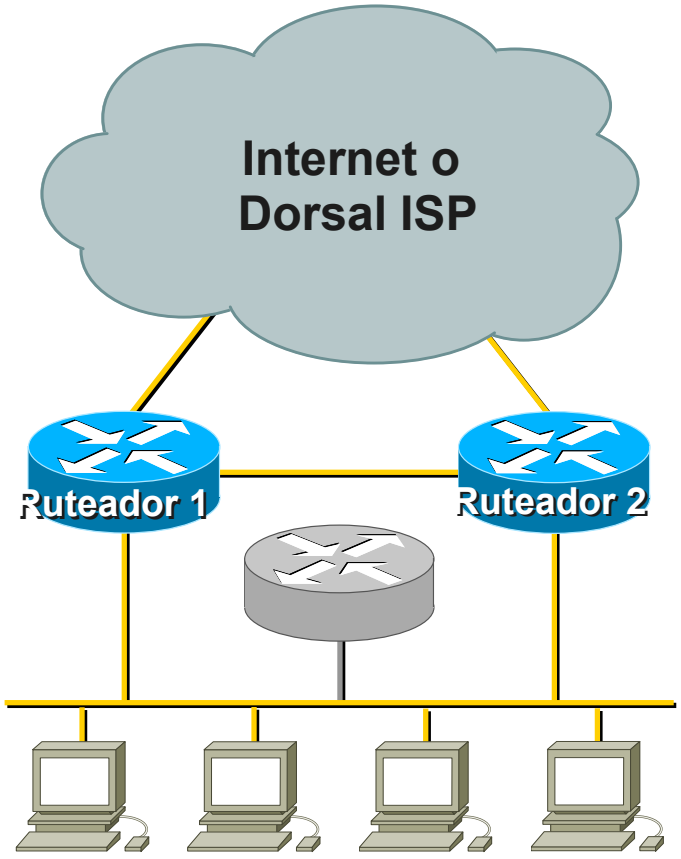


Technology

HSRP

```
Router1:
interface ethernet 0/0
ip address 169.223.10.1 255.255.255.0
standby 10 ip 169.223.10.254


Router2:
interface ethernet 0/0
ip address 169.223.10.2 255.255.255.0
standby 10 priority 150 pre-empt delay 10
standby 10 ip 169.223.10.254
standby 10 track serial 0 60
```



Diseño de Red Redundante

Disponibilidad de
Red de Área Amplia

Direversidad de Circuitos

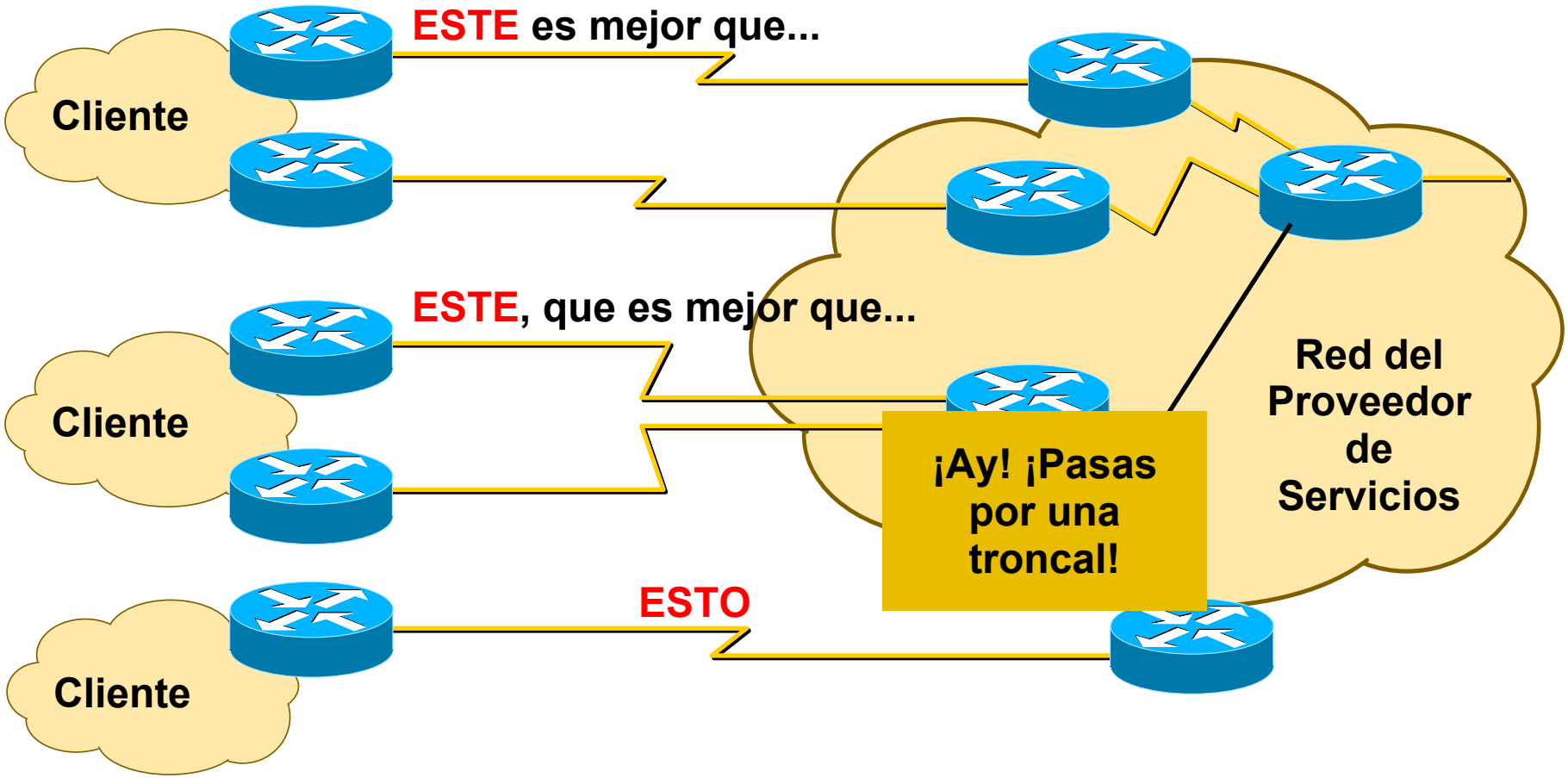
- Tener PVCs de respaldo a través del mismo puerto físico logra poco o nada
  - Es más probable que falle el puerto que un PVC individual
  - Use puertos separados
- Tener conexiones redundantes en el mismo ruteador no da independencia del ruteador
  - Use ruteador separados
- Use diferente proveedor de circuitos (si está disponible)
  - Problemas con un proveedor no significan un problema para su red

Diversidad de Circuitos

- Asegure que las instalaciones tiene diversos caminos para los circuitos para uno o más proveedores
- Asegure que las trayectorias de respaldo terminen en equipos separados del proveedor de servicio
- Asegure que las líneas no se unen en una troncal a través de trayectorias iguales para atravesar la red
- Pruebe y escriba esto en los contratos de Nivel de Servicio (SLA) con los proveedores

Tecnología

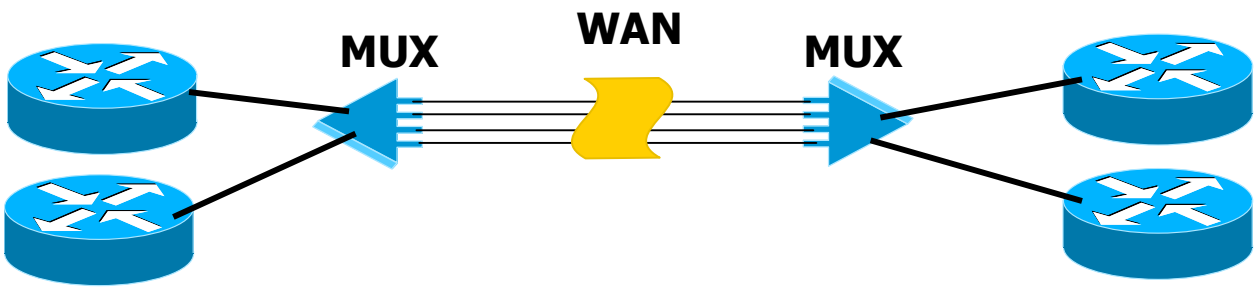
Diversidad de Circuitos



Tecnología

Unión de Circuitos – MUX

- Utilize hardware MUX
  - MUX en hardware pueden unir múltiples circuitos, proveen redundancia L1
  - Necesita un MUX similar del otro lado del enlace
  - Ruteador ve circuitos como un enlace
    - El MUX se encarca de las fallas
    - Usando ruteadores redudantes ayuda



Balanceo

- El balanceo ocurre cuando un ruteador tiene dos (o más) caminos de costos iguales a un mismo destino
- EIGRP permite balanceo con costos-desiguales
- Balanceo puede ser por-paquete o por-destino (omisión: por-destino)
- El balanceo puede ser una técnica poderosa para redundancia, dado que provee un camino alterno en caso de falla de un ruteador

Balanceo

- OSPF balancea en trayectorias de costos iguales por omisión
- EIGRP balancea con trayectorias de costos iguales por omisión, y puede ser configurado para usar costos desiguales:

```
router eigrp 111
 network 10.1.1.0
 variance 2
```

- Balanceo con costos desiguales no es recomendado; puede crear muchos problemas de sincronización y retransmisiones.

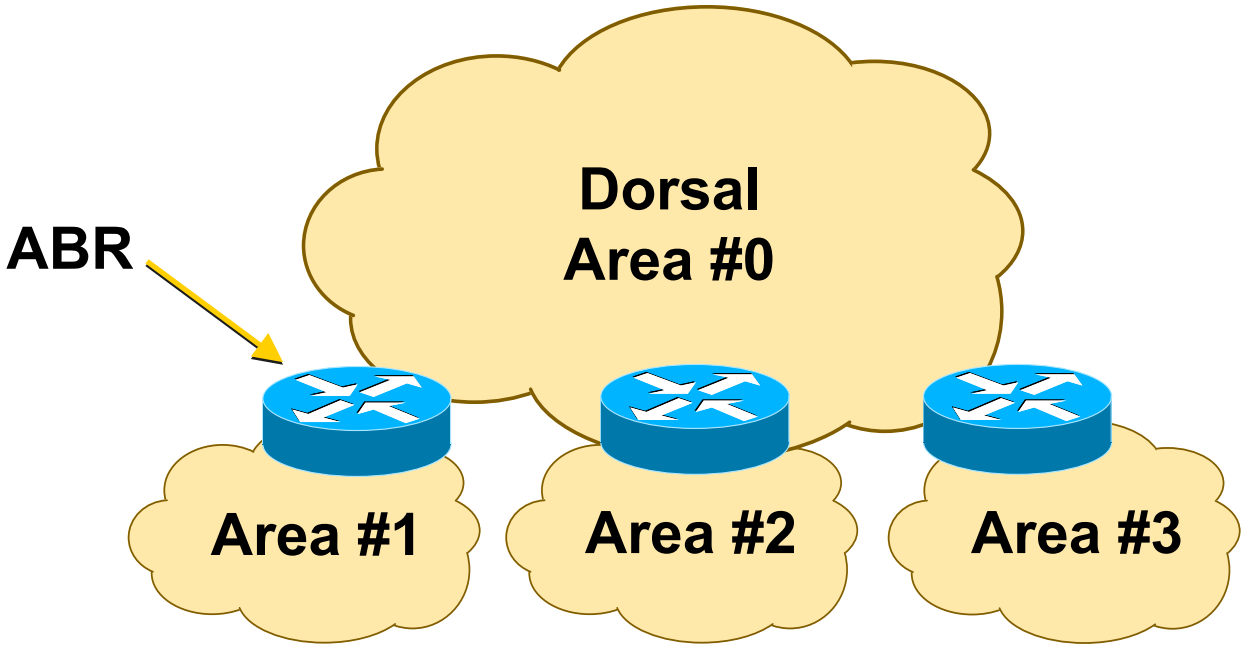
Convergencia

- El tiempo de convergencia de un protocolo de enrutamiento seleccionado puede afectar la disponibilidad general de la red de área amplia
- Un área a examinar es el impacto del diseño de L2 en la eficacia de L3

Factores que Determinan la Convergencia del Protocolo

- Tamaño de la red
- Limitaciones del número de saltos
- Arreglos de Peering (borde, dorsal)
- Velocidad de detección de cambios
- Propagación de información de cambio
- Diseño de Red: jerarquía, resumen (summarization) y redundancia

OSPF – Estructura Jerarquíca



- La topología de un área es invisible fuera del área
  - LSA flooding es limitado al área
  - Calculo de SPF se realiza separadamente en cada área

Factores que Asisten en la Convergencia del Protocolo

- Mantenga el número pequeño de dispositivos de ruteo en cada área (entre 15 – 20)
  - Reduce el tiempo requerido para convergencia
- Evite mallas complejas de dispositivos en un área
  - Usualmente dos enlaces son todo lo necesario
- Mantenga pequeña la cuenta de prefijos en los protocolos de ruteo interno
  - Números más grandes significan más tiempo para calcular el camino más corto
- Utilice valores por omisión del fabricante para protocolos de ruteo a menos que entienda el impacto de "mover las perillas"
  - Perillas están ahí para mejorar el desempeño en ciertas condiciones

Diseño de Red Redundante

Disponibilidad del Internet

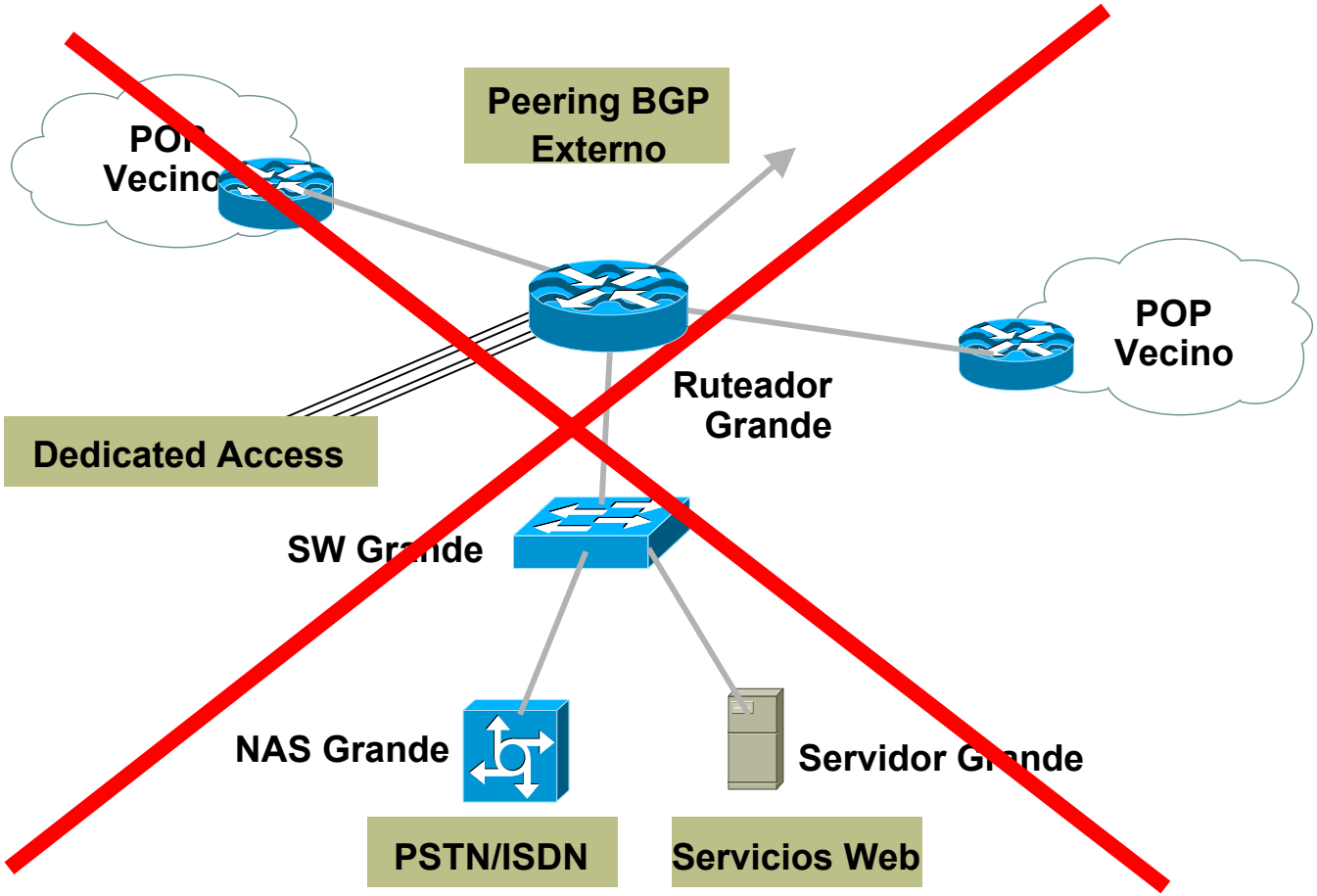
Diseño del PoP (Punto de Presencia)

- Un ruteador no puede hacer todo
- Redundancia redundancia redundancia
- Los ISPs exitosos construyen dos de todo
- Dos dispositivos pequeños en vez de uno grande:
  - Dos ruteadores para una función
  - Dos switches para una función
  - Dos enlaces para una función

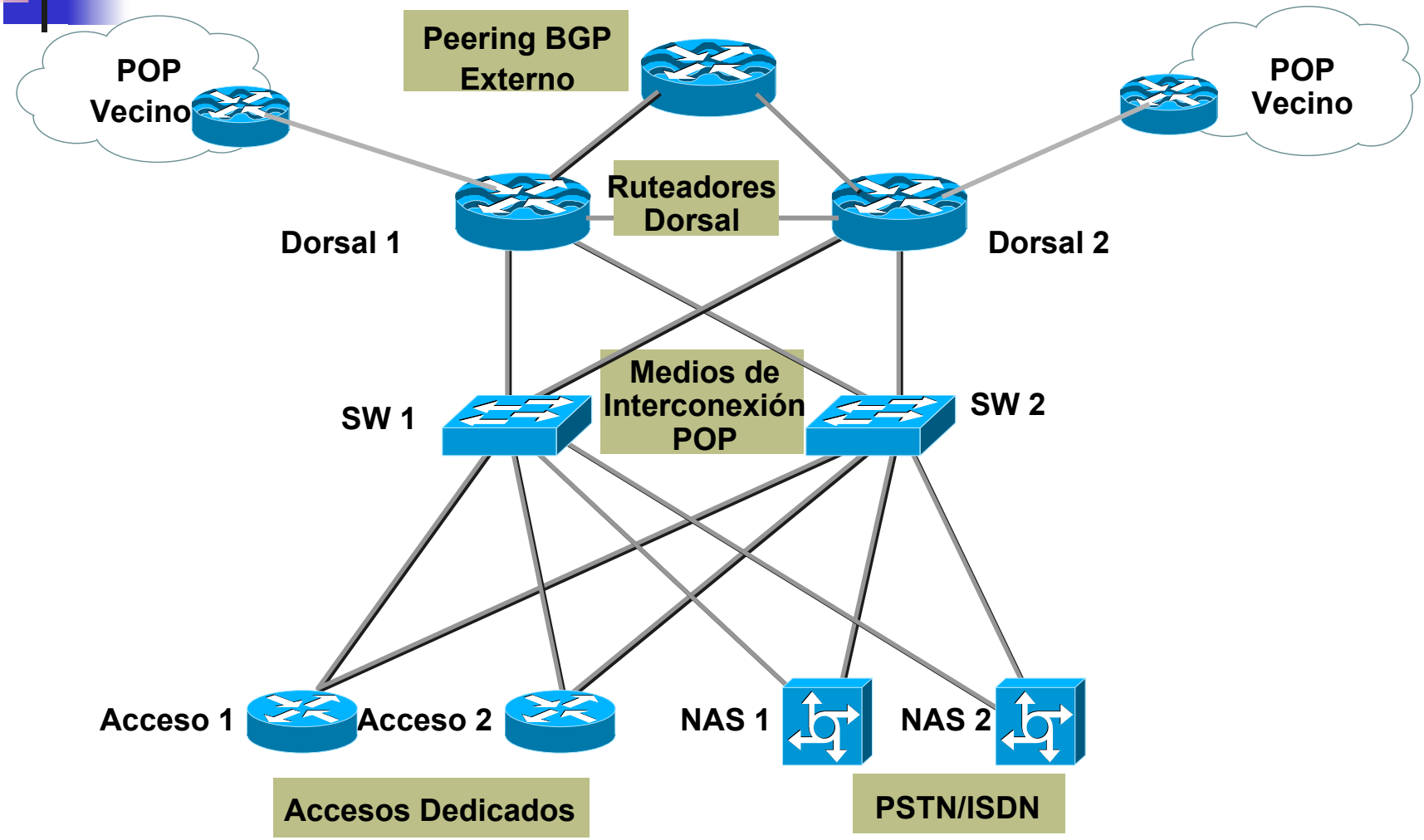
Diseño del PoP

- Dos de todo no significa complejidad
- Evita diseños de redes complejas con mucha malla
  - Difícil de operar
  - Difícil de depurar
  - Difícil de escalar
  - Usualmente demuestra desempeño pobre

Diseño del PoP – Incorecto



Diseño del PoP – Corecto



Hubs vs. Switches

- Hubs (Repetidor)
  - Estos son obsoletos
    - Switches cuestan un poco más
  - El tráfico en un hub es visible en todos los puertos
    - Realmente es un sustitución del cable coaxial Ethernet
    - ¿¡Seguridad!?
  - El desempeño es muy bajo
    - 10Mbps compartido entre todos los dispositivos de LAN
    - Tráfico alto en un dispositivo impacta a todos los demas
  - Usualmente no existe administración

Hubs vs. Switches

- Switches
  - Cada puerto es invisible a los demás
  - Alto desempeño
    - 10/100Mbps por puerto
    - Tráfico cargado en un puerto no impacta a los demás puertos
  - Switches 10/100 son comunes y baratos
  - Utilice un switch sin bloqueo en la dorsal
    - Paquetes no tienen que esperar para conmutar
  - Capacidad de Administración (SNMP vía IP, CLI)
  - Fuentes de electricidad redundantes son útiles

Cuidado con IP Estático en Dial

- Problemas
  - NO escala
  - Rutas /32 de clientes en el IGP – IGP no escala
  - Mas clientes, convergencia del IGP mas lento
  - El soporte se hace mas costoso
- Soluciones
  - Enruta los clientes “Dial Estático” al mismo RAS o grupo de RAS detrás del ruteador de distribución
  - Utilice bloque de direcciones continuas
  - Hazlo muy caro – te cuesta dinero implantar y soportar

Diseño de Red Redundante

¡Operaciones!

Centro de Operaciones de Red (NOC)

- El NOC es necesario par una red ISP pequeña
  - Puede ser una PC llamado NOC, con un UPS, en el cuarto del equipo.
  - El último recurso para acceso a la red
  - Captura bitácoras (logs) de información de la red
  - Tiene acceso remoto de afuera
    - Dialup, SSH,…
  - Entrena la gente para operarlo
  - Escala de una PC y soporte conforme crece la organización

Operaciones

- El NOC es esencial para todos los ISPs
- Los procedimientos operacionales son necesarios
  - Monitorea los circuitos fijos, dispositivos de acceso, servidores
  - Si algo falla, alguien debe ser notificado
- Caminos de escalación son necesarios
  - Ignorar un problema no ayuda a arreglarlo
  - Decide sobre el tiempo a reparar, escala arriba la cadena de reporte hasta que alguien lo arregla

Proceso

Operaciones

- Modificaciones a la red
- Una red bien diseñada corre tan bien como los que la operan
  - Decida y publique los itinerarios de mantenimiento
  - Y utilícelos como SON
  - No haga cambios fuera del período de mantenimiento, no importa que tan triviales puedan parecer

En Resumen

- Implantar una red IP altamente resistente requiere una combinación de procesos, diseño y tecnología
- "y ahora obedece diseño, tecnología y proceso; pero el más grande de estos es proceso"
- ¡Y no olvide "mantégalo sencillo"! (KISS)
  - Keep It Simple & Stupid!

Diseño

Tecnología

Proceso