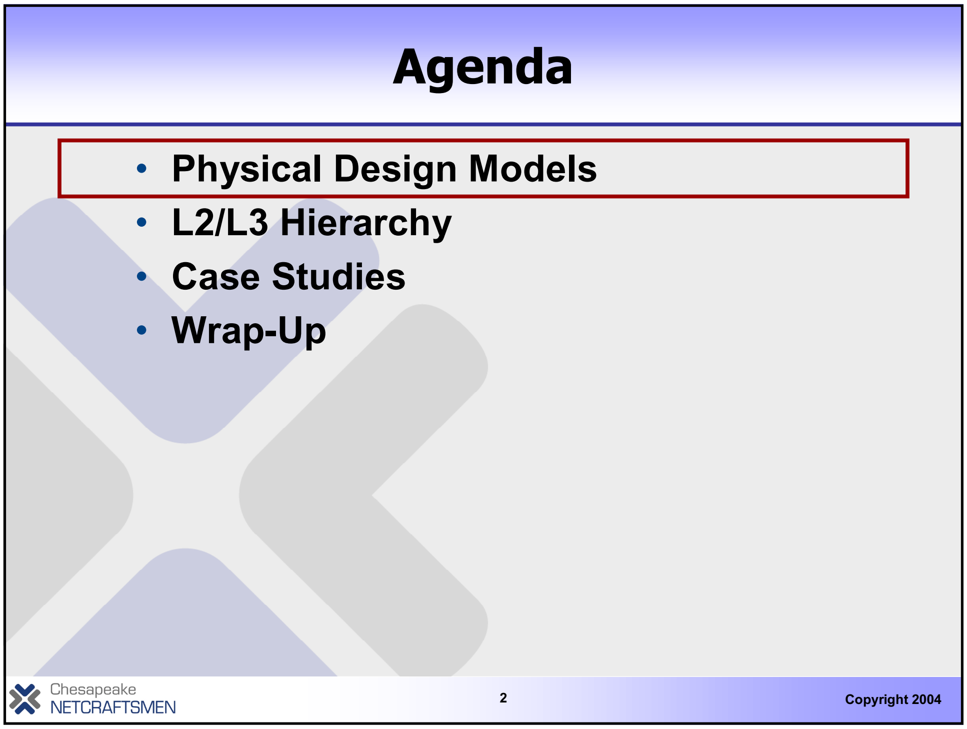
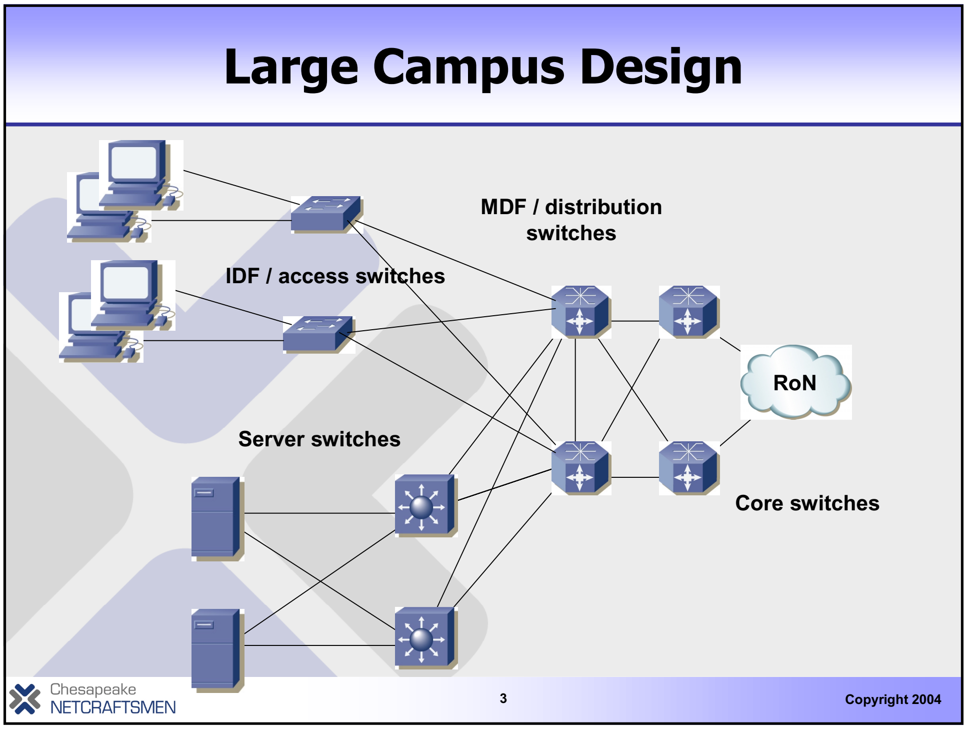
Switching Design Best Practices and Case Studies

Presented by
Dr. Peter Welcher
Chesapeake NetCraftsmen



Agenda

- Physical Design Models
- L2/L3 Hierarchy
- Case Studies
- Wrap-Up



Large Campus Design

MDF / distribution switches

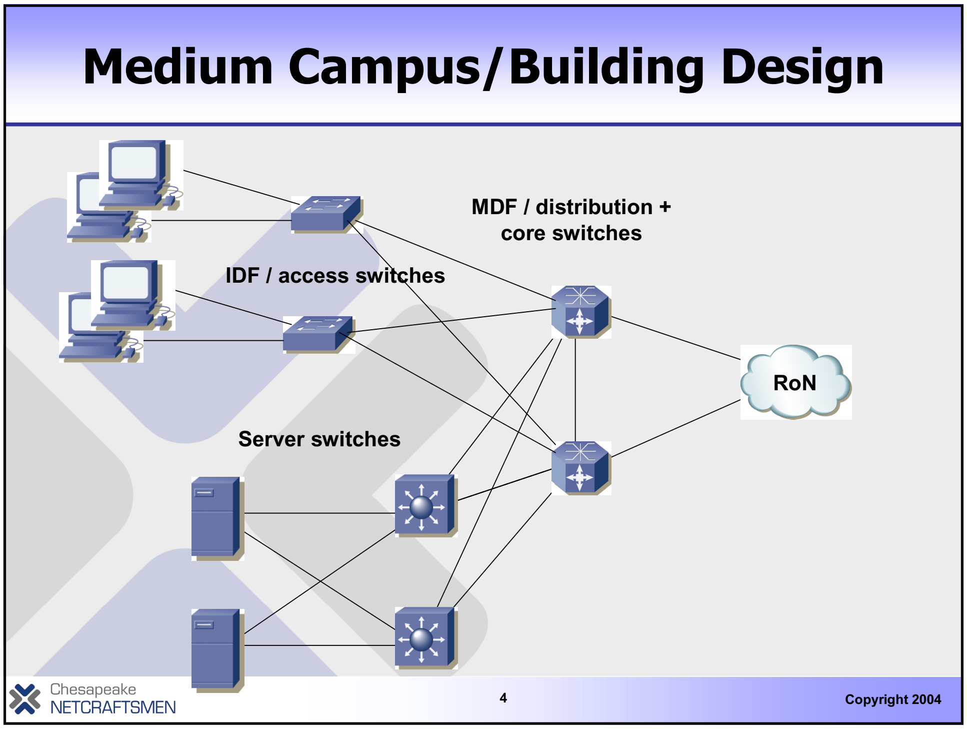
IDF / access switches

RoN

Server switches

Core switches

Chesapeake NETCRAFTSMEN 3 Copyright 2004



Medium Campus/Building Design

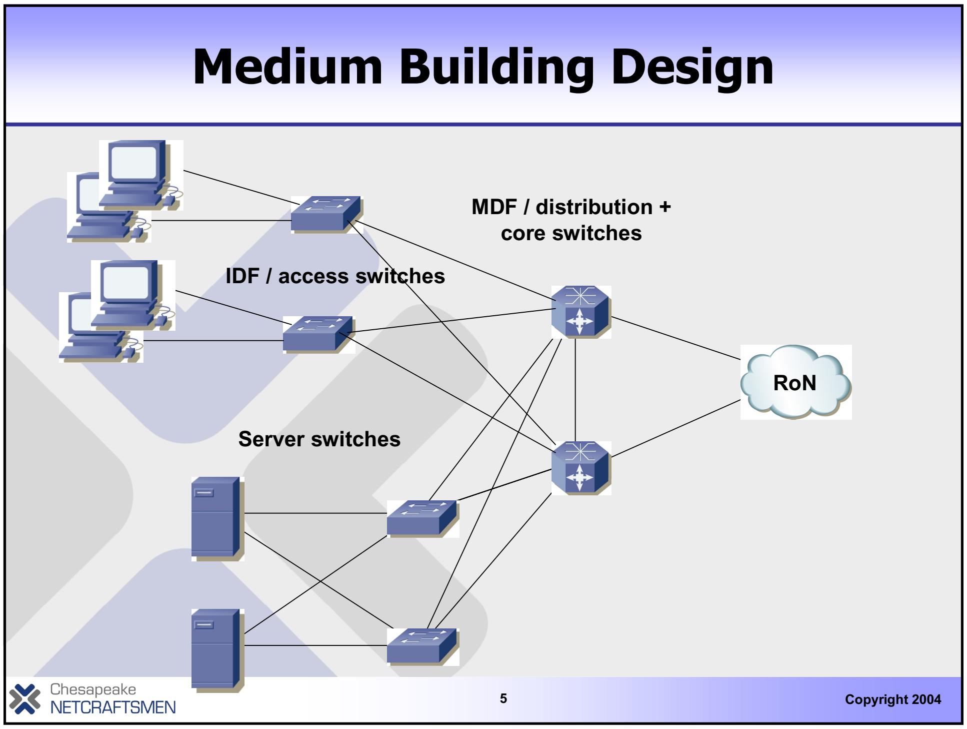
MDF / distribution + core switches

IDF / access switches

RoN

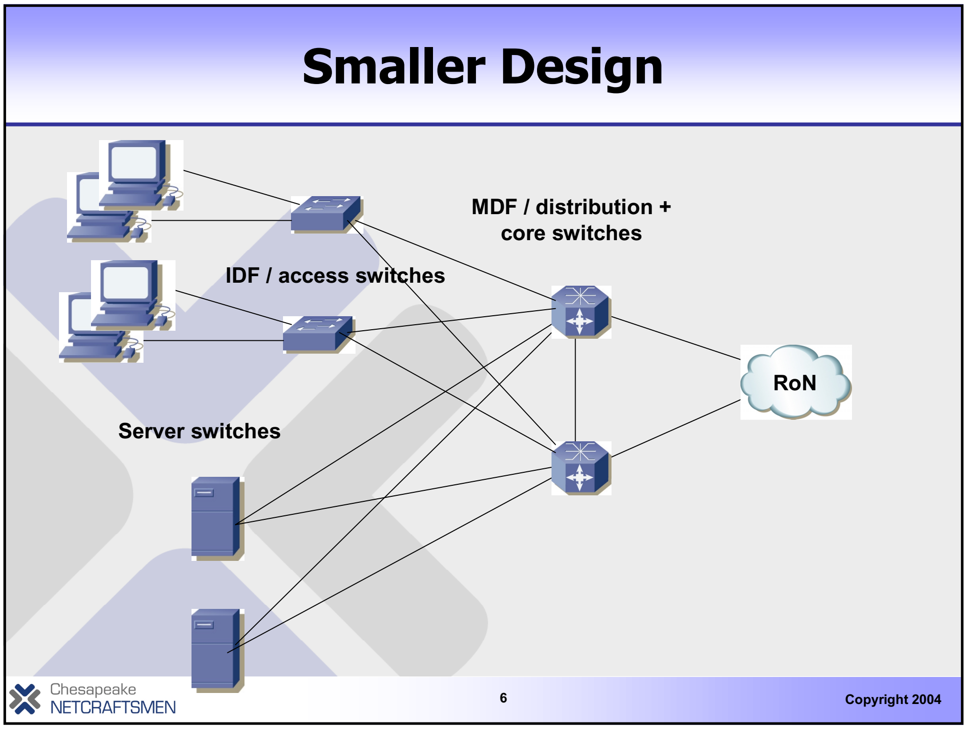
Server switches

Chesapeake NETCRAFTSMEN 4 Copyright 2004



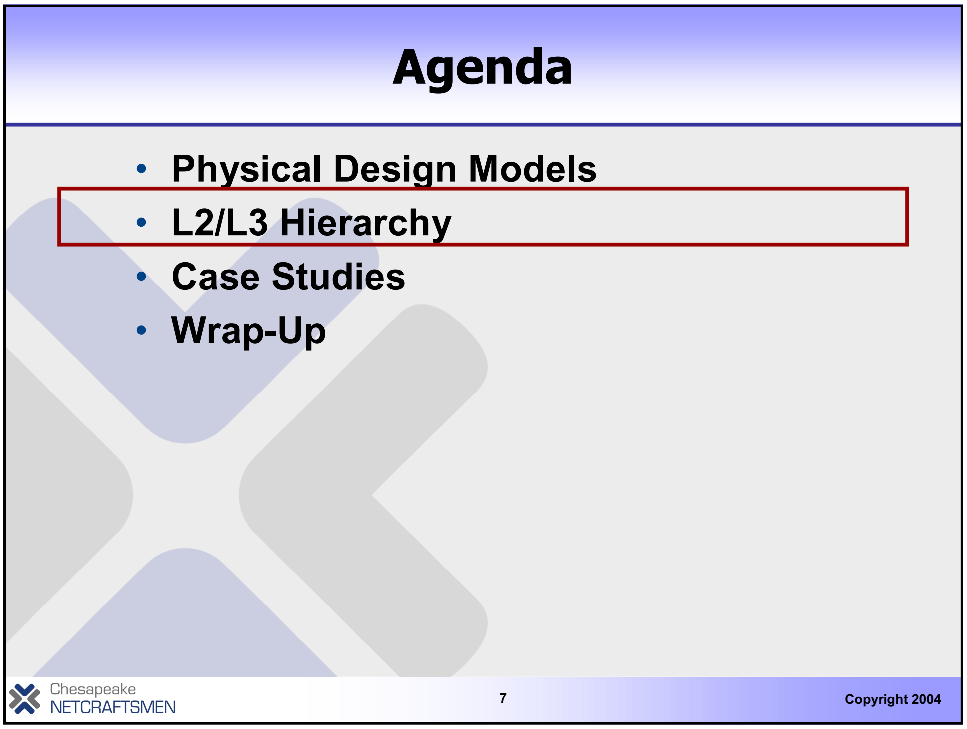
Medium Building Design

Chesapeake NETCRAFTSMEN 5 Copyright 2004



Smaller Design

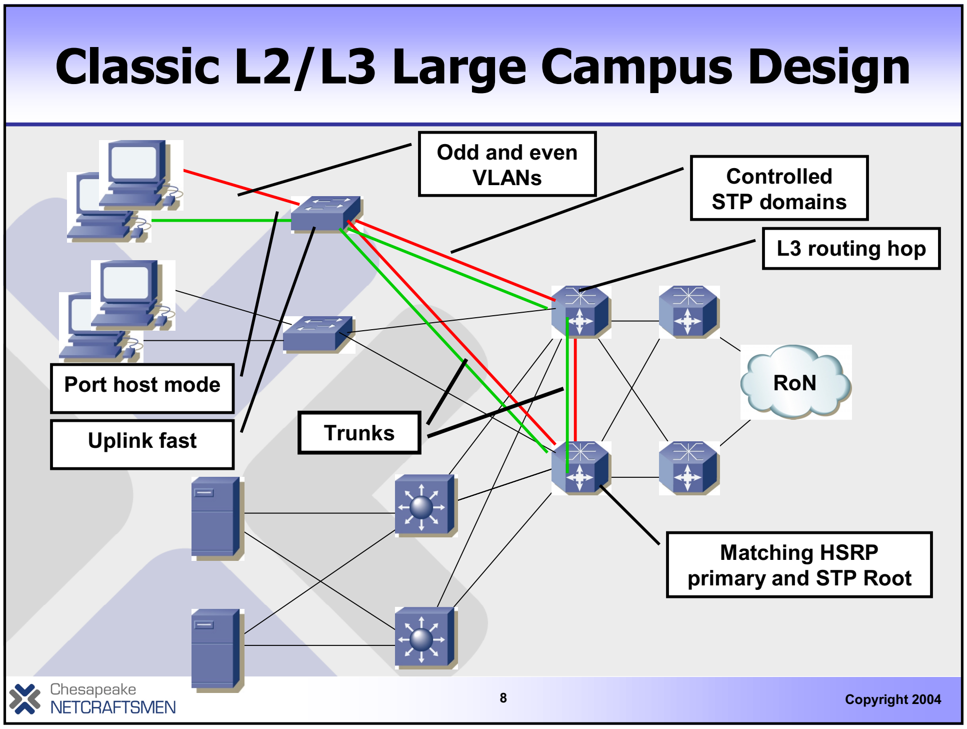
Chesapeake NETCRAFTSMEN 6 Copyright 2004



Agenda

- Physical Design Models
- L2/L3 Hierarchy
- Case Studies
- Wrap-Up

Chesapeake NETCRAFTSMEN 7 Copyright 2004



Classic L2/L3 Large Campus Design

Odd and even VLANs

Controlled STP domains

L3 routing hop

RoN

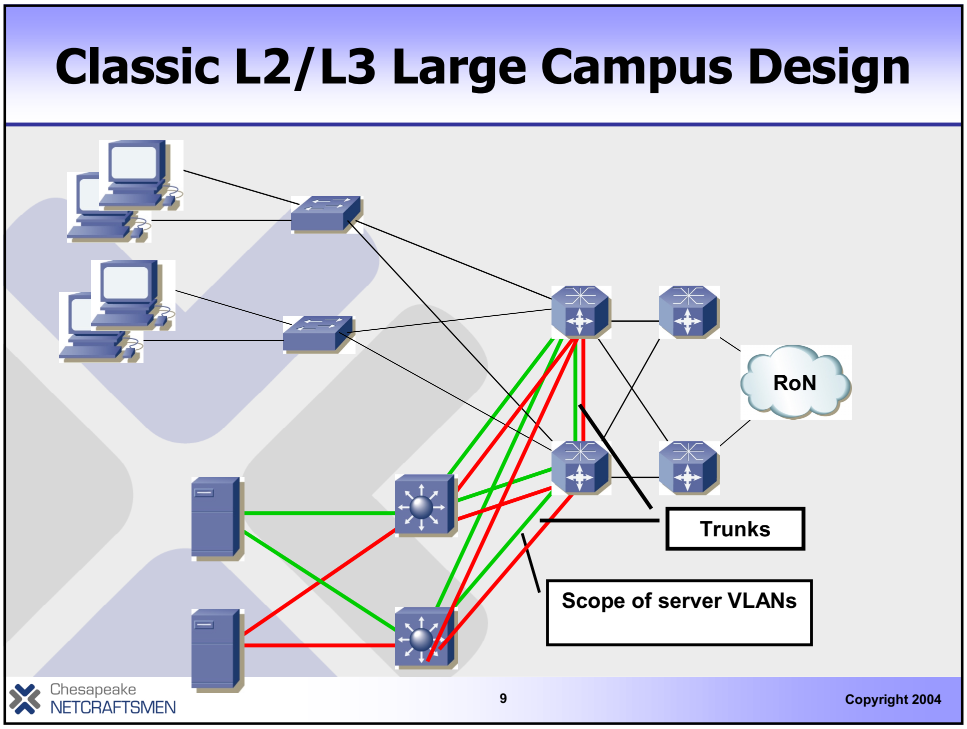
Port host mode

Uplink fast

Trunks

Matching HSRP primary and STP Root

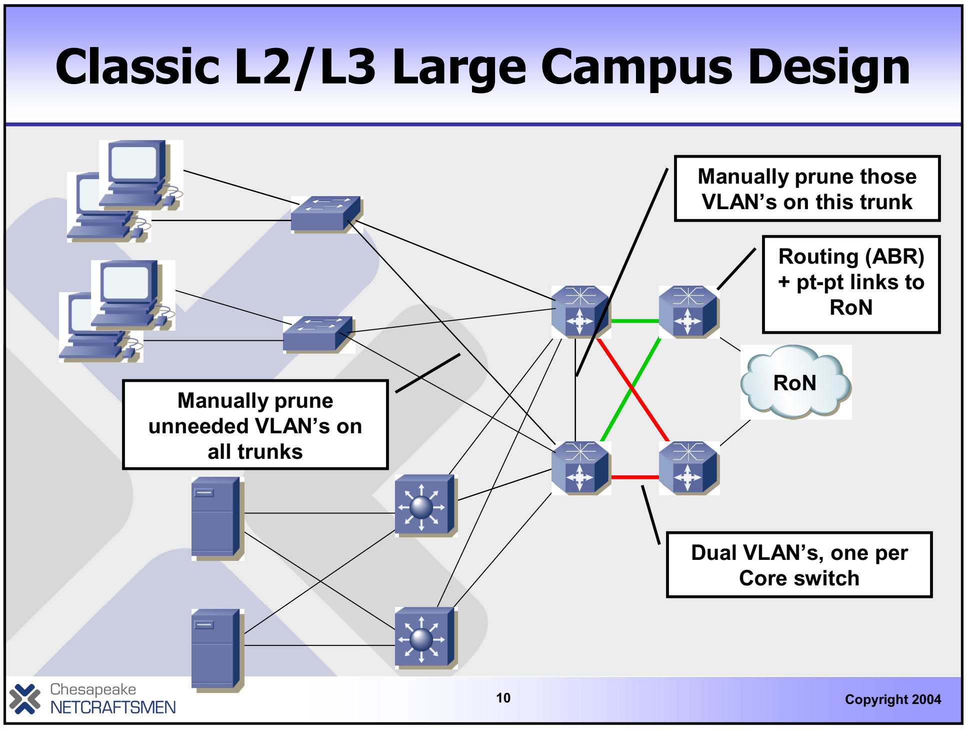
Chesapeake NETCRAFTSMEN 8 Copyright 2004



Classic L2/L3 Large Campus Design

Trunks

Scope of server VLANs



Classic L2/L3 Large Campus Design

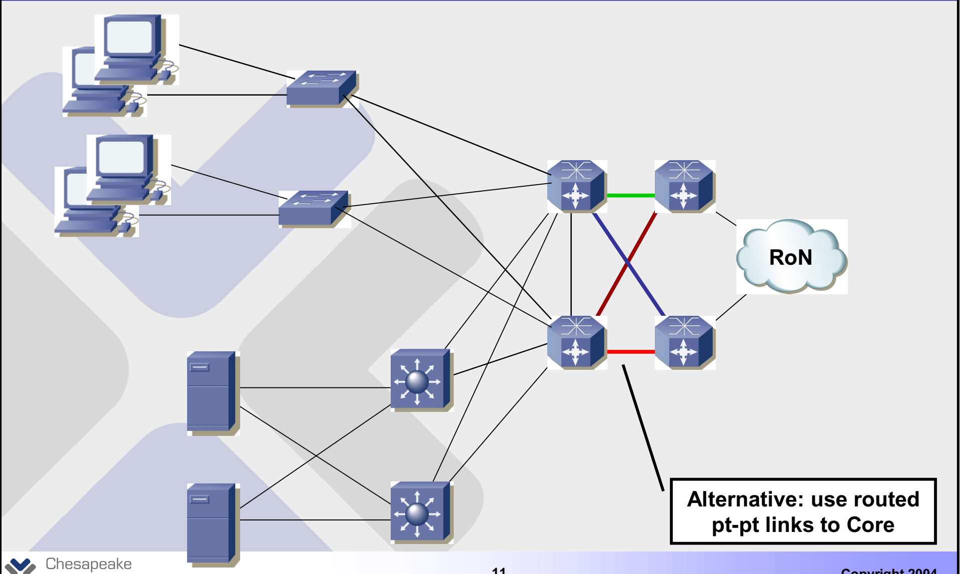
Manually prune those VLAN's on this trunk

Routing (ABR) + pt-pt links to RoN

Manually prune unneeded VLAN's on all trunks

Dual VLAN's, one per Core switch

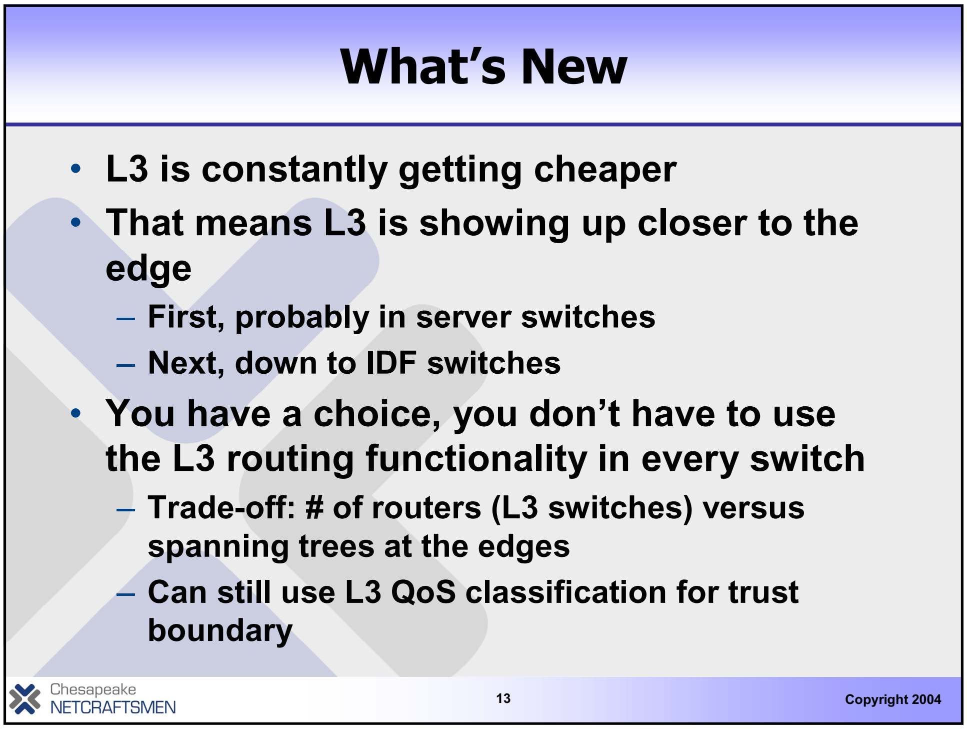
Classic L2/L3 Large Campus Design



Alternative: use routed pt-pt links to Core

Design Thoughts

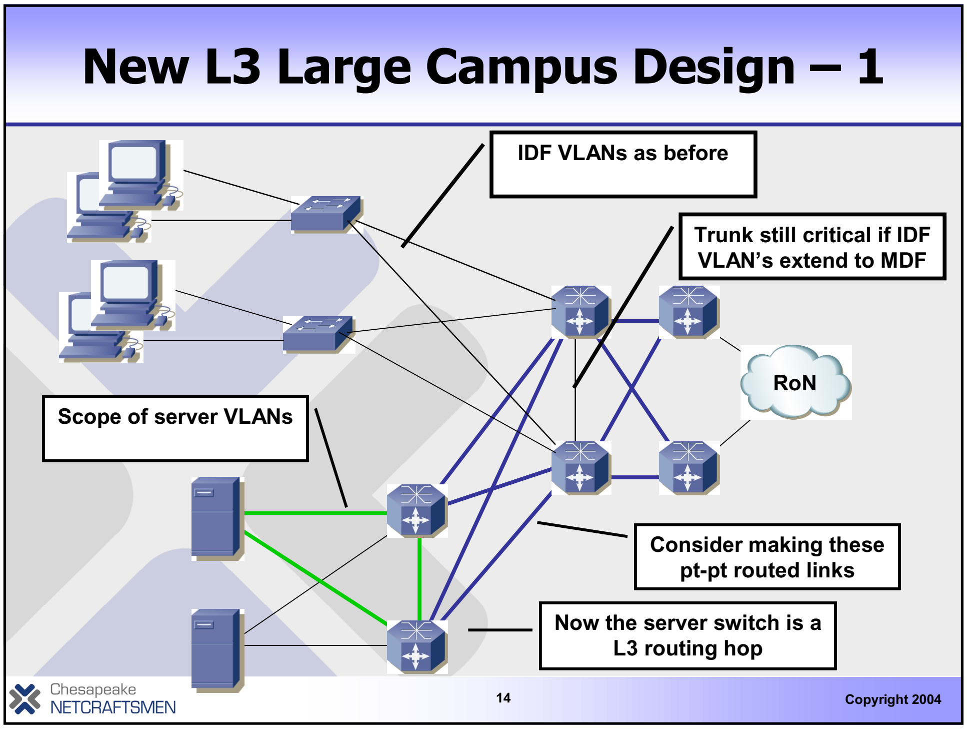
- STP can get ugly
  - Troubleshooting a ST problem is time-consuming and hard
  - Routing hop isolates L2 problems
- Advantages of small VLANs
  - IP address tells you where the device is
  - ST topology and failover behavior are known
  - MAC-based L2 attacks affect only that VLAN
- Disadvantages
  - Still some STP
  - ST loop may well affect MDF-MDF trunks and MDF CPU’s
  - That can cause big problems



What's New

- L3 is constantly getting cheaper
- That means L3 is showing up closer to the edge
  - First, probably in server switches
  - Next, down to IDF switches
- You have a choice, you don't have to use the L3 routing functionality in every switch
  - Trade-off: # of routers (L3 switches) versus spanning trees at the edges
  - Can still use L3 QoS classification for trust boundary

Chesapeake NETCRAFTSMEN 13 Copyright 2004



New L3 Large Campus Design – 1

IDF VLANs as before

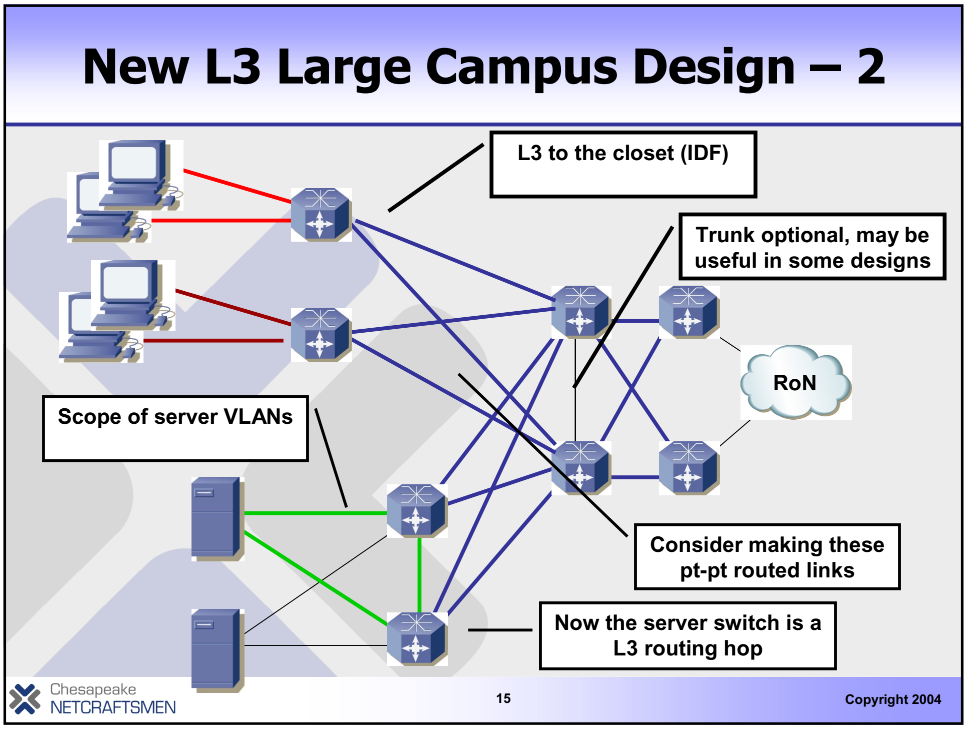
Trunk still critical if IDF VLAN's extend to MDF

Scope of server VLANs

Consider making these pt-pt routed links

Now the server switch is a L3 routing hop

Chesapeake NETCRAFTSMEN 14 Copyright 2004



New L3 Large Campus Design – 2

L3 to the closet (IDF)

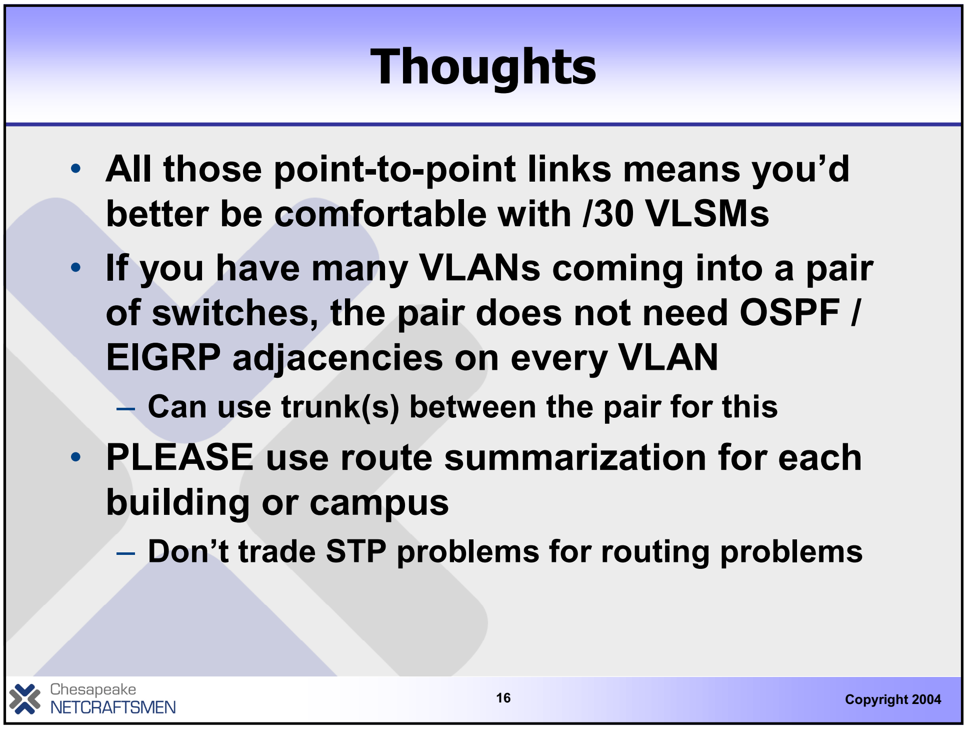
Trunk optional, may be useful in some designs

Scope of server VLANs

Consider making these pt-pt routed links

Now the server switch is a L3 routing hop

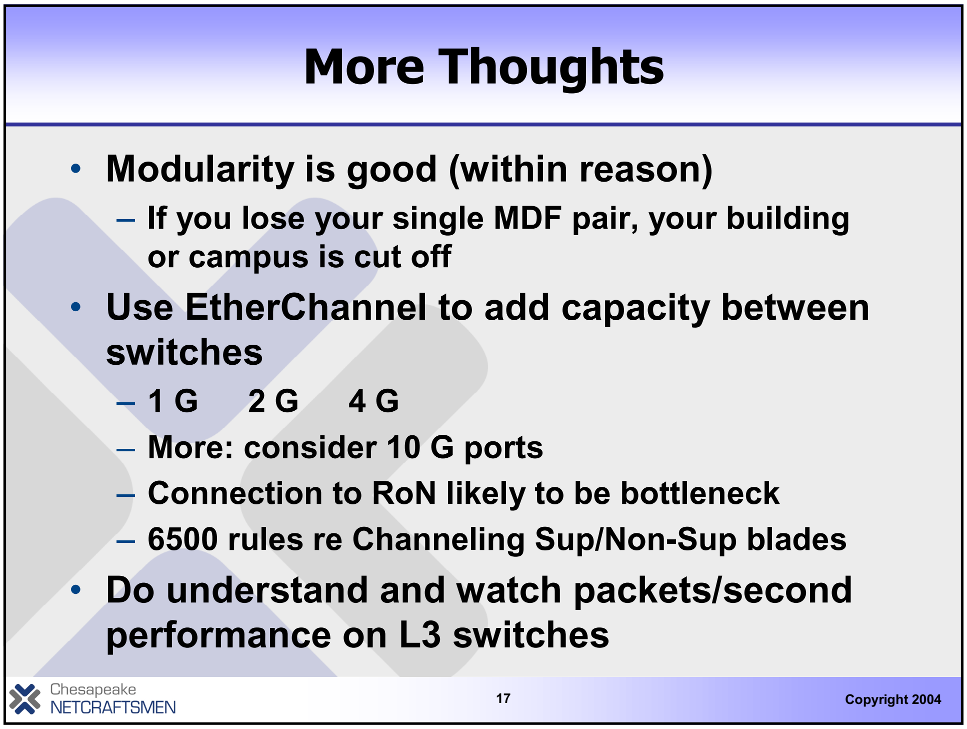
Chesapeake NETCRAFTSMEN 15 Copyright 2004



Thoughts

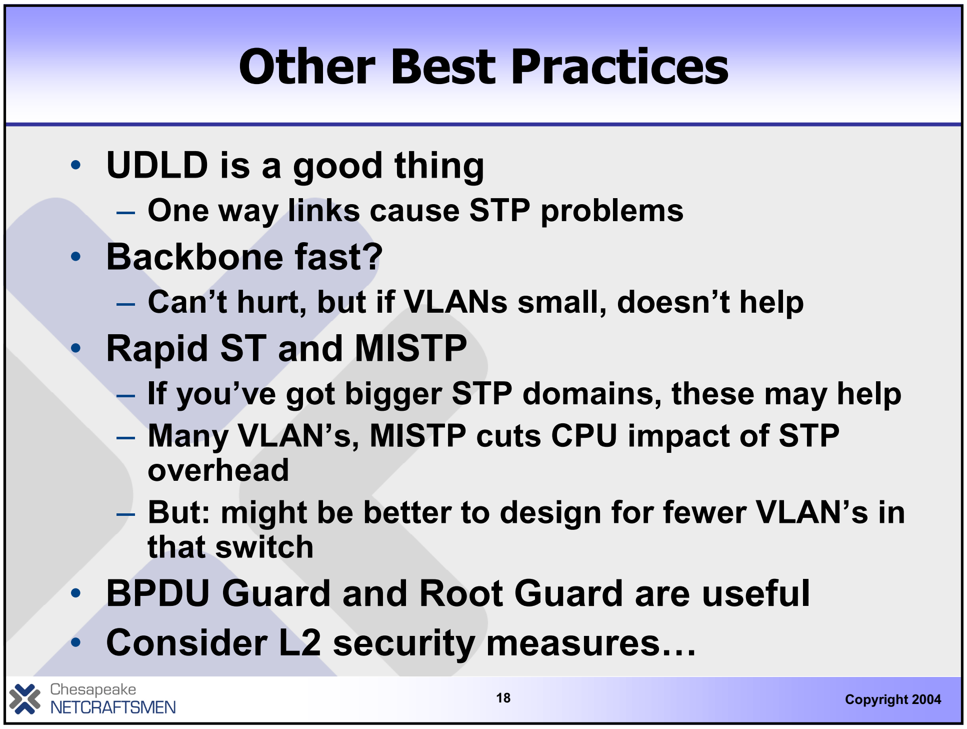
- All those point-to-point links means you'd better be comfortable with /30 VLSMs
- If you have many VLANs coming into a pair of switches, the pair does not need OSPF / EIGRP adjacencies on every VLAN
  - Can use trunk(s) between the pair for this
- PLEASE use route summarization for each building or campus
  - Don't trade STP problems for routing problems

Chesapeake NETCRAFTSMEN 16 Copyright 2004



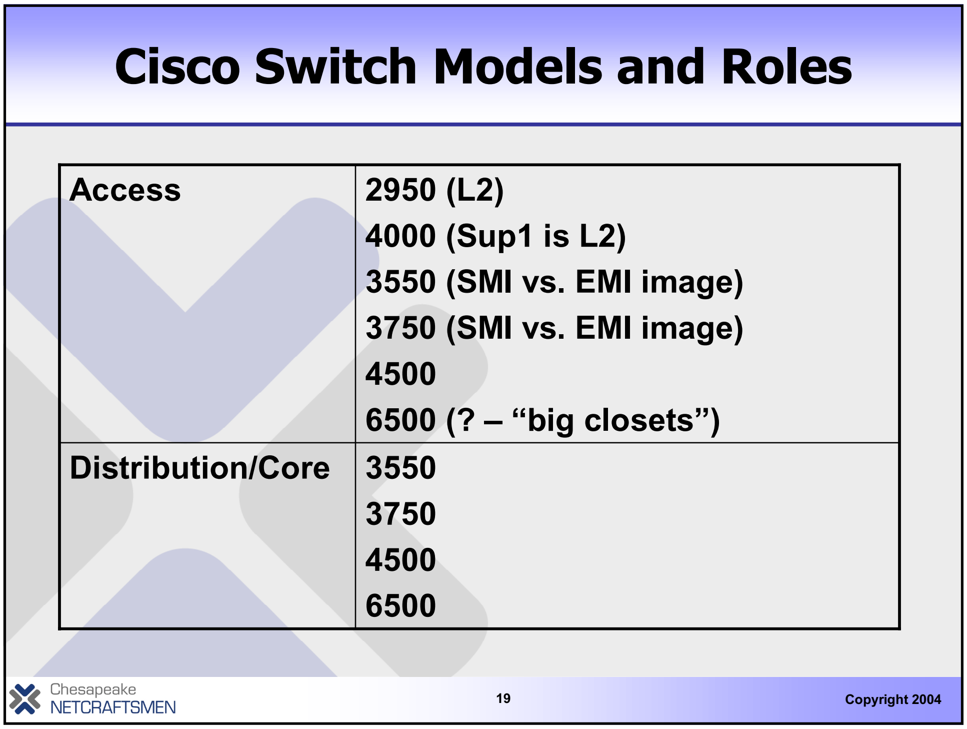
More Thoughts

- Modularity is good (within reason)
  - If you lose your single MDF pair, your building or campus is cut off
- Use EtherChannel to add capacity between switches
  - 1 G 2 G 4 G
  - More: consider 10 G ports
  - Connection to RoN likely to be bottleneck
  - 6500 rules re Channeling Sup/Non-Sup blades
- Do understand and watch packets/second performance on L3 switches



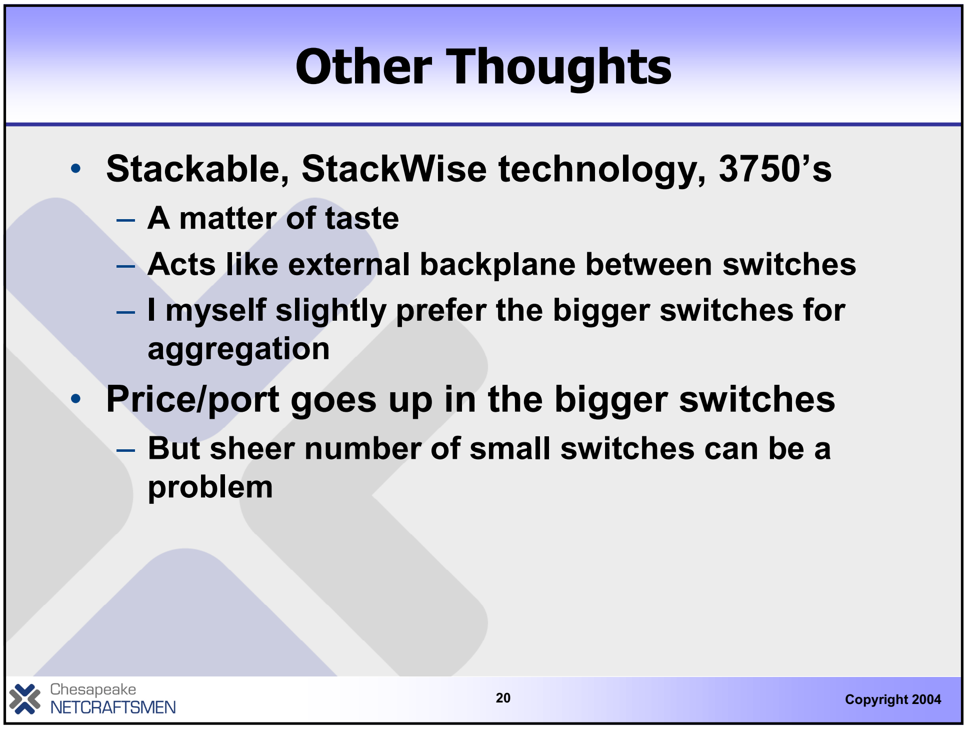
Other Best Practices

- UDLD is a good thing
  - One way links cause STP problems
- Backbone fast?
  - Can’t hurt, but if VLANs small, doesn’t help
- Rapid ST and MISTP
  - If you’ve got bigger STP domains, these may help
  - Many VLAN’s, MISTP cuts CPU impact of STP overhead
  - But: might be better to design for fewer VLAN’s in that switch
- BPDU Guard and Root Guard are useful
- Consider L2 security measures…



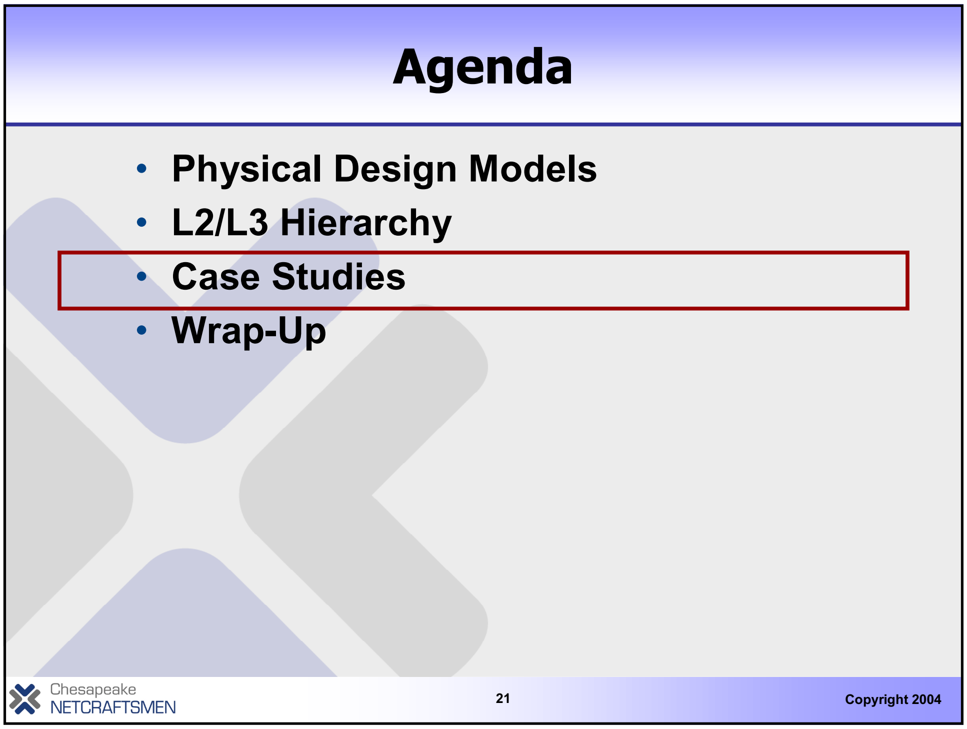
Cisco Switch Models and Roles

Access	2950 (L2) 4000 (Sup1 is L2) 3550 (SMI vs. EMI image) 3750 (SMI vs. EMI image) 4500 6500 (? – “big closets”)
Distribution/Core	3550 3750 4500 6500



Other Thoughts

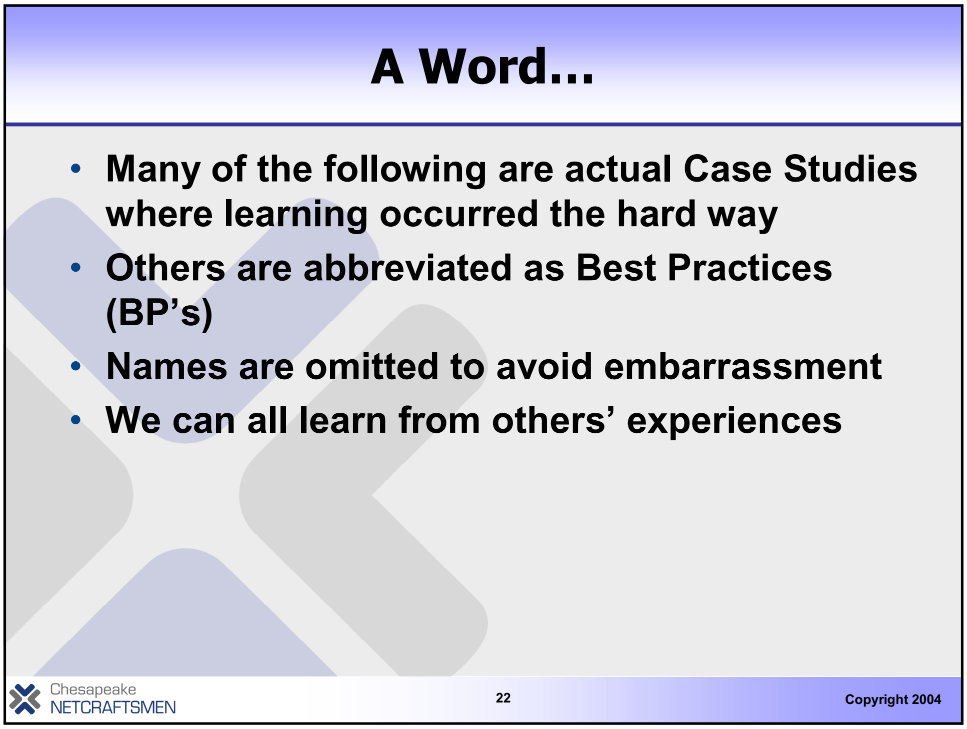
- Stackable, StackWise technology, 3750’s
  - A matter of taste
  - Acts like external backplane between switches
  - I myself slightly prefer the bigger switches for aggregation
- Price/port goes up in the bigger switches
  - But sheer number of small switches can be a problem



Agenda

- Physical Design Models
- L2/L3 Hierarchy
- Case Studies
- Wrap-Up

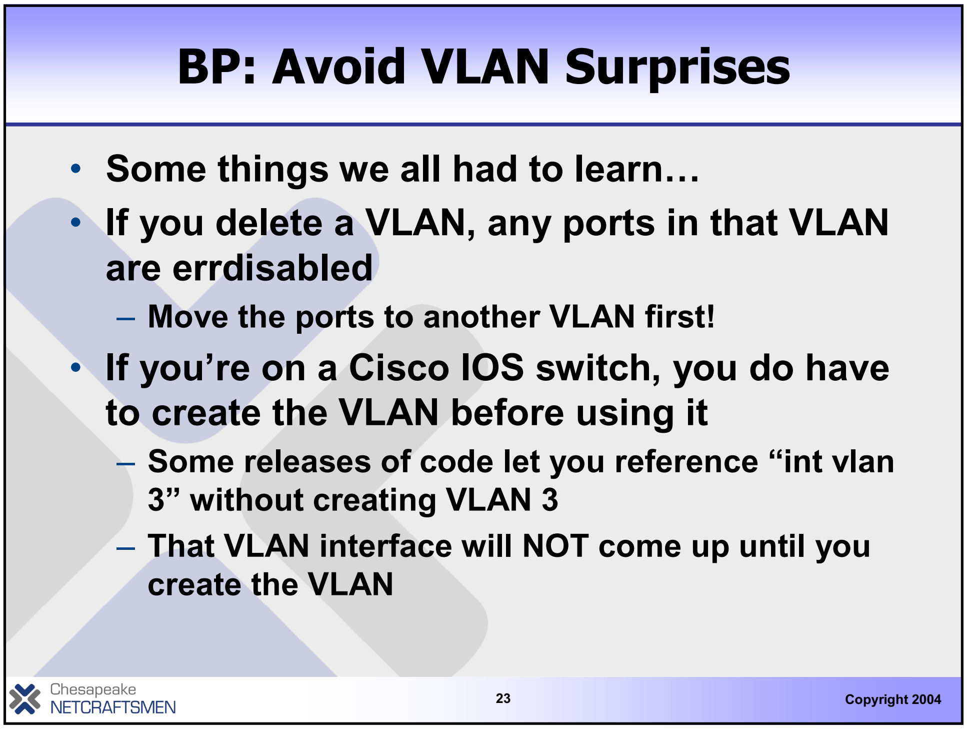
Chesapeake NETCRAFTSMEN 21 Copyright 2004



A Word…

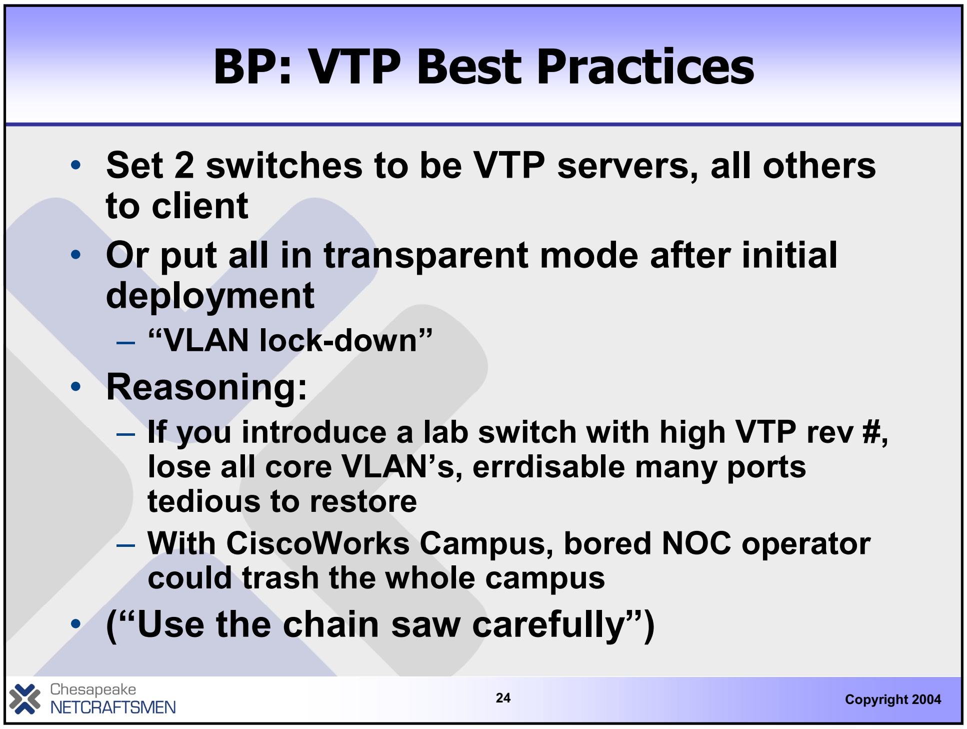
- Many of the following are actual Case Studies where learning occurred the hard way
- Others are abbreviated as Best Practices (BP’s)
- Names are omitted to avoid embarrassment
- We can all learn from others’ experiences

Chesapeake NETCRAFTSMEN 22 Copyright 2004



BP: Avoid VLAN Surprises

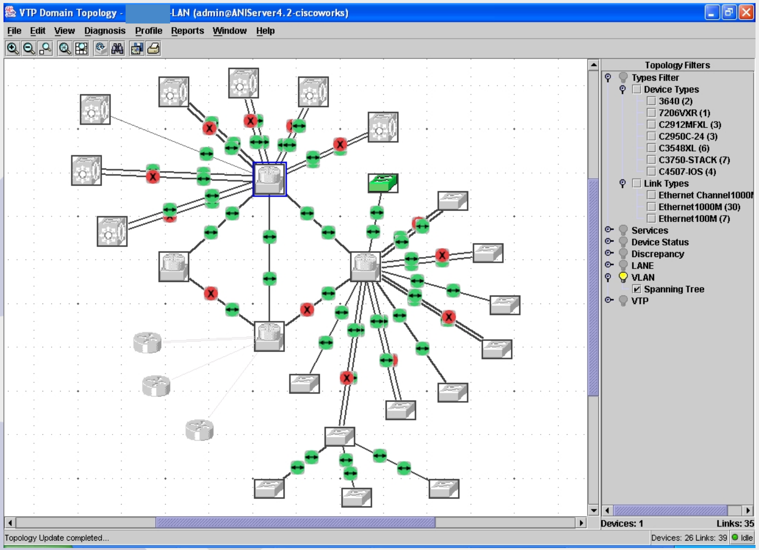
- Some things we all had to learn…
- If you delete a VLAN, any ports in that VLAN are errdisabled
  - Move the ports to another VLAN first!
- If you're on a Cisco IOS switch, you do have to create the VLAN before using it
  - Some releases of code let you reference "int vlan 3" without creating VLAN 3
  - That VLAN interface will NOT come up until you create the VLAN



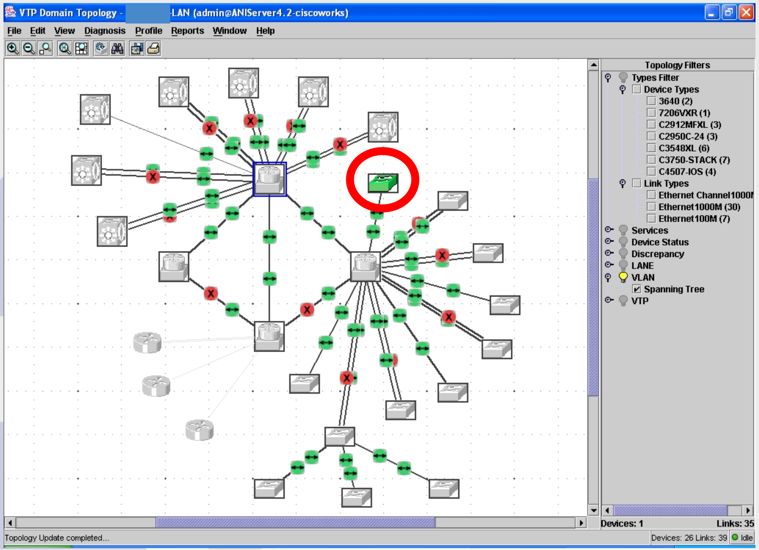
BP: VTP Best Practices

- Set 2 switches to be VTP servers, all others to client
- Or put all in transparent mode after initial deployment
  - "VLAN lock-down"
- Reasoning:
  - If you introduce a lab switch with high VTP rev #, lose all core VLAN's, errdisable many ports tedious to restore
  - With CiscoWorks Campus, bored NOC operator could trash the whole campus
- ("Use the chain saw carefully")

#1: Where’s The Root Bridge?



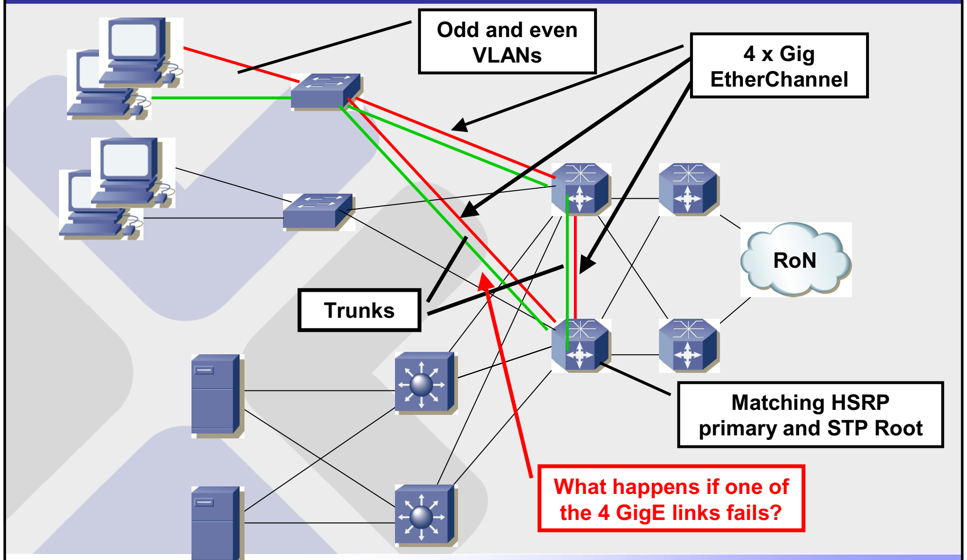
Where’s The Root Bridge?



Lessons Learned

- Older edge switch had lowest MAC address and became root bridge
  - This is a fairly common problem
  - It is not terrible in the above topology
- Better to KNOW and CONTROL where your root bridge is
  - "set spantree root …"
  - "spanning-tree vlan XX priority YY"
- Thought: much as we like it, CiscoWorks may not be able to find this out for you when your network is having problems

#2: Load Balancing



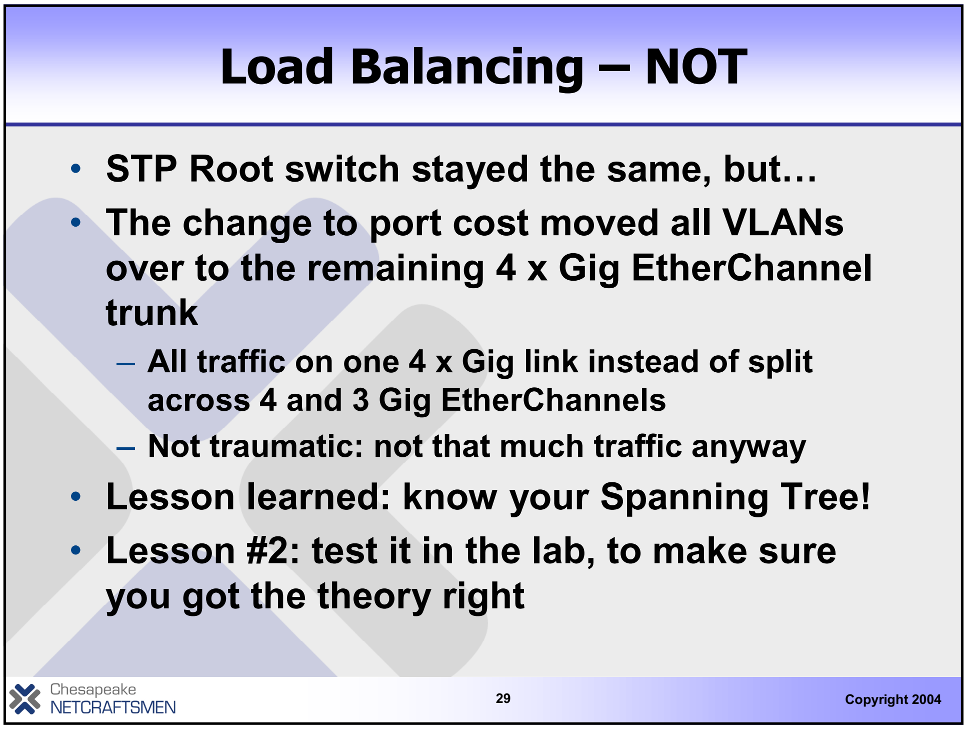
Odd and even VLANs

4 x Gig EtherChannel

Trunks

Matching HSRP primary and STP Root

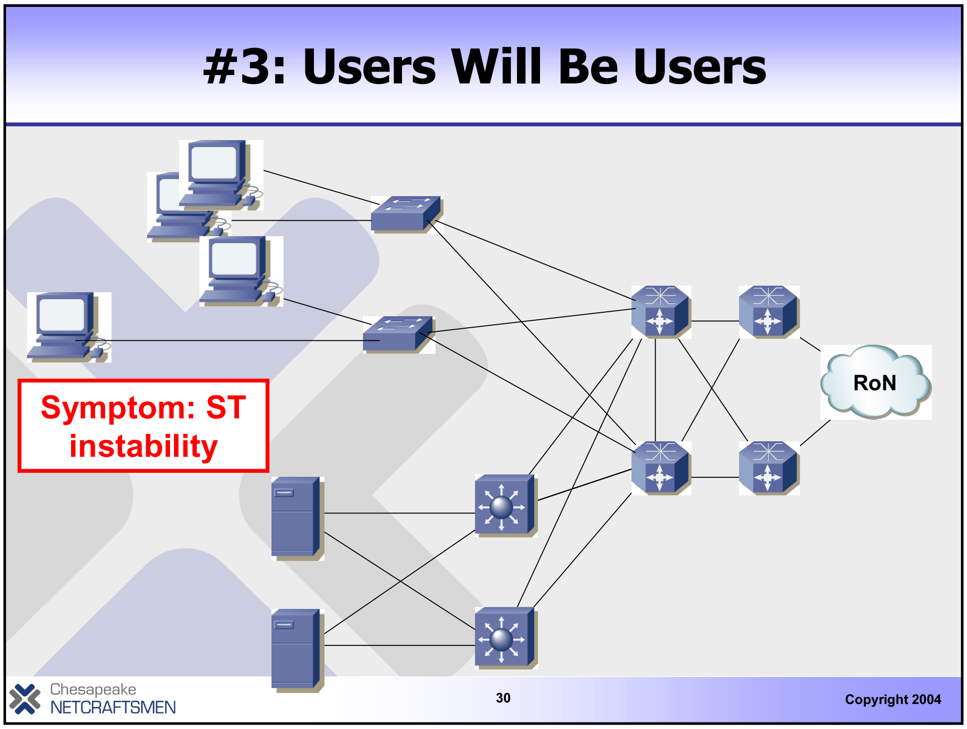
What happens if one of the 4 GigE links fails?



Load Balancing – NOT

- STP Root switch stayed the same, but…
- The change to port cost moved all VLANs over to the remaining 4 x Gig EtherChannel trunk
  - All traffic on one 4 x Gig link instead of split across 4 and 3 Gig EtherChannels
  - Not traumatic: not that much traffic anyway
- Lesson learned: know your Spanning Tree!
- Lesson #2: test it in the lab, to make sure you got the theory right

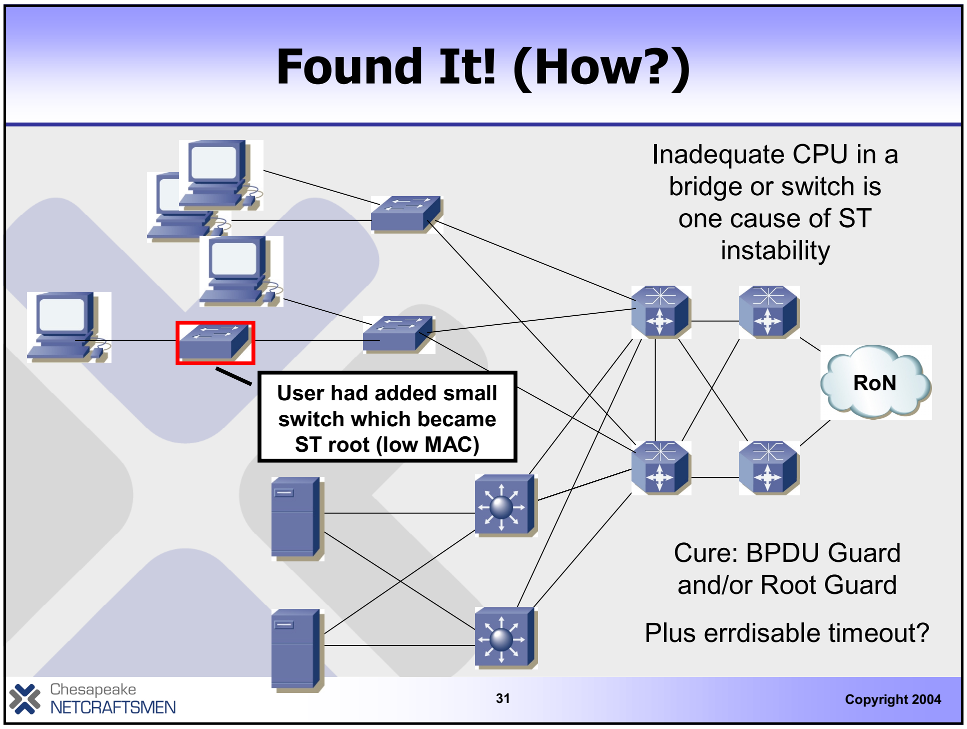
Chesapeake NETCRAFTSMEN 29 Copyright 2004



#3: Users Will Be Users

Symptom: ST instability

Chesapeake NETCRAFTSMEN 30 Copyright 2004



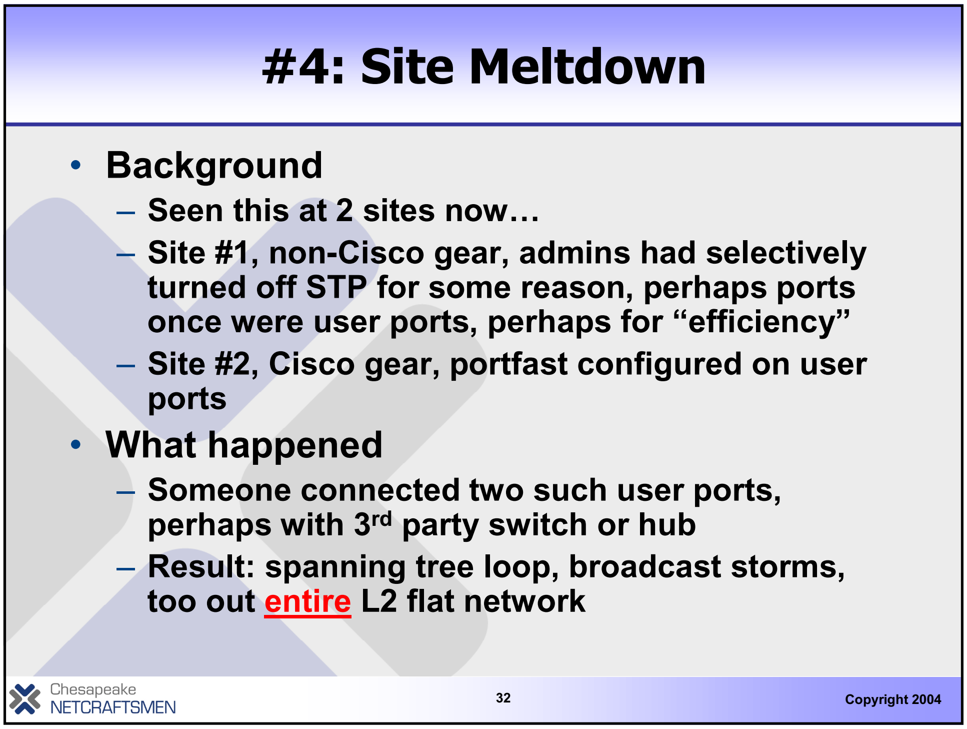
Found It! (How?)

Inadequate CPU in a bridge or switch is one cause of ST instability

User had added small switch which became ST root (low MAC)

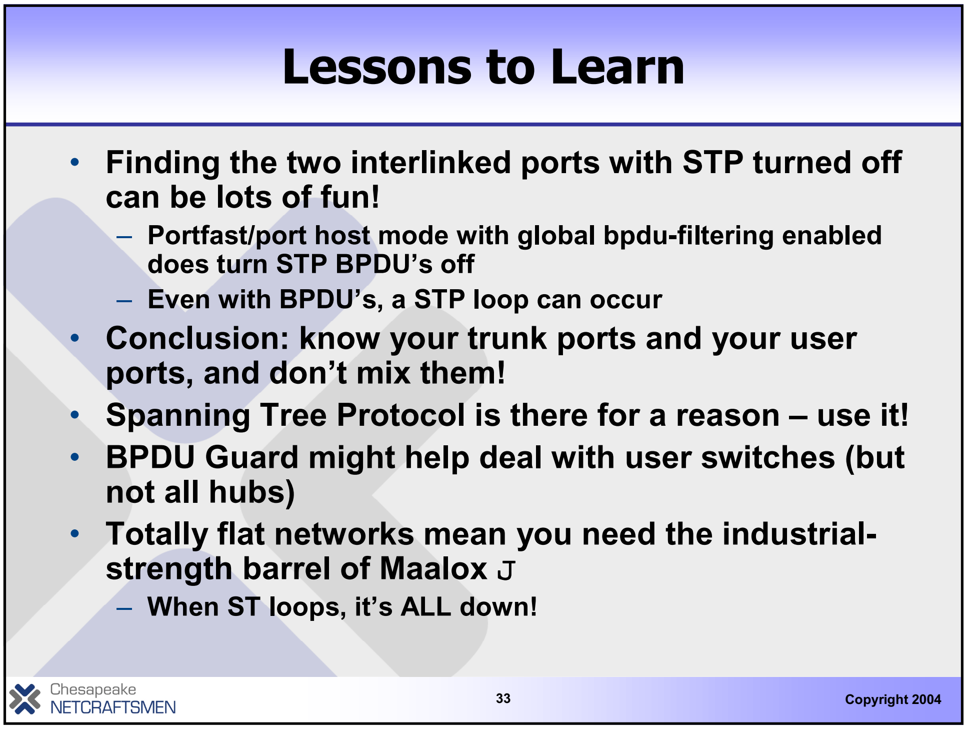
Cure: BPDU Guard and/or Root Guard

Plus errdisable timeout?



#4: Site Meltdown

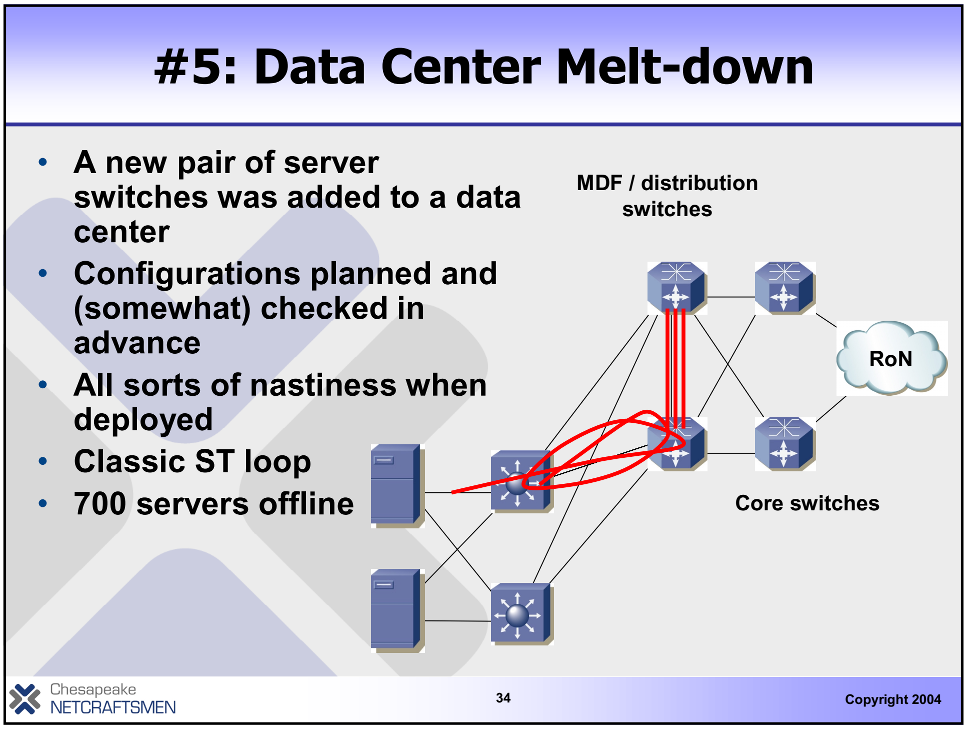
- Background
  - Seen this at 2 sites now…
  - Site #1, non-Cisco gear, admins had selectively turned off STP for some reason, perhaps ports once were user ports, perhaps for “efficiency”
  - Site #2, Cisco gear, portfast configured on user ports
- What happened
  - Someone connected two such user ports, perhaps with 3rd party switch or hub
  - Result: spanning tree loop, broadcast storms, too out entire L2 flat network



Lessons to Learn

- Finding the two interlinked ports with STP turned off can be lots of fun!
  - Portfast/port host mode with global bpdu-filtering enabled does turn STP BPDU’s off
  - Even with BPDU’s, a STP loop can occur
- Conclusion: know your trunk ports and your user ports, and don’t mix them!
- Spanning Tree Protocol is there for a reason – use it!
- BPDU Guard might help deal with user switches (but not all hubs)
- Totally flat networks mean you need the industrial-strength barrel of Maalox J
  - When ST loops, it’s ALL down!

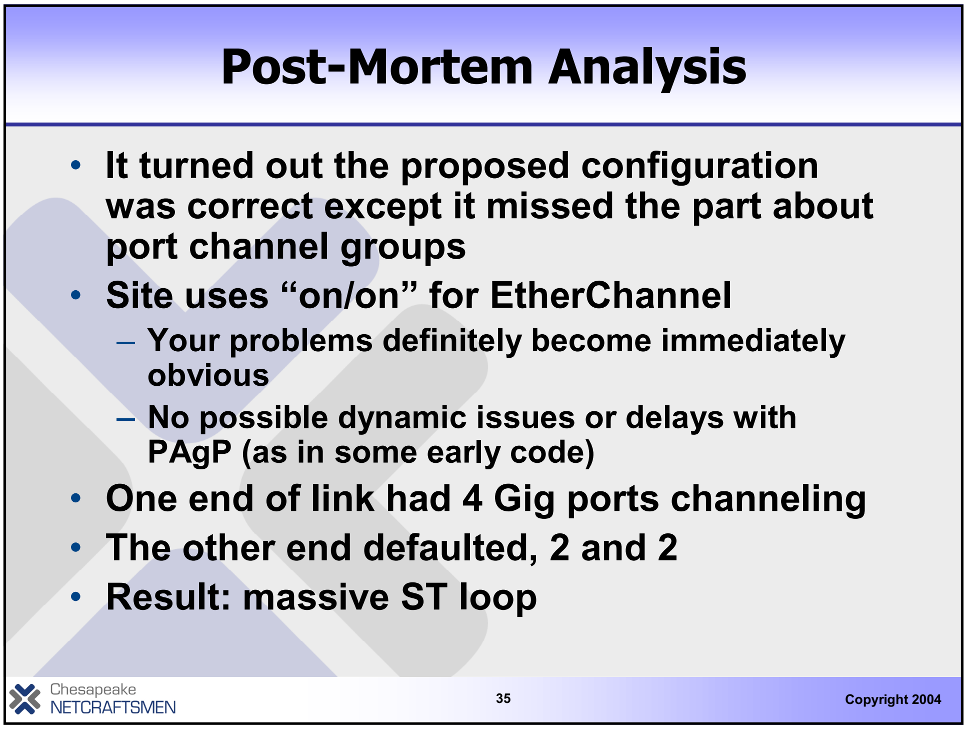
Chesapeake NETCRAFTSMEN 33 Copyright 2004



#5: Data Center Melt-down

- A new pair of server switches was added to a data center
- Configurations planned and (somewhat) checked in advance
- All sorts of nastiness when deployed
- Classic ST loop
- 700 servers offline

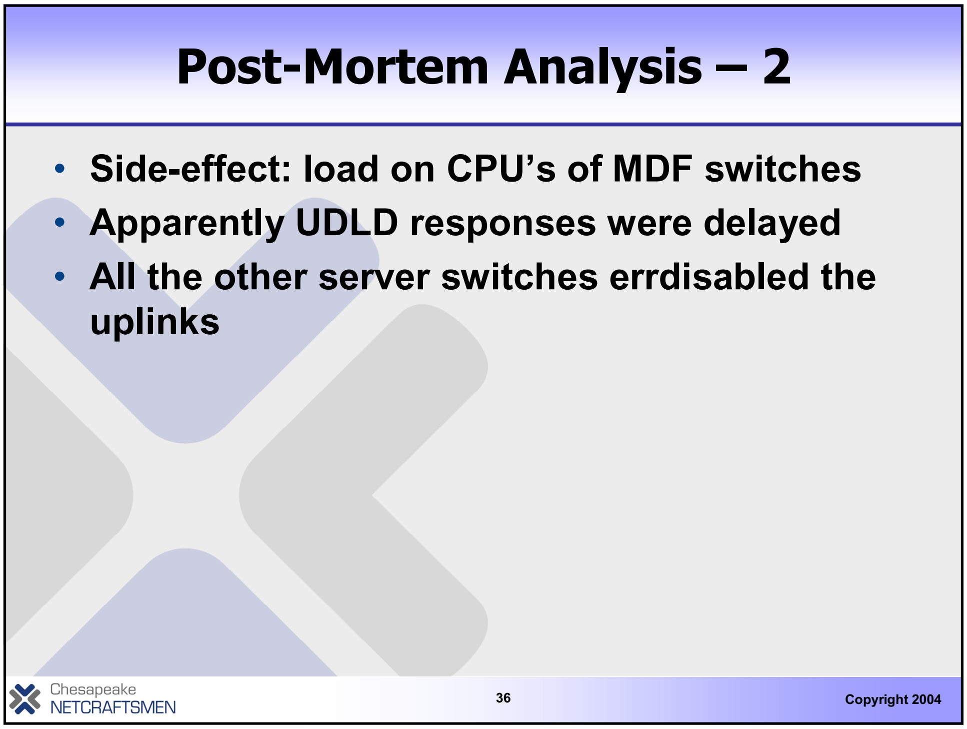
Chesapeake NETCRAFTSMEN 34 Copyright 2004



Post-Mortem Analysis

- It turned out the proposed configuration was correct except it missed the part about port channel groups
- Site uses “on/on” for EtherChannel
  - Your problems definitely become immediately obvious
  - No possible dynamic issues or delays with PAgP (as in some early code)
- One end of link had 4 Gig ports channeling
- The other end defaulted, 2 and 2
- Result: massive ST loop

Chesapeake NETCRAFTSMEN 35 Copyright 2004



Post-Mortem Analysis – 2

- Side-effect: load on CPU’s of MDF switches
- Apparently UDLD responses were delayed
- All the other server switches errdisabled the uplinks

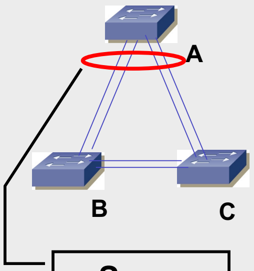
Chesapeake NETCRAFTSMEN 36 Copyright 2004

Lessons Learned

- Change control: check it before you do it
- Understand assumptions
  - Channel on/on means be careful on the rare occasions you touch the EtherChannels)
  - Desirable/desirable is safer for most sites
- Some other lessons learned along the way:
  - Don’t deploy untested hardware
  - Don’t deploy the CatOS that happens to be in the Sup blade on the shelf, think about what you want
  - Do replace blades with failed ports
  - A lab used for spare parts is NOT a lab!

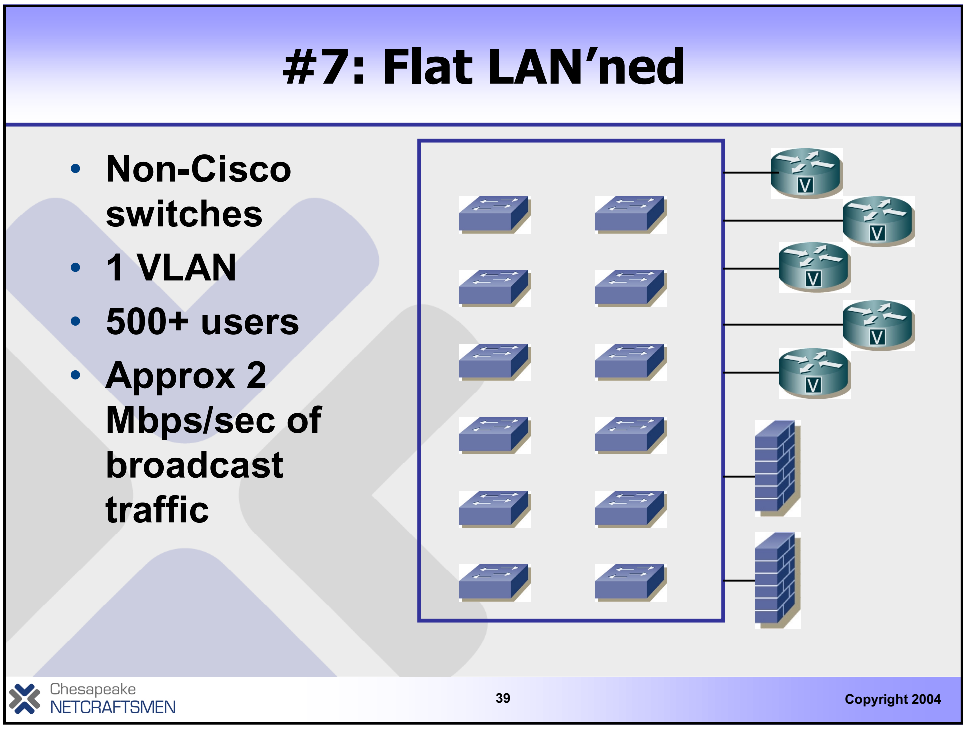
#6: EtherChannel No Go

- For some reason, the EtherChannel just wouldn’t come up!



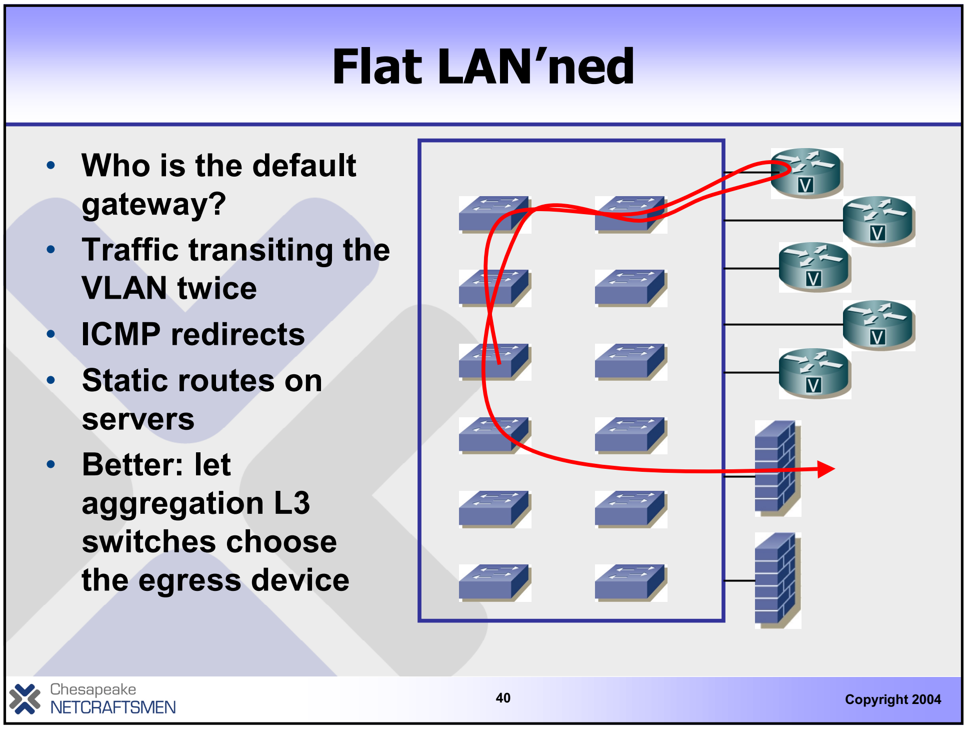
Same port channel group

```
%PAGP-5-PORTFROMSTP:Port 2/01 left bridge port 2/01
%PAGP-5-PORTTOSTP:Port 2/01 joined bridge port 2/01
%PAGP-5-PORTFROMSTP:Port 2/01 left bridge port 2/01
%PAGP-5-PORTTOSTP:Port 2/01 joined bridge port 2/01
%PAGP-5-PORTFROMSTP:Port 2/03 left bridge port 2/03
%PAGP-5-PORTTOSTP:Port 2/03 joined bridge port
```



#7: Flat LAN’ned

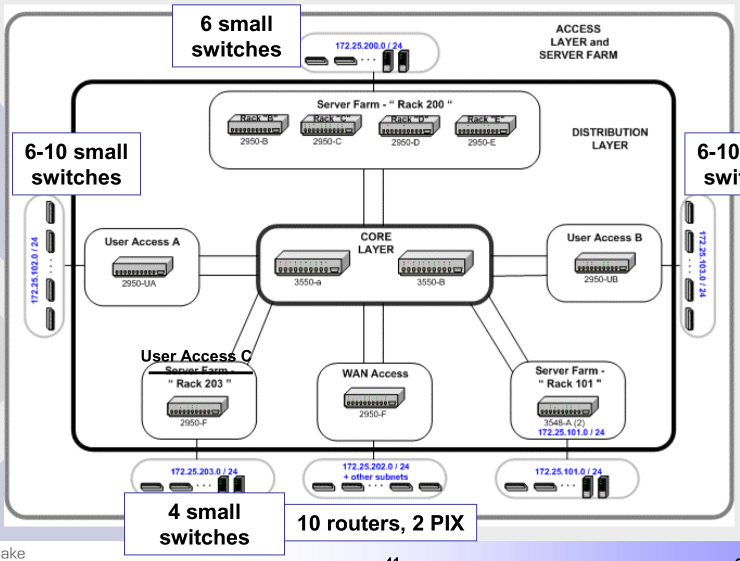
- Non-Cisco switches
- 1 VLAN
- 500+ users
- Approx 2 Mbps/sec of broadcast traffic



Flat LAN’ned

- Who is the default gateway?
- Traffic transiting the VLAN twice
- ICMP redirects
- Static routes on servers
- Better: let aggregation L3 switches choose the egress device

#8: Incremental Growth



6 small switches

6-10 small switches

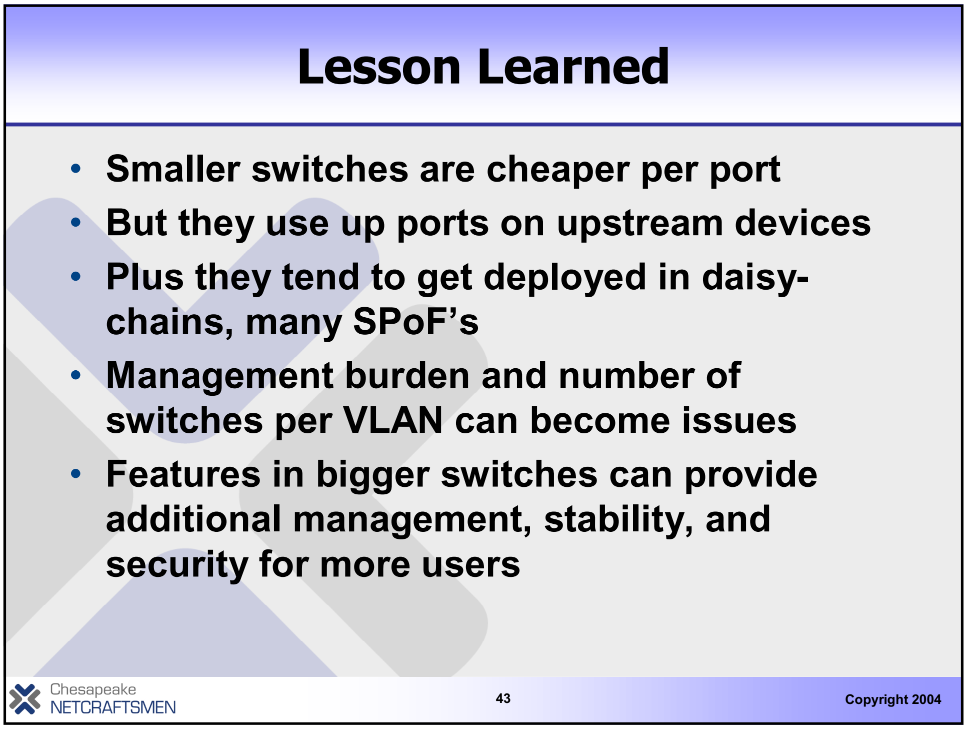
6-10 small switches

4 small switches

10 routers, 2 PIX

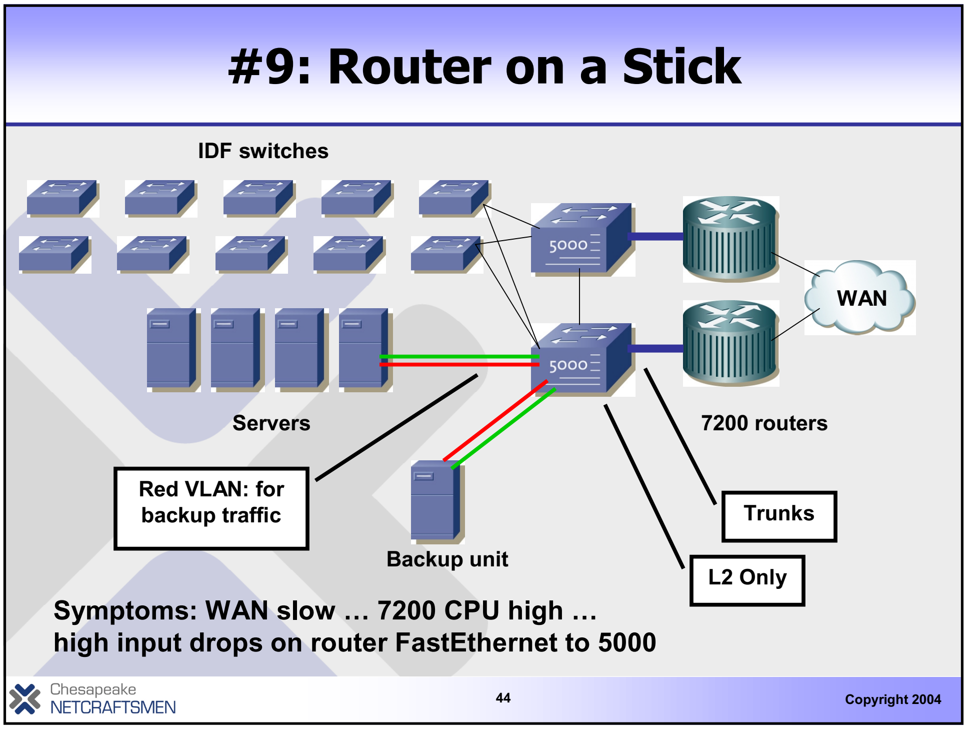
Improvements

- Good features
  - Core is L3
  - 2 Gig EtherChannel trunks: plenty of bandwidth
- Less good
  - Too many small switches, too little time
  - Not enough Gig aggregation ports in Core
  - Possibly large server traffic flows across core (Citrix to apps and DB’s in the other server farm)
- Recommendations
  - Replace Core with bigger switches
  - Remove Single Points of Failure
    - § Consider dual distribution layer switches
    - § Dual connection to WAN routers
  - Configuration audit / cleanup / consistency



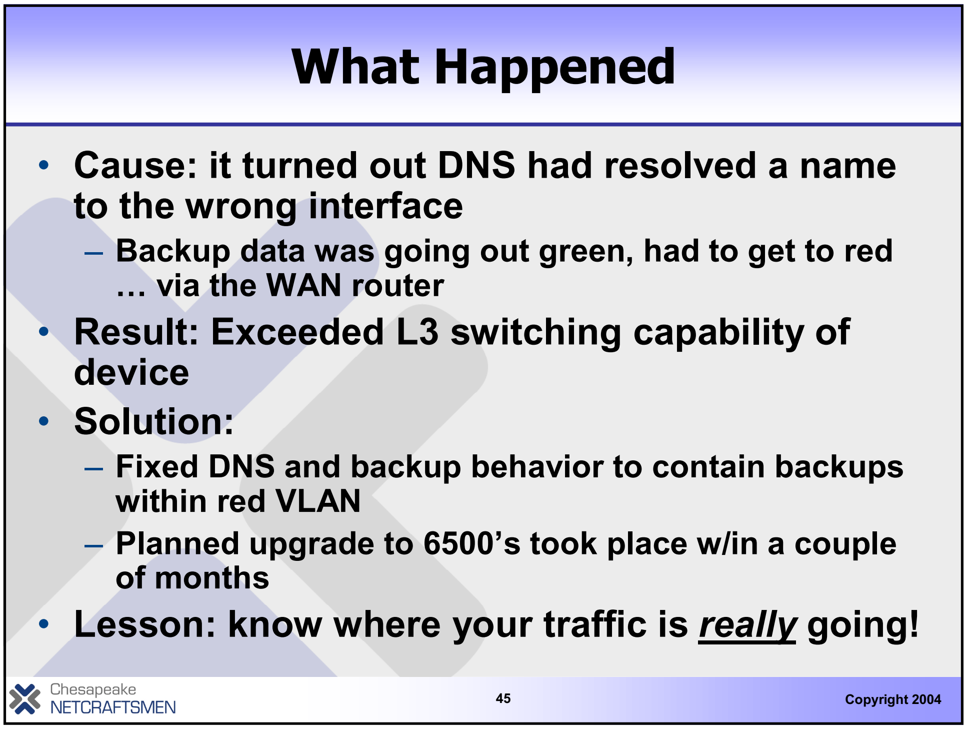
Lesson Learned

- Smaller switches are cheaper per port
- But they use up ports on upstream devices
- Plus they tend to get deployed in daisy-chains, many SPoF's
- Management burden and number of switches per VLAN can become issues
- Features in bigger switches can provide additional management, stability, and security for more users



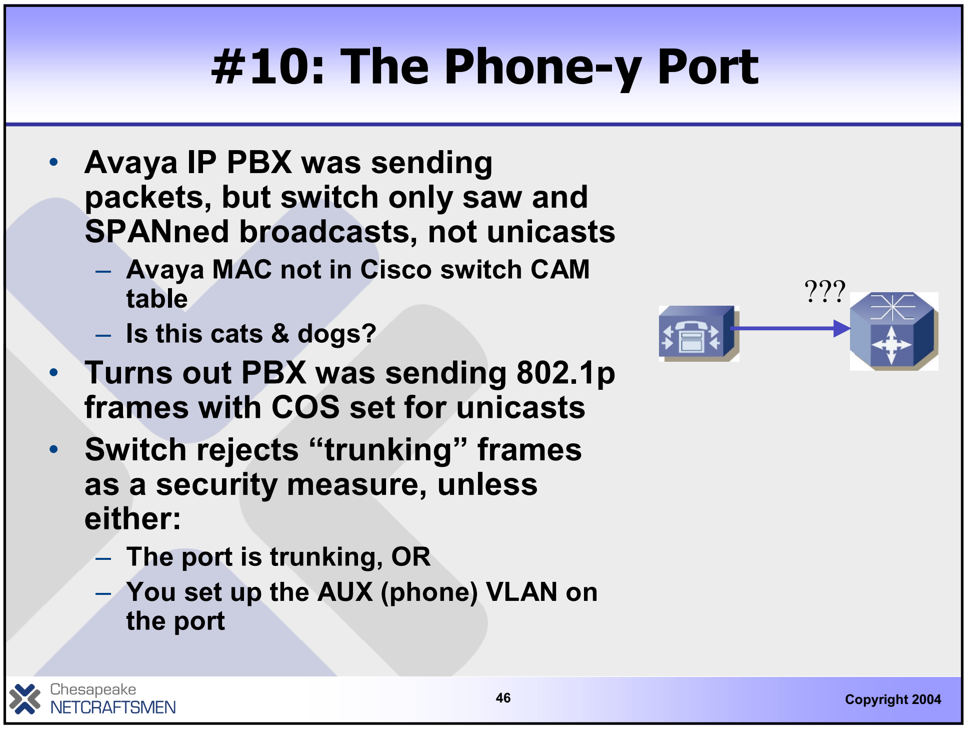
#9: Router on a Stick

Symptoms: WAN slow … 7200 CPU high …
high input drops on router FastEthernet to 5000



What Happened

- Cause: it turned out DNS had resolved a name to the wrong interface
  - Backup data was going out green, had to get to red … via the WAN router
- Result: Exceeded L3 switching capability of device
- Solution:
  - Fixed DNS and backup behavior to contain backups within red VLAN
  - Planned upgrade to 6500's took place w/in a couple of months
- Lesson: know where your traffic is ***really*** going!



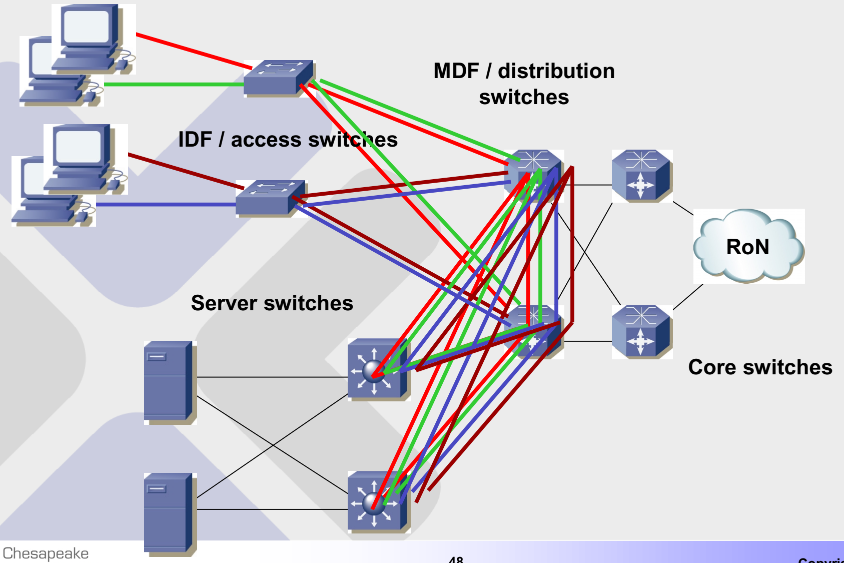
#10: The Phone-y Port

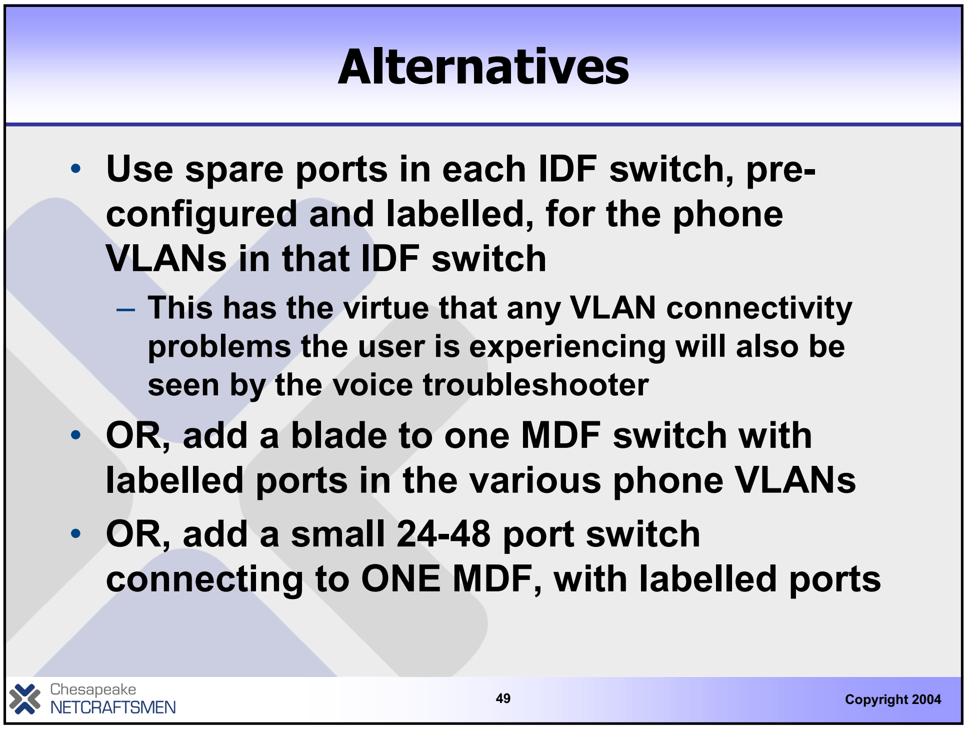
- Avaya IP PBX was sending packets, but switch only saw and SPANned broadcasts, not unicasts
  - Avaya MAC not in Cisco switch CAM table
  - Is this cats & dogs?
- Turns out PBX was sending 802.1p frames with COS set for unicasts
- Switch rejects "trunking" frames as a security measure, unless either:
  - The port is trunking, OR
  - You set up the AUX (phone) VLAN on the port

#11: IP Telephony VLANs

- Large site deploying Nortel-based IPT
- Using Cisco-style phone VLANs, 2 per IDF switch
- SE recommended having a blade with ports pre-configured, one per phone VLAN, for ease of troubleshooting by voice team
- Not a bad idea on the surface
- Pros/cons of doing this at the server switch?

IPT VLANs

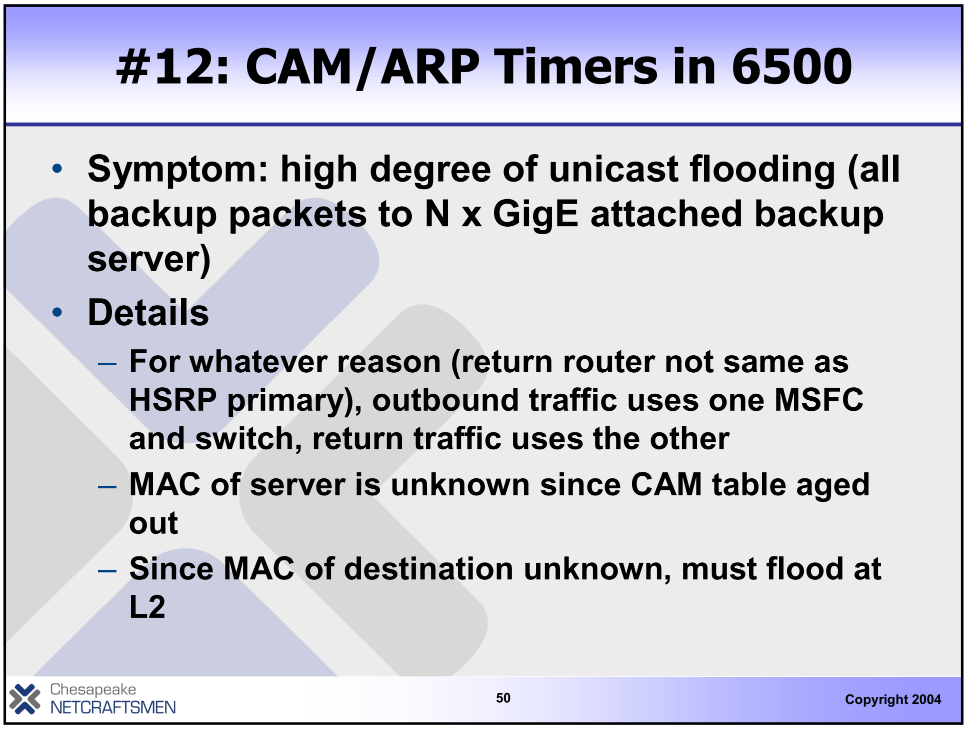




Alternatives

- Use spare ports in each IDF switch, pre-configured and labelled, for the phone VLANs in that IDF switch
  - This has the virtue that any VLAN connectivity problems the user is experiencing will also be seen by the voice troubleshooter
- OR, add a blade to one MDF switch with labelled ports in the various phone VLANs
- OR, add a small 24-48 port switch connecting to ONE MDF, with labelled ports

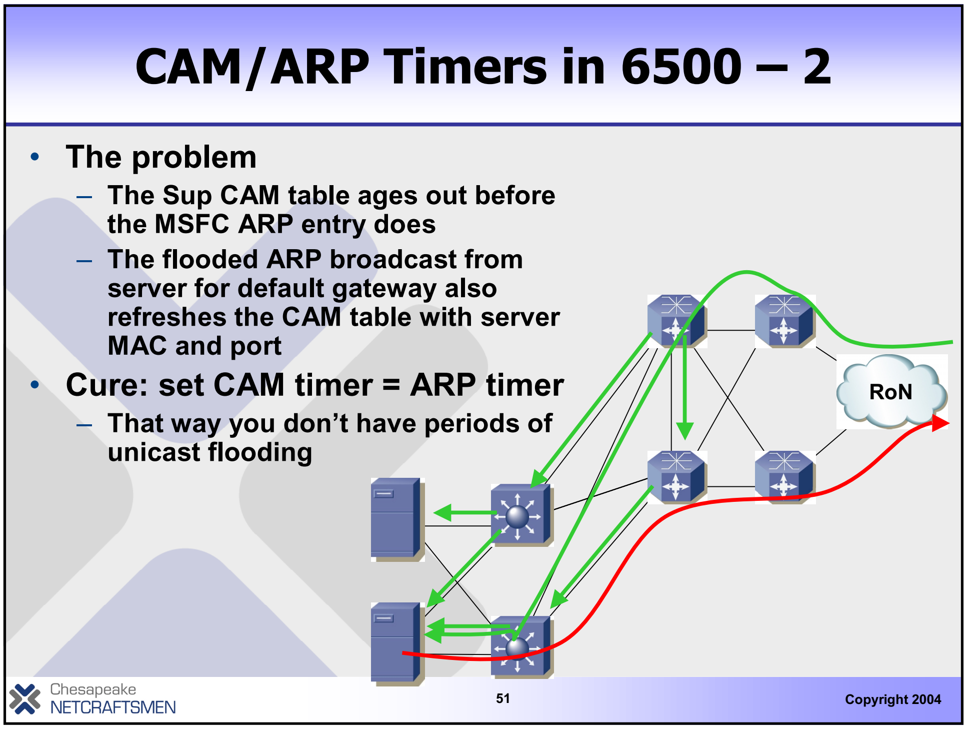
Chesapeake NETCRAFTSMEN 49 Copyright 2004



#12: CAM/ARP Timers in 6500

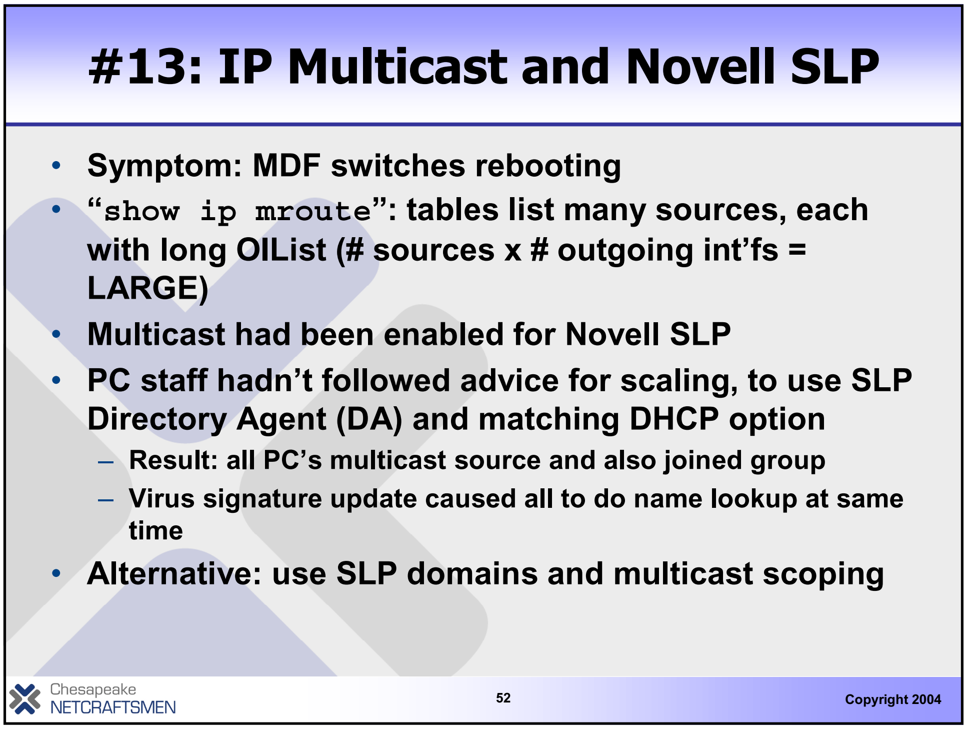
- Symptom: high degree of unicast flooding (all backup packets to N x GigE attached backup server)
- Details
  - For whatever reason (return router not same as HSRP primary), outbound traffic uses one MSFC and switch, return traffic uses the other
  - MAC of server is unknown since CAM table aged out
  - Since MAC of destination unknown, must flood at L2

Chesapeake NETCRAFTSMEN 50 Copyright 2004



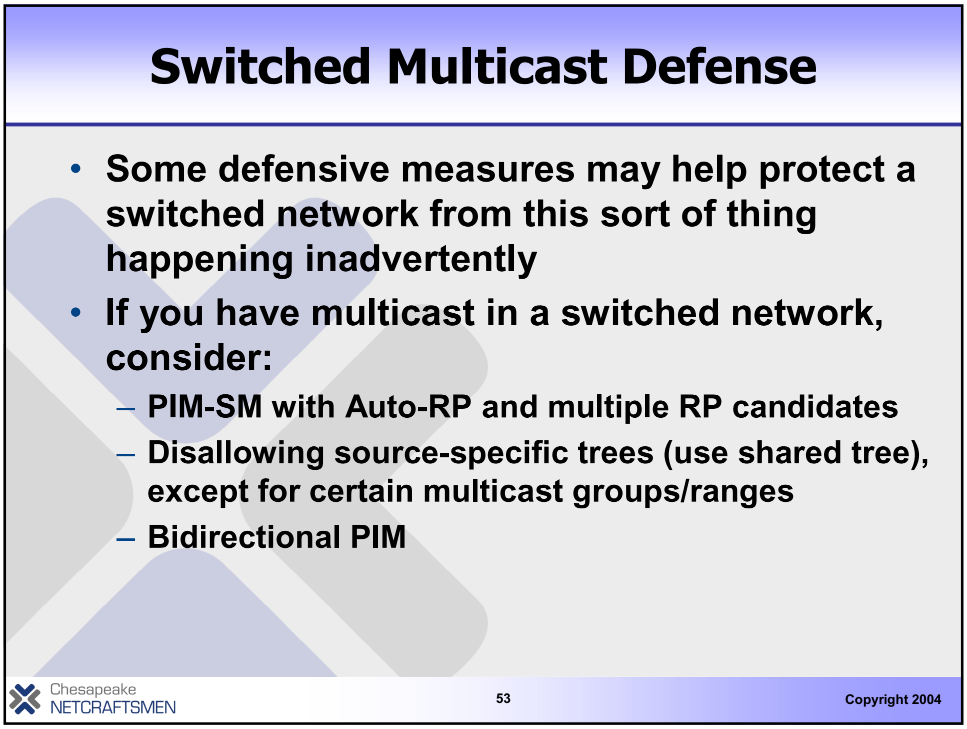
CAM/ARP Timers in 6500 – 2

- The problem
  - The Sup CAM table ages out before the MSFC ARP entry does
  - The flooded ARP broadcast from server for default gateway also refreshes the CAM table with server MAC and port
- Cure: set CAM timer = ARP timer
  - That way you don't have periods of unicast flooding



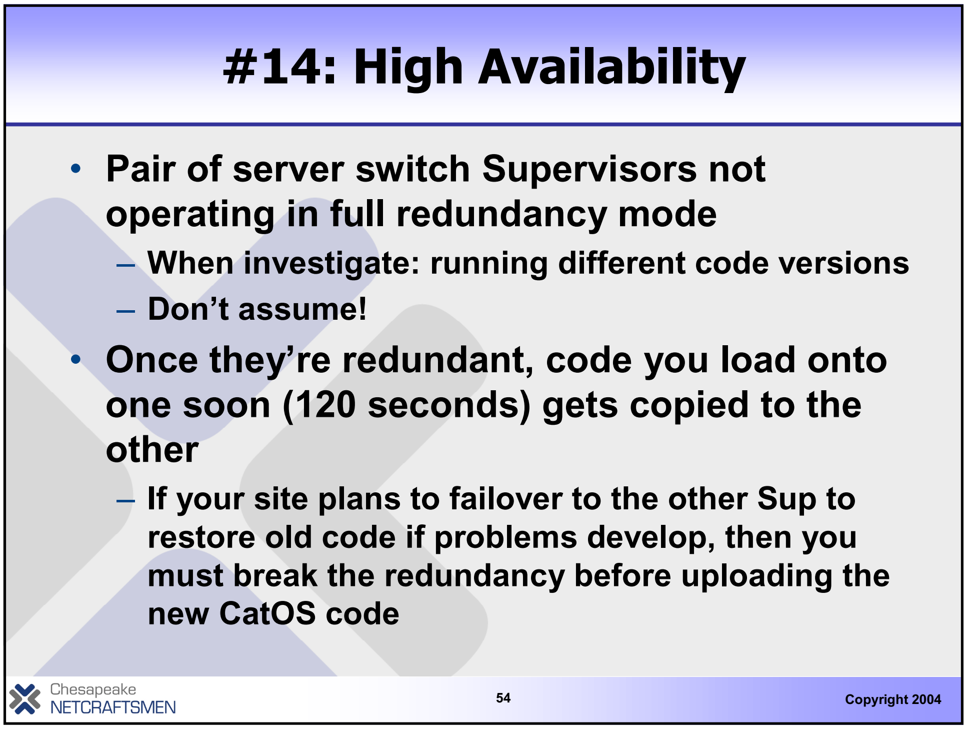
#13: IP Multicast and Novell SLP

- Symptom: MDF switches rebooting
- "`show ip mroute`": tables list many sources, each with long OIList (# sources x # outgoing int'fs = LARGE)
- Multicast had been enabled for Novell SLP
- PC staff hadn't followed advice for scaling, to use SLP Directory Agent (DA) and matching DHCP option
  - Result: all PC's multicast source and also joined group
  - Virus signature update caused all to do name lookup at same time
- Alternative: use SLP domains and multicast scoping



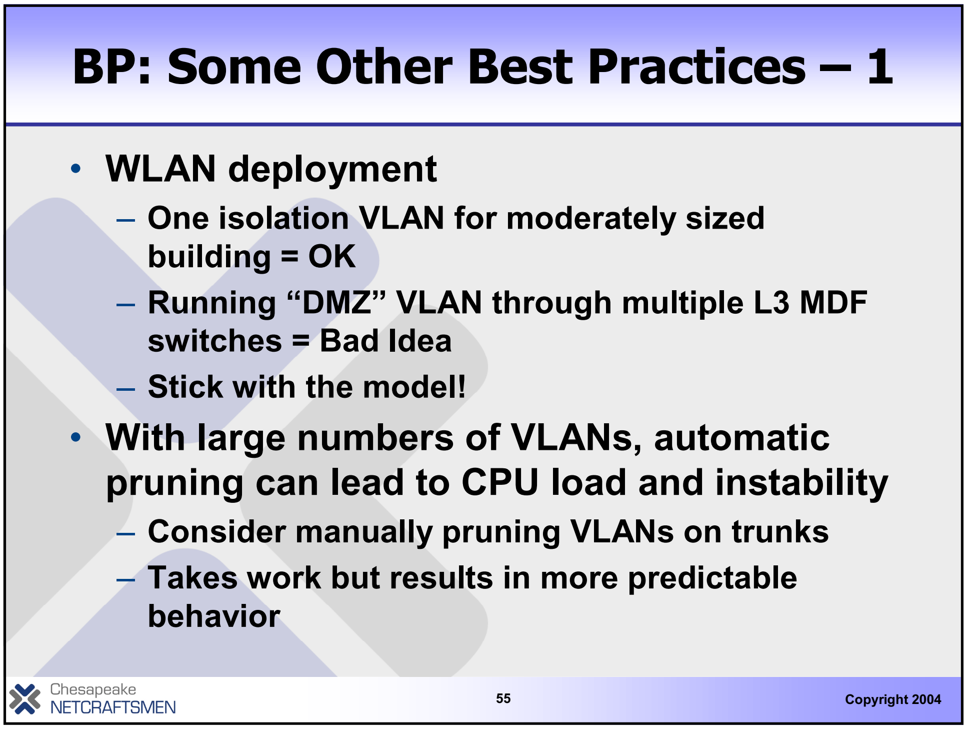
Switched Multicast Defense

- Some defensive measures may help protect a switched network from this sort of thing happening inadvertently
- If you have multicast in a switched network, consider:
  - PIM-SM with Auto-RP and multiple RP candidates
  - Disallowing source-specific trees (use shared tree), except for certain multicast groups/ranges
  - Bidirectional PIM



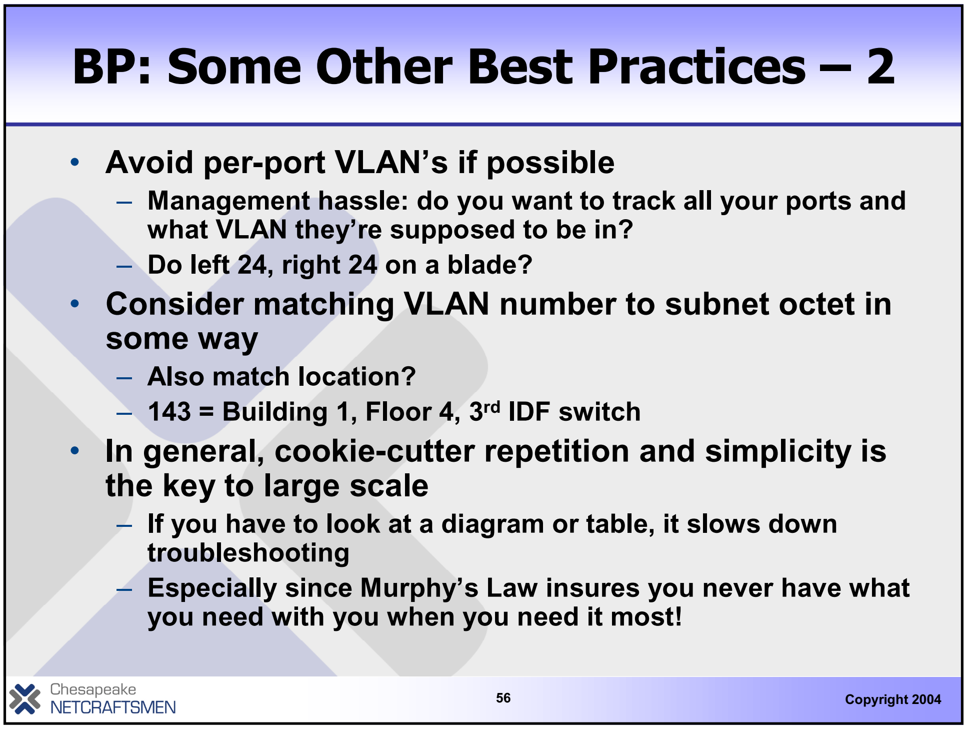
#14: High Availability

- Pair of server switch Supervisors not operating in full redundancy mode
  - When investigate: running different code versions
  - Don’t assume!
- Once they’re redundant, code you load onto one soon (120 seconds) gets copied to the other
  - If your site plans to failover to the other Sup to restore old code if problems develop, then you must break the redundancy before uploading the new CatOS code



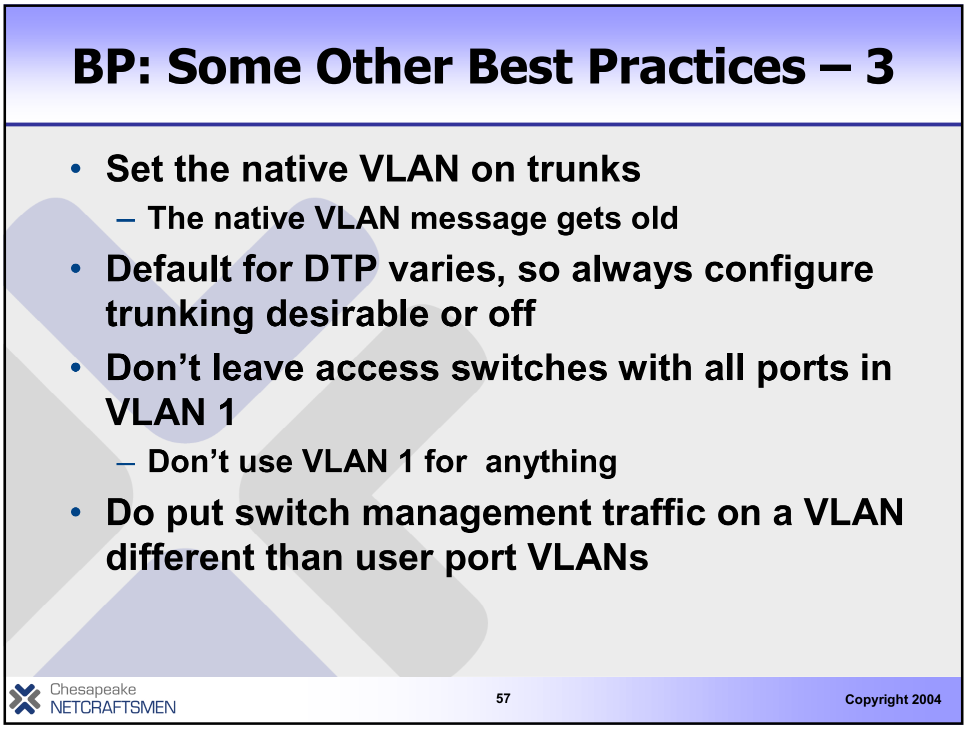
BP: Some Other Best Practices – 1

- WLAN deployment
  - One isolation VLAN for moderately sized building = OK
  - Running “DMZ” VLAN through multiple L3 MDF switches = Bad Idea
  - Stick with the model!
- With large numbers of VLANs, automatic pruning can lead to CPU load and instability
  - Consider manually pruning VLANs on trunks
  - Takes work but results in more predictable behavior



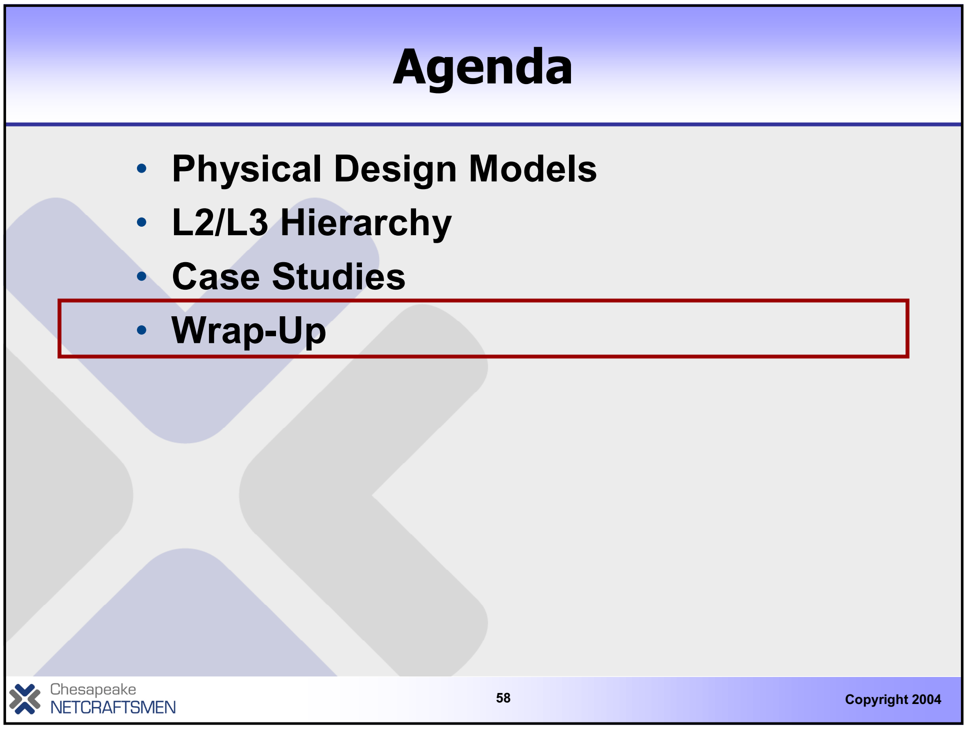
BP: Some Other Best Practices – 2

- Avoid per-port VLAN’s if possible
  - Management hassle: do you want to track all your ports and what VLAN they’re supposed to be in?
  - Do left 24, right 24 on a blade?
- Consider matching VLAN number to subnet octet in some way
  - Also match location?
  - 143 = Building 1, Floor 4, 3rd IDF switch
- In general, cookie-cutter repetition and simplicity is the key to large scale
  - If you have to look at a diagram or table, it slows down troubleshooting
  - Especially since Murphy’s Law insures you never have what you need with you when you need it most!



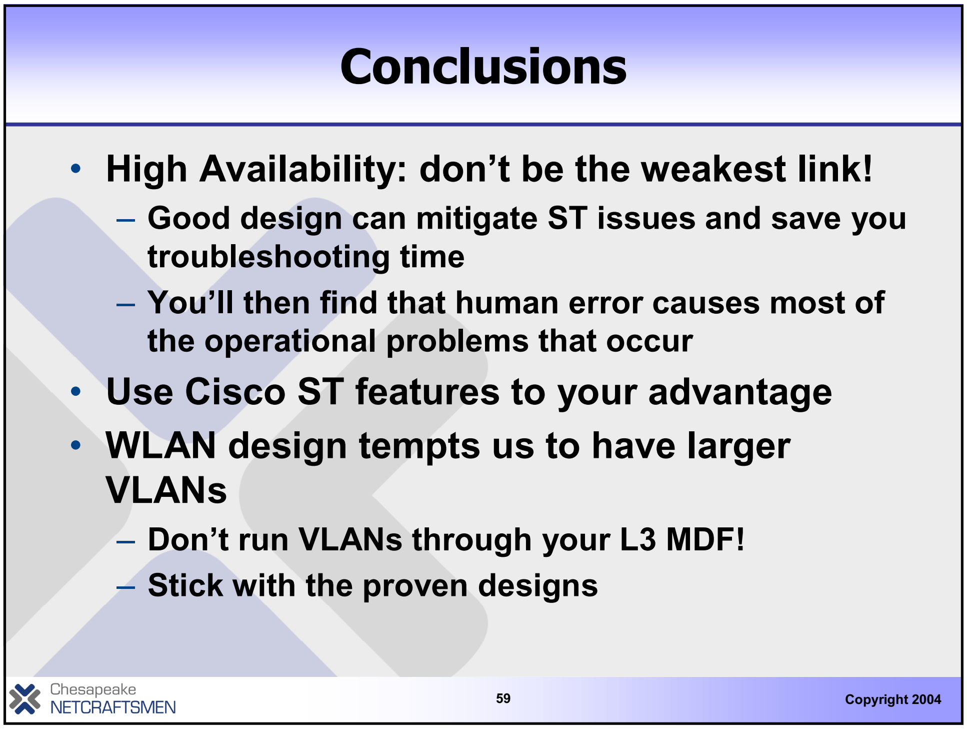
BP: Some Other Best Practices – 3

- Set the native VLAN on trunks
  - The native VLAN message gets old
- Default for DTP varies, so always configure trunking desirable or off
- Don't leave access switches with all ports in VLAN 1
  - Don't use VLAN 1 for anything
- Do put switch management traffic on a VLAN different than user port VLANs



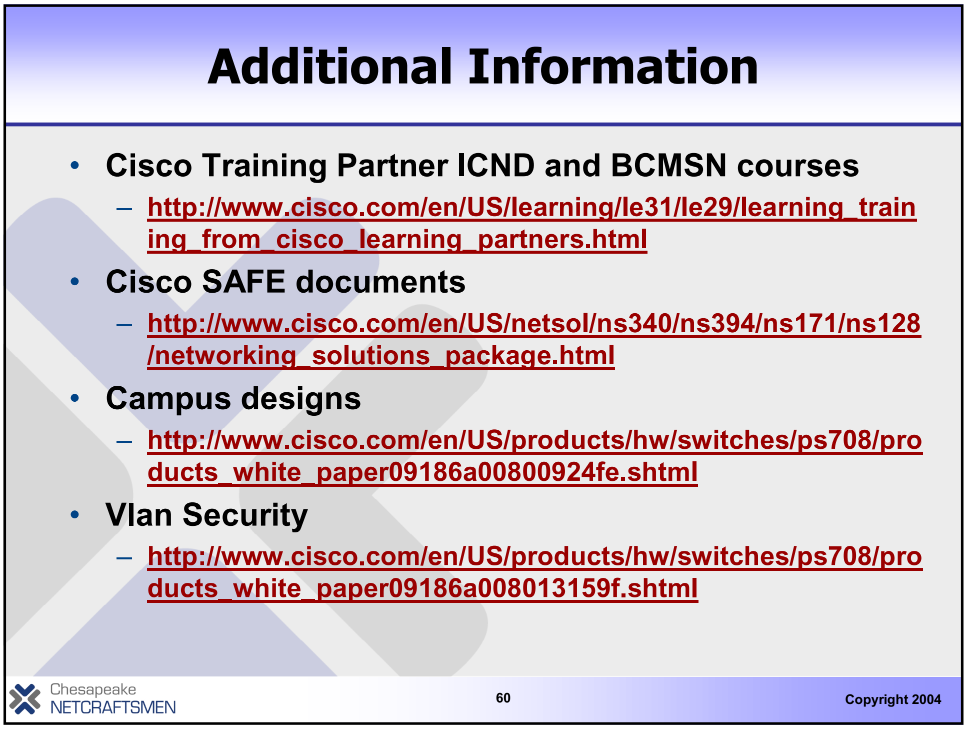
Agenda

- Physical Design Models
- L2/L3 Hierarchy
- Case Studies
- Wrap-Up



Conclusions

- High Availability: don’t be the weakest link!
  - Good design can mitigate ST issues and save you troubleshooting time
  - You’ll then find that human error causes most of the operational problems that occur
- Use Cisco ST features to your advantage
- WLAN design tempts us to have larger VLANs
  - Don’t run VLANs through your L3 MDF!
  - Stick with the proven designs



Additional Information

- Cisco Training Partner ICND and BCMSN courses
  - http://www.cisco.com/en/US/learning/le31/le29/learning_training_from_cisco_learning_partners.html
- Cisco SAFE documents
  - http://www.cisco.com/en/US/netsol/ns340/ns394/ns171/ns128/networking_solutions_package.html
- Campus designs
  - http://www.cisco.com/en/US/products/hw/switches/ps708/products_white_paper09186a00800924fe.shtml
- Vlan Security
  - http://www.cisco.com/en/US/products/hw/switches/ps708/products_white_paper09186a008013159f.shtml

Questions?

Chesapeake
NETCRAFTSMEN

61

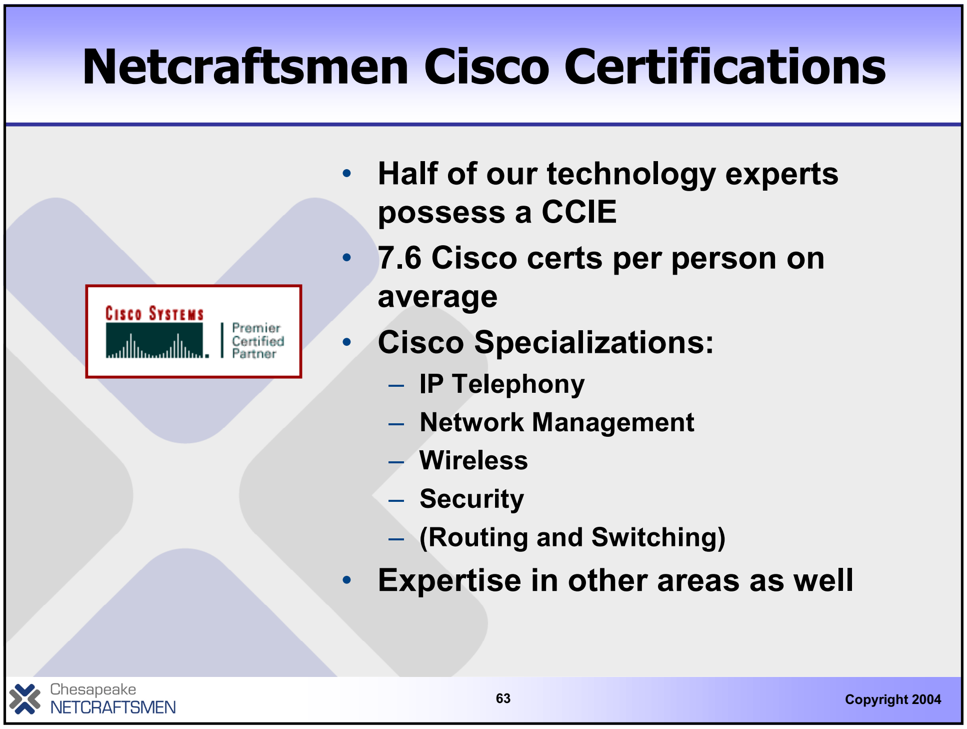
Copyright 2004

About the Speaker

- Dr. Pete Welcher
  - Cisco CCIE #1773, CCSI #94014, CCIP
  - Network design & management consulting
    - Stock quotation firm, 3000 routers, TCP/IP
    - Second stock quotation firm, 2000 routers, UDP broadcasts
    - Hotel chain, 1000 routers, SNA
    - Government agency, 1500 routers
  - Teach many of the Cisco courses
- Enterprise Networking Magazine articles
  - http://www.netcraftsmen.net/welcher/papers

Chesapeake
NETCRAFTSMEN

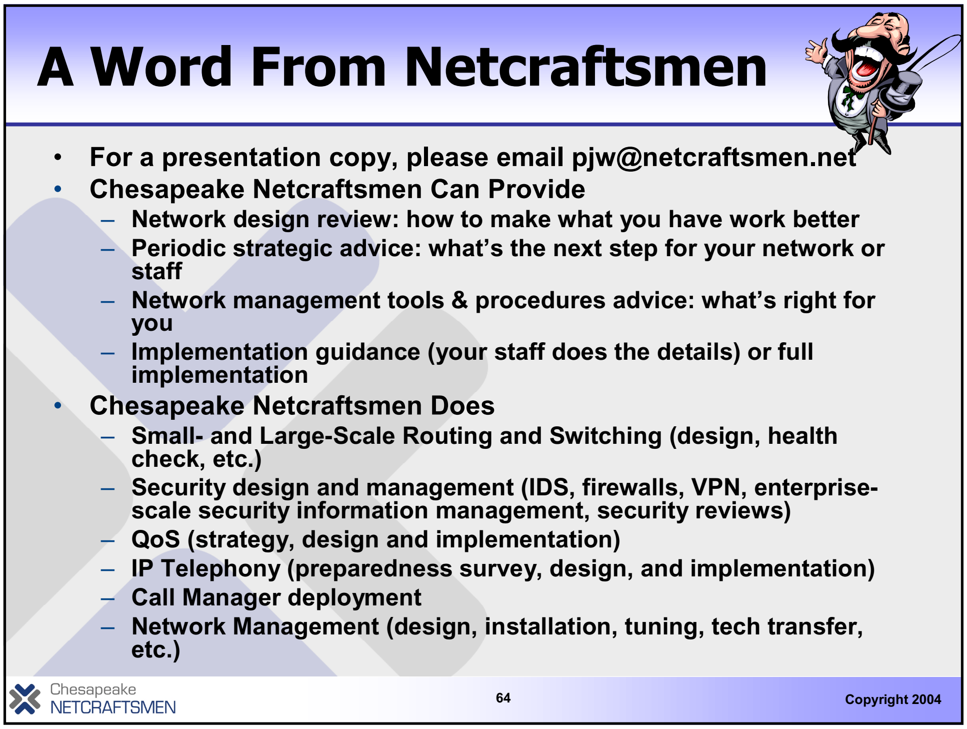
Copyright 2004



Netcraftsmen Cisco Certifications

Cisco Systems | Premier Certified Partner

- Half of our technology experts possess a CCIE
- 7.6 Cisco certs per person on average
- Cisco Specializations:
  - IP Telephony
  - Network Management
  - Wireless
  - Security
  - (Routing and Switching)
- Expertise in other areas as well



A Word From Netcraftsmen

- For a presentation copy, please email pjw@netcraftsmen.net
- Chesapeake Netcraftsmen Can Provide
  - Network design review: how to make what you have work better
  - Periodic strategic advice: what's the next step for your network or staff
  - Network management tools & procedures advice: what's right for you
  - Implementation guidance (your staff does the details) or full implementation
- Chesapeake Netcraftsmen Does
  - Small- and Large-Scale Routing and Switching (design, health check, etc.)
  - Security design and management (IDS, firewalls, VPN, enterprise-scale security information management, security reviews)
  - QoS (strategy, design and implementation)
  - IP Telephony (preparedness survey, design, and implementation)
  - Call Manager deployment
  - Network Management (design, installation, tuning, tech transfer, etc.)