



Every Choice Affects Availability! **Determination or Flexibility?**

- **Would you support 27 desktop environments?**
- **Would you support 13 network vendors?**
- **Would you use 57 varieties of Cisco IOS®?**

Design for High Availability

- **The concepts**
- **The rules**
- **Design building block**
- **Design backbone**
- **Determinism and failure analysis**

What Is an HA Network?

	Availability Requirement	Unplanned Downtime	Redundancy	H/W MTTR Replacement	Service Mgmt
Reliable Network	99.9%	8 Hours 46 Minutes	No	Up to 24 Hours	No
HA Network	99.99%	53 Minutes	Yes	Up to 4 Hours	Yes
Non-Stop Network	99.9999%	32.6 Sec	Yes	2 Hours	Yes

HA Design Reality Check! Assume Things Fail—Then What?

- Networks are complex
- Things break, people make mistakes
- MTBF is too theoretical—what happens if a failure occurs?
- Simple, structured, deterministic design required for fast recovery
- The “trade-off's”—your choices are important

Variables in the High Availability Function

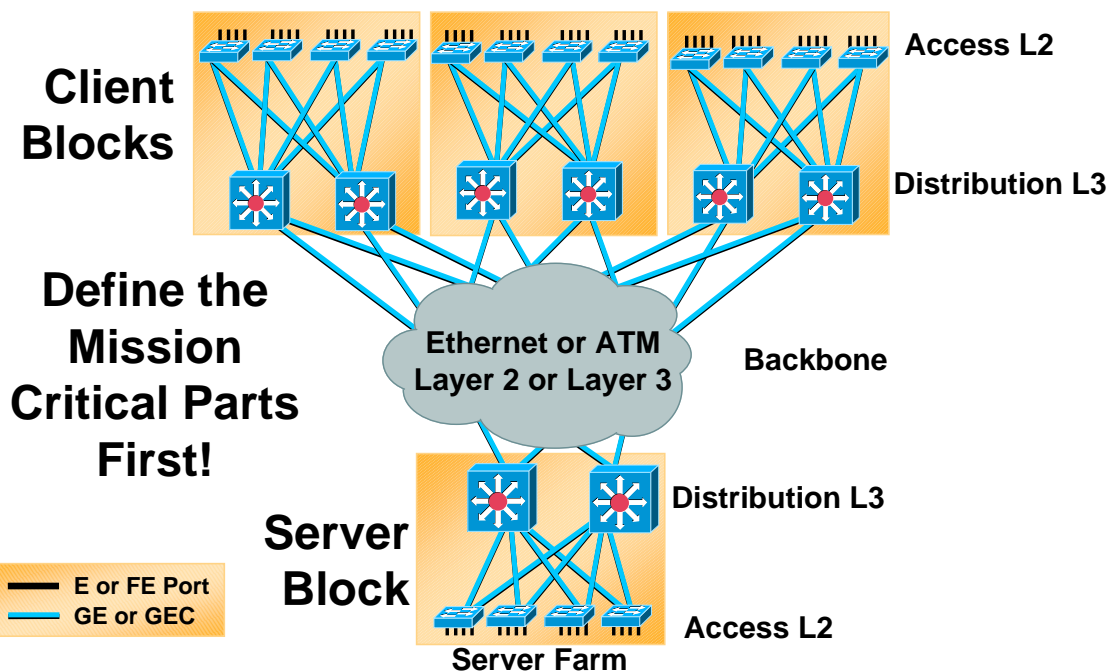
- **Server and application**
Redundancy features
- **Outside the box**
The logical design is critical
Network features and protocols
Geophysical diversity is powerful
- **Inside the box**
HA features and MTBF
- **Administration and troubleshooting**

2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

7

HA Gigabit Campus Architecture Survivable Modules and Survivable Backbone

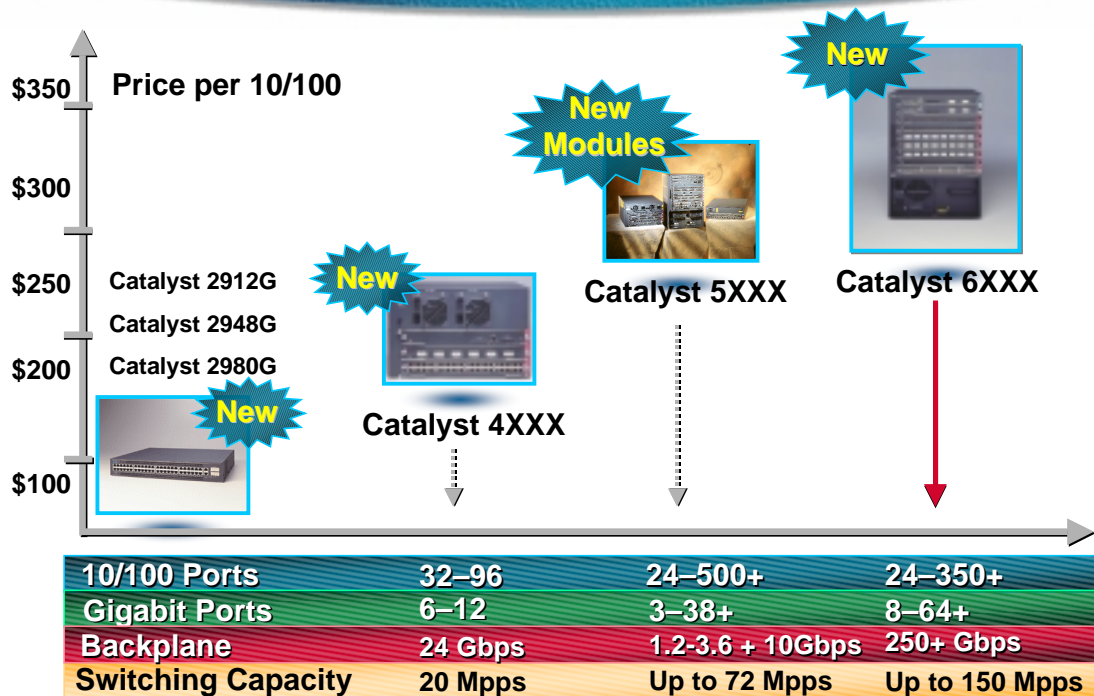


2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

8

QoS Enabled Campus Switching Design Solution then **Pick** the Product



2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

9

HA Network Design Concepts **Think Outside the Box**

- 1) Collapse the sandwich
- 2) Spanning tree failure domain
- 3) Map L3 to L2 to L1
- 4) Scaling and hierarchy
- 5) ABC's of module and backbone design

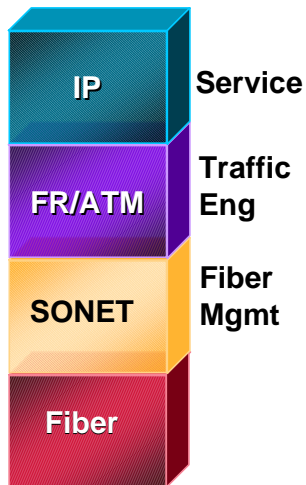
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

10

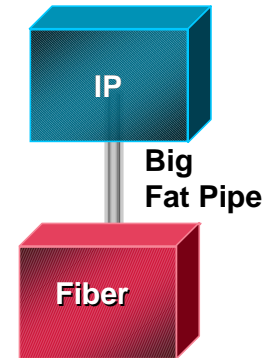
Concept 1—Collapse the Sandwich Route IP over Glass

Traditional Model



- Lower equipment cost
- Lower operational cost
- Simplified architecture
- Scalable capacity

Optical Internetworking



2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

11

DWDM Fat Pipes Cisco Metro 1500

- 32 channel system
- 8 channels/shelf
- Point-to-point and point-to-multipoint
- Up to 2.5 Gbps/channel, bit rate and protocol independent
- 1:1 line level protection
- SNMP manageable



2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

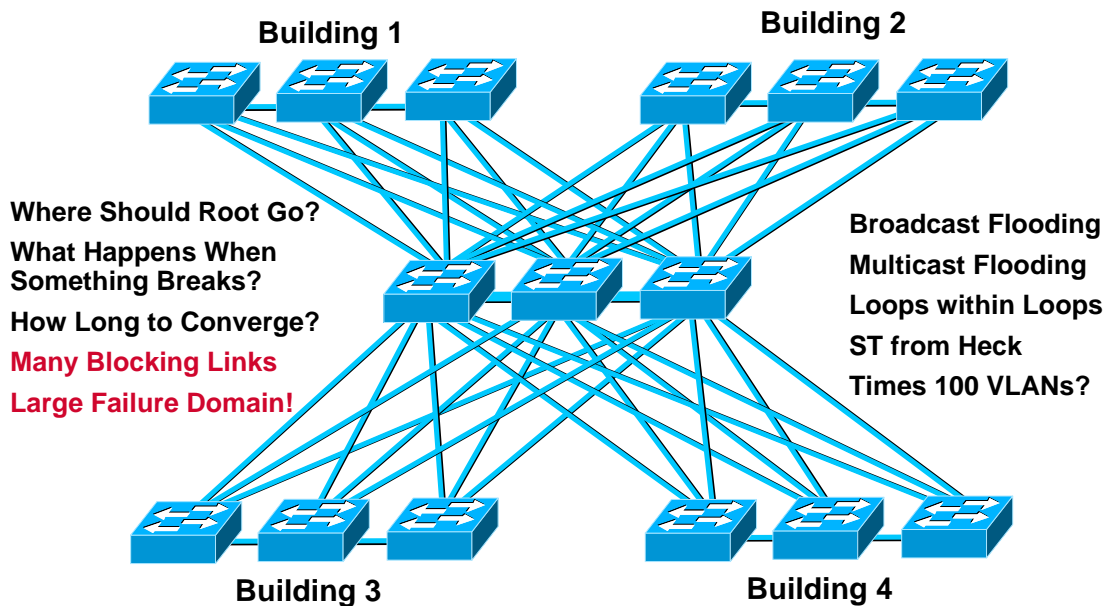
Cisco.com

12

Concept 2—Minimize STP Failure Domain

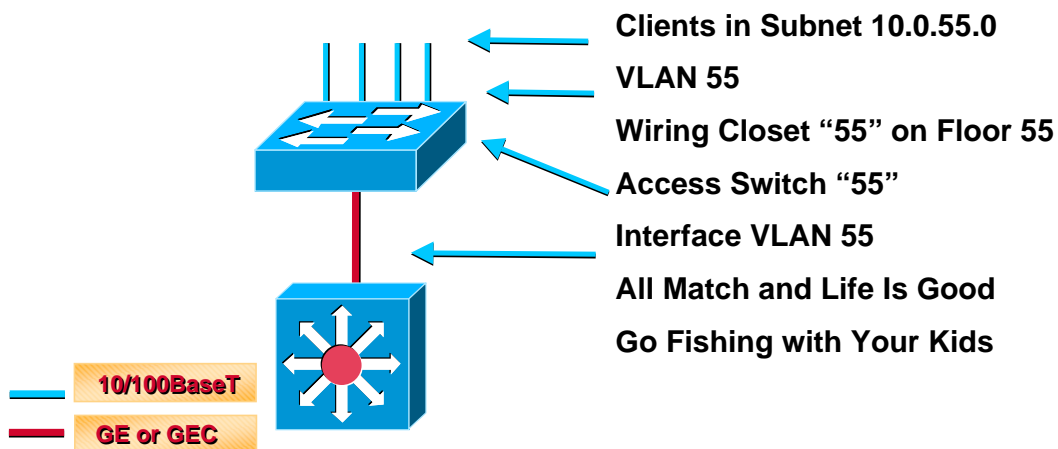
Public Enemy Number One

Avoid Highly Meshed, Non-Deterministic Large Scale L2 = VLAN Topology



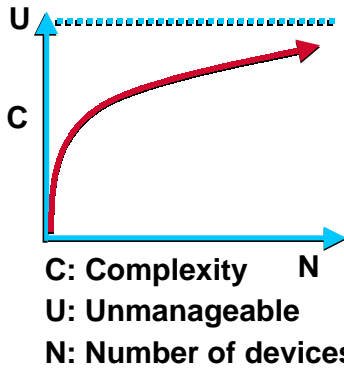
Concept 3—Map L3 to L2 to L1

- Easy administration and troubleshooting

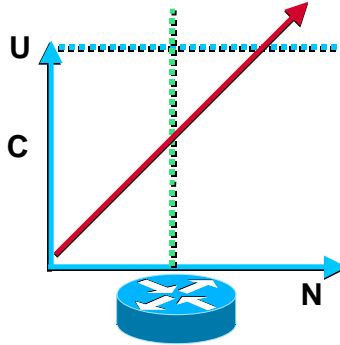


Concept 4—Scaling and Hierarchy

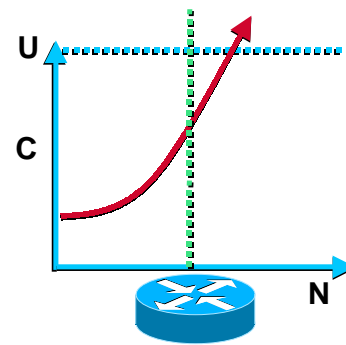
Strong Hierarchies Like Telephone System and Internet Segment Addressing and Therefore Scale



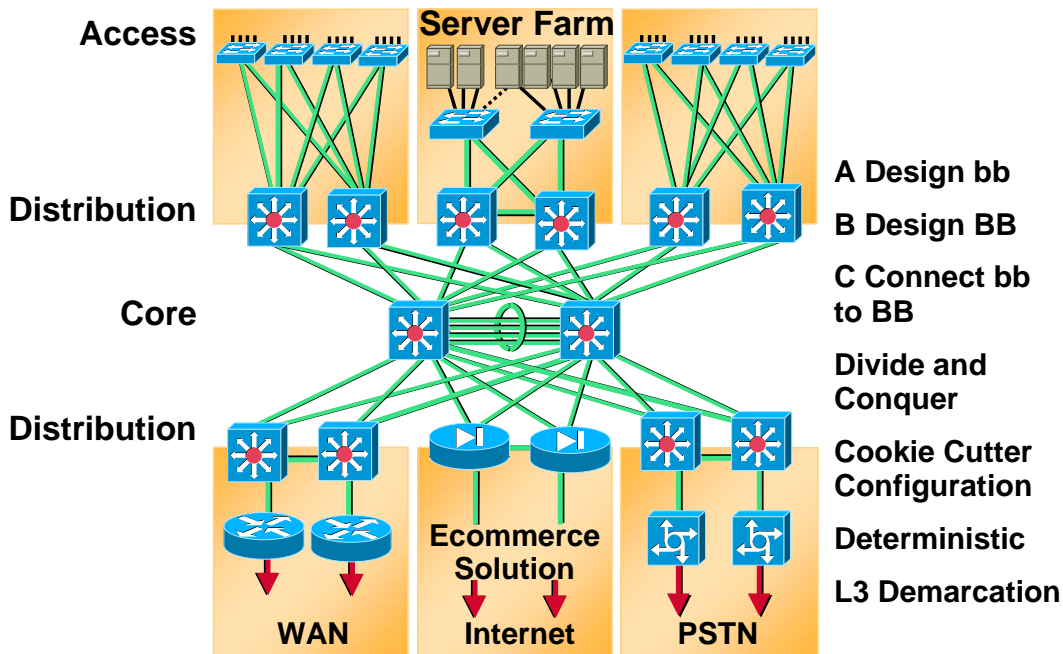
Flat L2 Ethernet Is Easy but Does Not Scale



ATM LANE Is Logically Flat, Scales as N Squared



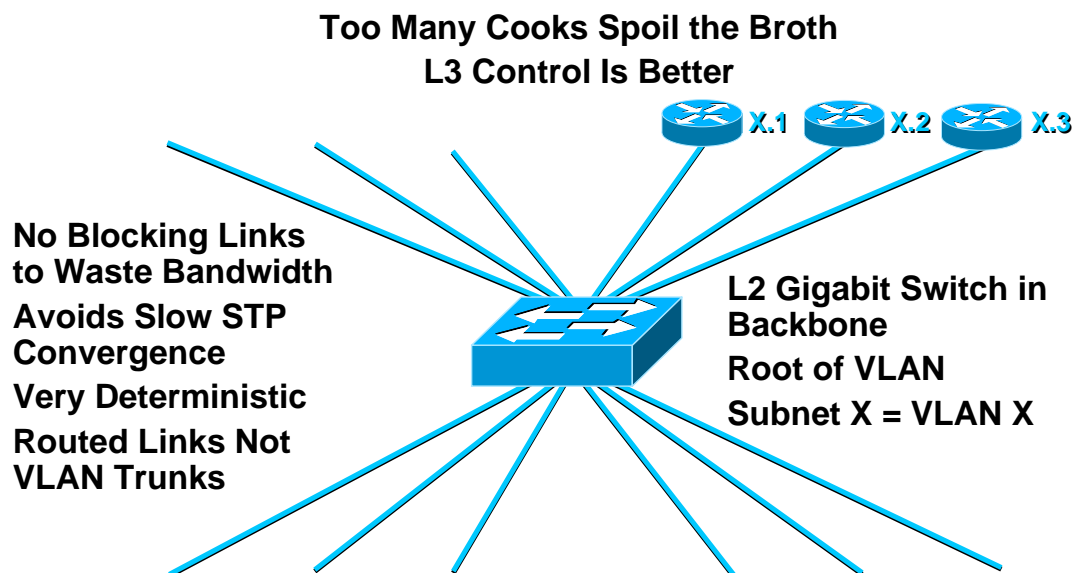
Concept 5—Building Block and Backbone Design ABCs



Ten Rules for High Availability

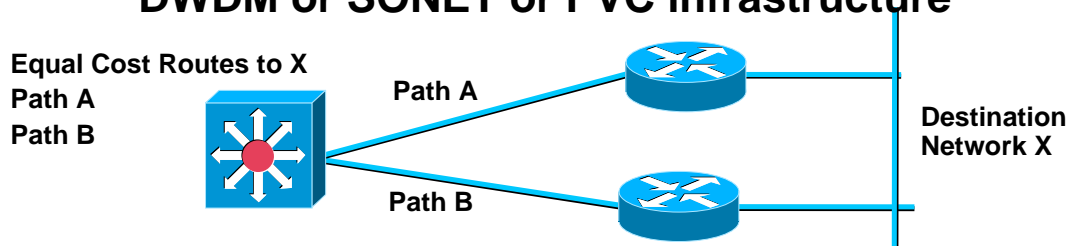
- 1) Eliminate STP loops
- 2) L3 dual-path design
- 3) EtherChannel across cards
- 4) Workgroup servers
- 5) Use HSRP track
- 6) Passive interfaces
- 7) Leave escape route
- 8) Oversubscription guidelines
- 9) HA for single attached servers
- 10) Features and philosophy

Rule 1—Eliminate STP Loops in the Backbone and Mission Critical Points



Rule 2—Dual Equal-Cost Path L3

- **Load balance-don't waste bandwidth**
Unlike L1 and L2 redundancy
- **Fast recovery to remaining path**
Detect L1 down and purge—1 to 2 seconds
- **Works with any routed fat pipes**
Gigabit Ethernet or EtherChannel
DWDM or SONET or PVC infrastructure



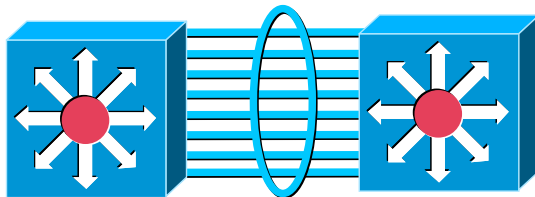
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

19

Rule 3—EtherChannel® Across Cards

- **Increased availability**
- **Sub second recovery**
- **Single L2 STP link**
- **Single L3 subnet**
- **Small complexity increase**
Less if channel set “on”



Catalyst 6500
Up to 8 Ports in Channel
L2 Switched
VLAN Trunk
L3 Switched Subnet

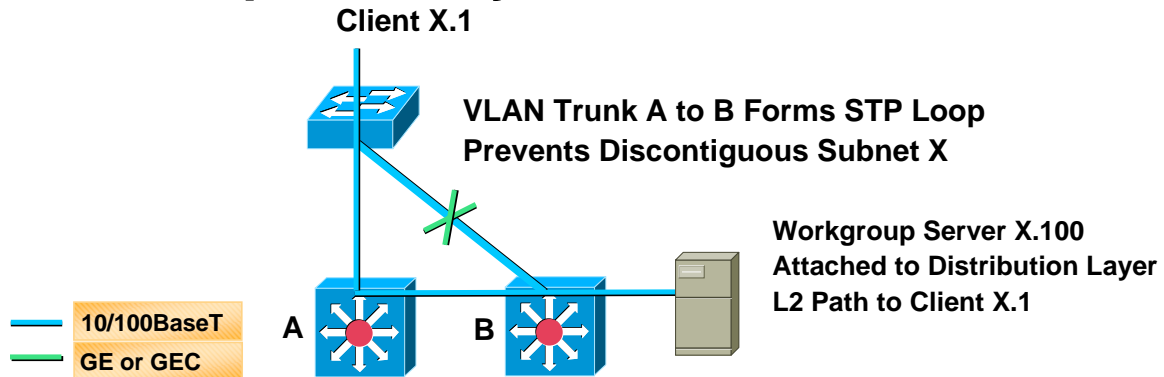
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

20

Rule 4—Connect Workgroup Server

- Enterprise server farms are better
- L3 demarcation is better
- L2 recovery path prevents black hole
- Example of why extended L2 is bad



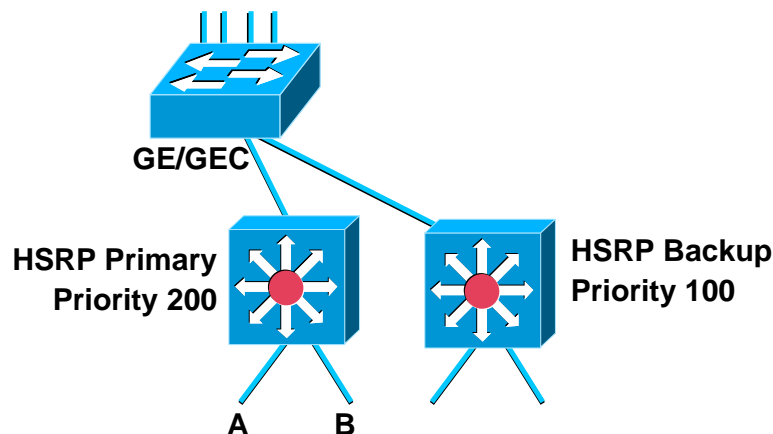
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

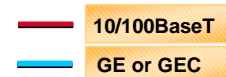
21

Rule 5—Use HSRP Track

- Shortest path-best outbound gateway



Track Interface A: Lower Priority 75
Track Interface B: Lower Priority 75
HSRP Fails over If Both A and B Lost



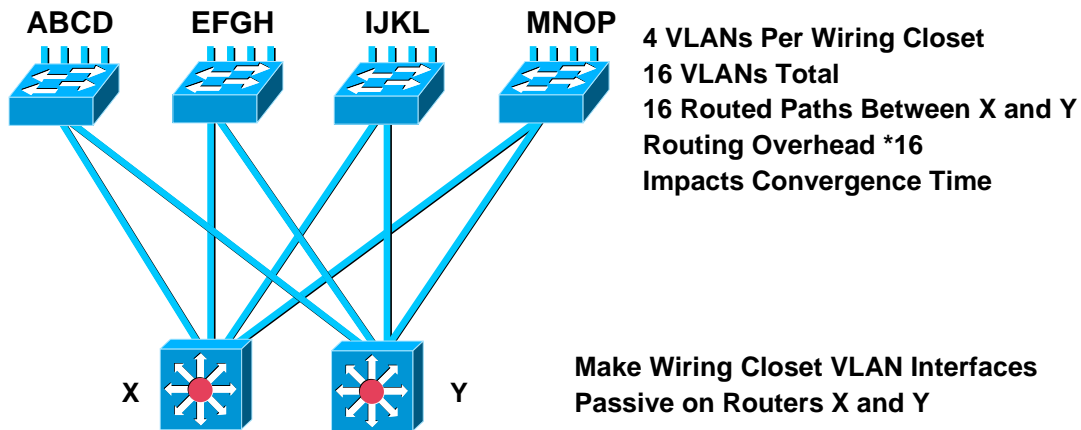
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

22

Rule 6—Use Passive Interfaces

- Routers peer across each wiring closet VLAN
- Passive interfaces reduce overhead
- CDP, STP, VTP still exchanged



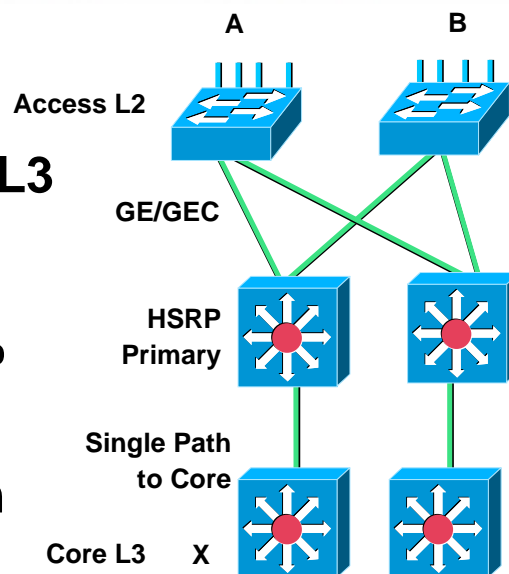
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

23

Rule 7—Leave Escape Routes

- Problem with single path designs
- Routing black hole if L3 engine on X reloads
- Lights are on but nobody home—HSRP does not recover
- Remove passive from two wiring closet subnets A, B for backup routes



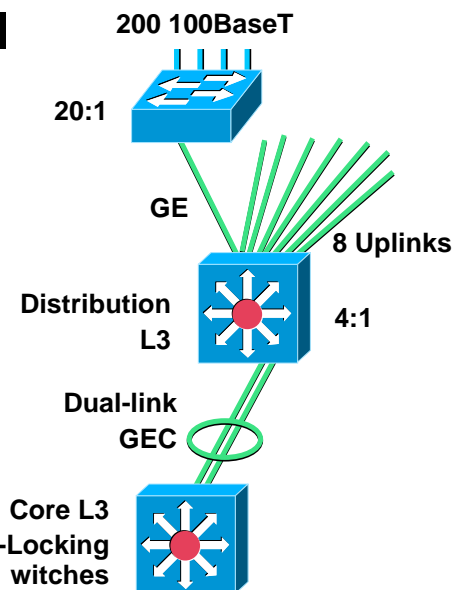
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

24

Rule 8—Oversubscription Guidelines

- Oversubscription part of all networks—not bad
- QoS required IFF congestion occurs
- Protect real time flows at congested points
- Over-subscription rules of thumb work well
- 20:1 at wiring closet
- Less in distribution and server farm



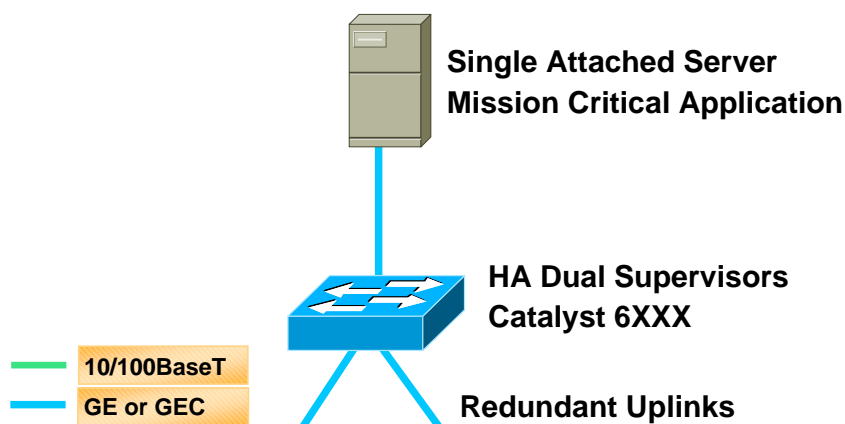
2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

25

Rule 9—Dual Supervisors—HA for Single Attached Servers

- Single point of failure
- Dual supervisors-fast stateful recovery
- No increase in complexity



2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

26

Building Block Templates

1) Standard model

Simple, structured

2) VLAN model

More flexible

3) Large scale server farm model

Accommodate dual NIC

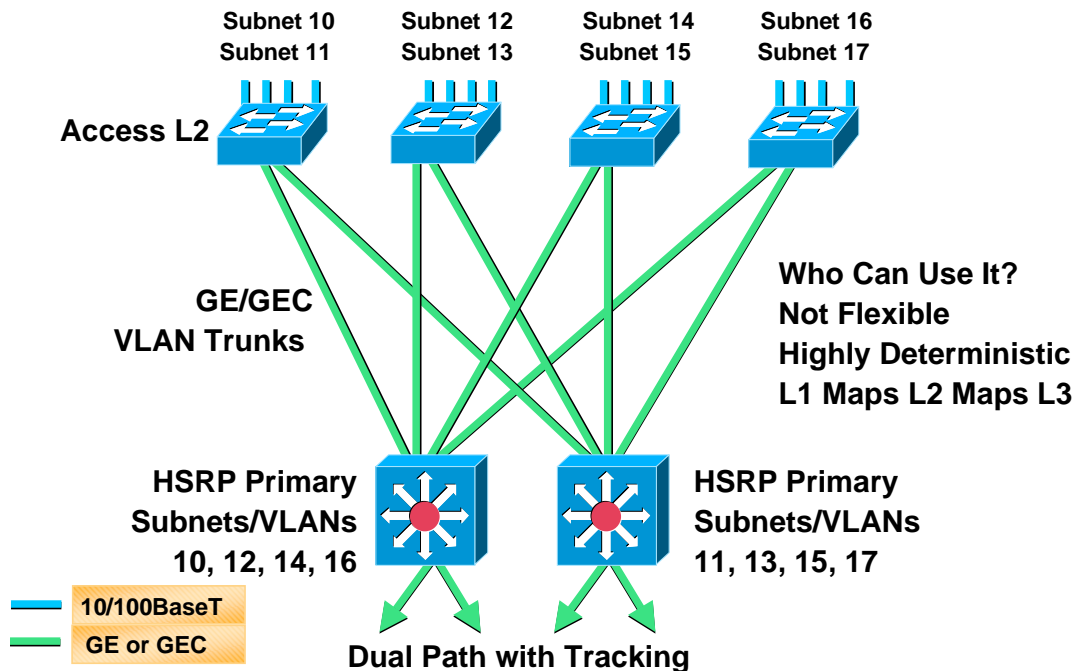
4) Small scale server farm model

Accommodate dual NIC

5) E-commerce front end

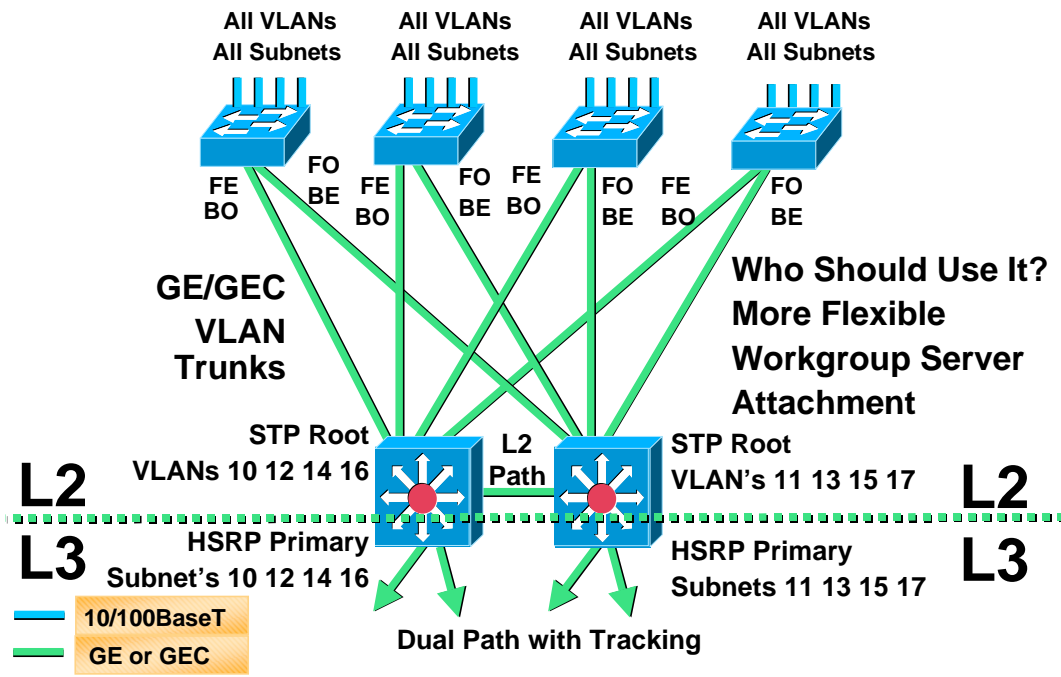
Web hosting B-C

1—Standard Building Block No Loops—No STP Complexity



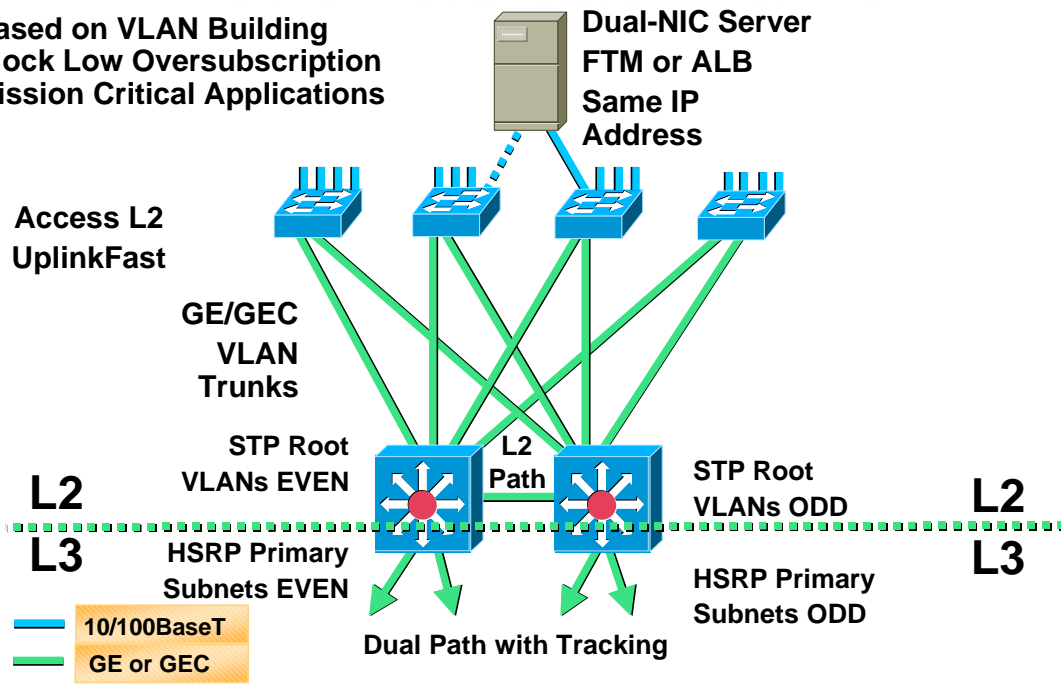
2—VLAN Building Block

Make L2 Design Match L3 Design All VLANs Terminate at L3 Boundary



3—Large-Scale Server Farm Building Block

Based on VLAN Building Block
Low Oversubscription
Mission Critical Applications



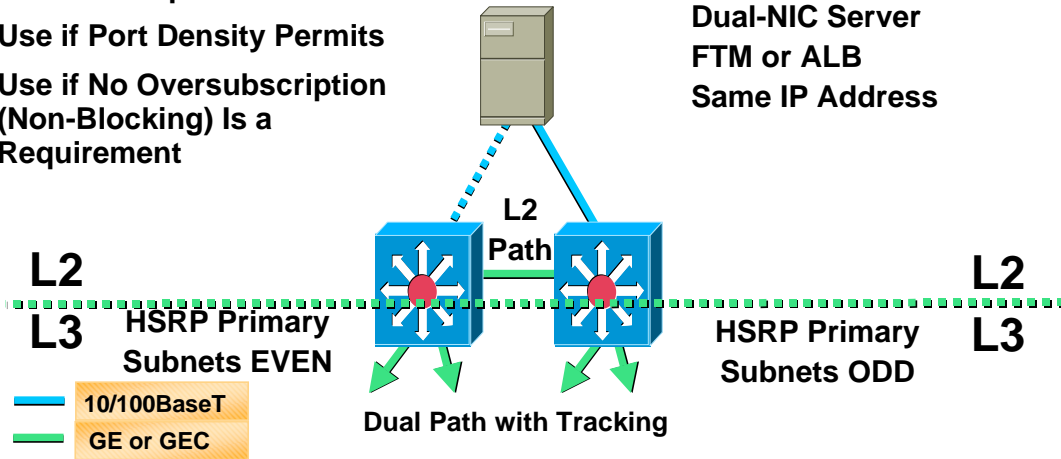
4—Small-Scale Server Farm Building Block

Simplified Building Block with No STP Loops

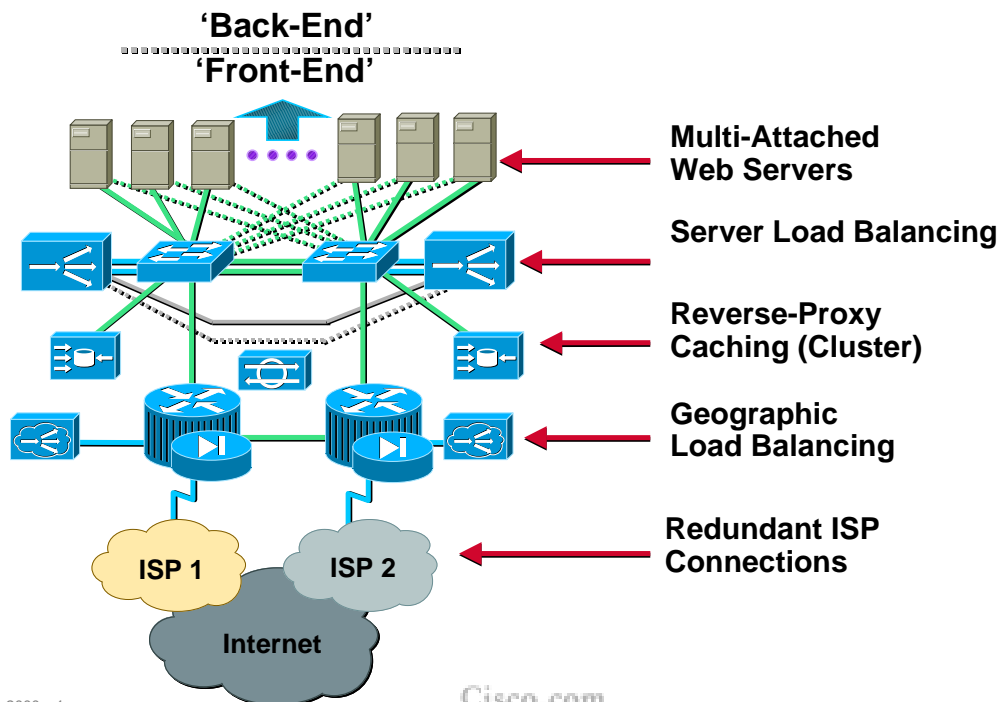
Use if Port Density Permits

Use if No Oversubscription (Non-Blocking) Is a Requirement

Dual-NIC Server
FTM or ALB
Same IP Address



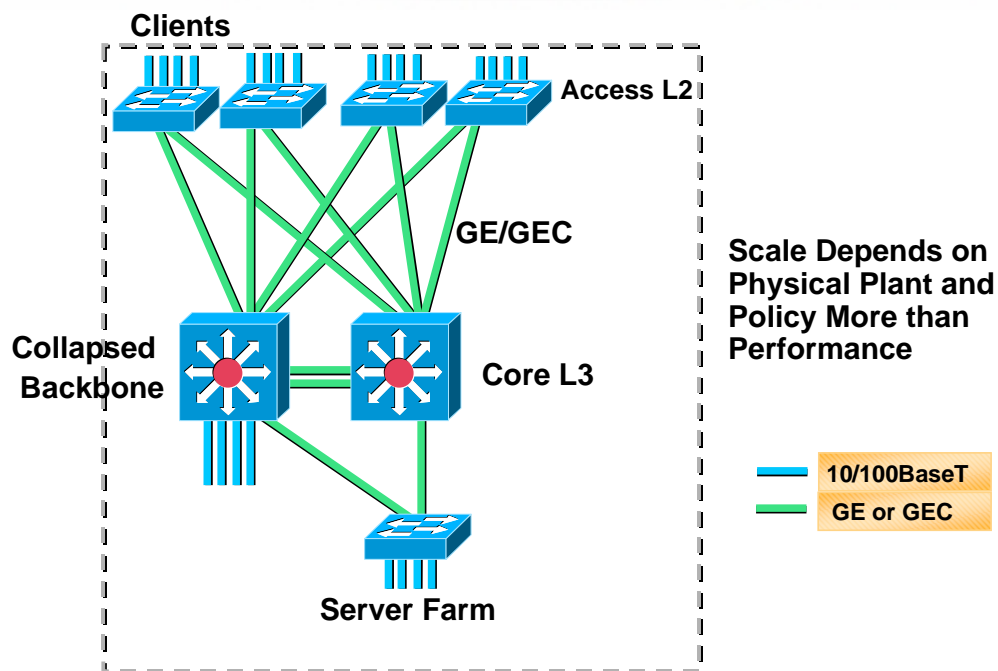
5—E-Commerce Front End



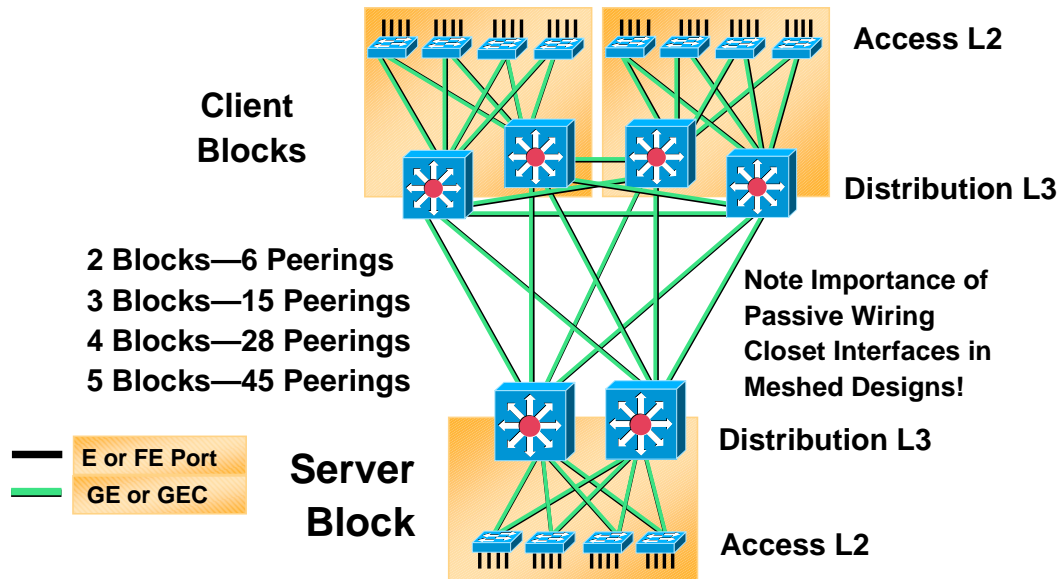
Redundant Backbone Models Increasing Scale

- 1) Collapsed L3 backbone
- 2) Full mesh
- 3) Partial mesh
- 4) Dual-path L2 switched
- 5) Dual-path L3 switched

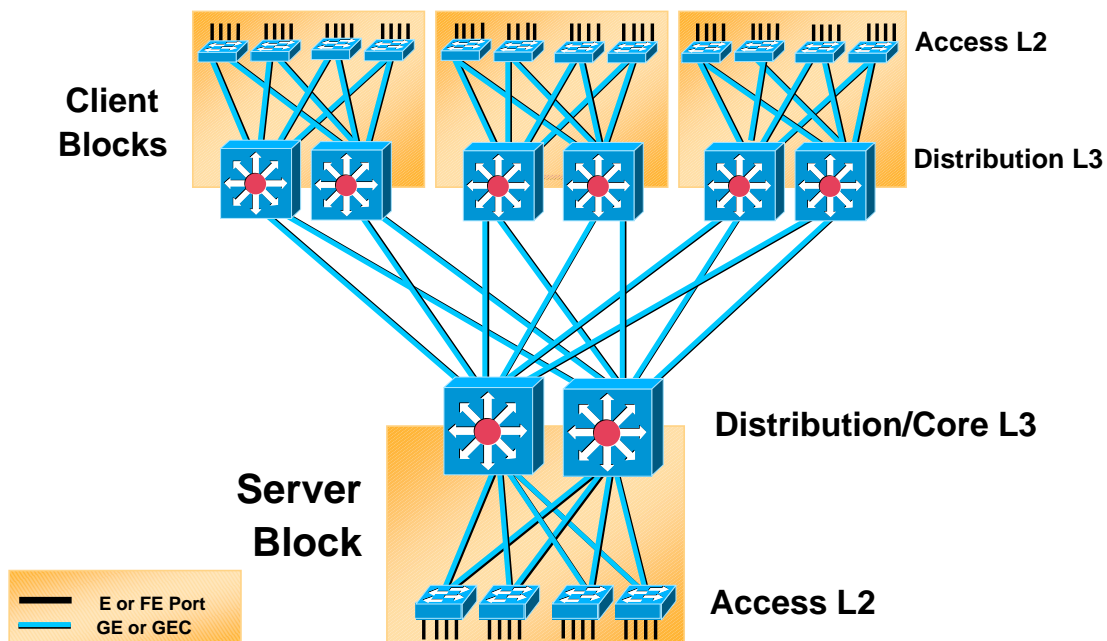
1—Collapsed L3 Backbone Large Building or Small Campus



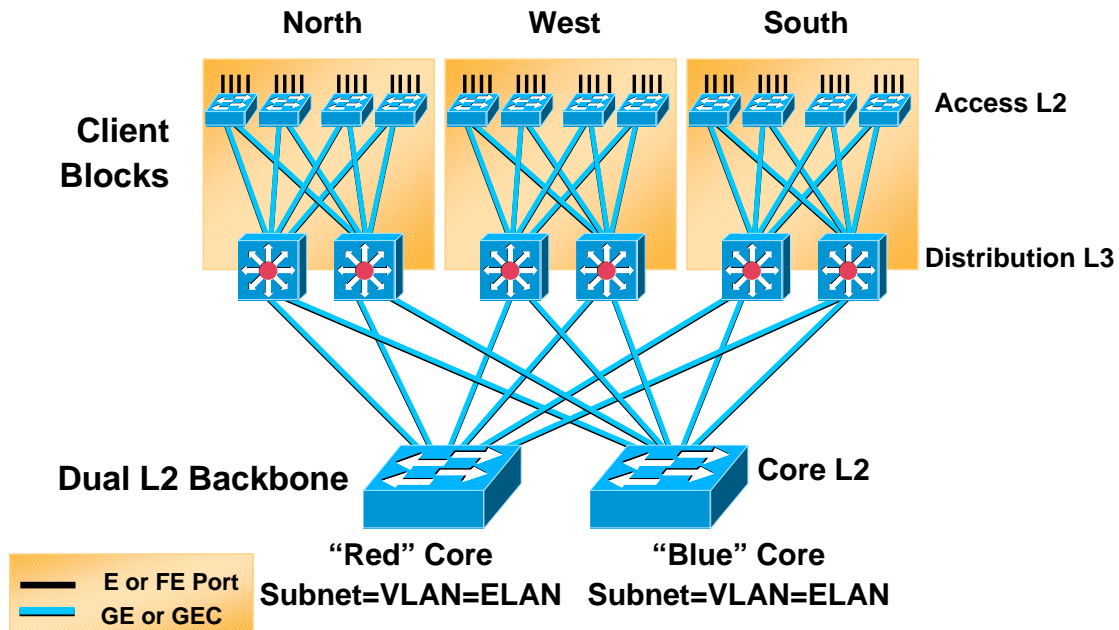
2—Full Mesh Backbone Small Campus—N Squared Limitation



3—Partial Mesh Backbone Medium Campus—Traffic Flow to Server Farm



4—Dual-Path L2 Switched Backbone No STP Loops or VLAN Trunks in Core



2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

39

Benefits of a L3 Backbone

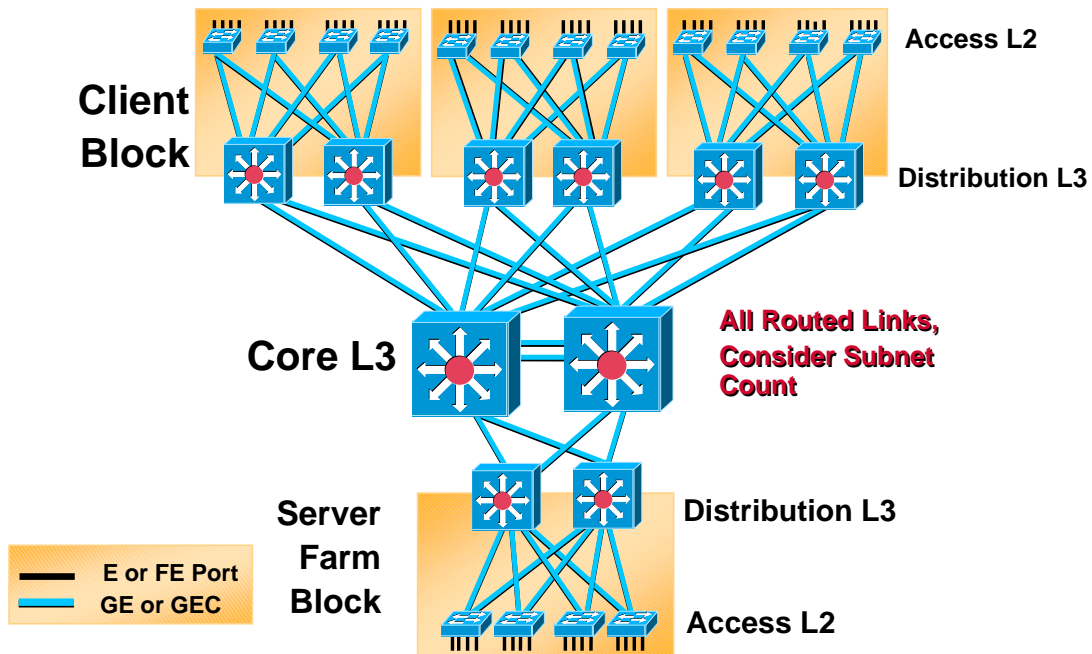
- Multicast PIM routing control
- Load balancing
- No blocked links
- Fast convergence EIGRP/OSPF
- Greater scalability overall
- Router peering reduced
- Cisco IOS features in the backbone

2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

40

5—Dual-Path L3 Backbone Largest Scale, Intelligent Multicast



2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

41

QoS Enabled Gigabit Campus Backbone

- **Catalyst 6000 and 6500**

Classify by port, any
L3/L4 information

Three queues—one for
real time

Two data queues with
two WRED thresholds

WRR scheduling

Mark and drop per flow

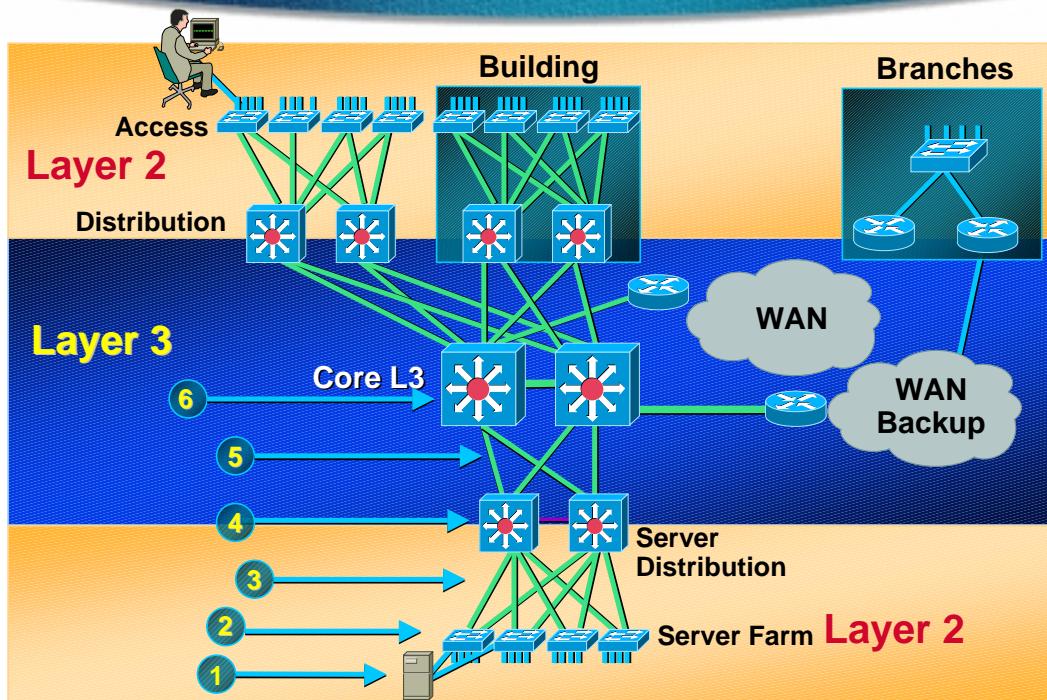


2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc.

Cisco.com

42

Failure Analysis Exercise



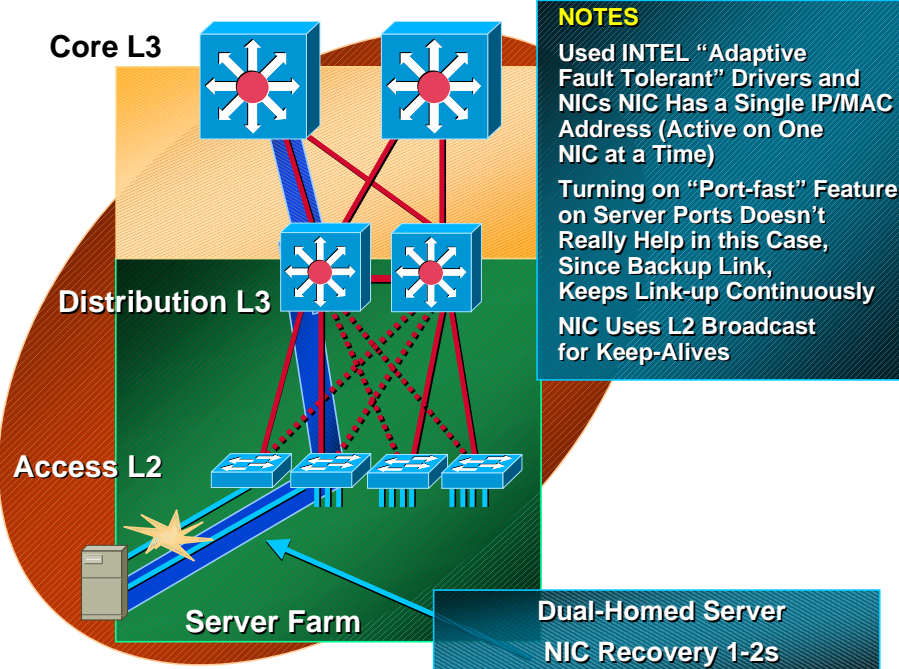
Network Recovery Modes

1, 2	Server NIC	<2 seconds
3, 4	UplinkFast L2 HSRP L3	2 sec + 3 sec (tune)
5, 6	HSRP track	3 sec (tune)
dual-path channel	dual-path L3 EtherChannel	<2 seconds <1 second
L3 general	EIGRP/OSPF	10-20 s (tune)
L2 general	STP	30s or 50s

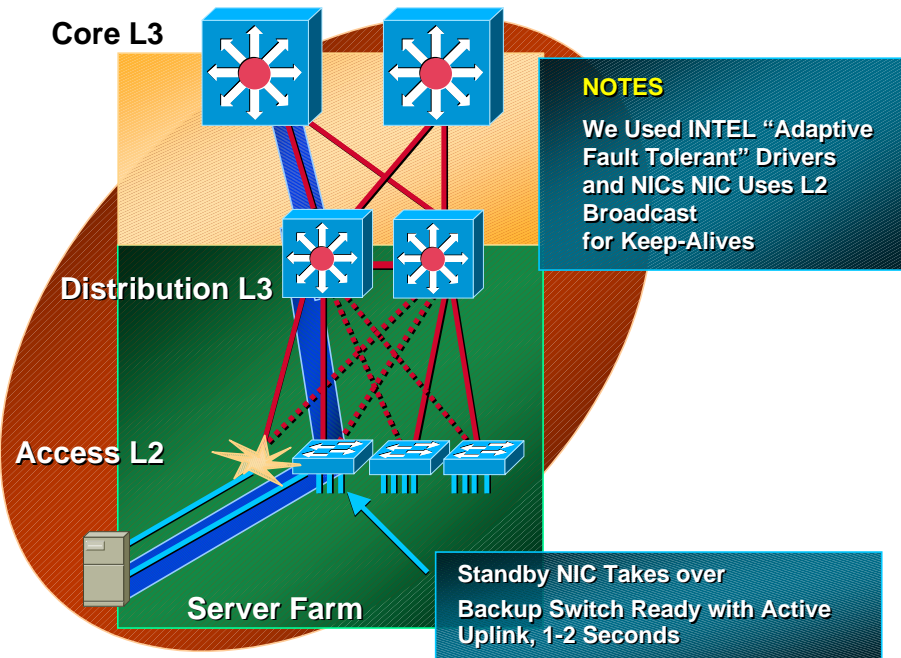
Restore Considerations

- Restoring can take longer in some cases—more complex-schedule
- When powering up L1 may come up before L3 builds routing table-temporary black hole for HSRP
- Use “preempt delay” for HSRP

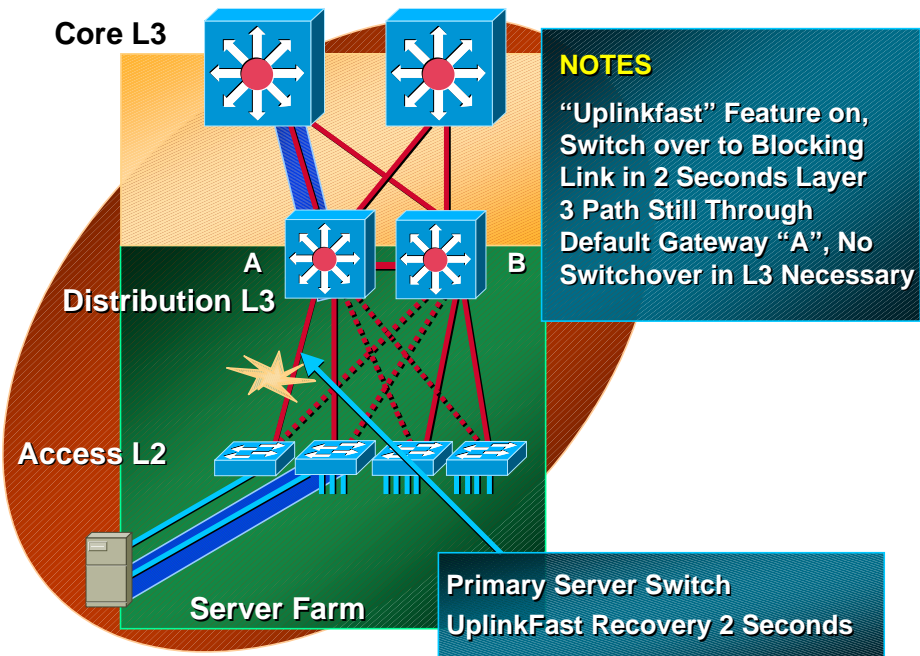
Failure Analysis Exercise Server NIC



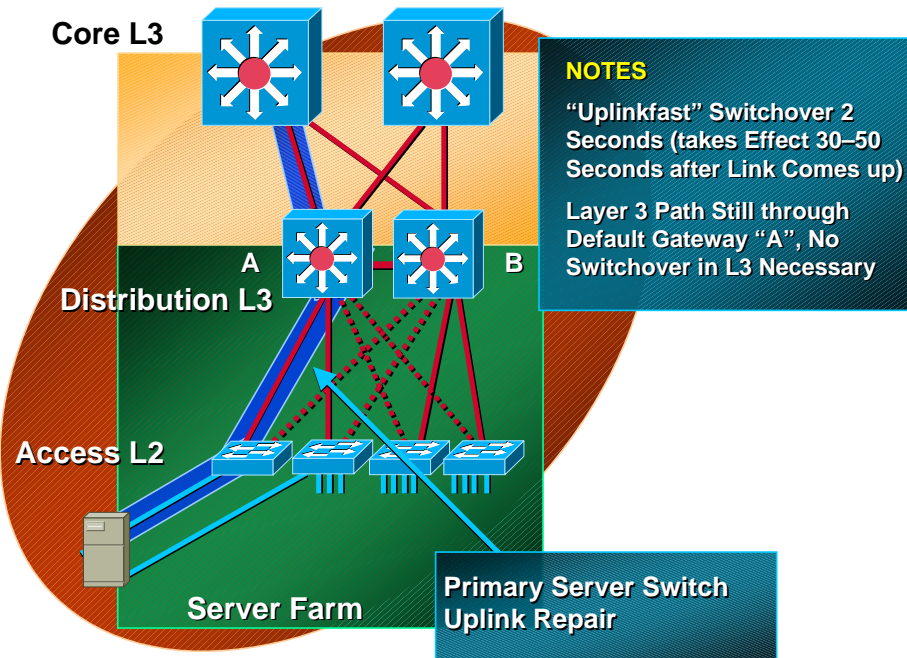
Failure Analysis Exercise Primary Server Switch



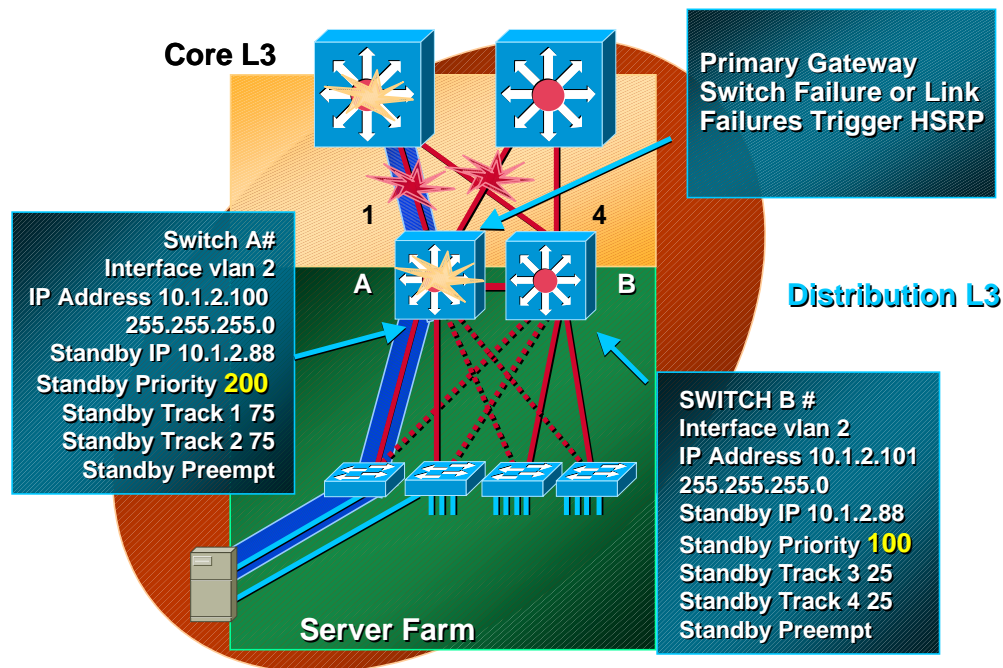
Failure Analysis Exercise Server Switch Uplink



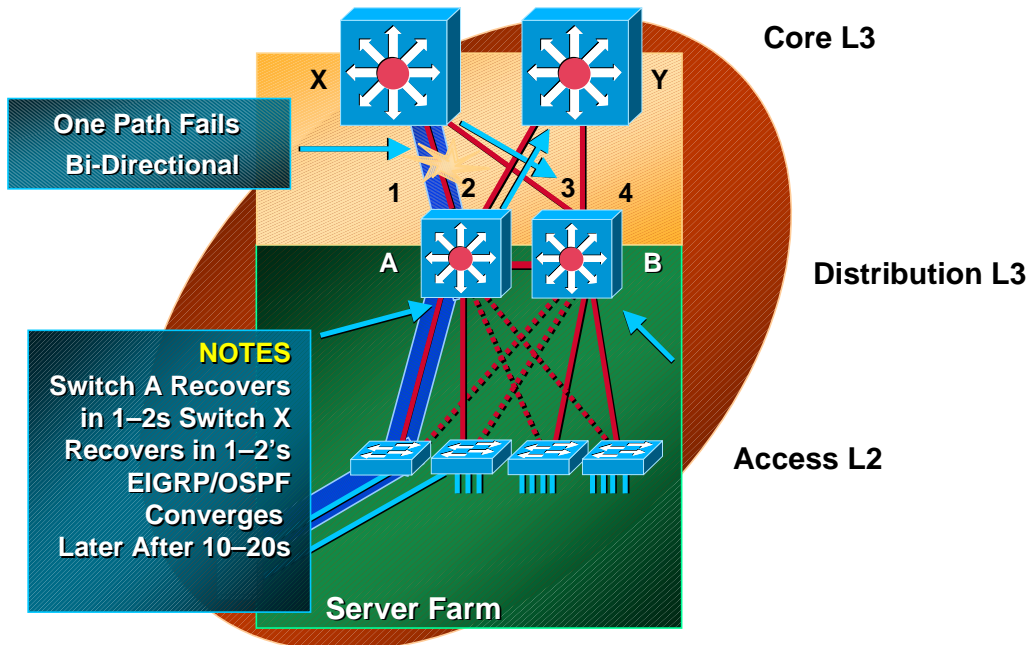
Failure Analysis Exercise Server Switch Restore



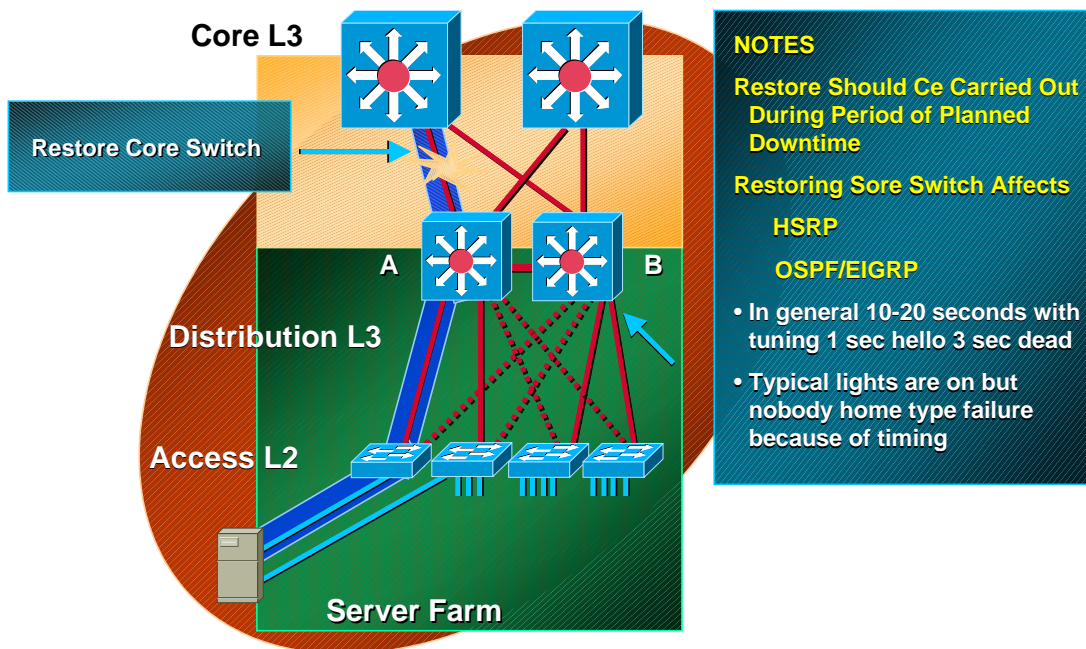
Failure Analysis Primary Gateway Router HSRP



Failure Analysis Primary Gateway Router HSRP



Failure Analysis Restore Core Switch



Campus Fail-Over Layer 2 Recovery and Tuning

- **STP**
Tune 'diameter' to 2 on root switch
Improves recovery by reducing ST to a triangle
- **PortFast**
Access/desktop ports only 1 s
Move directly from linkup into forwarding
- **UplinkFast**
No tuning—2 seconds—wiring closet only
Only applies to VLANs with loop (triangle)
- **Backbonefast**
Improves convergence (1–2 sec + 2xFwd_delay)
for indirect link failures
Eliminates maxage timeout

Campus Fail-Over Layer 3 Recovery and Tuning

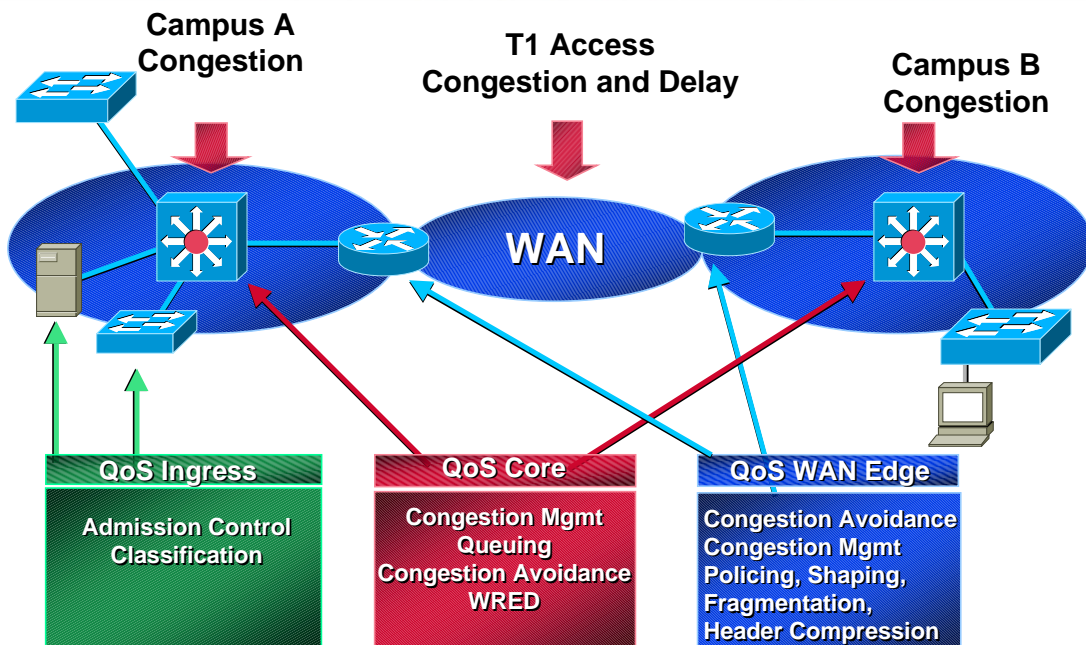
- **Caution with aggressive tuning**
Good when stable, highly summarized
- **HSRP (fast LAN links)**
Tune hello timer 1 sec, dead timer 3 sec
<4s to converge
- **OSPF (fast LAN links)**
Tune hello timer 1 sec, dead timer 3 sec
<4s to recognize problem, then converge
- **EIGRP (fast LAN links)**
Tune hello timer 1 sec, hold timer 3 sec
<4s to recognize problem, then converge

QoS in the Campus

HA at Micro Level for Applications

- Required if end-to-end packet VoIP delay > 150 ms
- Required if congestion causes high packet loss for mission critical data applications
- Requires system approach

Architect for Guaranteed QoS



Guaranteed QoS for VoIP Delay Budget 150 ms

- **Tx delay inverse to link speed**
 - For 1500B packet
 - 8 ms at T1 speed
 - 12 ms at GE speed
- **Requires separate queue for real-time flows**
- **Real time traffic not bursty—never fills up the queue**
- **Dequeue algorithm does not matter at fast campus link speeds**

QoS for Data in the Campus

- **Relevant to flows in the same queue**
- **Protect high-priority bursty flows from low priority bursty flows**
- **Congestion avoidance WRED does the job**

Keeping Networks **Available!**

- **KISS—eliminate complex L2**
- **ASU—building blocks**
- **Redundant backbone**
- **Redundant L3 paths**
- **L3 segments failure domain**

Related Sessions

- **2214: Introduction to IP Multicast**
- **2208: Deploying IGRP/EIGRP**
- **2205: Deploying OSPF**
- **2301: Deploying QoS**
- **2216: Deploying IP Multicast**
- **QoS for Campus Networks**
- **Network Management Campus**
- **Catalyst 6000 Product Update**



**Designing High-Availability
Campus Networks**

Session 2801

2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc. Cisco.com 61



**Please Complete Your
Evaluation Form**

Session 2801

2801
1213_05_2000_c1 © 2000, Cisco Systems, Inc. Cisco.com 62

CISCO SYSTEMS



EMPOWERING THE INTERNET GENERATIONSM