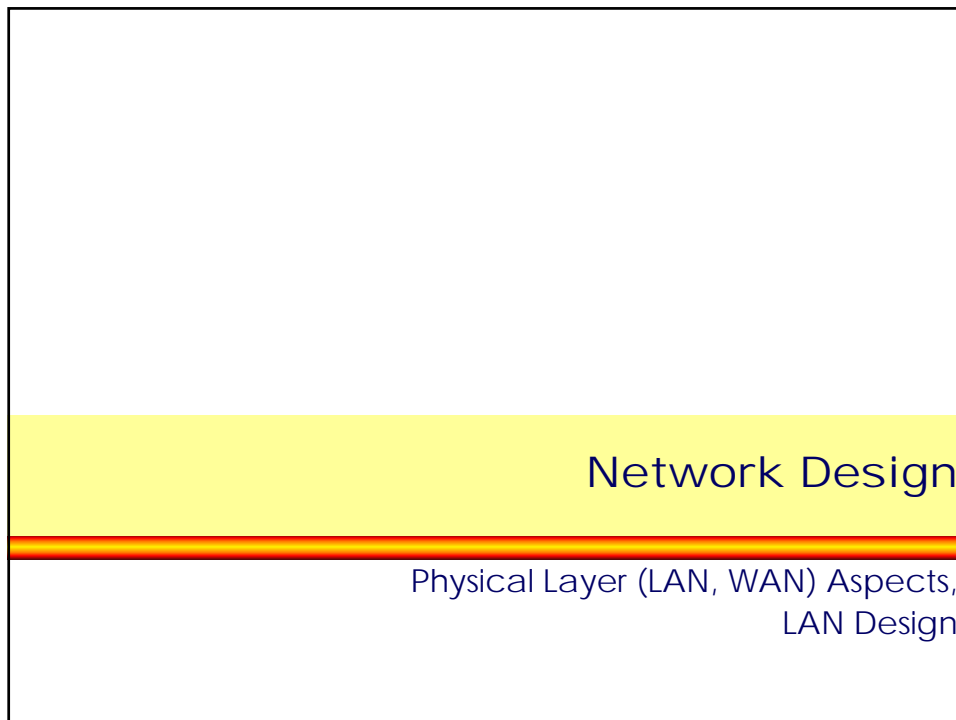


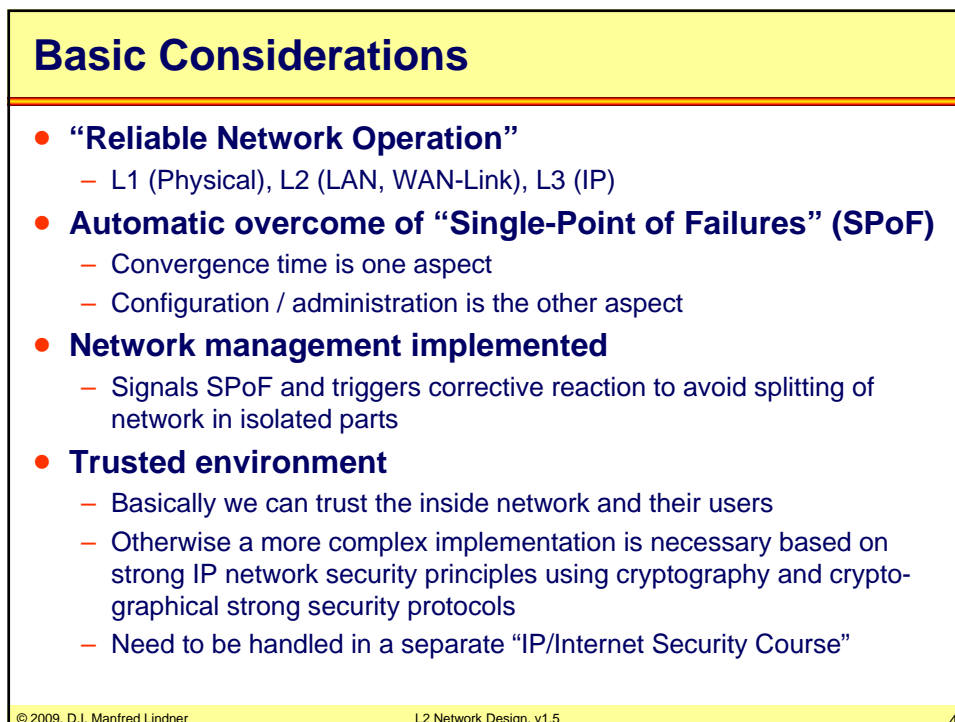
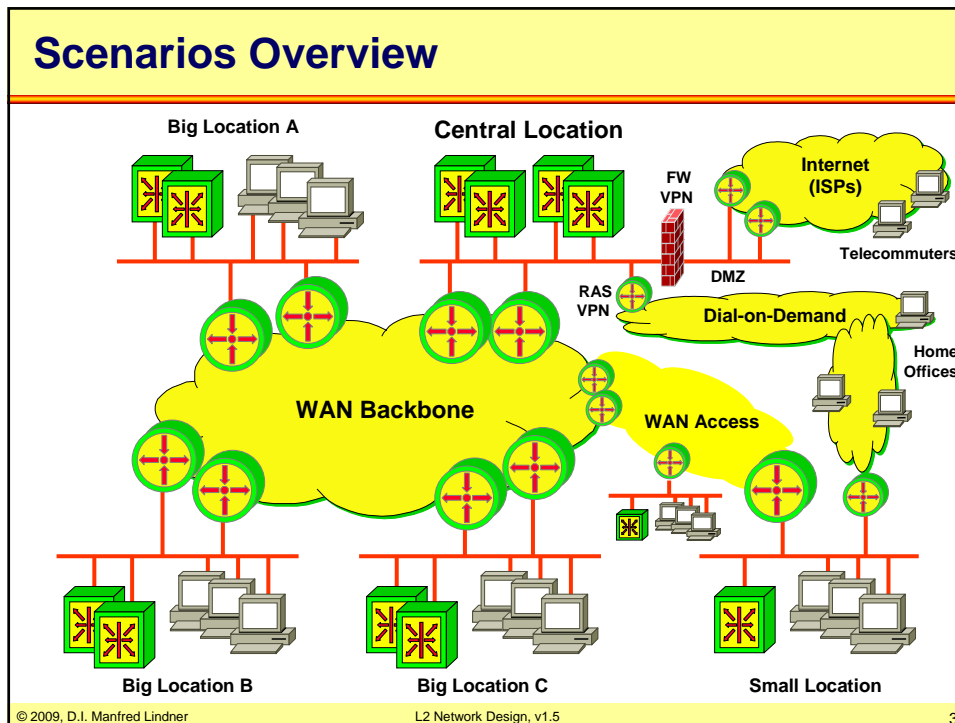
L101 - L2 Network Design



Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

L101 - L2 Network Design



L101 - L2 Network Design

Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

5

Network Infrastructure

- **Technical equipment needs a defined environment for physical environment parameters**
 - Electricity
 - e.g. UPS
 - Humidity, Temperature
 - e.g. air-condition, positive air flow
- **Upper layer redundancy need to follow the physical paths**
 - e.g. two physical links to two different switches or routers should take separated paths (LAN and WAN)
 - redundant network components should be located at different physical places

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

6

L101 - L2 Network Design

Network Infrastructure

- **Technical equipment needs conventional physical protection**
 - Access control for buildings, rooms
 - Guards, cards, passwords, ...
 - Technical equipment in locked environment to avoid unauthorized access like direct attachment via management console
 - Hubs, switches, routers, server
 - WLAN access points (?)
 - Must be monitored (camera) and should produce an alarm in case of manipulation especially in public areas

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

7

Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

8

L101 - L2 Network Design

Structured Cabling (LAN)

• Physical Wiring

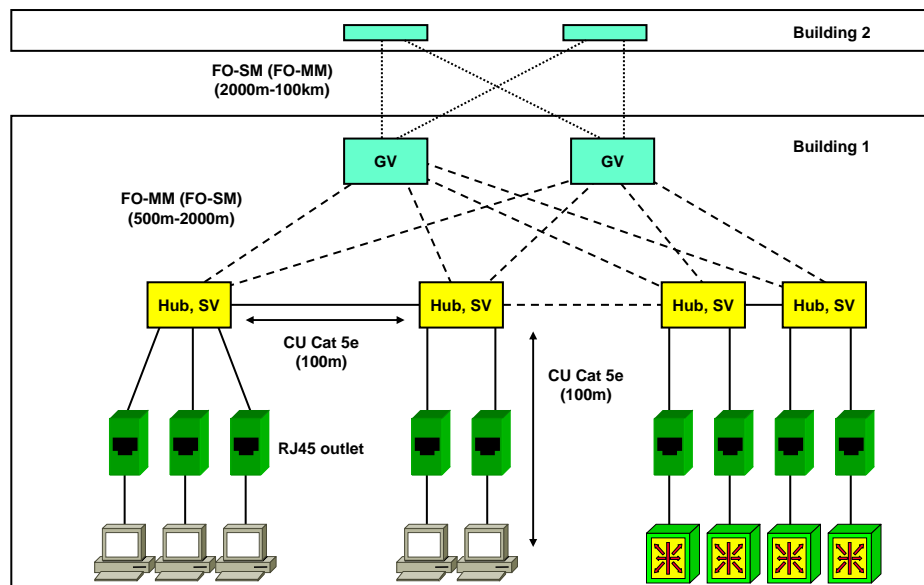
- Should follow the principle of structured cabling
- Primary
 - End system to first "Hub" (Repeater or L2 Switch)
 - "Stockwerkverteiler"
 - CU-UTP, Category 5e or better
 - FO for extreme conditions only
- Secondary
 - Hubs to central functions
 - "Gebäudeverteiler"
 - FO-MM (FO-SM)
- Tertiary
 - Interconnections of buildings
 - FO-MM (FO-SM)

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

9

Structured Cabling (LAN)



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

10

L101 - L2 Network Design

Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

11

WAN Alternatives

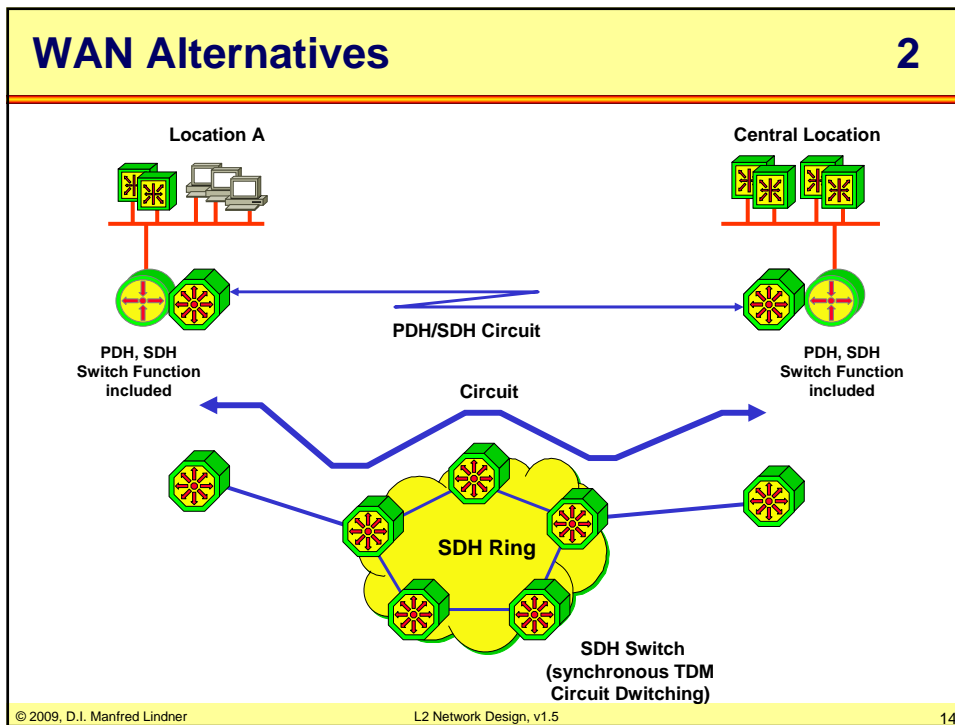
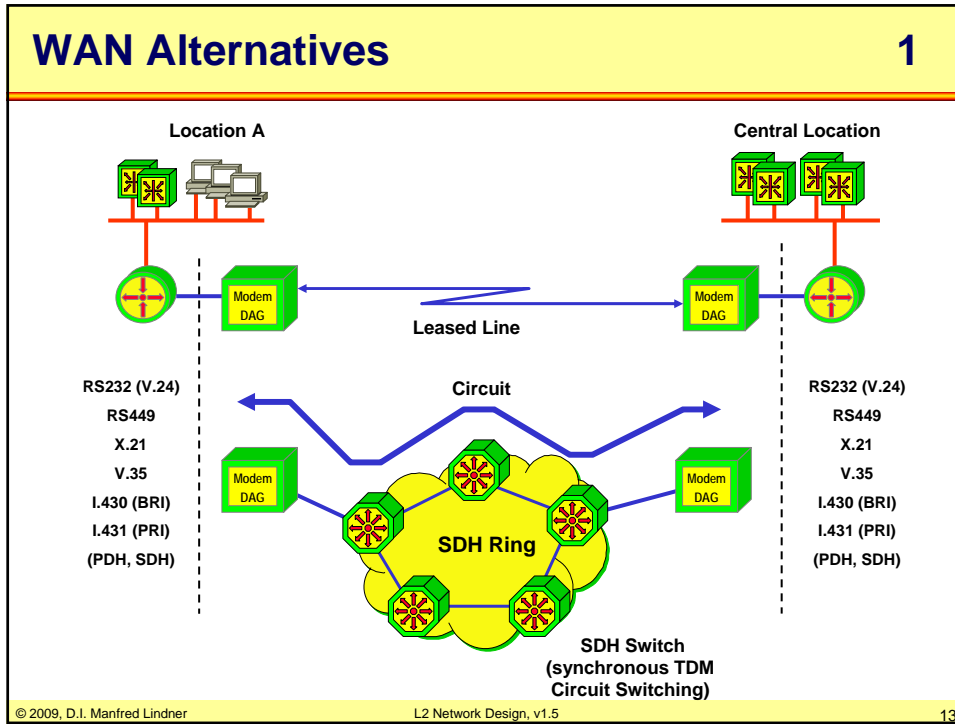
- **Leased line service**
 - Usually based on PDH, SDH or ISDN
 - “Standleitung”
 - “Circuit” with defined bandwidth and constant delay
- **Virtual circuit service**
 - X.25, Frame Relay, ATM
 - PVC or SVC
 - “Virtual circuit” with certain QoS (Quality of Service) guarantees (e.g. committed minimal throughput, bounded delay = worst case delay)
- **Backup service**
 - “Dial on Demand”, “Bandwidth on Demand”
 - ISDN (circuit), X.25-(Frame Relay)-ATM (virtual circuit in SVC operation mode)

© 2009, D.I. Manfred Lindner

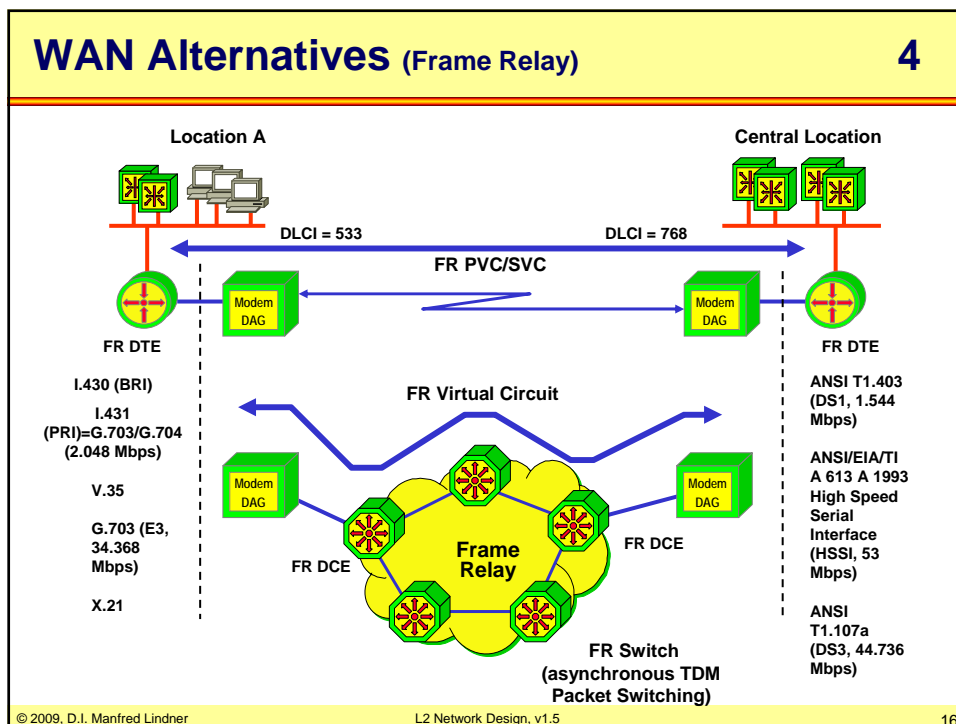
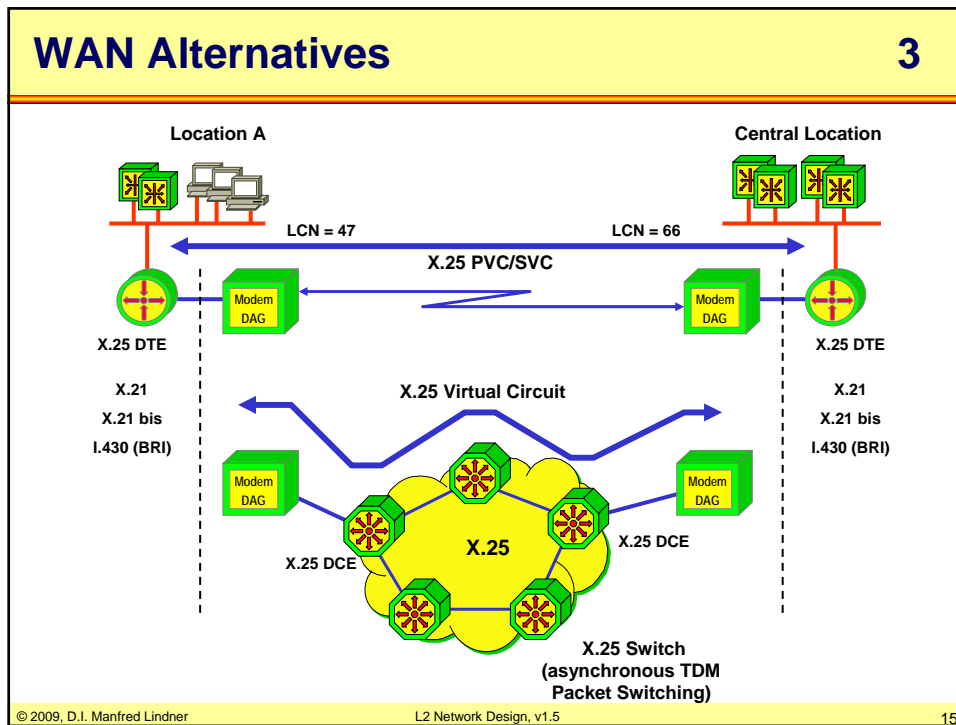
L2 Network Design, v1.5

12

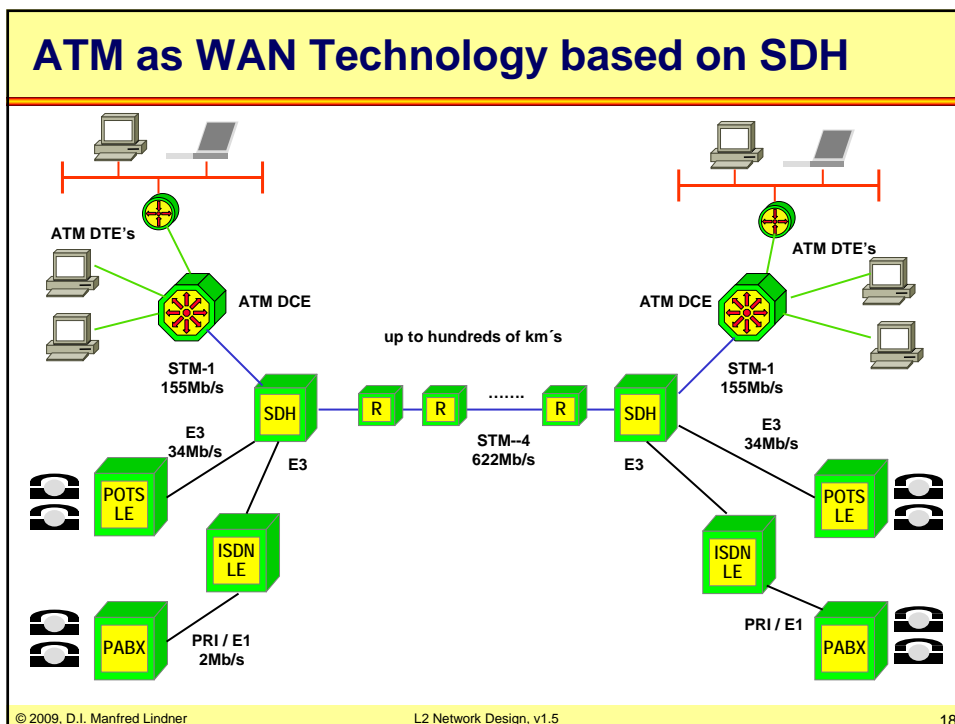
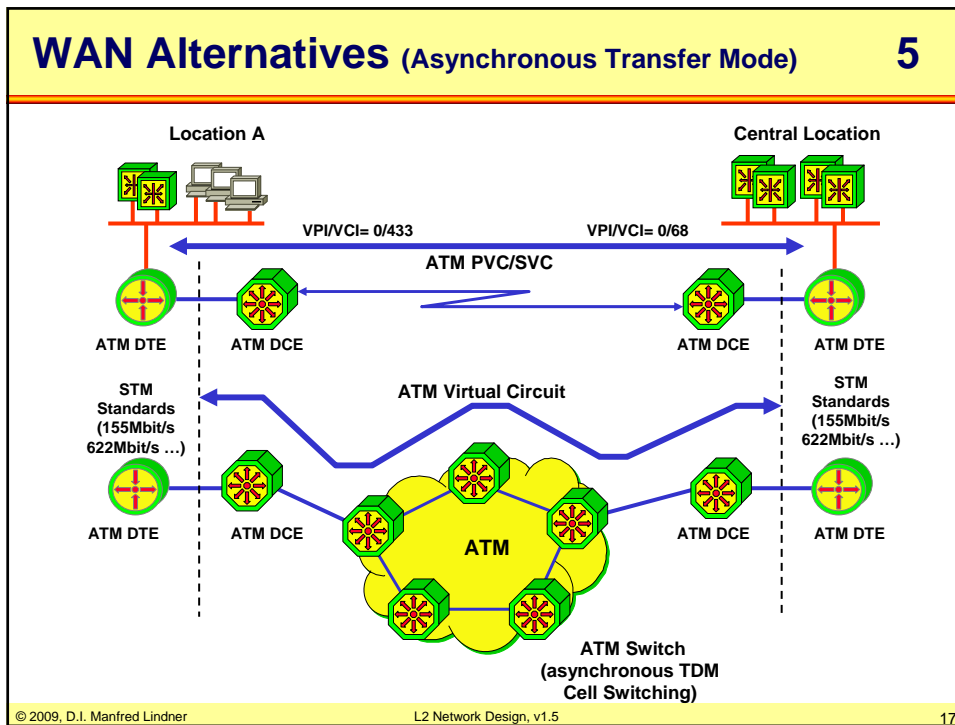
L101 - L2 Network Design



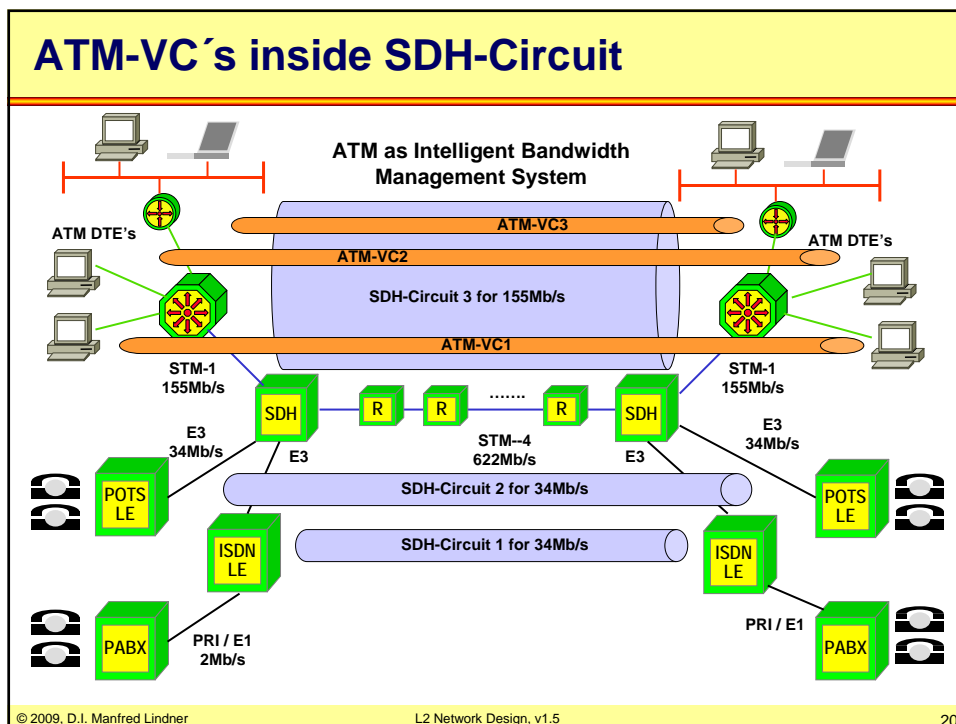
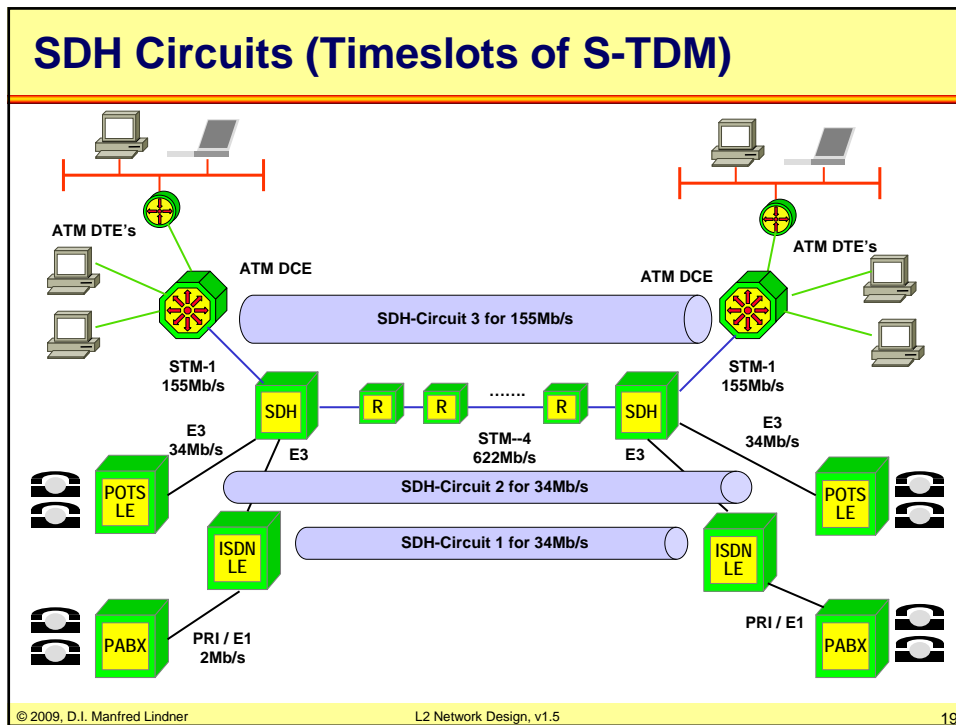
L101 - L2 Network Design



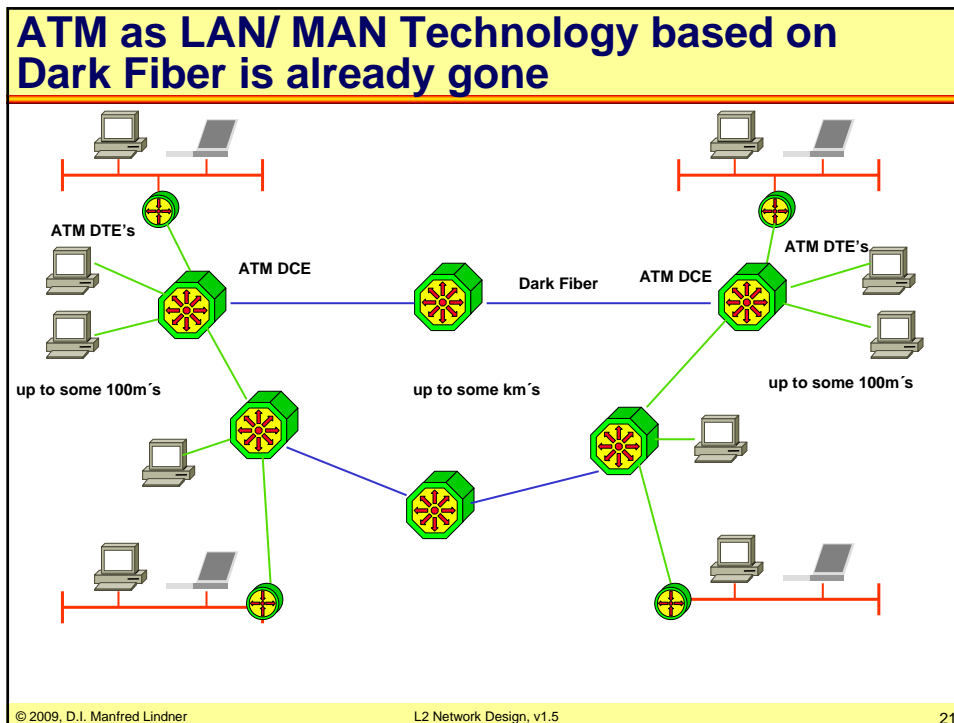
L101 - L2 Network Design



L101 - L2 Network Design



L101 - L2 Network Design



WAN Service Considerations

- **Who is responsible for providing the service?**
 - Service Provider or
 - Department of own company
 - Note: functions of configuration, implementation, management, operation, monitoring, maintenance need to be established
 - Service Level Agreement (SLA)
- **What about redundancy?**
 - Can a redundant line take a true different physical way end-to-end?
 - Should different service providers be used?

L101 - L2 Network Design

Classical VPN's

- X.25, Frame Relay or ATM in the core
- Dedicated physical switch ports for every customers CPE
 - router, bridge, computer
- Customer traffic separation in the core done by concept of virtual circuit
 - PVC service
 - management overhead
 - SVC service with closed user group feature
 - signaling overhead
- Separation of customers inherent to virtual circuit technique

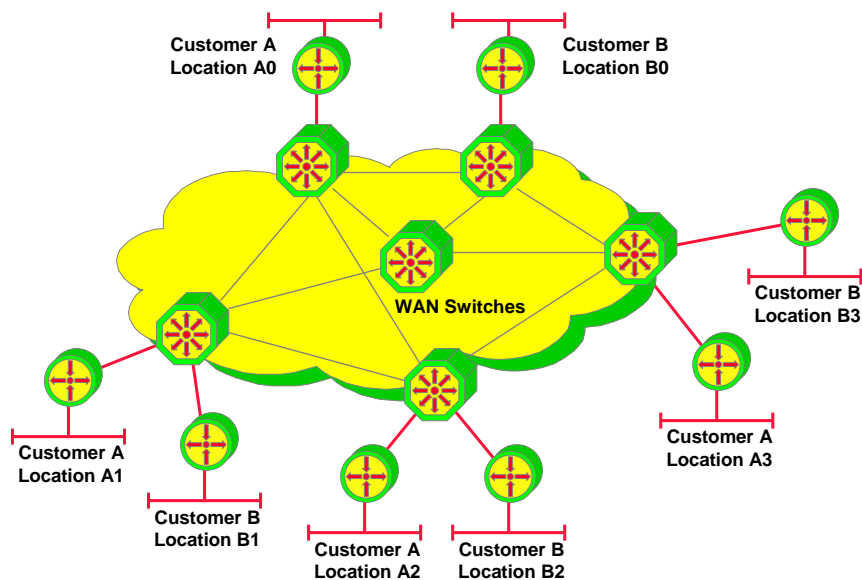
VPN's based on Overlay Model

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

23

Physical Topology of Classical VPN

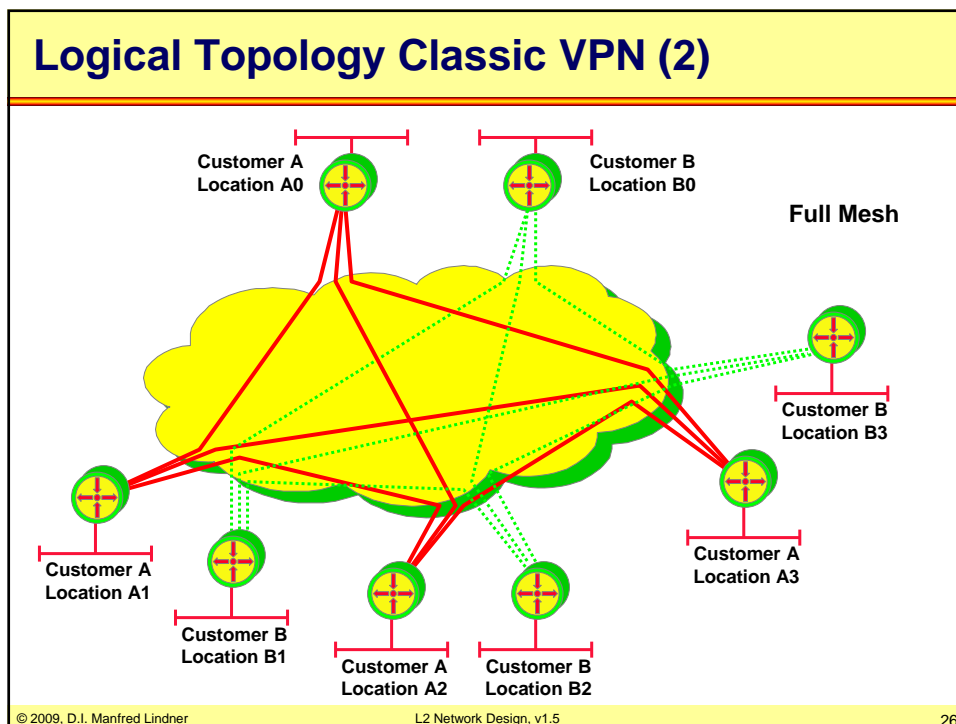
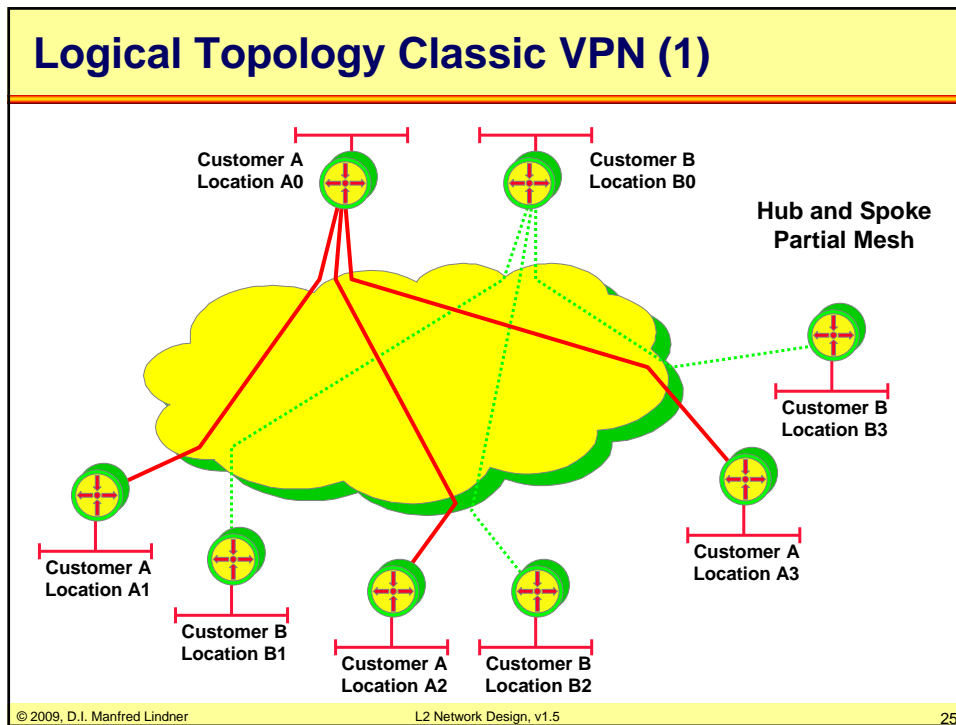


© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

24

L101 - L2 Network Design



L101 - L2 Network Design

Virtual Private Networks based on IP

- Single technology end-to-end
 - IP forwarding and IP routing
- No WAN switches in the core
 - Based on different technology (X.25, FR or ATM)
 - Administered by different management techniques
- Often private means control over separation but not privacy
 - Data are seen in clear-text in the core

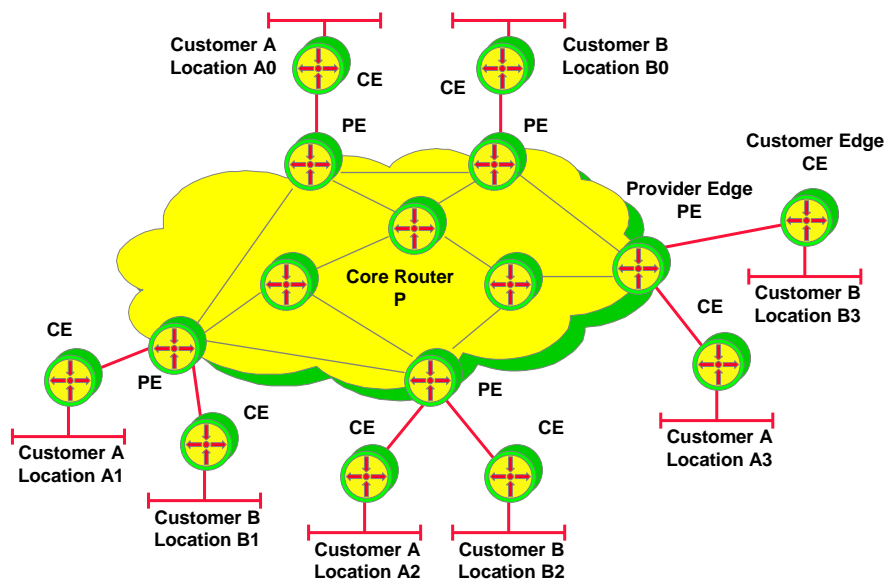
VPN's based on Peer Model

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

27

Physical Topology IP VPN



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

28

L101 - L2 Network Design

Possible Solutions for IP VPN's

- **IP addresses of customers non overlapping**
 - filtering and policy routing techniques can be used in order to guarantee separation of IP traffic
 - exact technique depends on who manages routes at the customer site
- **IP addresses of customers overlapping**
 - tunneling techniques must be used in order to guarantee separation of IP traffic
 - GRE
 - L2F, PPTP, L2TP
 - MPLS-VPN
- **If privacy is a topic**
 - encryption techniques must be used
 - SSL/TLS, IPsec

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

29

Tunneling Solutions for IP VPN's

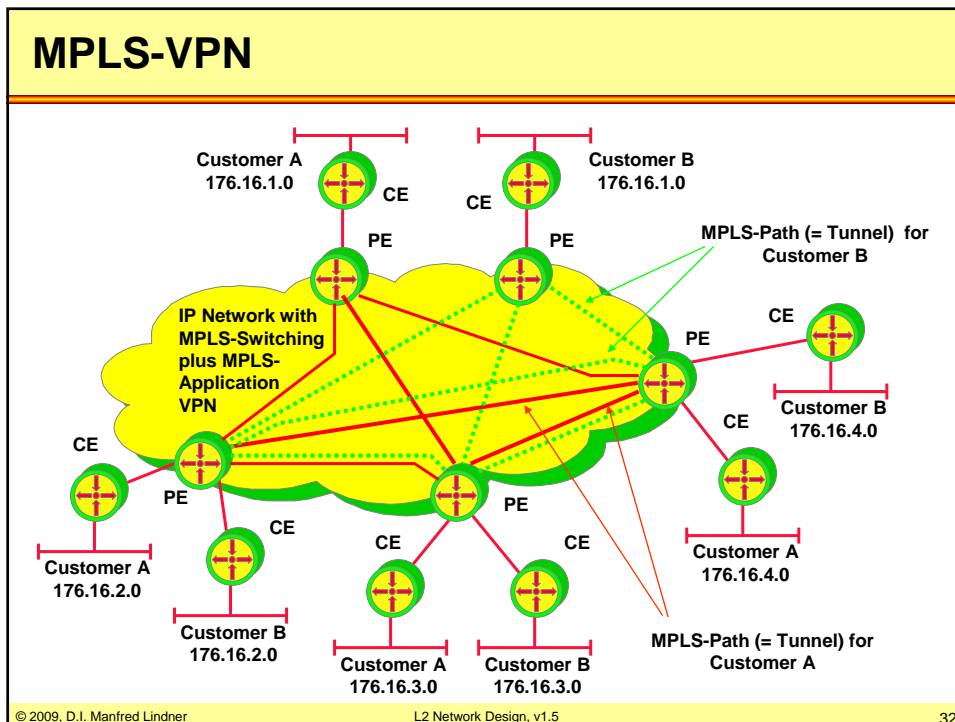
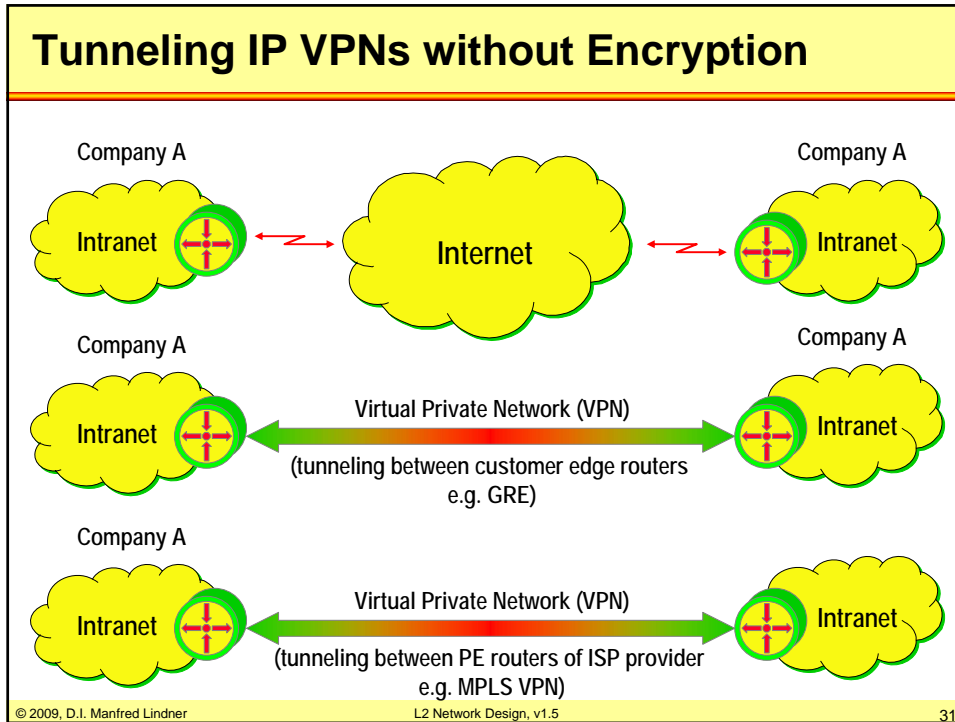
- **Tunneling techniques are used in order to guarantee separation of IP traffic**
 - IP in IP Tunneling or GRE (Generic Routing Encapsulations)
 - Bad performance on PE router
 - PPTP or L2TP for LAN to LAN interconnection
 - Originally designed for PPP Dial-up connections
 - LAN – LAN is just a special case
 - MPLS-VPN
 - Best performance on PE router
- **In all these cases**
 - Privacy still an aspect of the customer

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

30

L101 - L2 Network Design



L101 - L2 Network Design

MPLS VPN – Best of Both Worlds

- **Combines VPN Overlay model with VPN Peer model**
- **PE routers allow route isolation**
 - By using Virtual Routing and Forwarding Tables (VRF) for differentiating routes from the customers
 - Allows overlapping address spaces
- **PE routers participate in P-routing**
 - Hence optimum routing between sites
 - Label Switched Paths are used within the core network
 - Easy provisioning (sites only)
- **Overlapping VPNs possible**
 - By a simple (?) attribute syntax

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

33

What does MPLS VPN mean for the Provider?

- **Requires MPLS Transport within the core**
 - Using the label stack feature of MPLS
- **Requires MP-BGP among PE routers**
 - Supports IPv4/v6, VPN-IPv4, multicast
 - Default behavior: BGP-4
- **Requires VPN-IPv4 96 bit addresses**
 - 64 bit Route Distinguisher (RD)
 - 32 bit IP address
- **Every PE router uses one VRF for each VPN**
 - Virtual Routing and Forwarding Table (VRF)

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

34

L101 - L2 Network Design

Encryption Solutions for IP VPN's

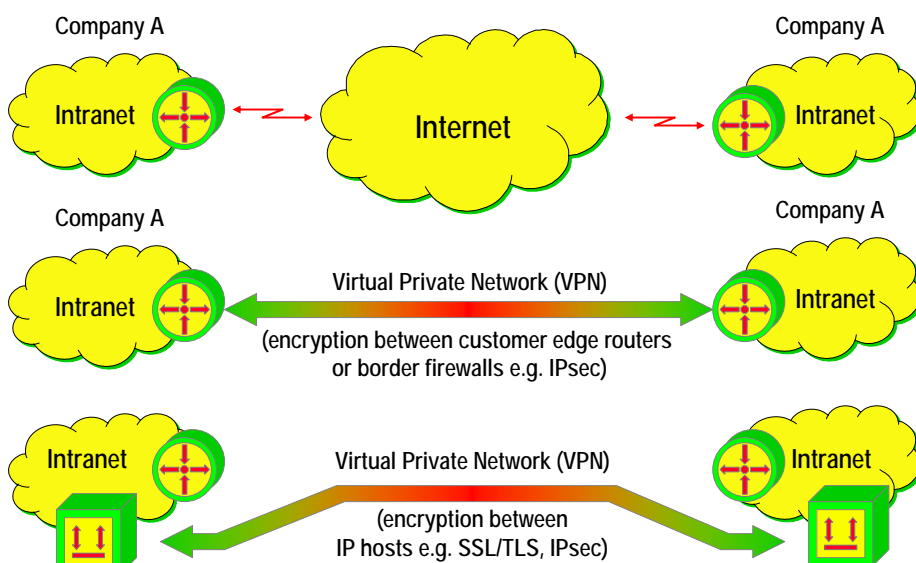
- **If privacy is a topic tunneling techniques with encryption are used in order to hide IP traffic**
 - SSL (secure socket layer)
 - Usually end-to-end
 - Between TCP and Application Layer
 - IPsec
 - Could be end-to-end
 - Could be between special network components (e.g. firewalls, VPN concentrators) only
 - Between IP and TCP/UDP Layer
 - PPTP and L2TP Tunnels
 - With encryption turned on via PPP option

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

35

Tunneling IP VPNs without Encryption

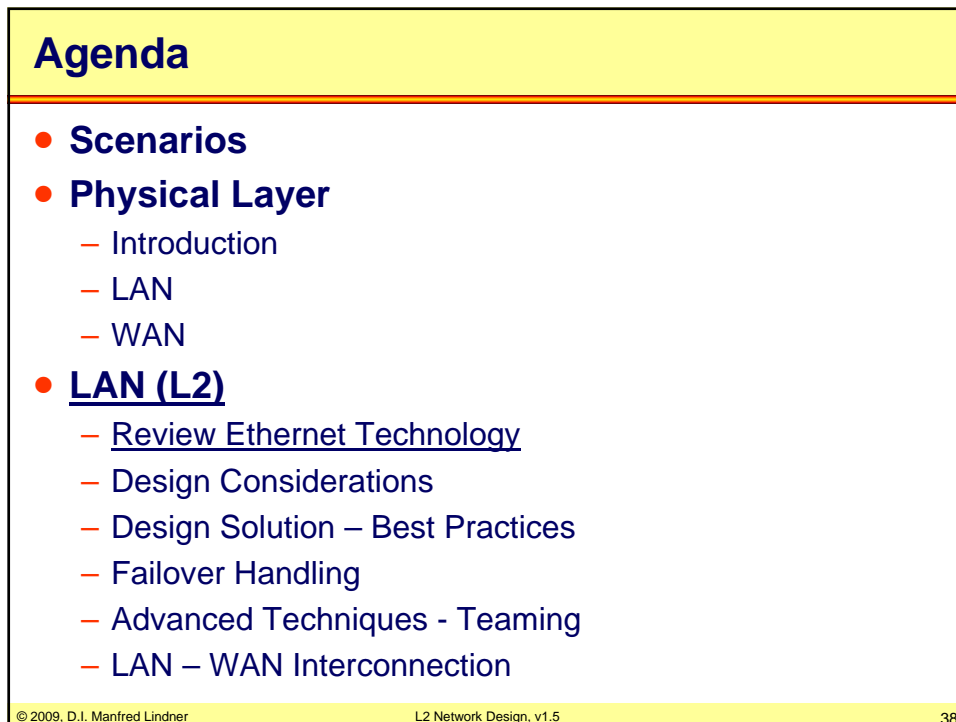
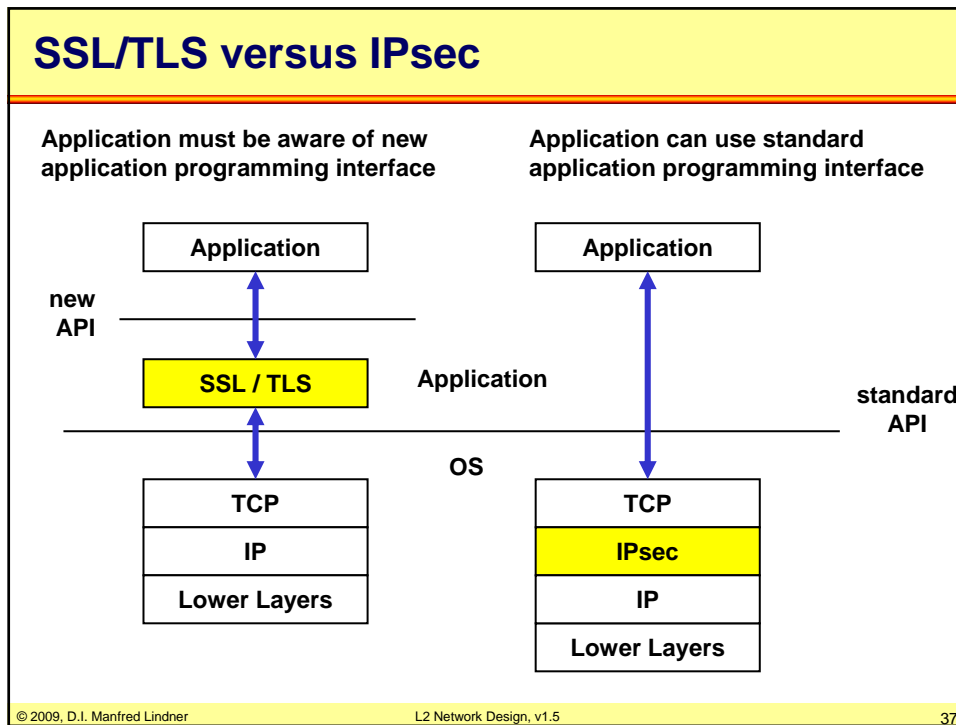


© 2009, D.I. Manfred Lindner

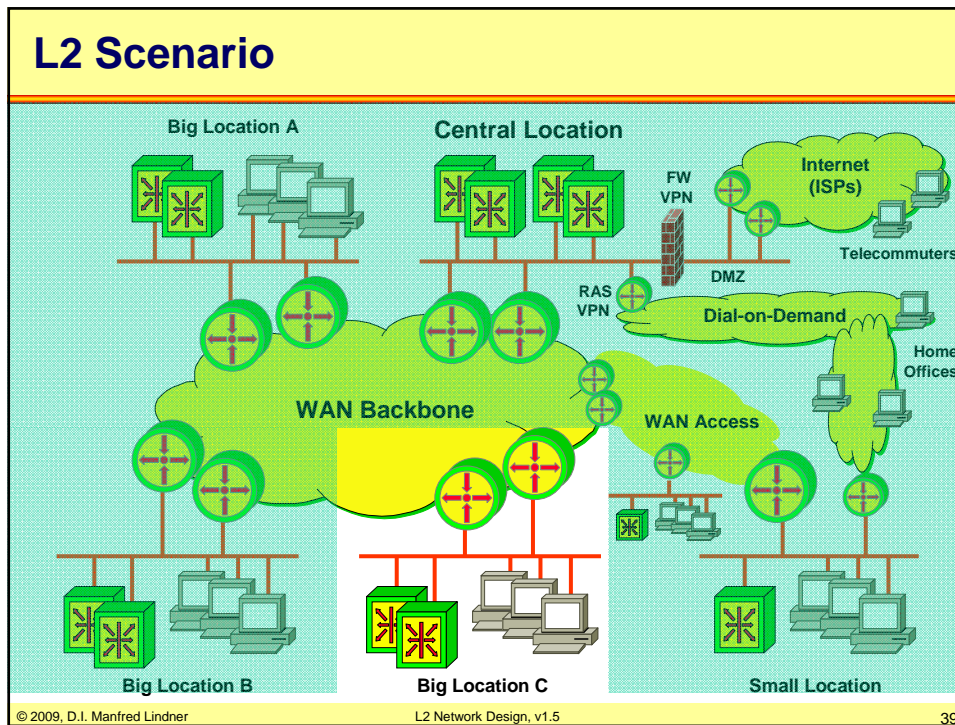
L2 Network Design, v1.5

36

L101 - L2 Network Design



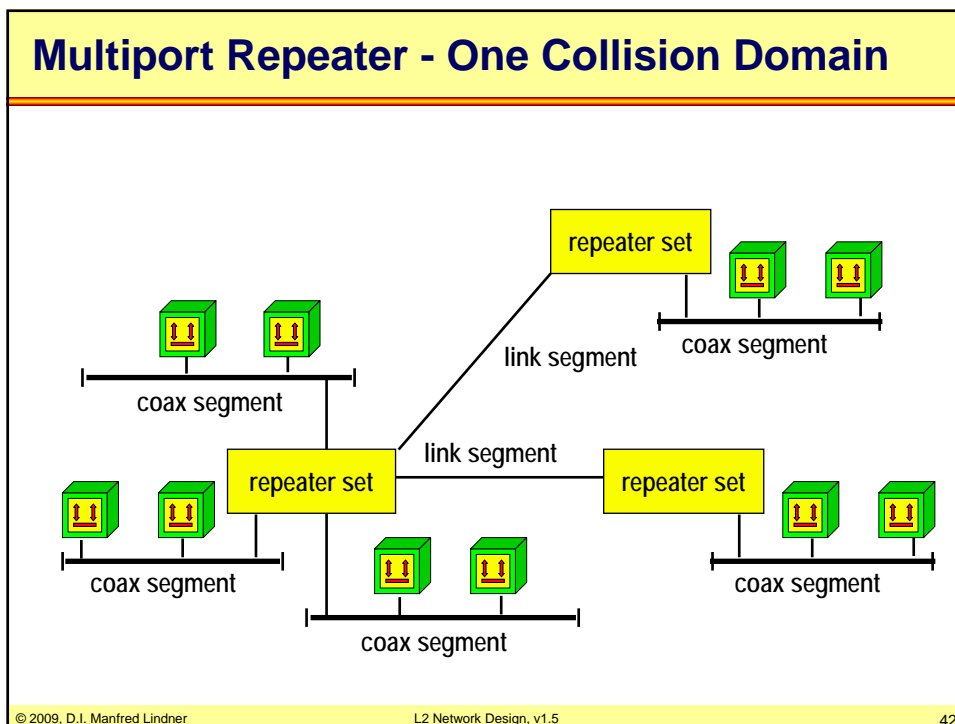
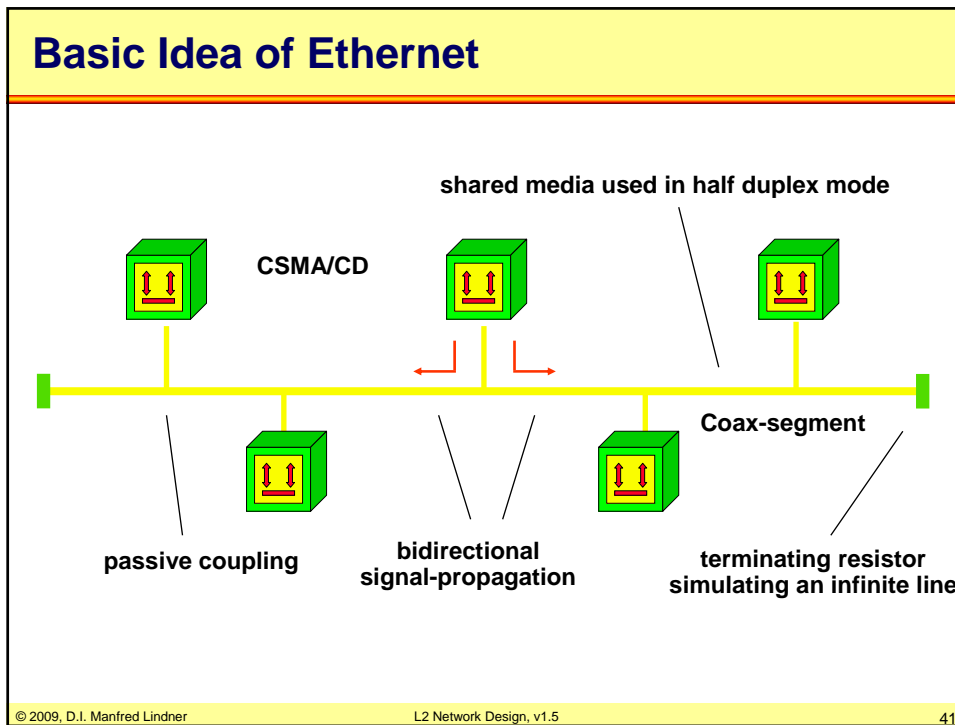
L101 - L2 Network Design



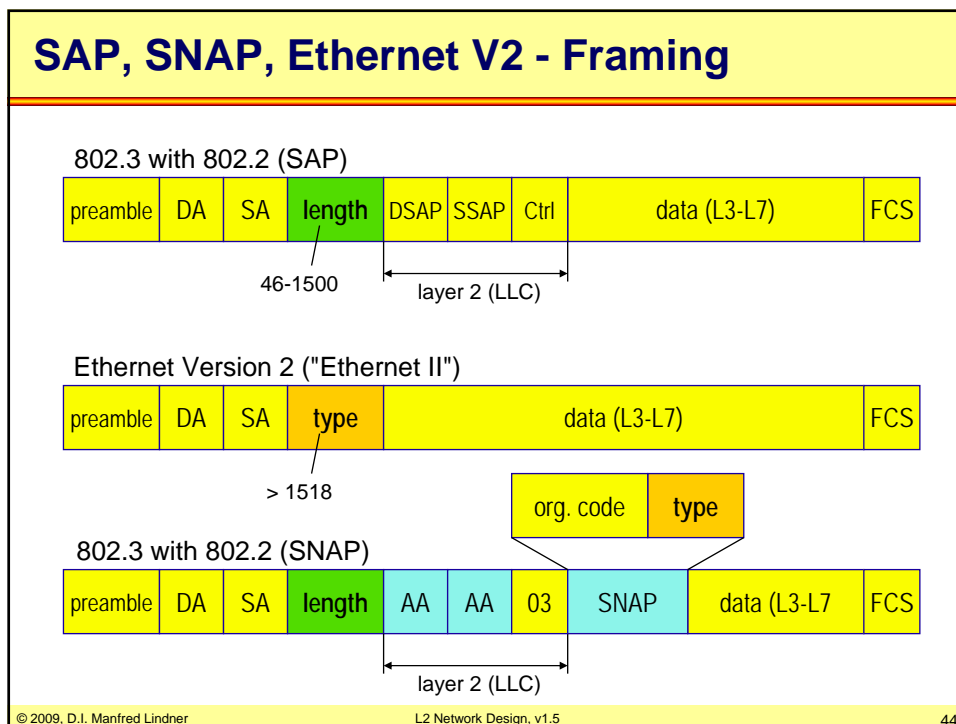
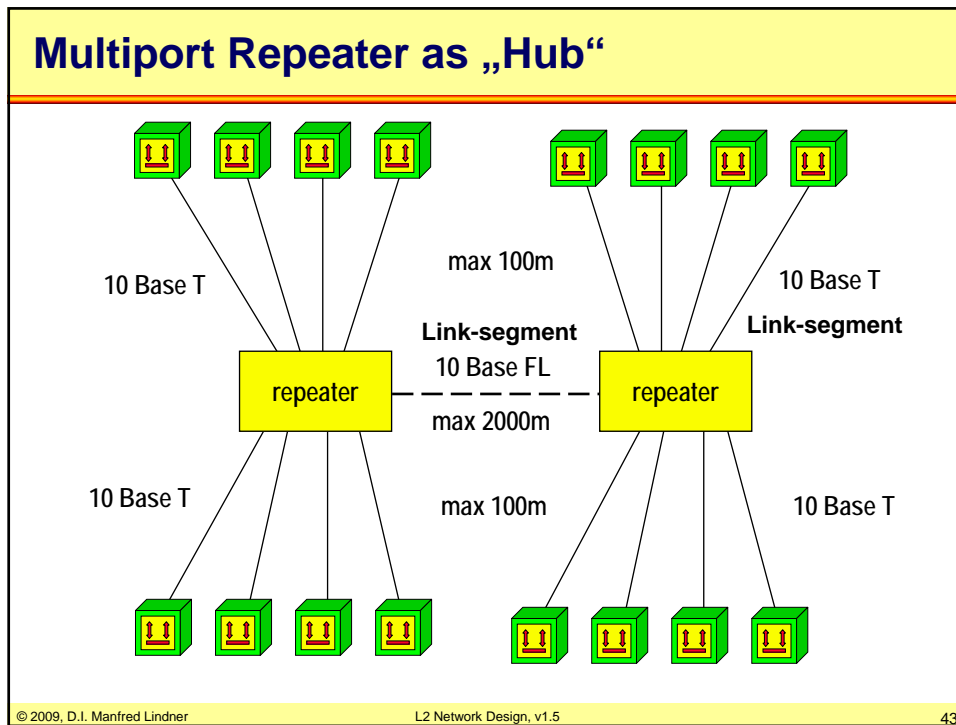
Review L2 Network Components (Basic)

- **Ethernet**
 - LAN
 - Originally shared media (cable)
 - Coax segment
 - CSMA/CD as conflict solution if more than one network station access the cable
 - Limited distance
- **Repeater**
 - Amplifier
 - Expansion of LAN
 - Link segment to connect remote repeaters
 - Collision domain

L101 - L2 Network Design



L101 - L2 Network Design



L101 - L2 Network Design

Review L2 Network Components (Bridge)

- **Bridge**

- Packet switch based on store and forward of frames
- Bridging table based on MAC addresses
- “Transparent Bridging”
 - Learning location of a system based on SA-MAC address with aging mechanism
 - Forwarding decision based on DA-MAC address with filtering and flooding
- Broadcast domain

- **“Ethernet Switch” (L2 switch)**

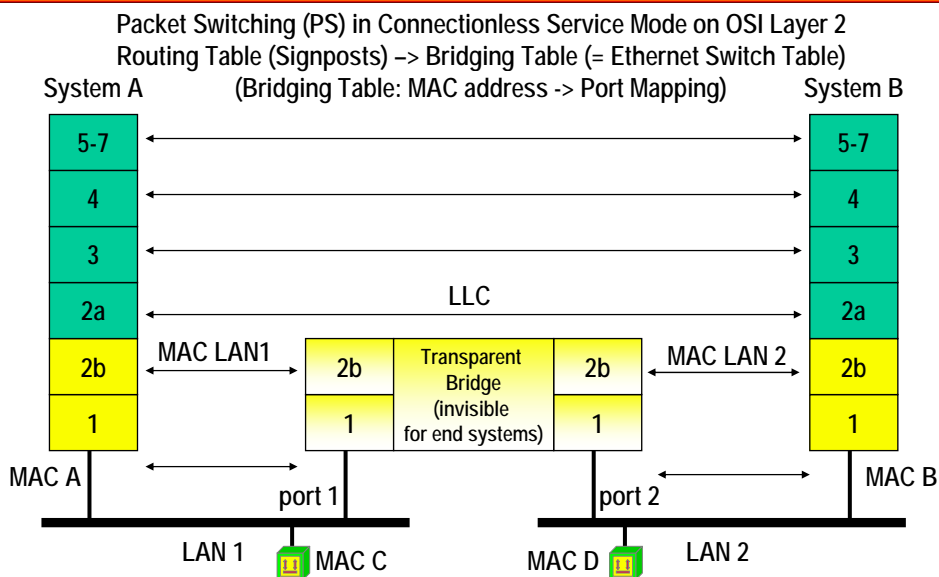
- Fast transparent bridge
- More ports
- VLAN capability

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

45

Transparent Bridge = Ethernet Switch

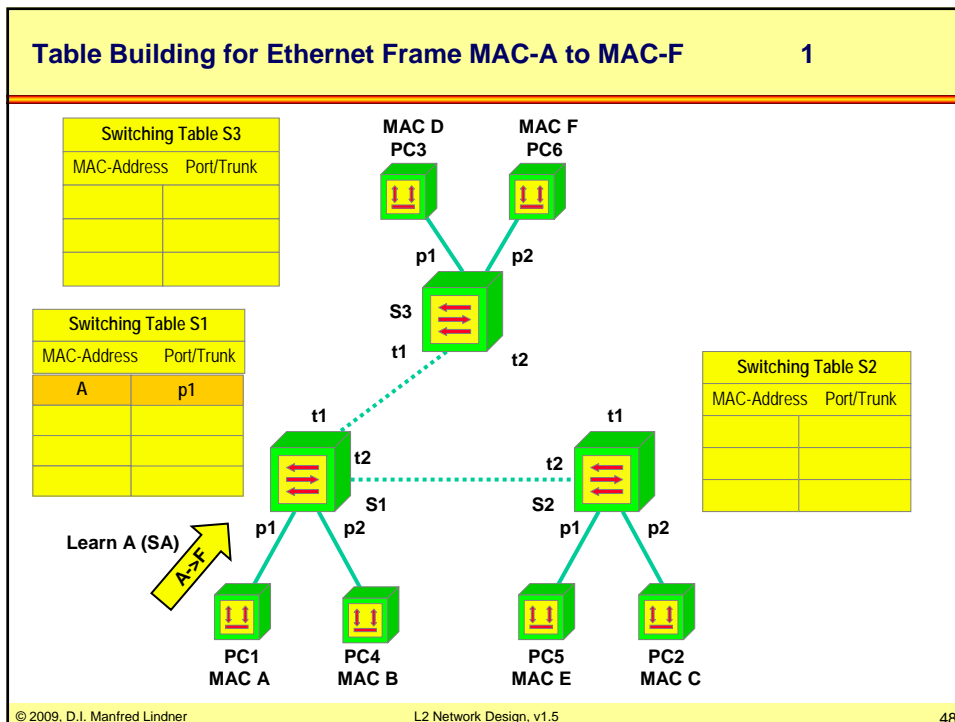
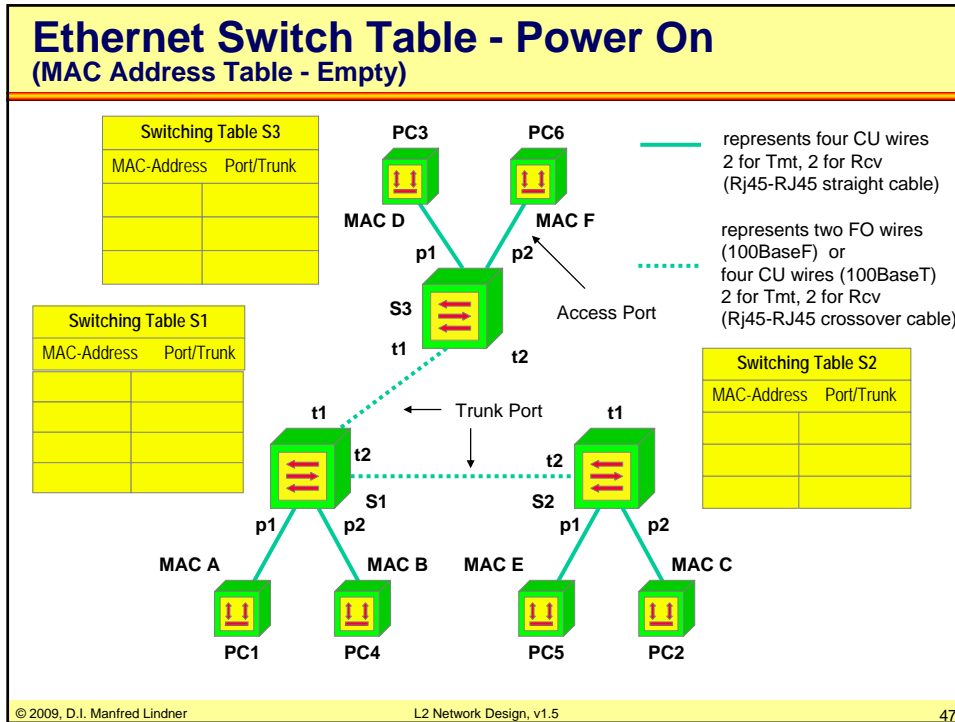


© 2009, D.I. Manfred Lindner

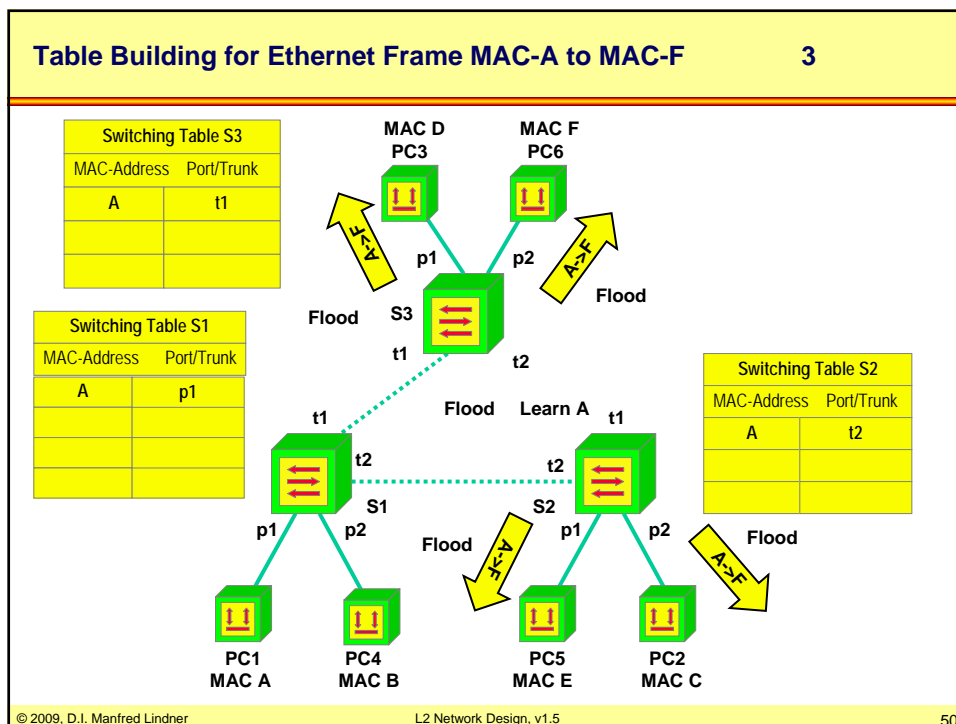
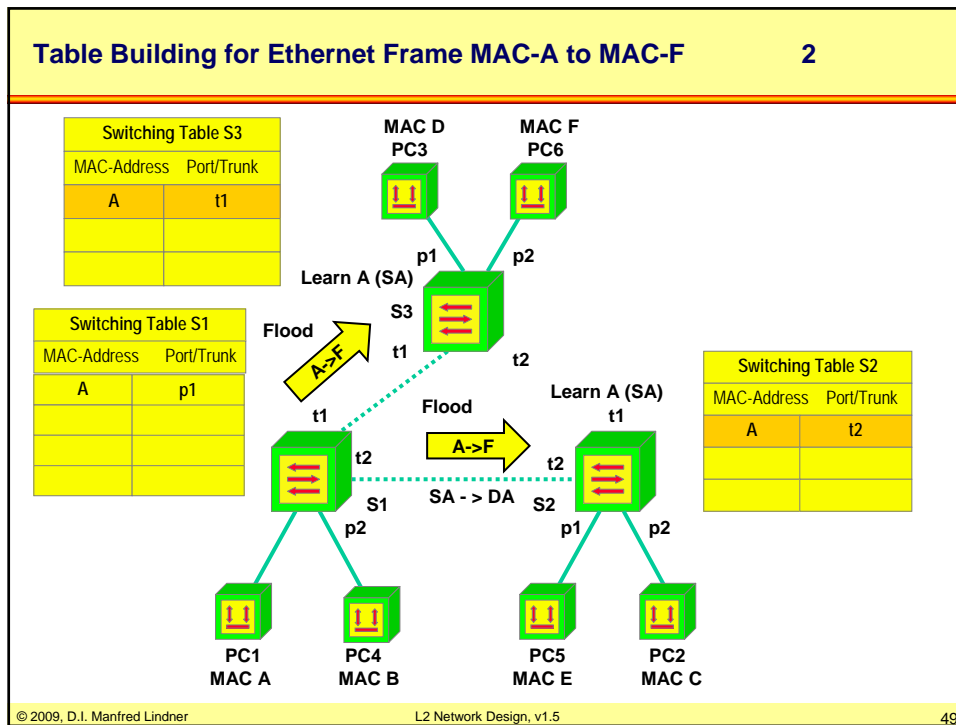
L2 Network Design, v1.5

46

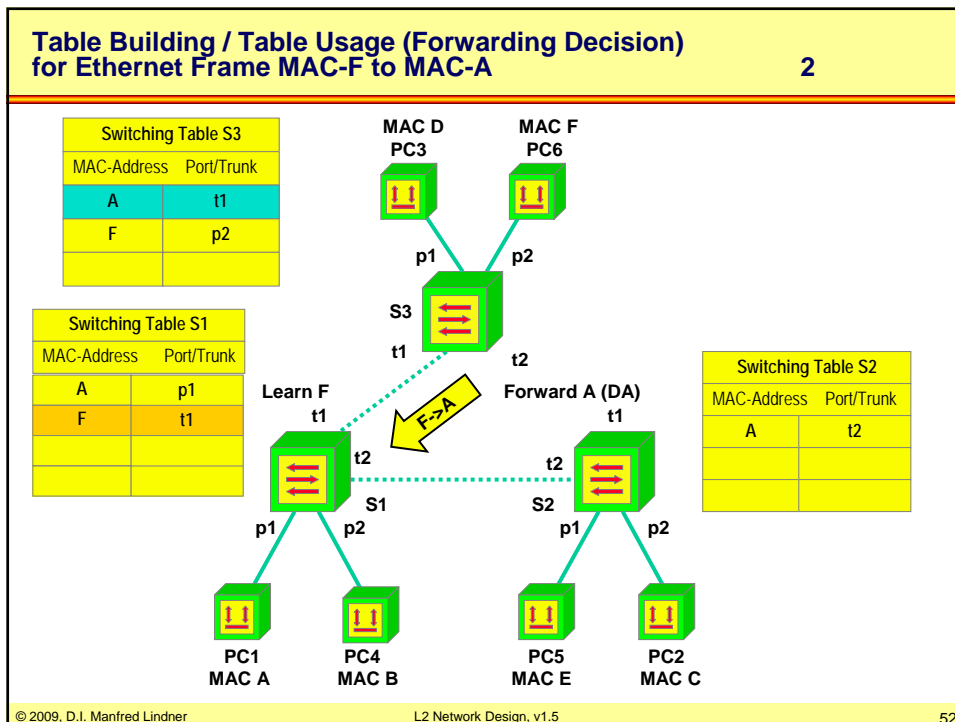
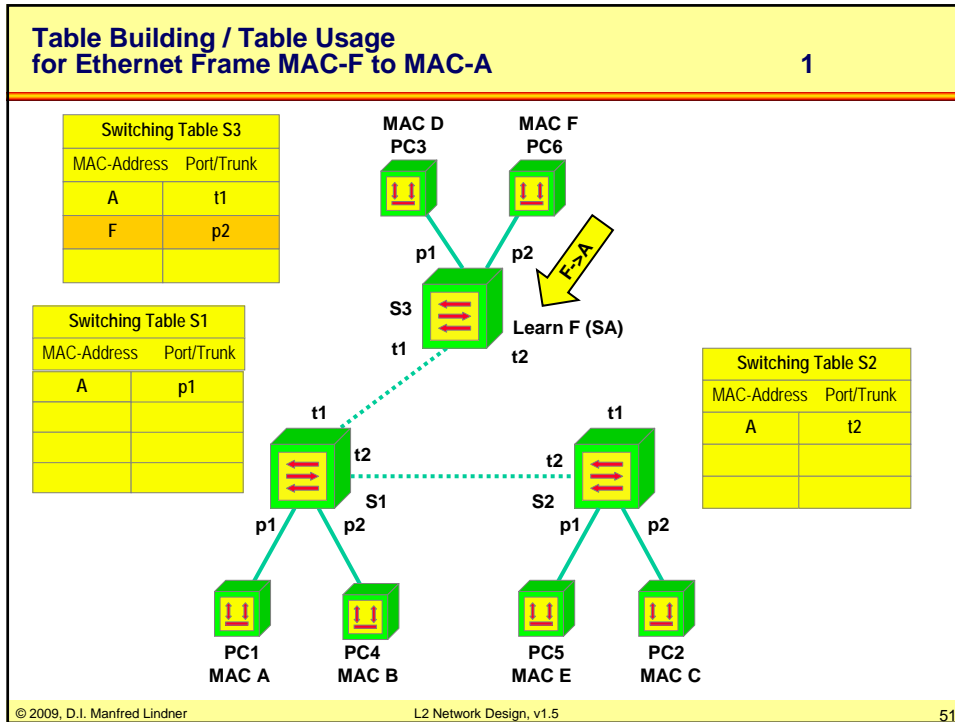
L101 - L2 Network Design



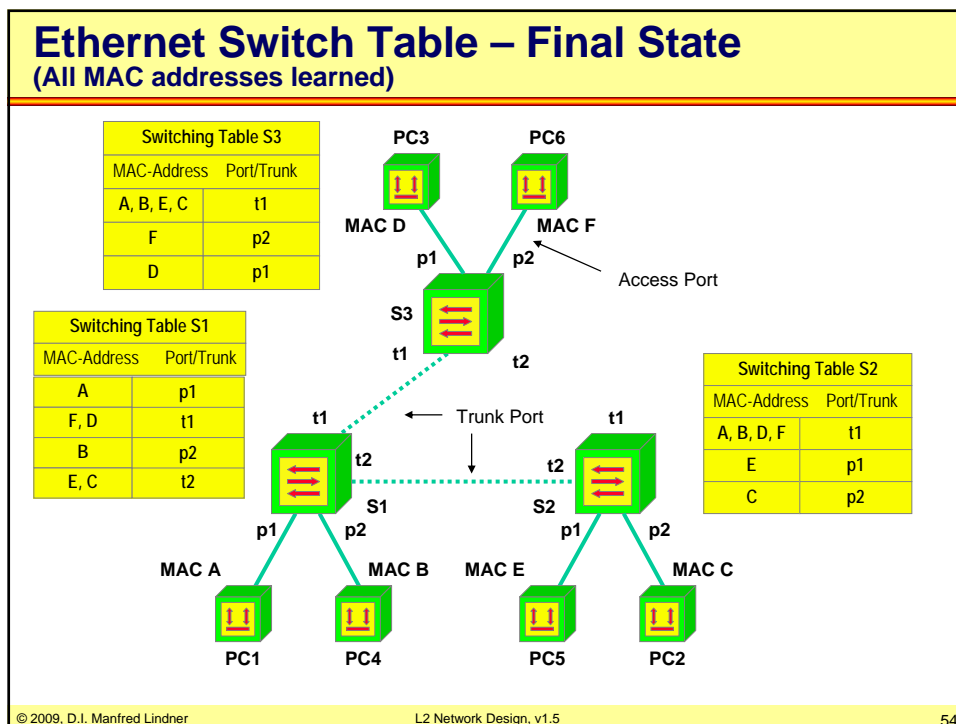
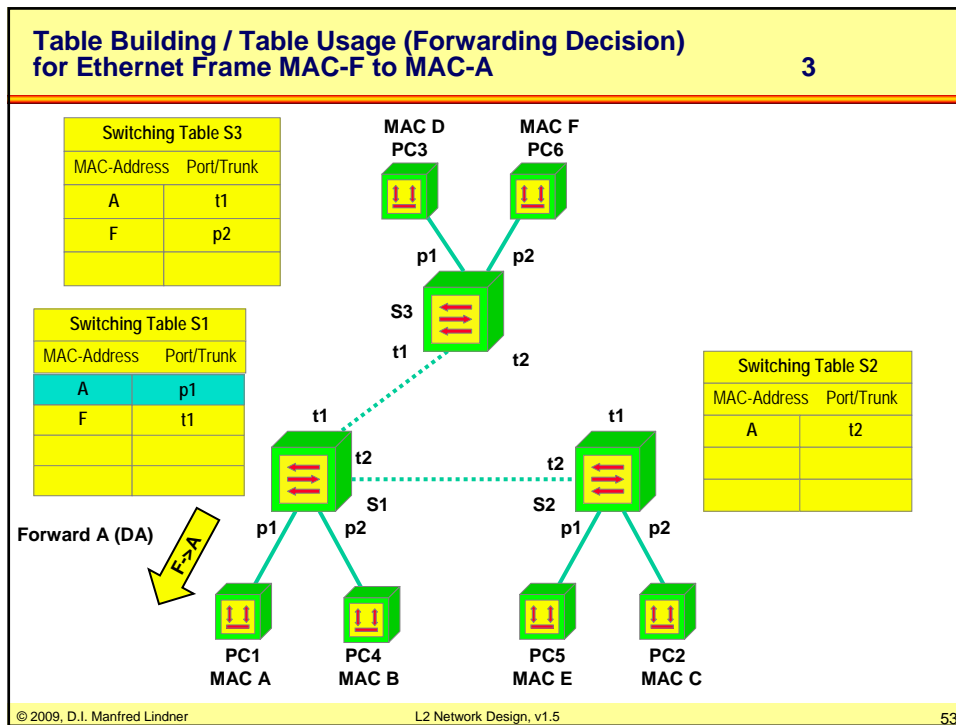
L101 - L2 Network Design



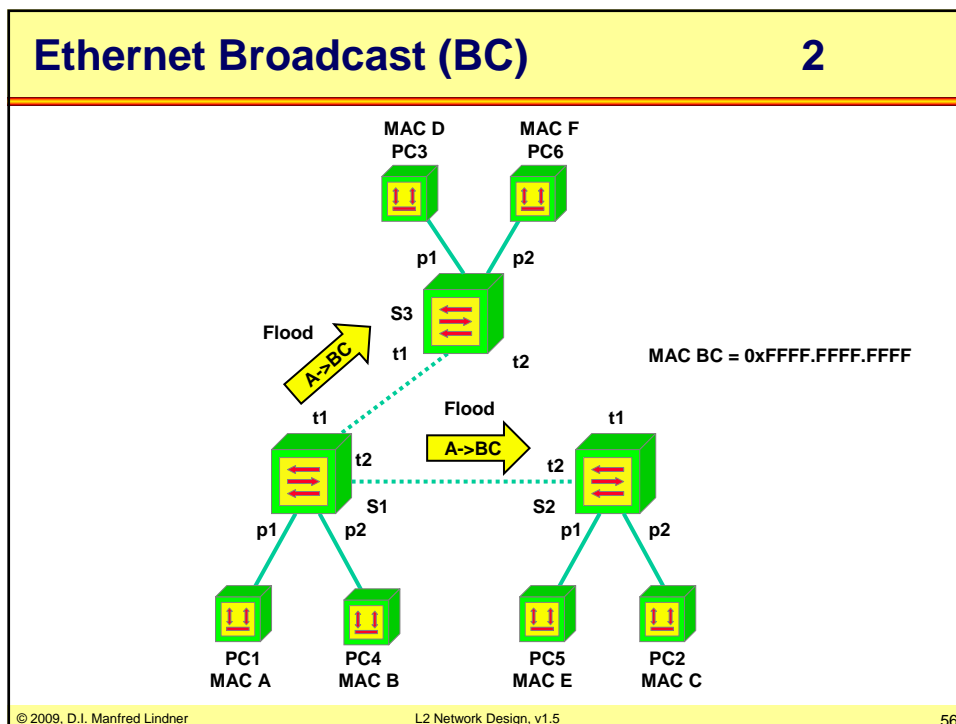
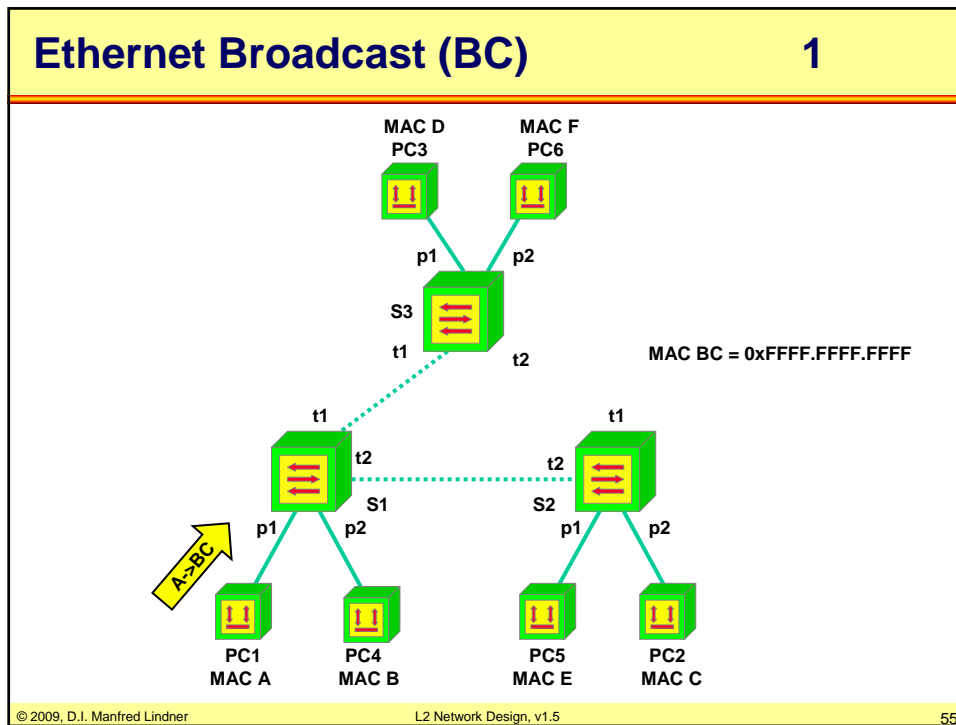
L101 - L2 Network Design



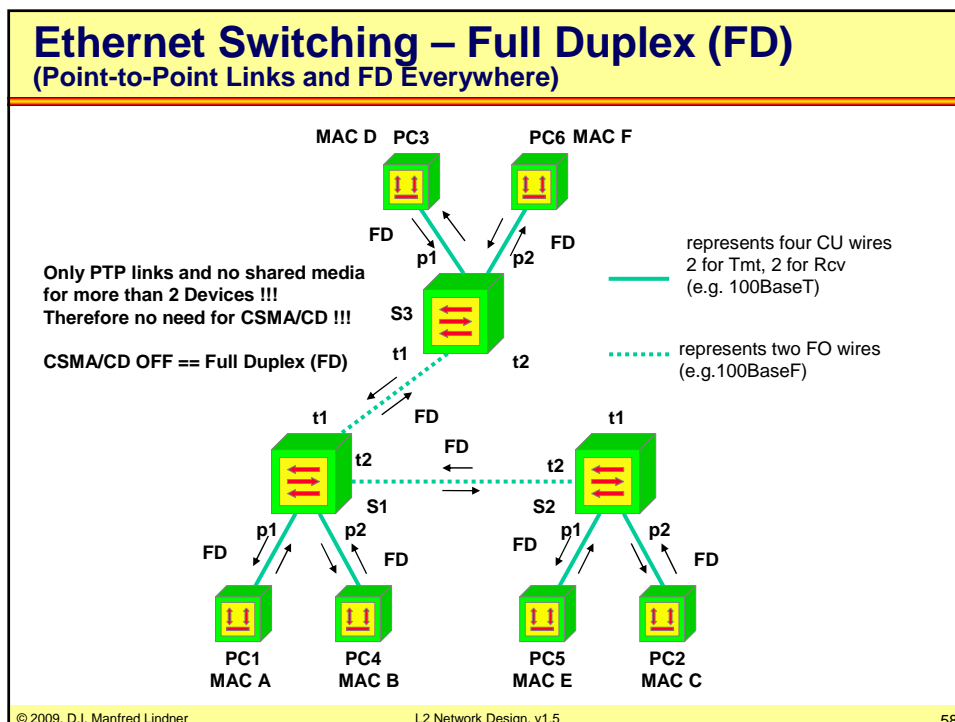
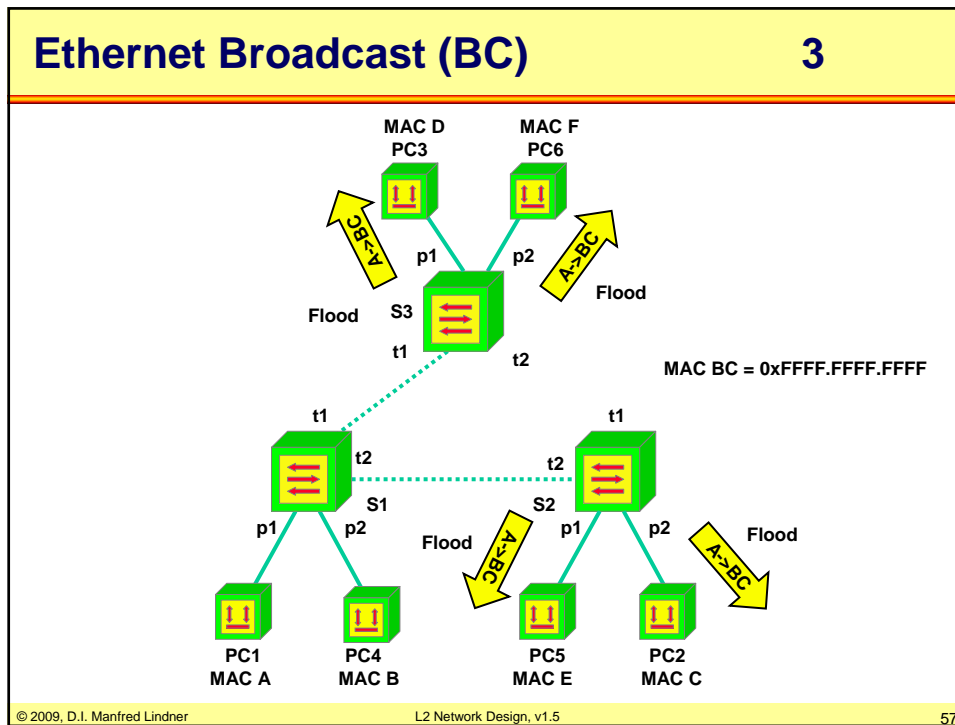
L101 - L2 Network Design



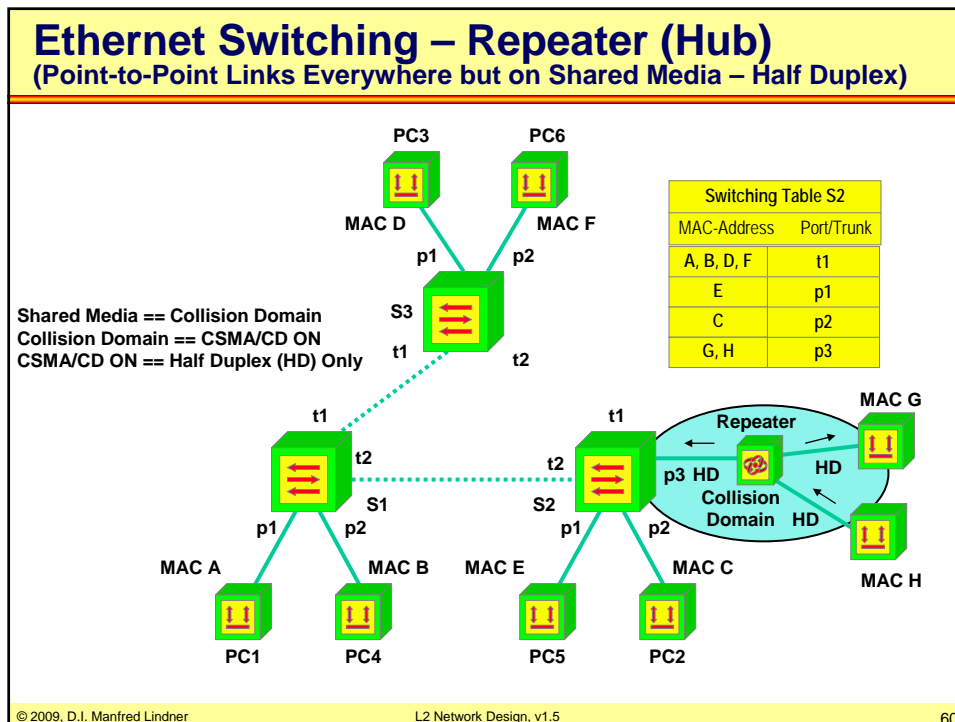
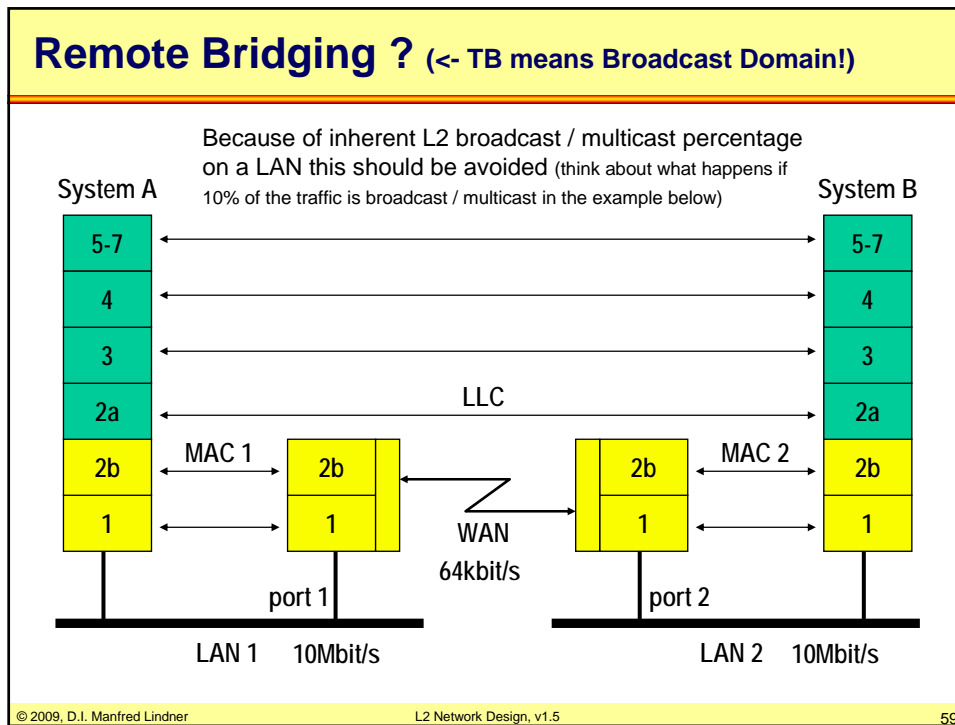
L101 - L2 Network Design



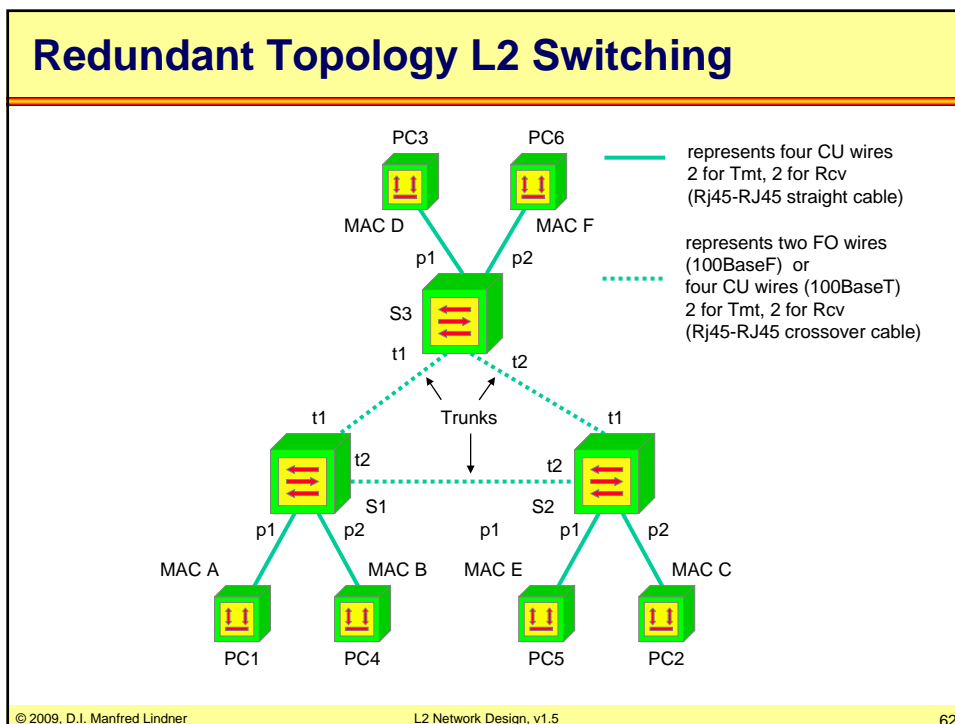
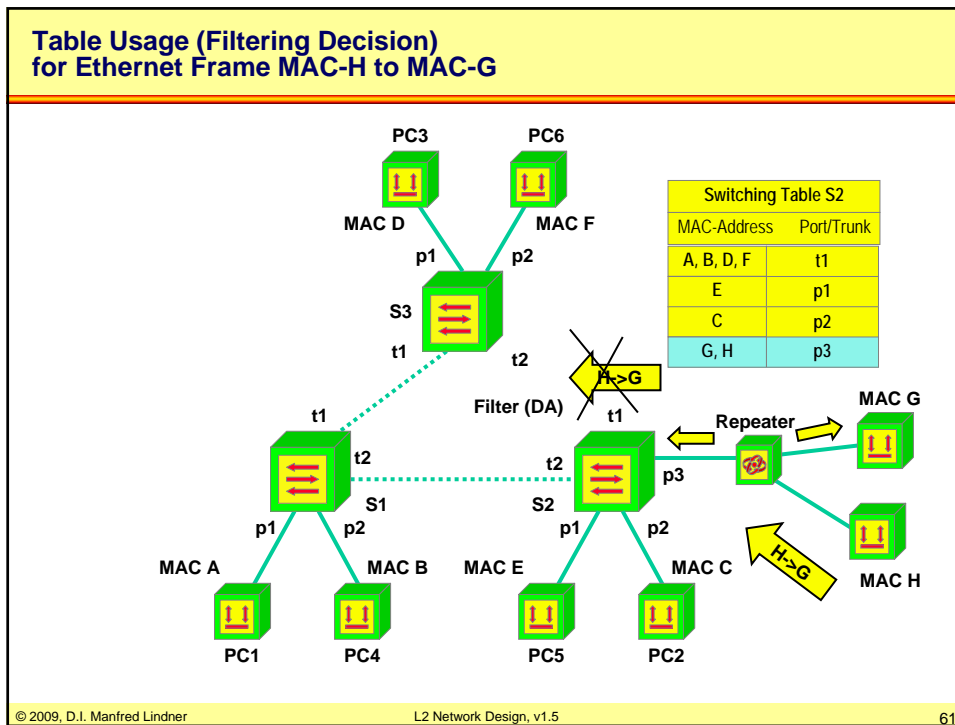
L101 - L2 Network Design



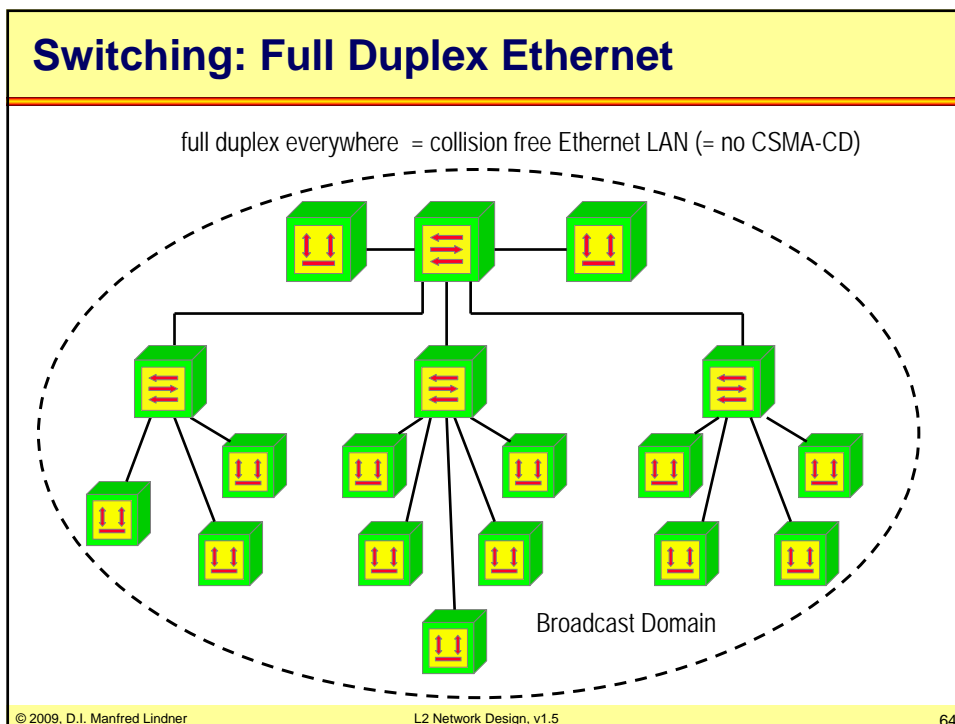
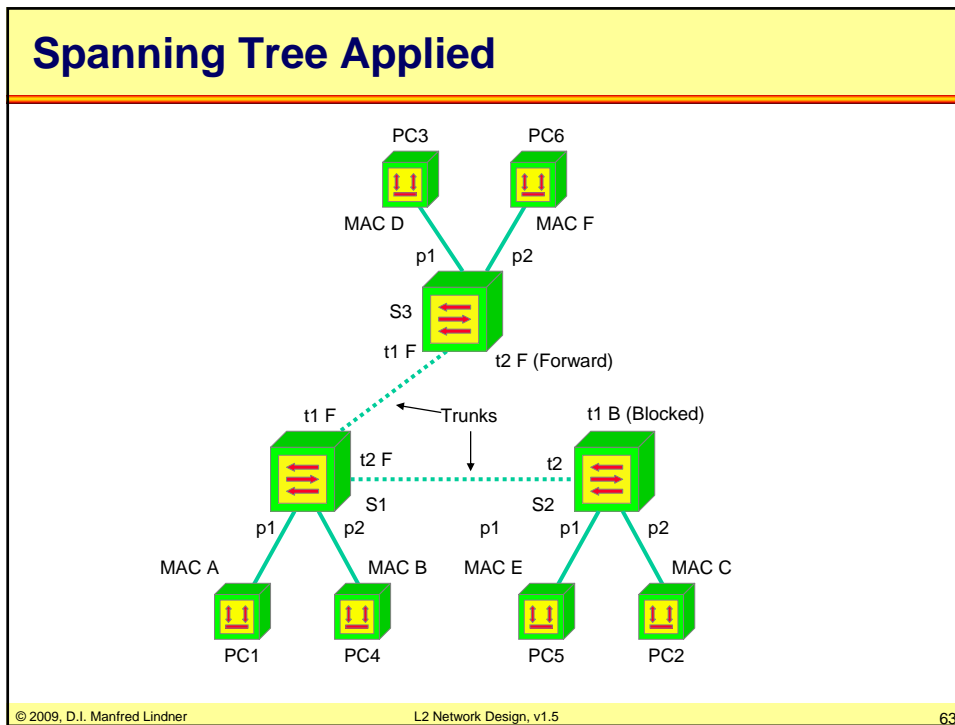
L101 - L2 Network Design



L101 - L2 Network Design

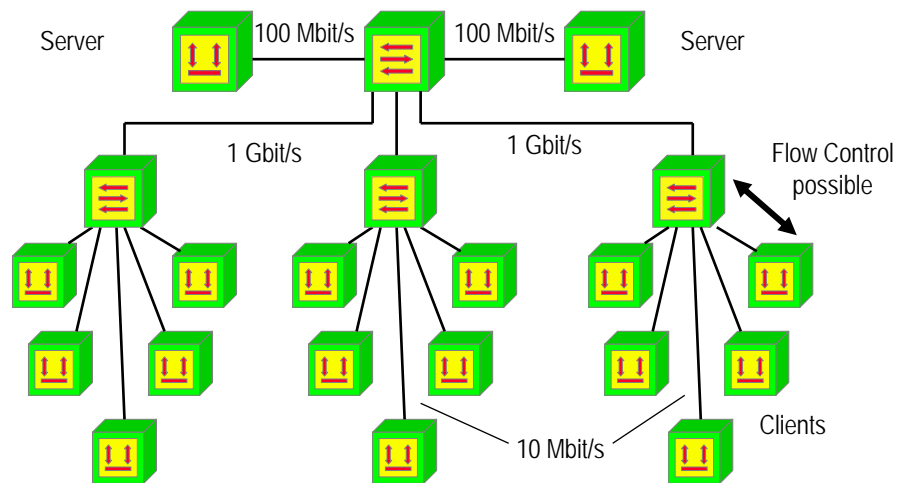


L101 - L2 Network Design



L101 - L2 Network Design

Switching: Variable Speed, QoS and Flow Control



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

65

Review L2 Network Components (STP)

- **“Spanning Tree Protocol” (STP)**
 - Parallel paths between two LAN segments would cause broadcast storm
 - STP takes care that there is always exact one active path between any 2 stations
 - implemented by a special bridge protocol which is used between the bridges for communication
 - using BPDU (Bridge Protocol Data Unit) packets with MAC-multicast address
 - failure of active path causes activation of a redundant path
 - Convergence time worst case 50 seconds with default parameters for hello time, max age and forward delay
 - $\text{time} = \text{max age} + 2 * (\text{forward delay})$

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

66

L101 - L2 Network Design

Parameters for STP

1

• Bridge Identifier (Bridge ID)

- combination of MAC-address and a priority number
 - typically, the lowest MAC-address of all ports is used for that
 - note: although bridge will not be seen by end systems, for bridge communication and management purposes a bridge will listen to one or more dedicated MAC addresses
 - Bridge-ID = priority#.mac#
- priority number can be configured by the administrator
 - default value is 32768
- lowest Bridge ID has highest priority
 - lowest configured priority number
- if you keep default values
 - the bridge with the lowest MAC address will have the highest priority

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

67

Parameters for STP

2

• Port Cost (C)

- costs in order to access local interface
- inverse proportional to the transmission rate
- default cost = 1000 / transmission rate in Mbit/s
 - so 10 Mbit/s Ethernet has a default Path Cost of 100
 - with occurrence of 1Gbit/s Ethernet rule was adapted
 - 100 Mbit/s = 19, 1Gbit/s = 4, 10Gbit/s = 2
- can be configured to a different value by the administrator

• Port Identifier (Port ID)

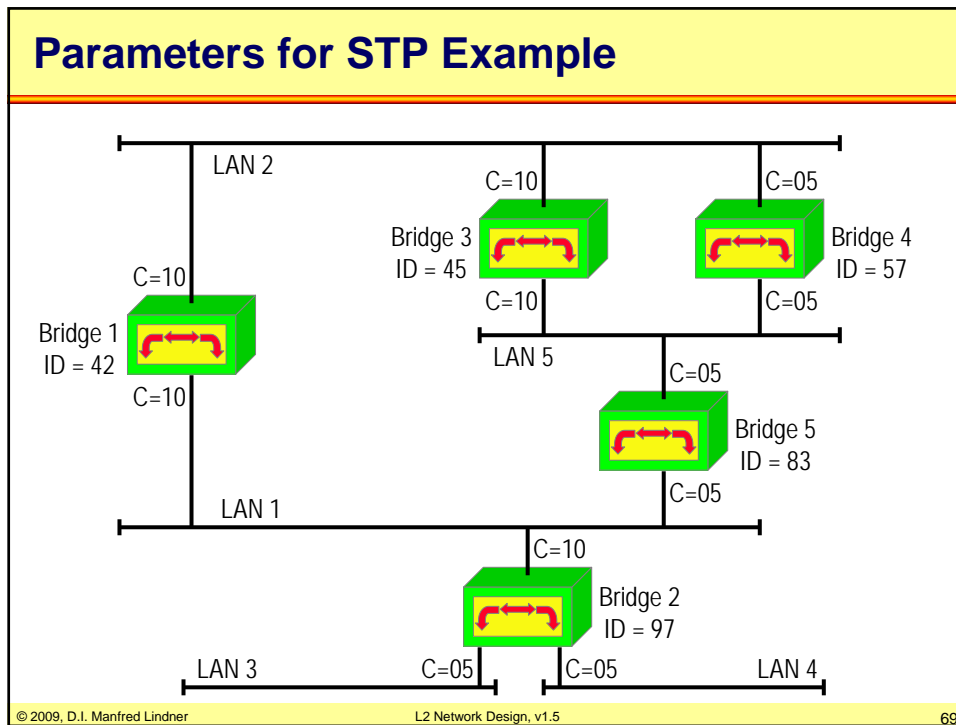
- combination of port number and a priority number
 - Port-ID = port priority#.port#
- configured by the administrator
 - default port priority = 128

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

68

L101 - L2 Network Design



Format of STP Messages - BPDU Format

Prot. ID	Prot. Vers.	BPDU Type	Flags	Root ID	Root Path Costs	Bridge ID	Port ID	Mess. Age	Max Age	Hello Time	Fwd. Delay
2 Byte	1 Byte	1 Byte	1 Byte	8 Byte	4 Byte	8 Byte	2 Byte	2 Byte	2 Byte	2 Byte	2 Byte

BPDU Bridge Protocol Data Unit (OSI term for this kind of message)

Root ID Who seems to be or who is the root bridge (R-ID)?

Root Path Cost How far is the root bridge away from me (RPC)?

Bridge ID ID of bridge transmitting this BPDU (O-ID)

Port ID port over which this BPDU was transmitted (P-ID)

© 2009, D.I. Manfred Lindner L2 Network Design, v1.5 70

L101 - L2 Network Design

BPDUs Fields		1
<ul style="list-style-type: none">- Protocol Identifier:<ul style="list-style-type: none">- 0000 (hex) for STP 802.1D- Protocol Version:<ul style="list-style-type: none">- 00 (hex) for actual version- BPDUs Type:<ul style="list-style-type: none">- 00 (hex) for Configuration BPDUs- 80 (hex) for Topology Change Notification (TCN) BPDUs- Root Identifier:<ul style="list-style-type: none">- 2 bytes for priority (default 32768)- 6 bytes for MAC-address- Root Path Costs in binary representation:<ul style="list-style-type: none">- range 1-65535- Bridge Identifier:<ul style="list-style-type: none">- structure like Root Identifier		
<small>© 2009, D.I. Manfred Lindner</small>	<small>L2 Network Design, v1.5</small>	<small>71</small>

BPDUs Fields		2
<ul style="list-style-type: none">- Port Identifier:<ul style="list-style-type: none">- 1 byte priority (default 128)- 1 byte port number- Message Age (range 1-10s):<ul style="list-style-type: none">- age of Configuration BPDUs- transmitted by root-bridge initially using zero value, each passing-on (by designated bridge) increases this number- Max Age (range 6-40s):<ul style="list-style-type: none">- aging limit for information obtained from Configuration BPDUs- basic parameter for detecting idle failures (e.g. root bridge = dead)- default 20 seconds- Hello Time (range 1-10s):<ul style="list-style-type: none">- time interval for generation of periodic Configuration BPDUs by root bridge- default 2 seconds		
<small>© 2009, D.I. Manfred Lindner</small>	<small>L2 Network Design, v1.5</small>	<small>72</small>

L101 - L2 Network Design

BPDUs Fields

3

- Forward Delay (range 4-30s):
 - time delay for putting a port in the forwarding state
 - default 15 seconds
 - but that means 15 seconds listening plus 15 seconds learning
- Hello Time, Max Age, Forward Delay are specified by Root-Bridge
- Flags (a "1" indicates the function):
 - bit 8 ... Topology Change Acknowledgement (TCA)
 - bit 1 ... Topology Change (TC)
 - used together with TCN BPDUs for signalling topology changes
 - note: in case of topology change MAC addresses may be found on an other port in the bridging table; therefore relearning of the dynamic bridging table is necessary; relearning is triggered by the root bridge

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

73

MAC Addresses / LSAP / Network Diameter

- **bridges use for STP-communication:**
 - multicast address:
 - 0180 C200 0000 hex
 - 0180 C200 0001 to 0180 C200 000F are reserved
 - 0180 C200 0010 hex All LAN Bridges Management Group Address
 - Note :
 - all addresses in Ethernet format
 - on the Token Ring the functional address: 0300 0000 8000 hex
 - the L-SAP of LLC header
 - 42 hex Bridge Spanning Tree Protocol
- **Maximum Bridge Diameter**
 - maximum number of bridges between any two end systems is 7 using default values for Hello Time, Forward Delay and Max Age

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

74

L101 - L2 Network Design

Spanning Tree applied

- **STP Algorithm summarized:**

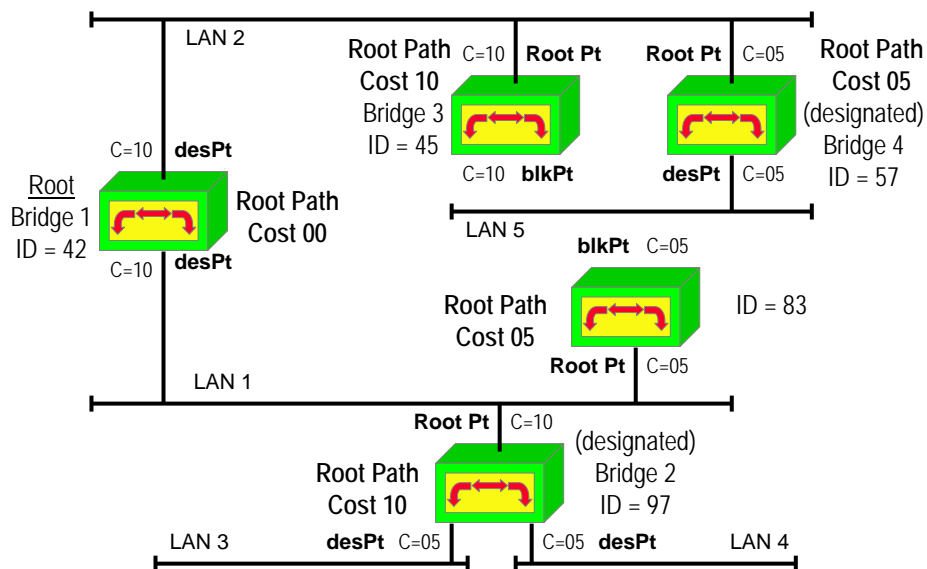
- select the root bridge
- select the root ports
 - by computation of the shortest path from any other bridge to the root bridge
 - root port points to the shortest path towards the root
- select one designated bridge for every LAN segment
 - corresponding port is called designated port
- set the designated and root ports in forwarding state
- set all other ports in blocking state
- creates single paths from the root to all leaves (LAN segments) of the network

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

75

Final STP Picture for Example

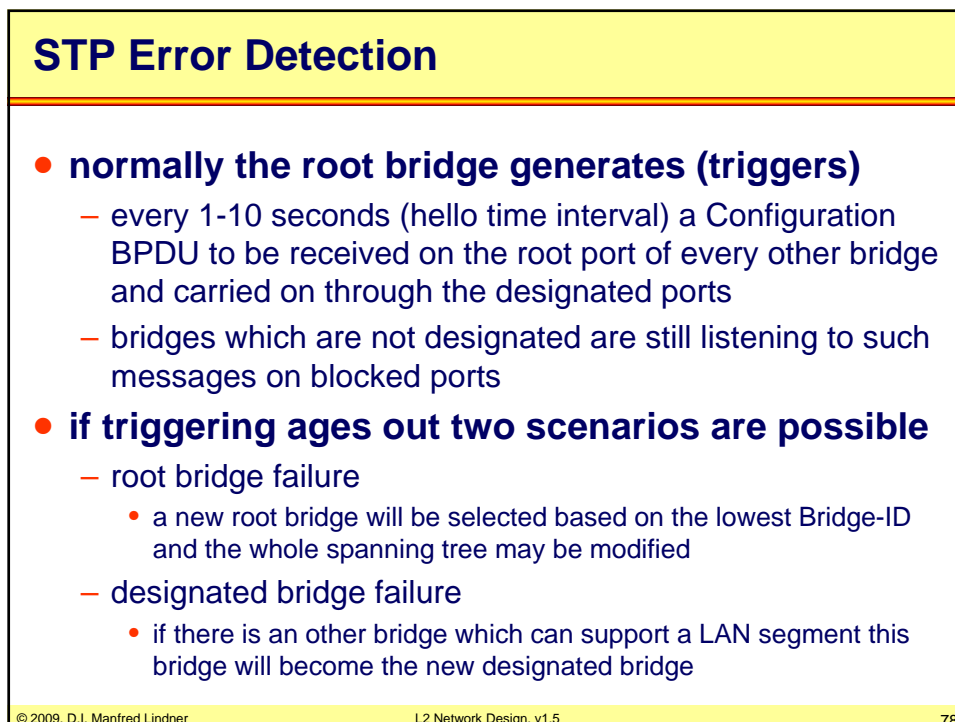
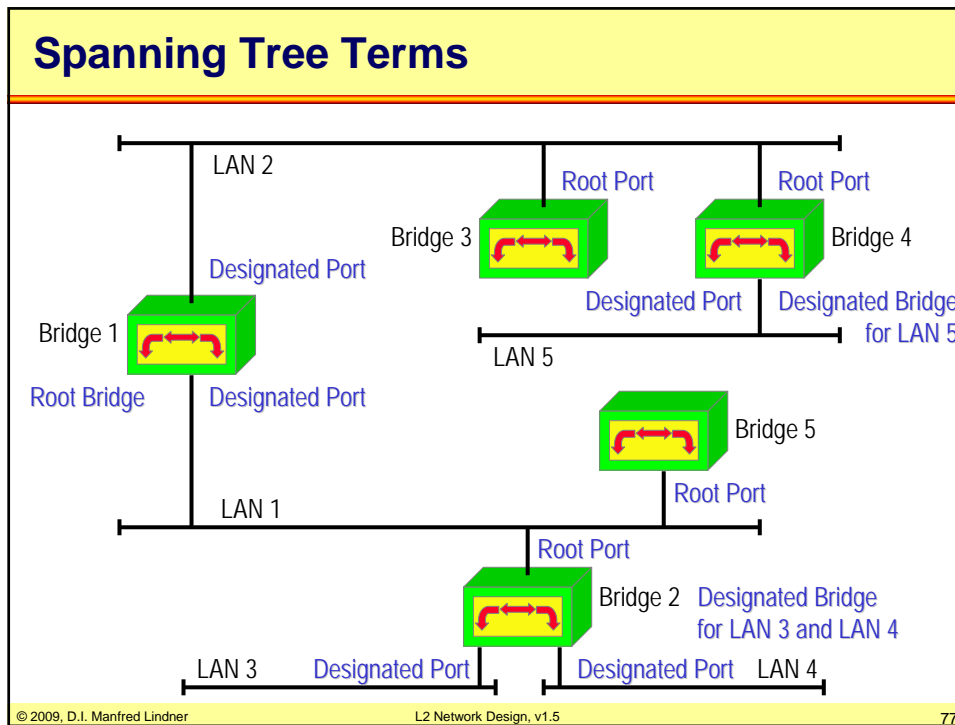


© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

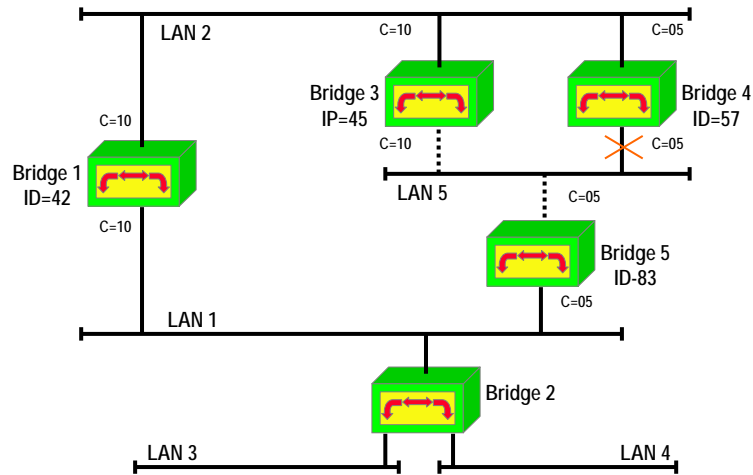
76

L101 - L2 Network Design



L101 - L2 Network Design

STP Convergence Time – Failure at Designated Bridge



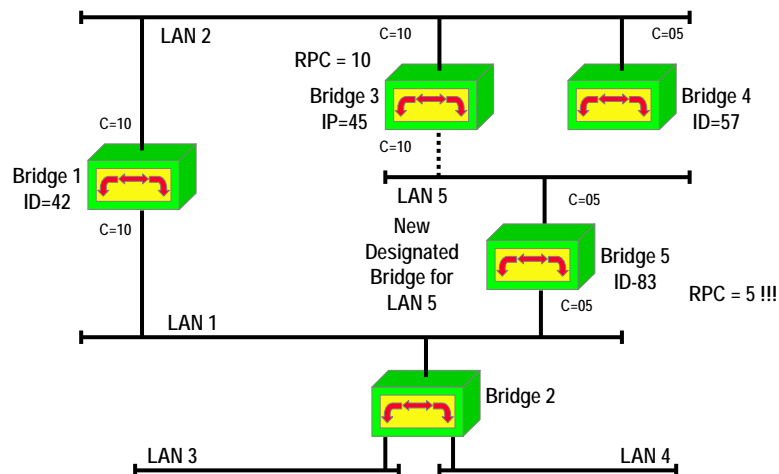
- Time = max age (20 sec) + 2*forward delay (15 sec Listening + 15 sec Learning) = 50 sec

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

79

STP Convergence Time – Failure at Designated Bridge – New Topology



- Time = max age (20 sec) + 2*forward delay (15 sec Listening + 15 sec Learning) = 50 sec

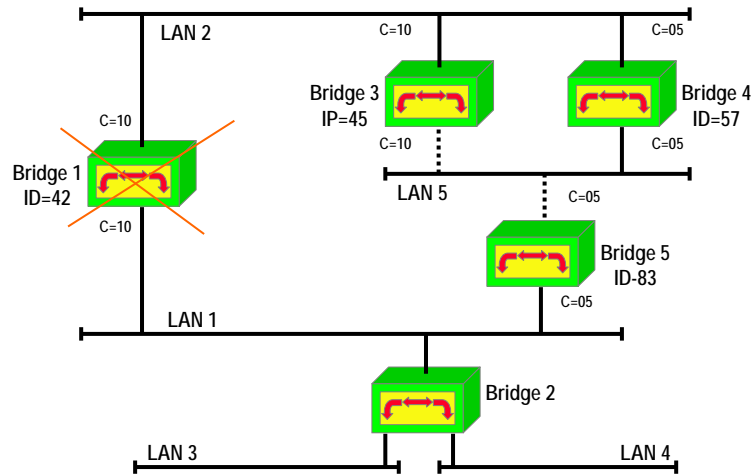
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

80

L101 - L2 Network Design

STP Convergence Time – Failure of Root Bridge



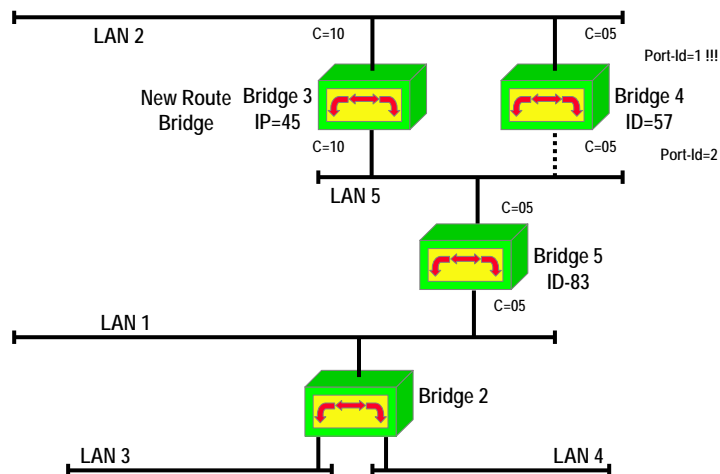
- Time = max age (20 sec) + 2*forward delay (15 sec Listening + 15 sec Learning) = 50 sec

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

81

STP Convergence Time – Failure of Root Bridge – New Topology



- Time = max age (20 sec) + 2*forward delay (15 sec Listening + 15 sec Learning) = 50 sec

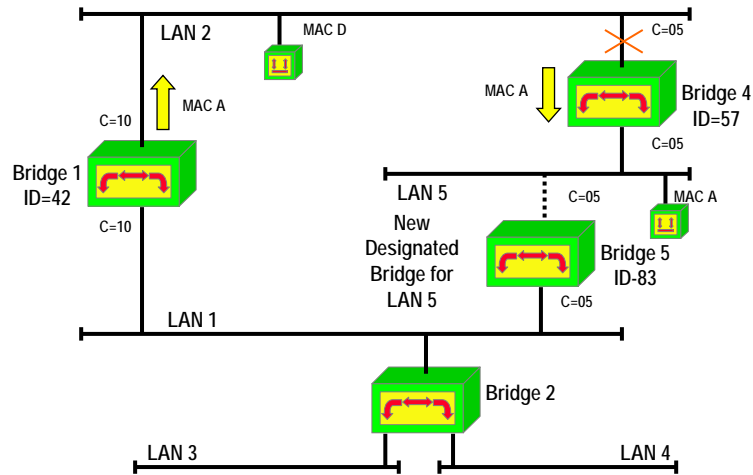
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

82

L101 - L2 Network Design

STP Convergence Time – Failure of Root Port



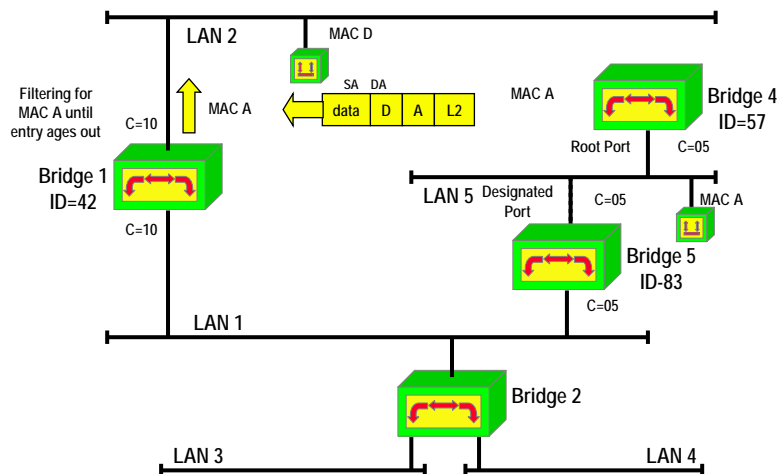
- Time = 2*forward delay (15 sec Listening + 15 sec Learning) = 30 sec

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

83

STP Convergence Time – Failure of Root Port - Interruption of Connectivity D->A



- Time = 2*forward delay (15 sec Listening + 15 sec Learning) = 30 sec

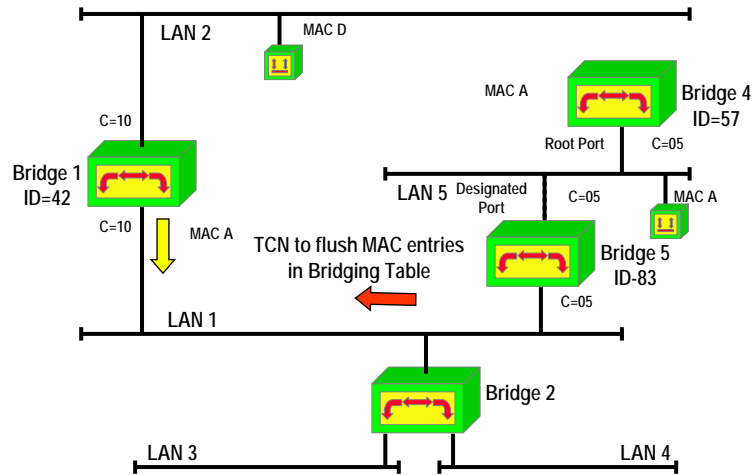
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

84

L101 - L2 Network Design

STP Convergence Time – Failure of Root Port – Topology Change Notification (TCN)



- **Time = 2*forward delay (15 sec Listening + 15 sec Learning) = 30 sec**

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

85

What has changed with Rapid Spanning Tree?

- **Rapid Spanning Tree (RSTP)**
 - IEEE 802.1D version 2004 (former IEEE 802.1w)
 - Can be seen as an evolution of the Spanning Tree Protocol (STP; IEEE 802.1D)
 - Capable of reverting back to 802.1D version 1998
 - Better to avoid it
 - Convergence time reduced to few seconds !!!
- **Terminology slightly changed**
 - Blocking port role is split into the Backup and Alternate port roles
 - Alternate port
 - Backup port
 - Root port and Designated port roles still remain the same
 - New port state
 - Discarding (see next slide for details)

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

86

L101 - L2 Network Design

Port States Comparison

STP (802.1d) Port State	RSTP (802.1w) Port State	Is Port included in active Topology?	Is Port learning MAC addresses?
disabled	discarding	No	No
blocking	discarding	No	No
listening	discarding	Yes	No
learning	learning	Yes	Yes
forwarding	forwarding	Yes	Yes

© 2009, D.I. Manfred Lindner

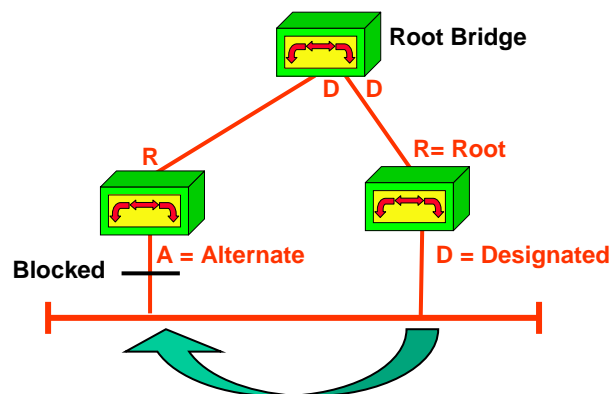
L2 Network Design, v1.5

87

New Port Roles

- **Alternate Port Roles**

- A port blocked by receiving BPDU's from a different bridge
- Provides an alternate path to the root bridge



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

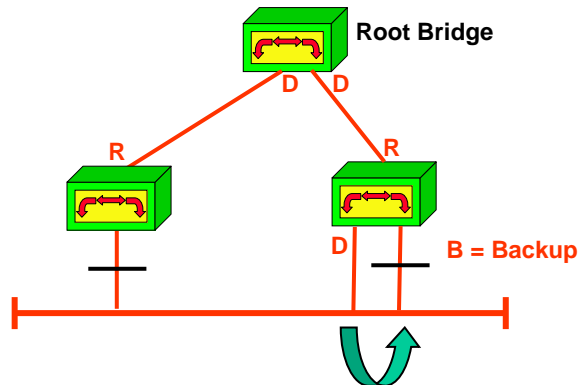
88

L101 - L2 Network Design

New Port Roles

- **Backup Port**

- A port blocked by receiving BPDU's from the same bridge
- Provides a redundant connectivity to the same segment



© 2009, D.I. Manfred Lindner

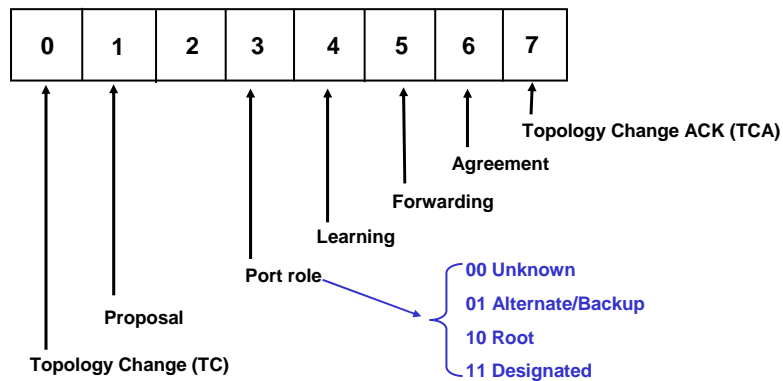
L2 Network Design, v1.5

89

BPDU Flag Field – New Values

- **Few changes have been introduced by RSTP**

- TC and TCA used by old STP
- RSTP also uses the 6 remaining bits



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

90

L101 - L2 Network Design

NEW BPDU Handling

- **Faster Failure Detection**

- BPDU's acting now as keepalives messages
 - Different to the 802.1D STP a bridge now sends a BPDU with its current information every <hello-time> seconds (2 by default), even if it does not receive any from the root bridge
- If hellos are not received for 3 consecutive times, port information is invalidated
 - because BPDU's are now used as keep-alive mechanism between bridges
- If a bridge fails to receive BPDU's from a neighbor, the connection has been lost
- No more max age and message age fields
 - Hop count is used instead

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

91

Proposal / Agreement

- **Explicit handshake between bridges**

- Upon link up event the bridge sends a proposal to become designated for that segment
- Remote bridge responds with an agreement if the port on which it received the proposal is the root port of the remote bridge
- As soon as receiving an agreement, bridge moves the port to the forwarding state

© 2009, D.I. Manfred Lindner

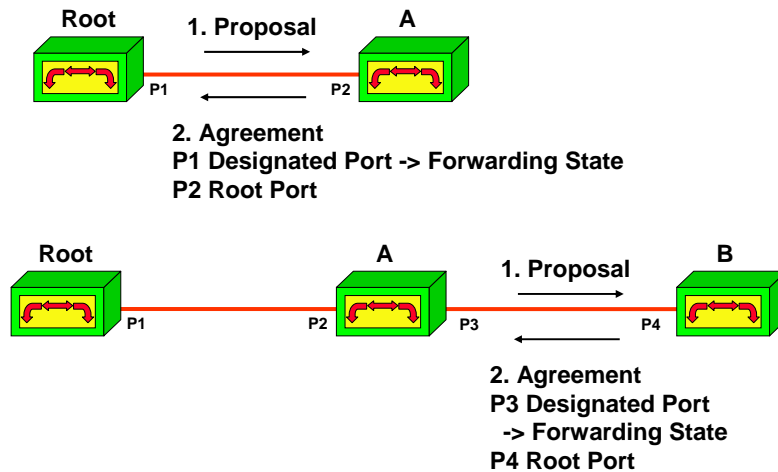
L2 Network Design, v1.5

92

L101 - L2 Network Design

Proposal/Agreement Sequence

- Suppose a new link is created between the root and switch A and a new switch B is inserted



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

93

Rapid Transition to Forwarding State

- Most important feature in 802.1w
- The legacy STP was passively waiting for the network to converge before turning a port into the forwarding state
- New RSTP is able to actively confirm that a port can safely transition to forwarding
- Real feedback mechanism, that takes place between RSTP-compliant bridges
- To achieve fast convergence on a port, the protocol relies upon 2 new variables
 - Edge ports
 - Link type

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

94

L101 - L2 Network Design

Rapid Transition to Forwarding State

- **RSTP can only achieve rapid transition to forwarding**
 - On edge ports
 - On point-to-point links (trunks between L2 switches)
- **Edge Ports**
 - Ports, which are directly connected to end stations cannot create bridging loops in the network and can thus directly perform on link setup transition to forwarding, skipping the listening and learning states
- **Link type**
 - Is automatically derived from the duplex mode of a port
 - A port operating in full-duplex will be assumed to be point-to-point
 - A port operating in half-duplex will be assumed to be a shared port

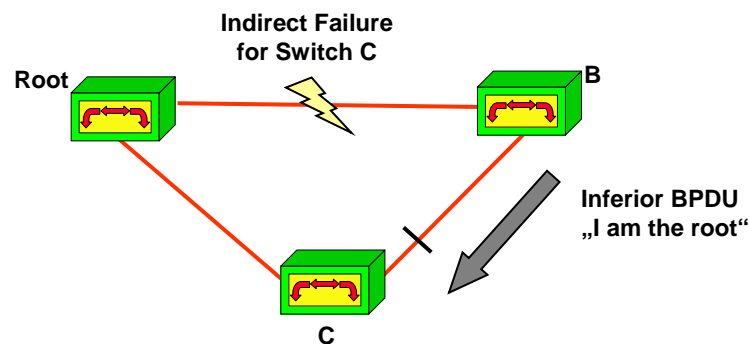
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

95

Accepting Inferior BPDU's

- B loses root port and sends BPDU claiming to be the root
- C immediately becomes designated for the blocked link between C and B and sends a proposal to B
- B sends an agreement and C set its port to forwarding
- Like Cisco's Backbone Fast



© 2009, D.I. Manfred Lindner

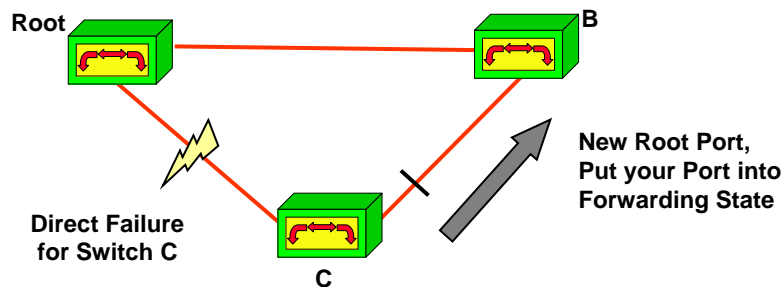
L2 Network Design, v1.5

96

L101 - L2 Network Design

Accepting New Root Port BPDUs

- C loses root port and sends BPDUs on the blocked link agreeing that this port is now root port
- C sets its port to forwarding
- Like Cisco's Uplink Fast



© 2009, D.I. Manfred Lindner

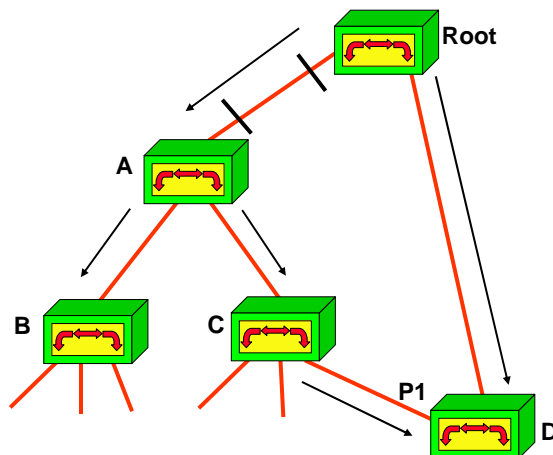
L2 Network Design, v1.5

97

Slow Convergence with Legacy STP

1

A new link between A and Root is being added to the bridged network



New ports coming up on the root and the switch A are immediately set in listening state, blocking traffic.

BPDUs from the root start propagating towards the leaves through A.

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

98

L101 - L2 Network Design

Slow Convergence with Legacy STP 2

Very quickly, the BPDU's from the root via A reach D that immediately blocks its port P1.
The topology has now converged, though, the network is disrupted for twice forward delay because A needs time for Listening and Learning on the new Port

© 2009, D.I. Manfred Lindner L2 Network Design, v1.5 99

Fast Convergence with RSTP 1

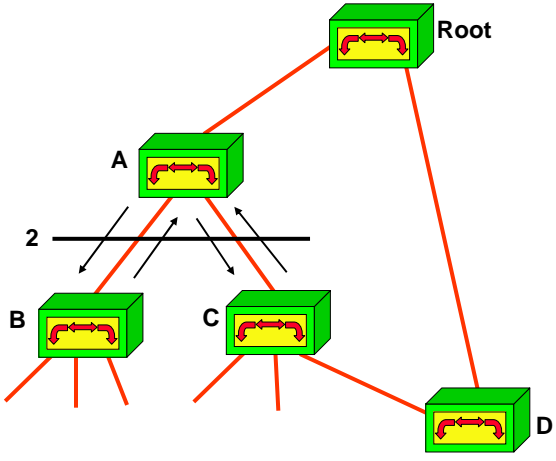
A new link between A and Root is being added to the bridged network

Both ports on link between A and the root are put in designated blocking as soon as they come up.
As soon as A receives the root's BPDU, it blocks its non-edge designated ports (= "sync"). A explicitly authorizes the root bridge to put its port in forwarding

© 2009, D.I. Manfred Lindner L2 Network Design, v1.5 100

L101 - L2 Network Design

Fast Convergence with RSTP
2

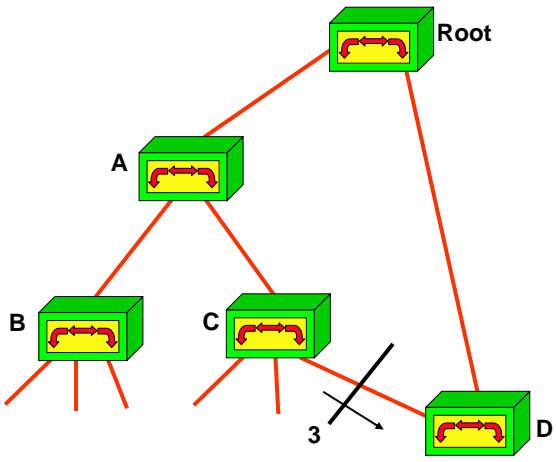


The link between Switch A and the root is put in forwarding state.

The network is now blocking below Switch A. This cut will travel down the tree along with the BPDU's originated by the root through Switch A. At that stage, the newly blocked ports on Switch A will also negotiate a quick transition to forwarding with their neighboring ports in Switch B and Switch C, that both in turn will initiate a sync operation (especially C to D in our case)

© 2009, D.I. Manfred Lindner
L2 Network Design, v1.5
101

Fast Convergence with RSTP
3



Switch C blocks its designated port to D.

We have reached the final topology, which means that port P1 on D ends up blocking. It's the same topology as for the STP example.

But we got this topology just in time necessary for the new BPDU's to travel down the tree. No timer has been involved in this quick convergence.

Convergence Time < 1 second

© 2009, D.I. Manfred Lindner
L2 Network Design, v1.5
102

L101 - L2 Network Design

Review L2 Network Components (VLAN)

- **VLAN Switch**

- Several virtual Ethernet switches in one physical box
- Each of them implementing a separate Virtual LAN to corresponding connected LAN user ports (access ports)
 - A separate bridging/switching table per VLAN
- Trunk ports used for interconnection of VLAN switches
 - VLAN Tagging (IEEE 802.1Q or Cisco ISL)
 - VLAN management protocols (like Cisco's VTP, DTP)
- Sometimes access ports
 - Use VLAN tagging to connect a network station (PC, router, firewall, etc...) to several VLAN's
- "Spanning Tree Protocol"
 - One Single STP for all VLAN's (802.1D)
 - Per VLAN STP (Cisco)
 - MIST (802.1w)

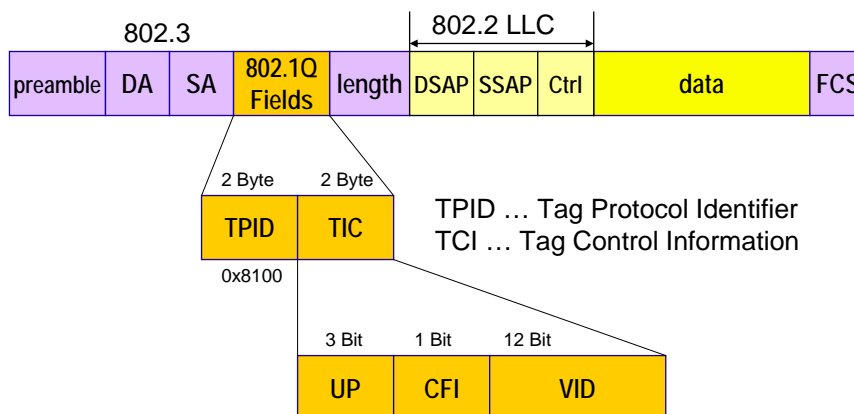
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

103

802.1Q VLAN Tagging

1



note: With tagging Ethernet's maximal frame length = 1522, minimal frame length = 68

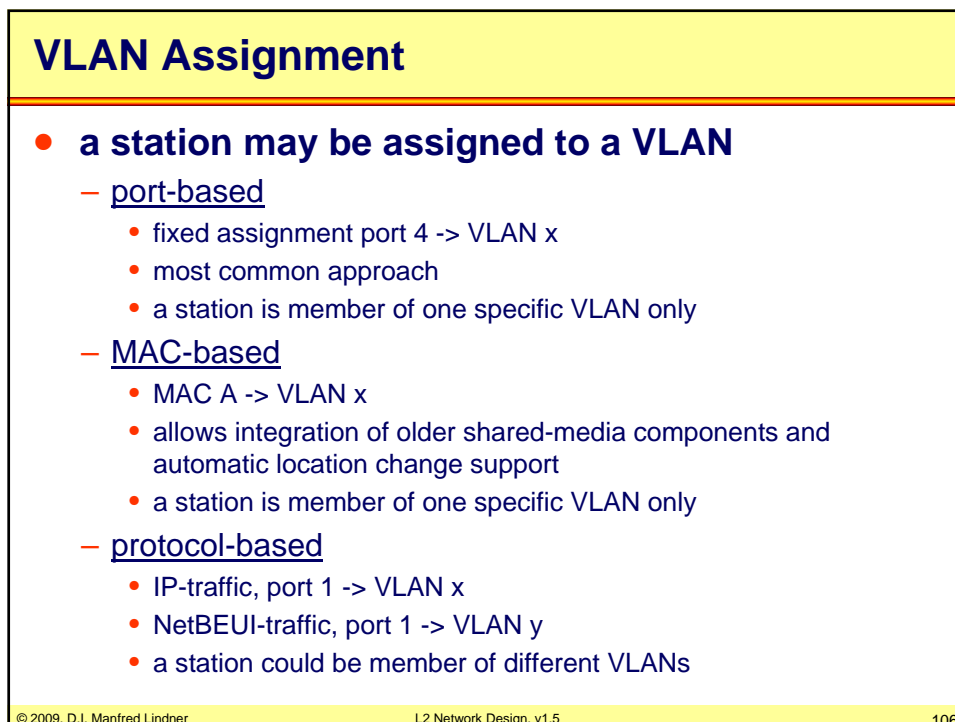
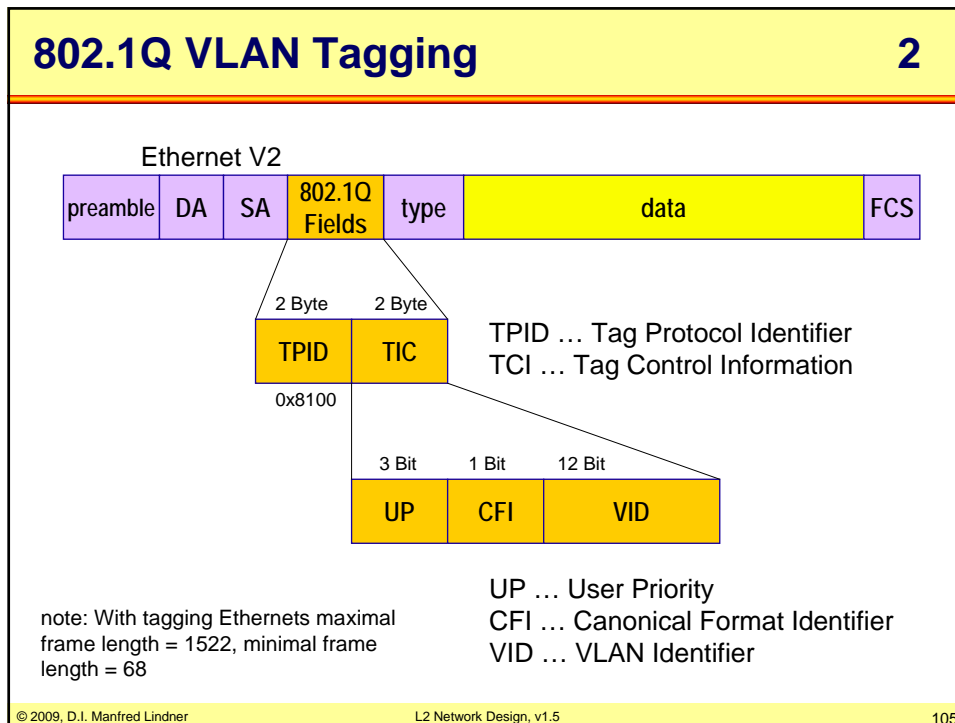
UP ... User Priority
CFI ... Canonical Format Identifier
VID ... VLAN Identifier

© 2009, D.I. Manfred Lindner

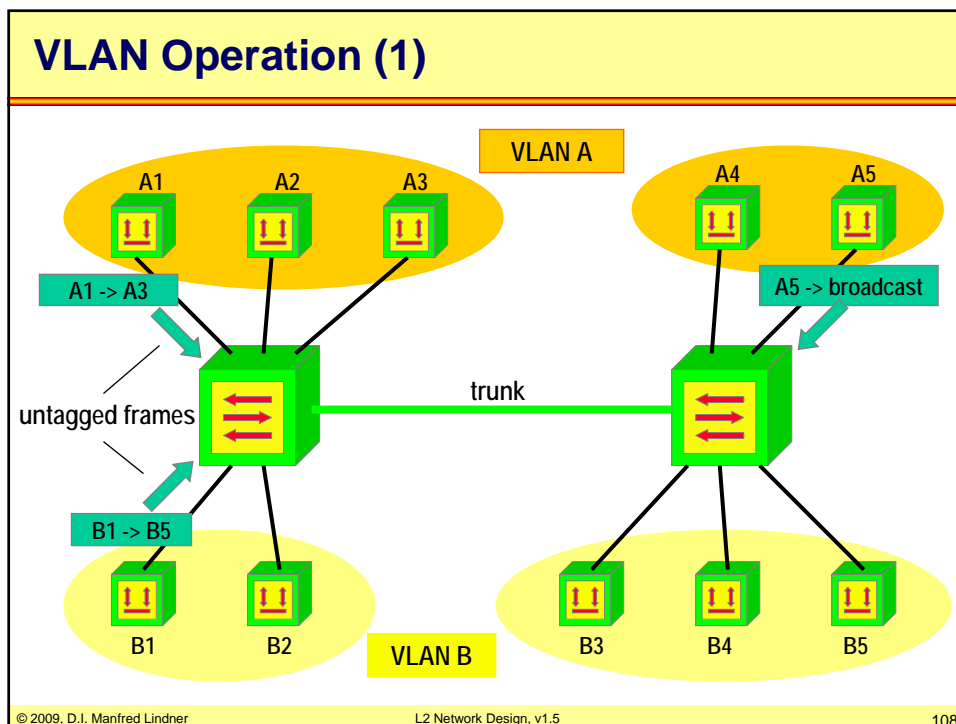
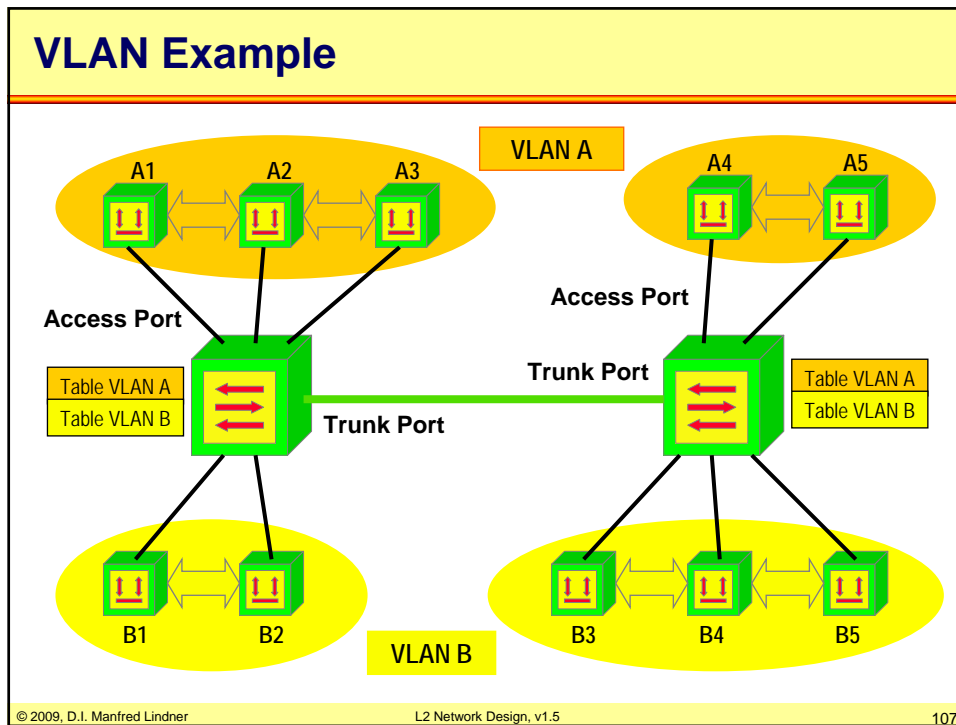
L2 Network Design, v1.5

104

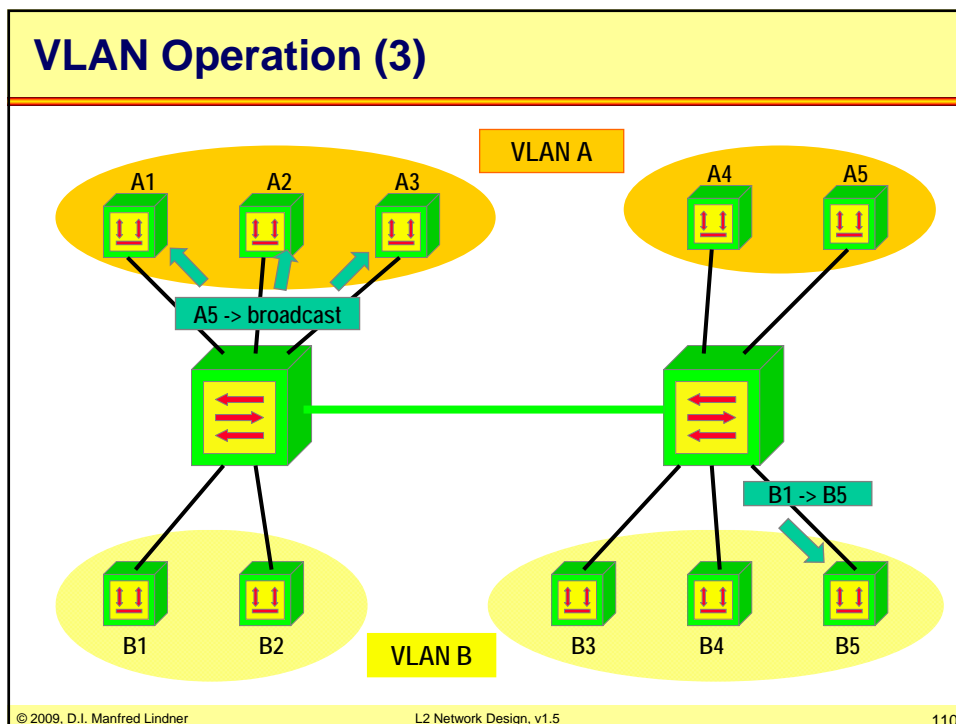
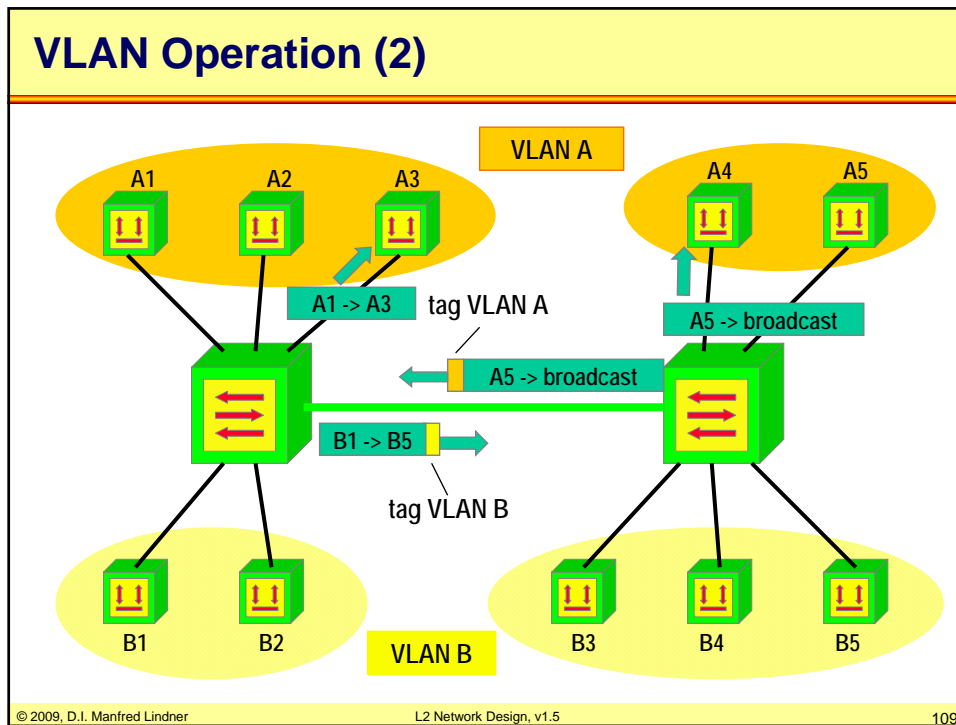
L101 - L2 Network Design



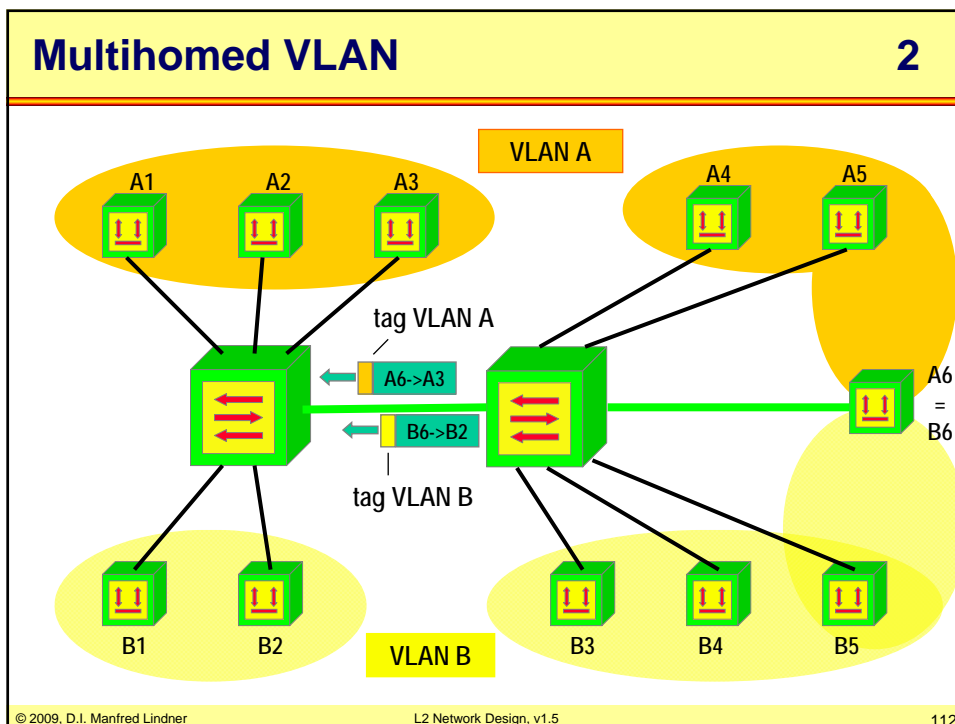
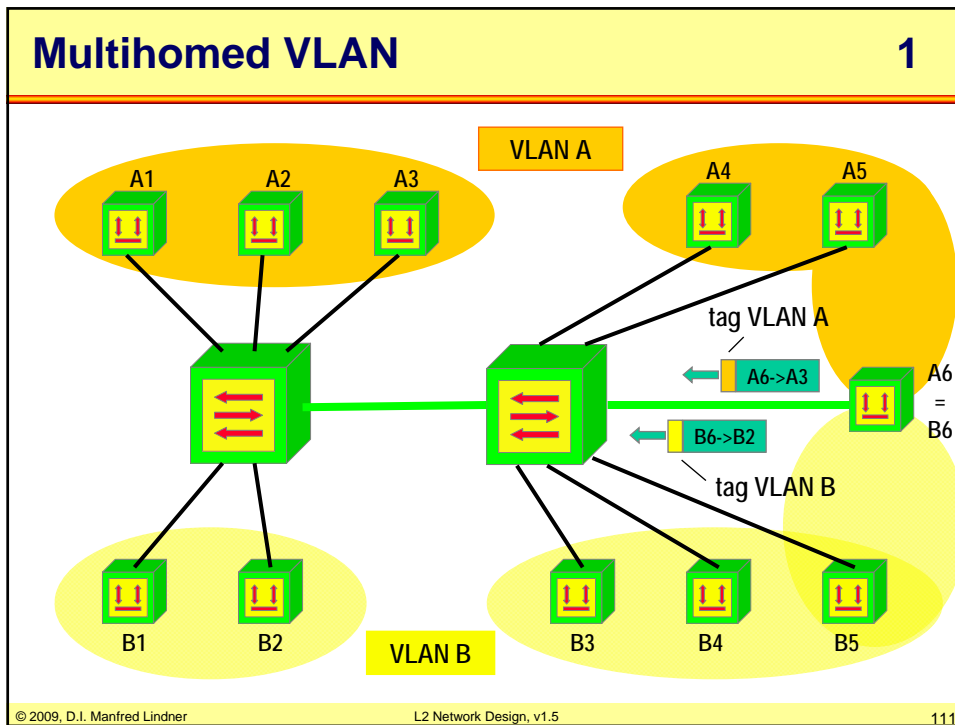
L101 - L2 Network Design



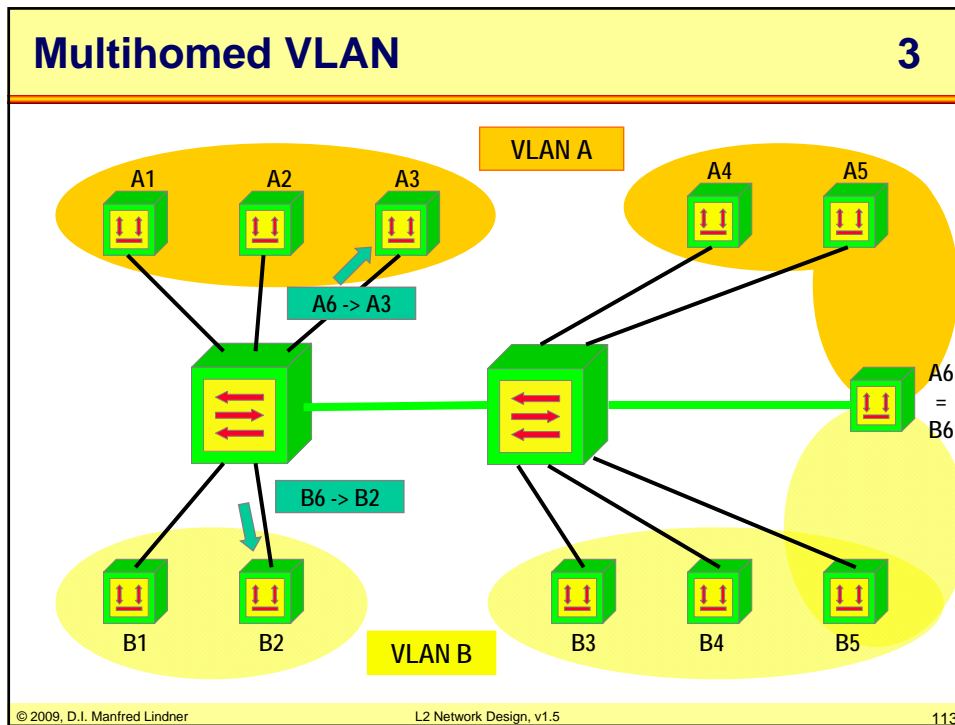
L101 - L2 Network Design



L101 - L2 Network Design



L101 - L2 Network Design

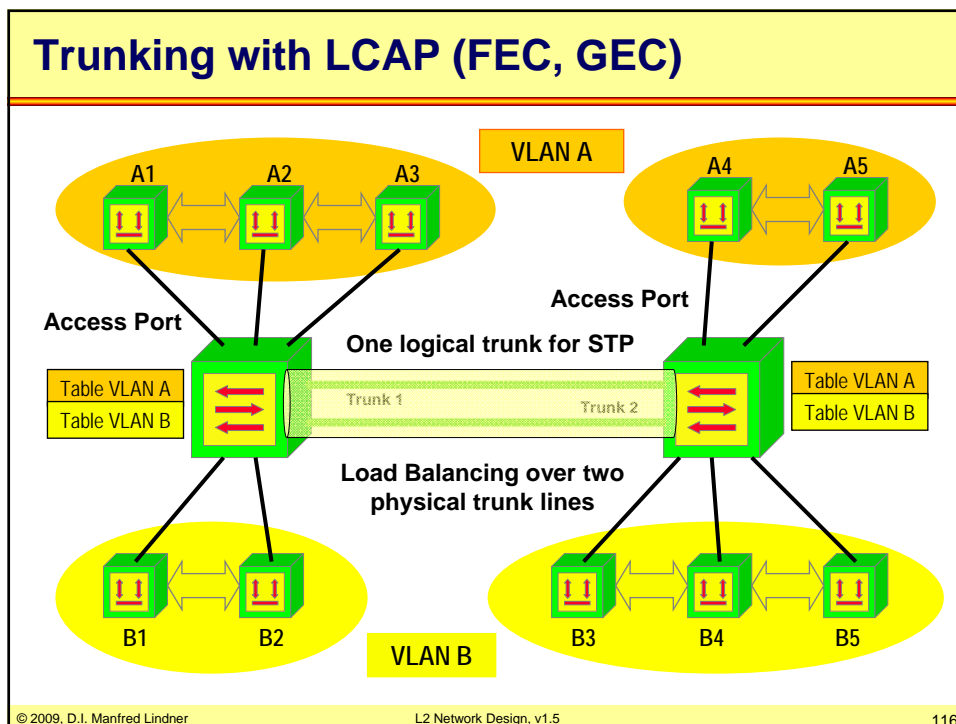
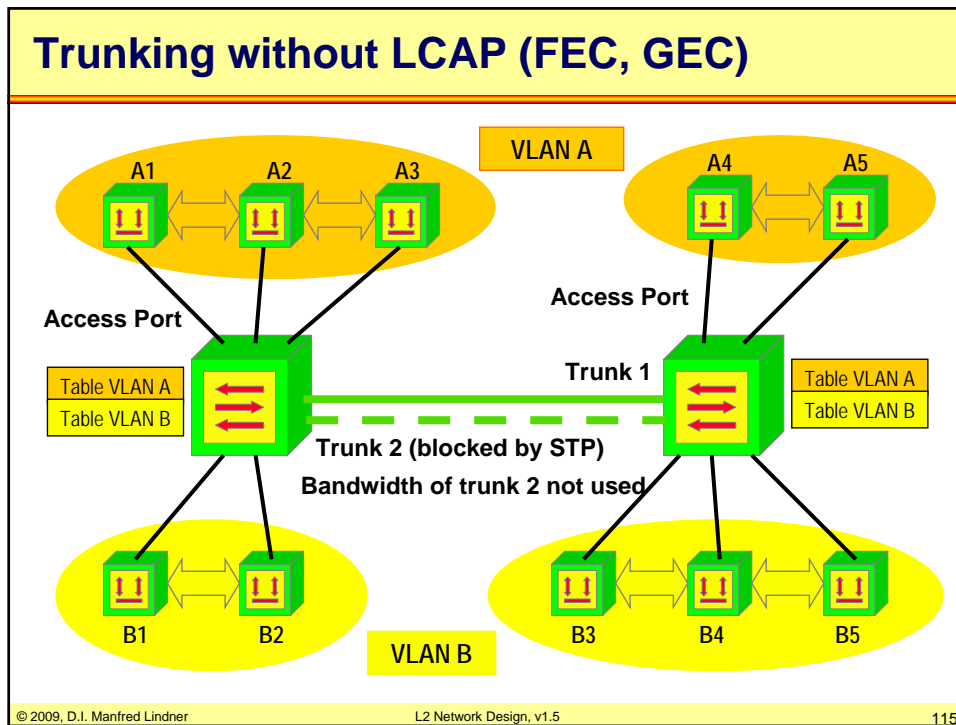


Trunking between L2 Switches

- **on trunks between multiport switches full duplex operation is possible**
 - hence "200 Mbit/s" with Fast Ethernet
 - hence "2 Gbit/s" with Gigabit Ethernet
- **on trunks bundling (aggregation) of physical links to one logical link is possible**
 - Fast Ethernet Channeling (Cisco)
 - 400 / 800 Mbit/s
 - Gigabit Ethernet Channeling (Cisco)
 - 4 / 8 Gbit/s
 - IEEE 802.3 (2002) LACP (Link Aggregation Control Protocol)

© 2009, D.I. Manfred Lindner
L2 Network Design, v1.5
114

L101 - L2 Network Design



L101 - L2 Network Design

Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

117

Design Principles

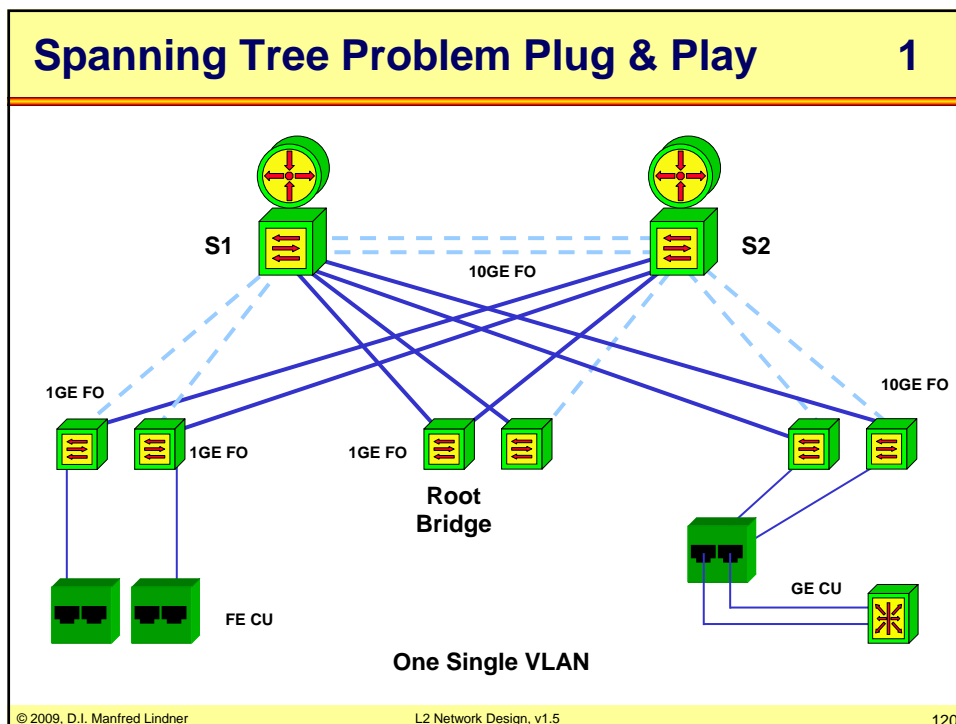
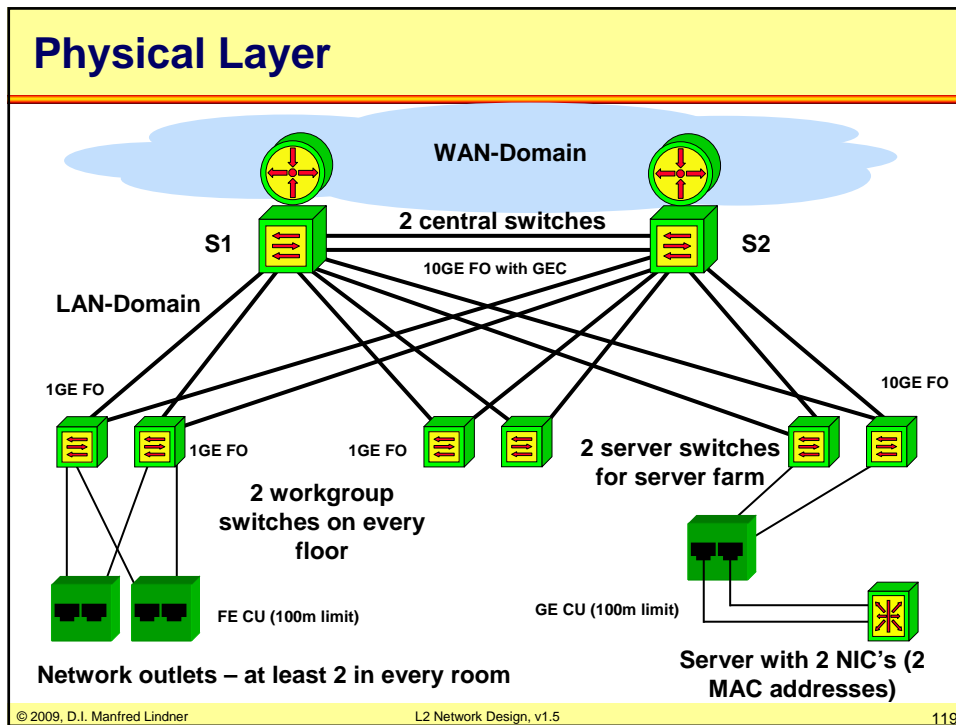
- **Avoiding of “Single Point of Failure”**
 - Physical link failure
 - Access link
 - Trunk link (LAN or WAN)
 - Network component failure
 - L2 Switch
 - Router, DHCP Server, DNS Server, Production Server
- **Load balancing in normal situations**
- **Server with two or more NIC's**
 - OS must support parallel operation and/or switch over between cards
- **Clients with two network outlets**
 - Two NIC's and special OS aspects may not economically be justified

© 2009, D.I. Manfred Lindner

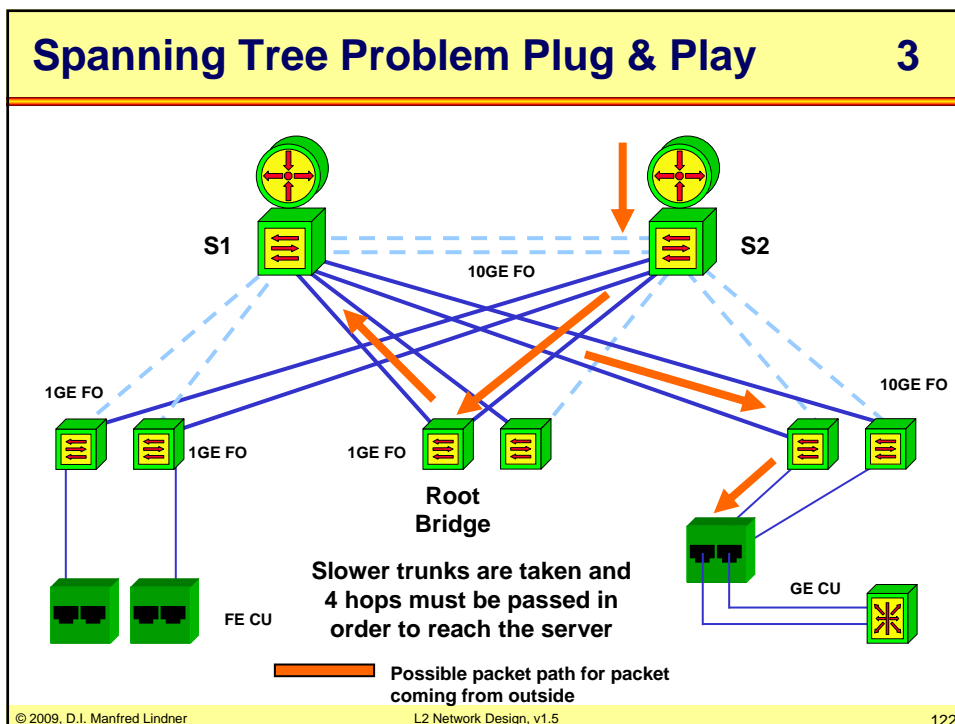
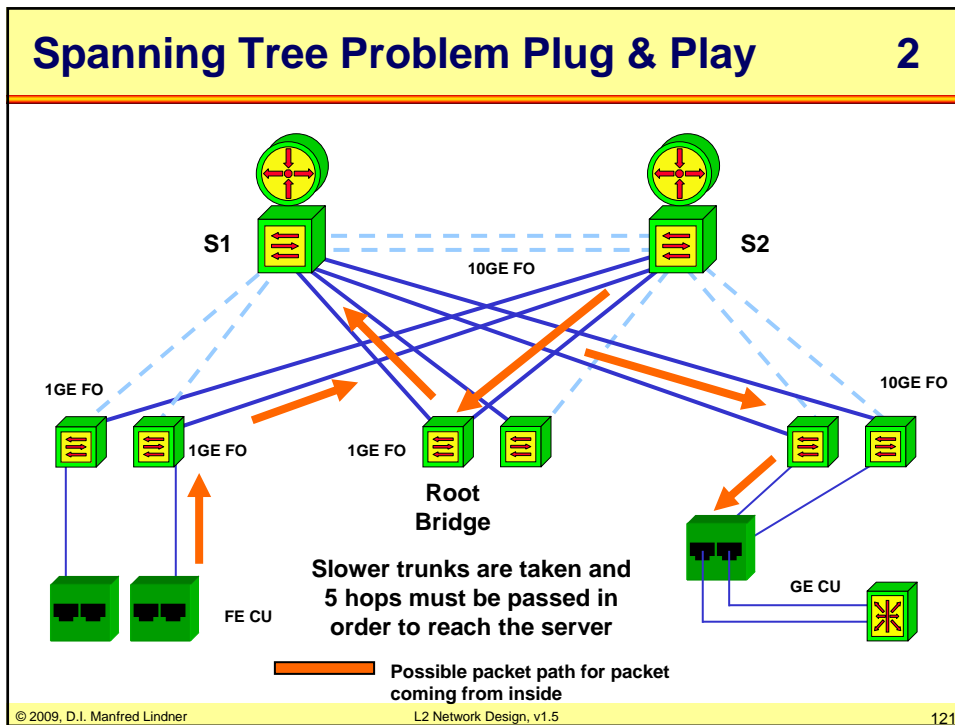
L2 Network Design, v1.5

118

L101 - L2 Network Design

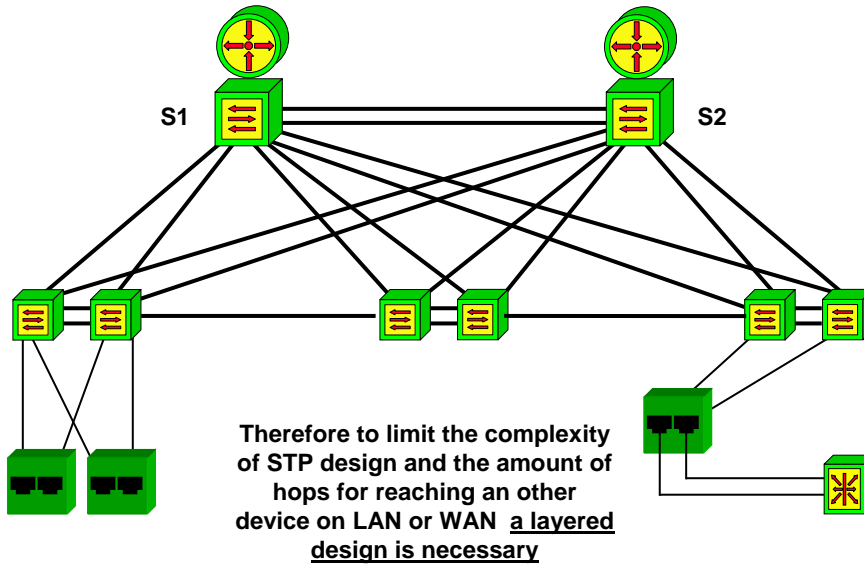


L101 - L2 Network Design



L101 - L2 Network Design

Figure Out STP in Complex Scenarios

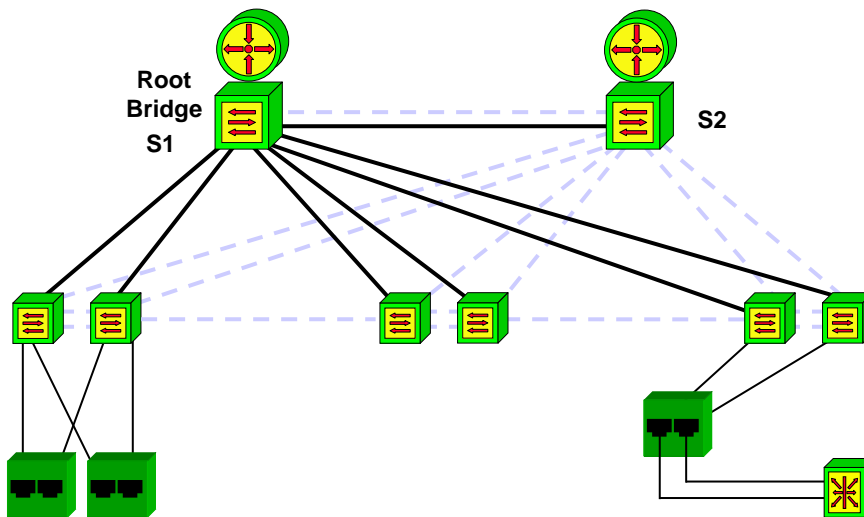


© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

123

Figure Out STP in Complex Scenarios Spanning Tree – Good Luck

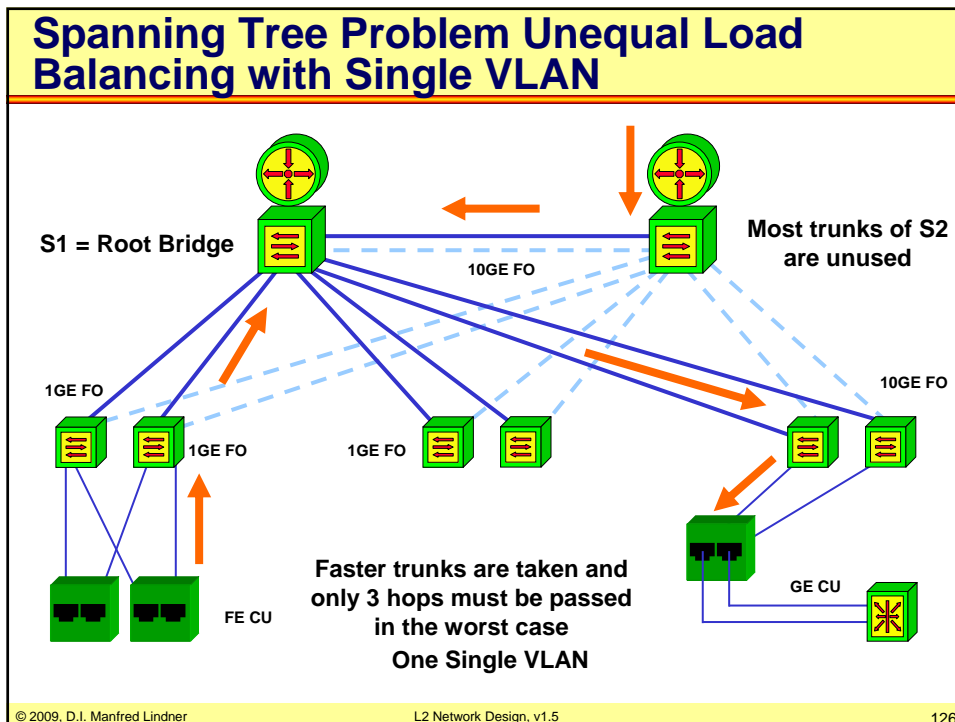
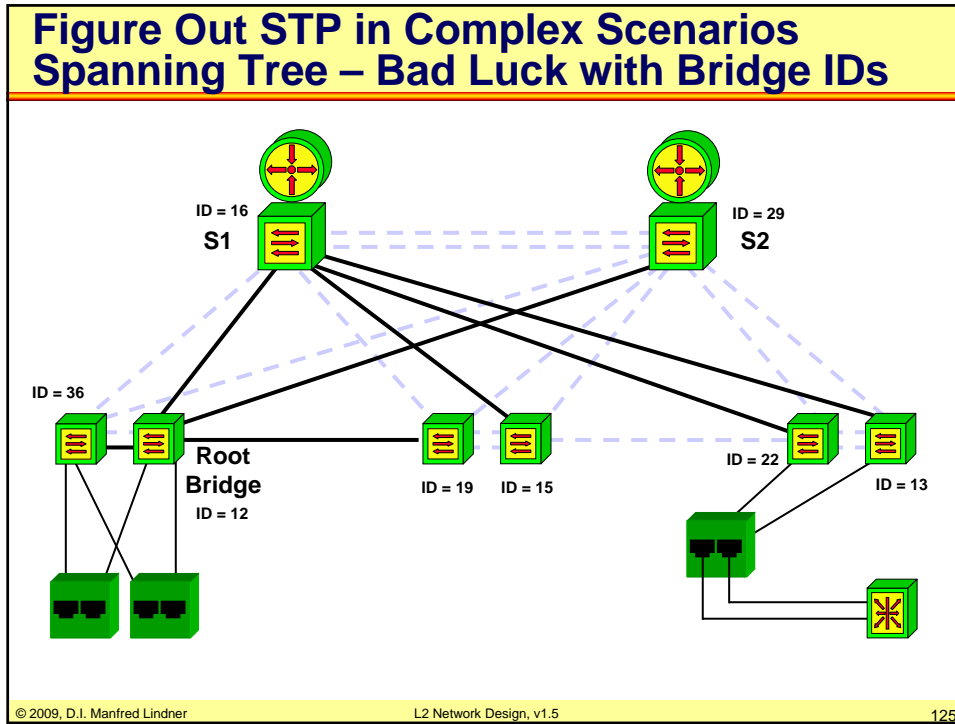


© 2009, D.I. Manfred Lindner

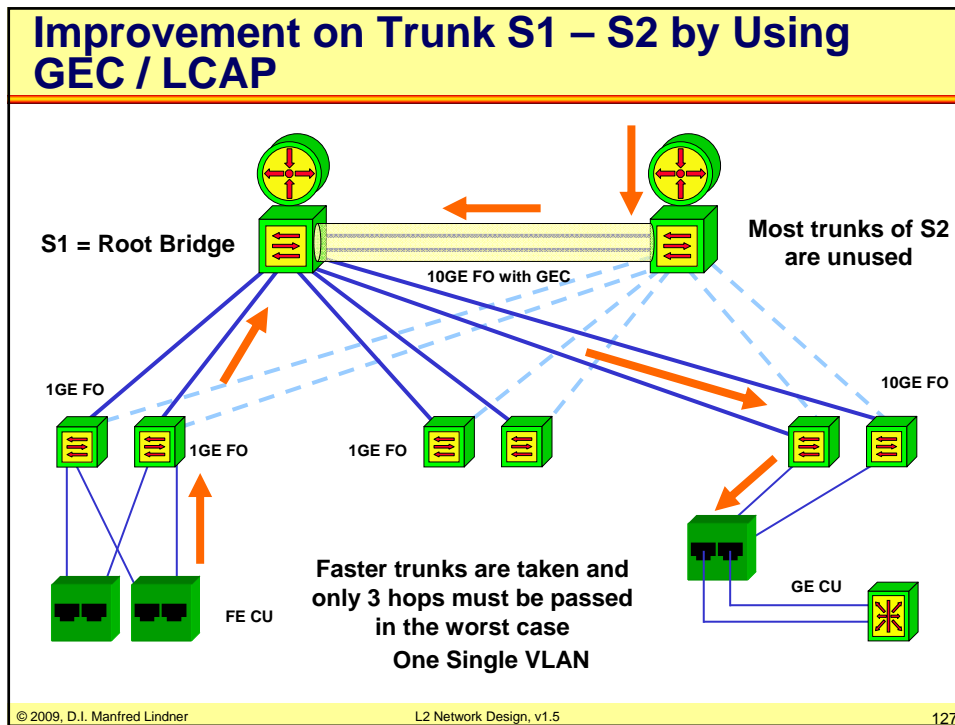
L2 Network Design, v1.5

124

L101 - L2 Network Design



L101 - L2 Network Design



Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

L101 - L2 Network Design

Best Practices

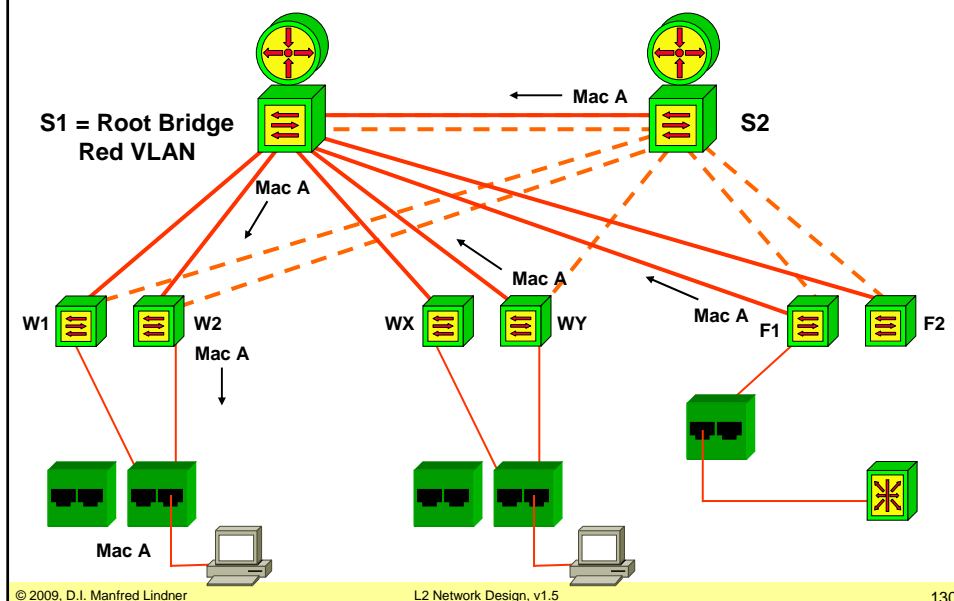
- **Build at least two separated VLAN's**
 - In case of IP that means two IP subnets
- **How to achieve?**
 - Per VLAN STP (Cisco)
 - MIST (Multiple Instances Spanning Tree)
 - IEEE 802.1d
- **Tune STP parameters**
 - In order to use all trunks and all switches in a similar way

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

129

Build STP for VLAN 1

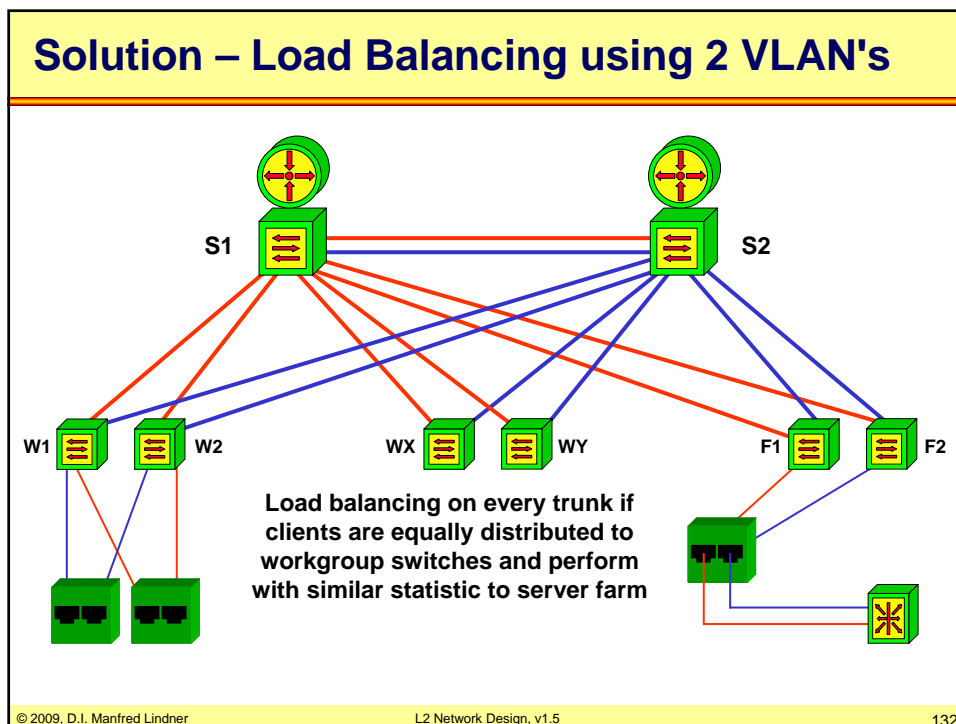
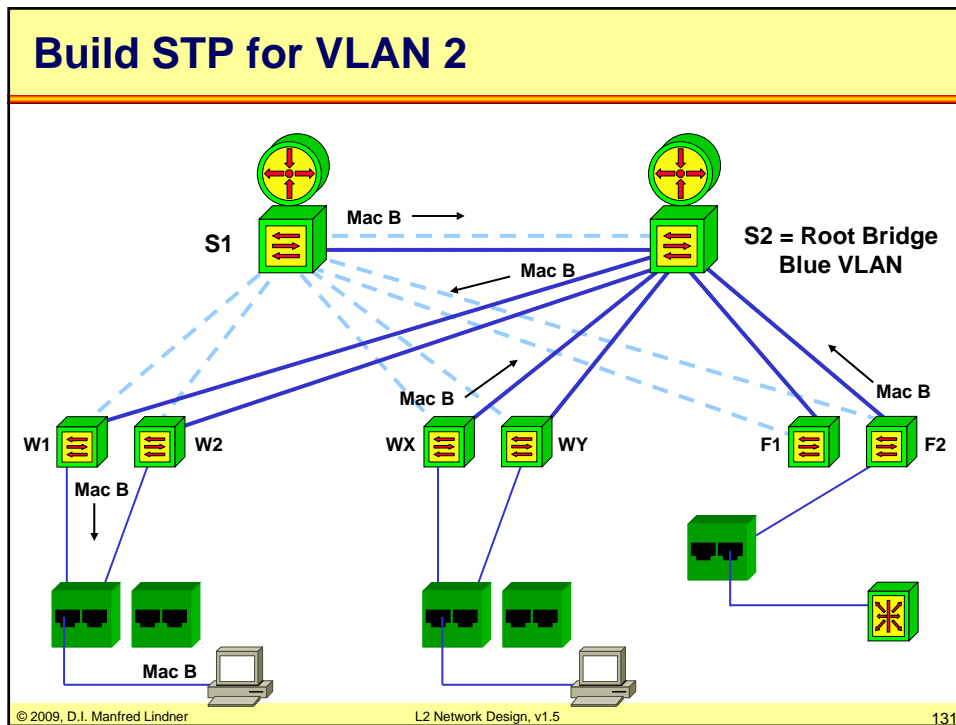


© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

130

L101 - L2 Network Design



L101 - L2 Network Design

Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

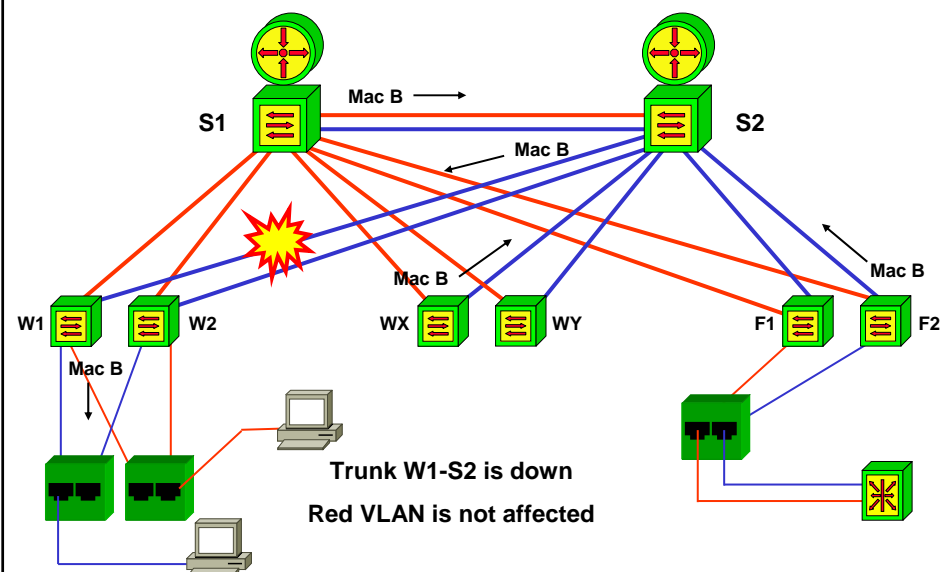
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

133

Link Failure (Trunk)

1

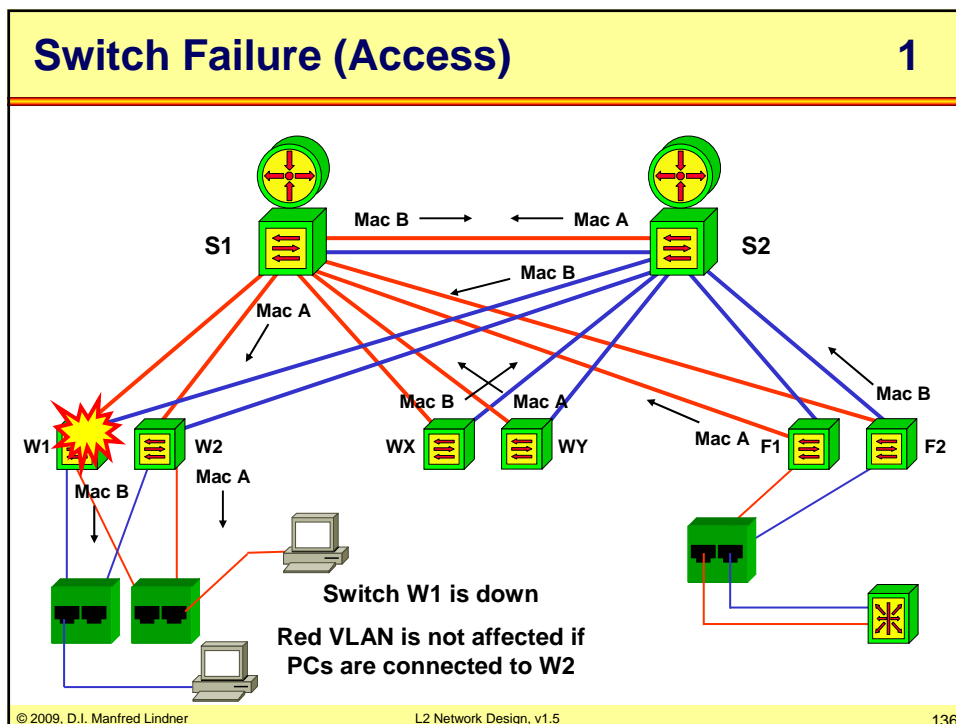
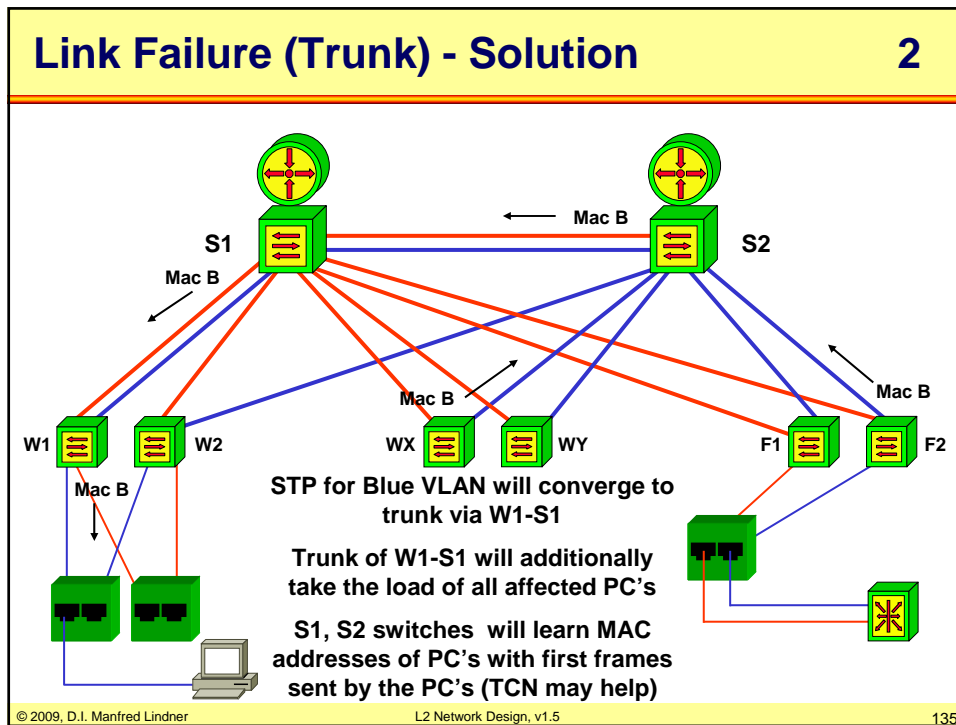


© 2009, D.I. Manfred Lindner

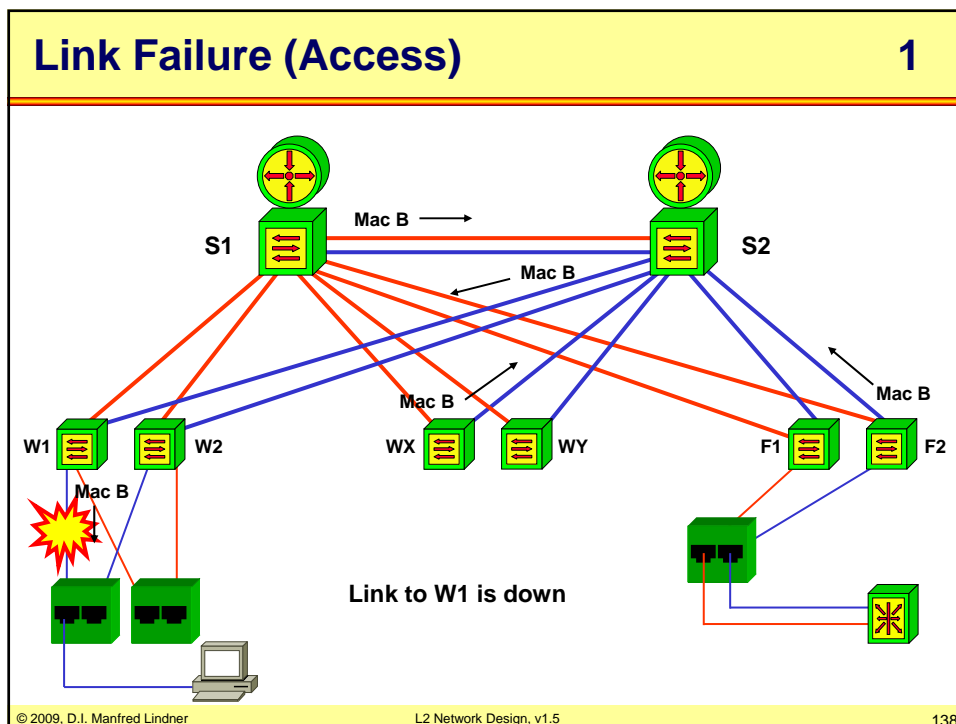
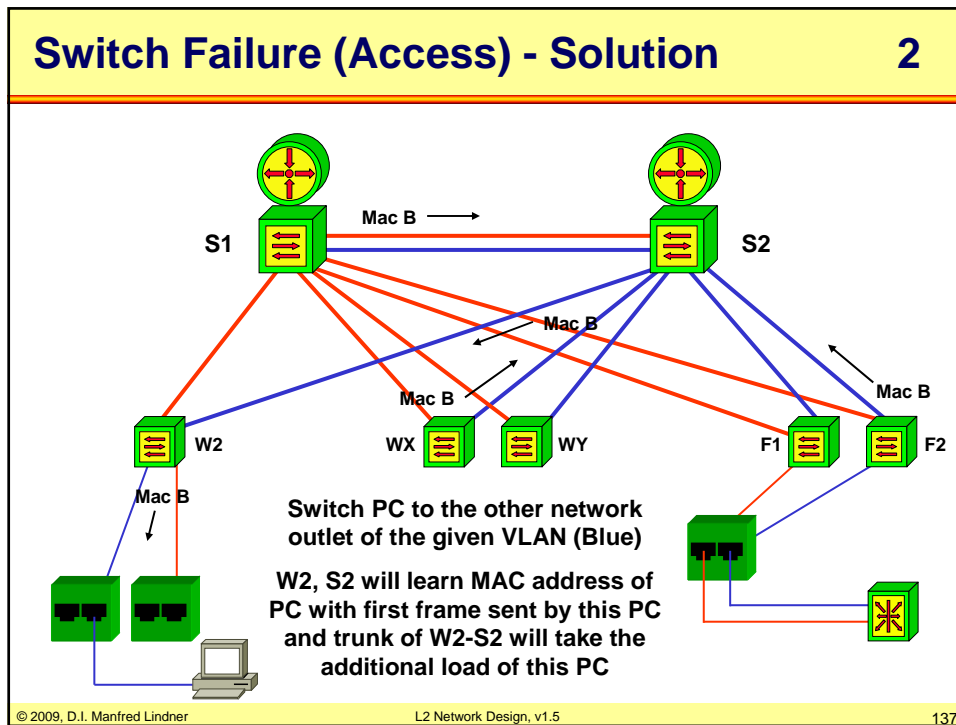
L2 Network Design, v1.5

134

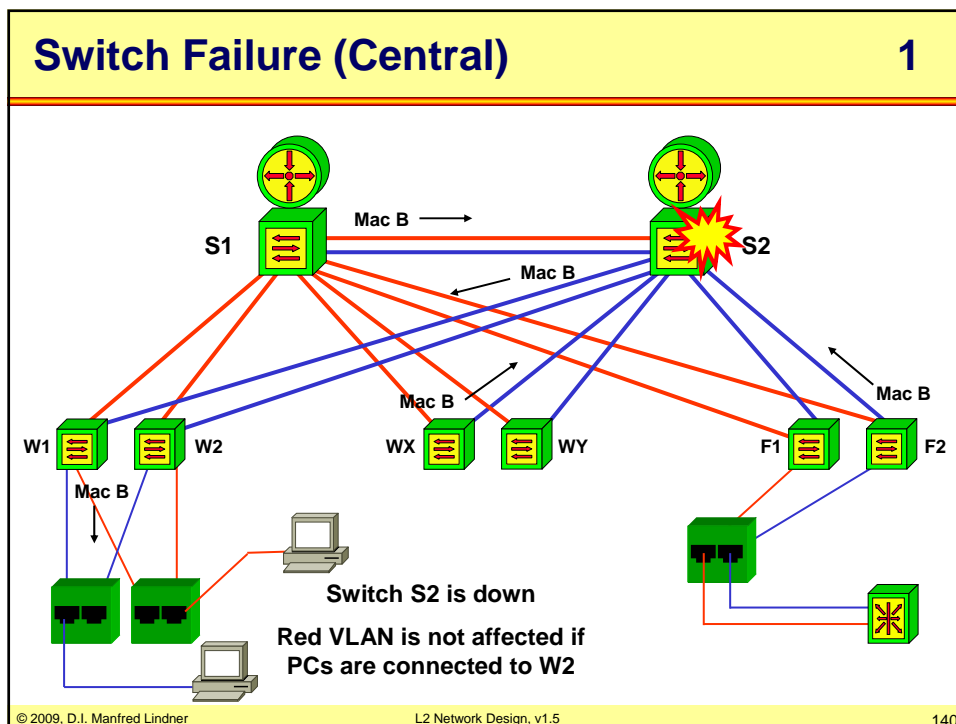
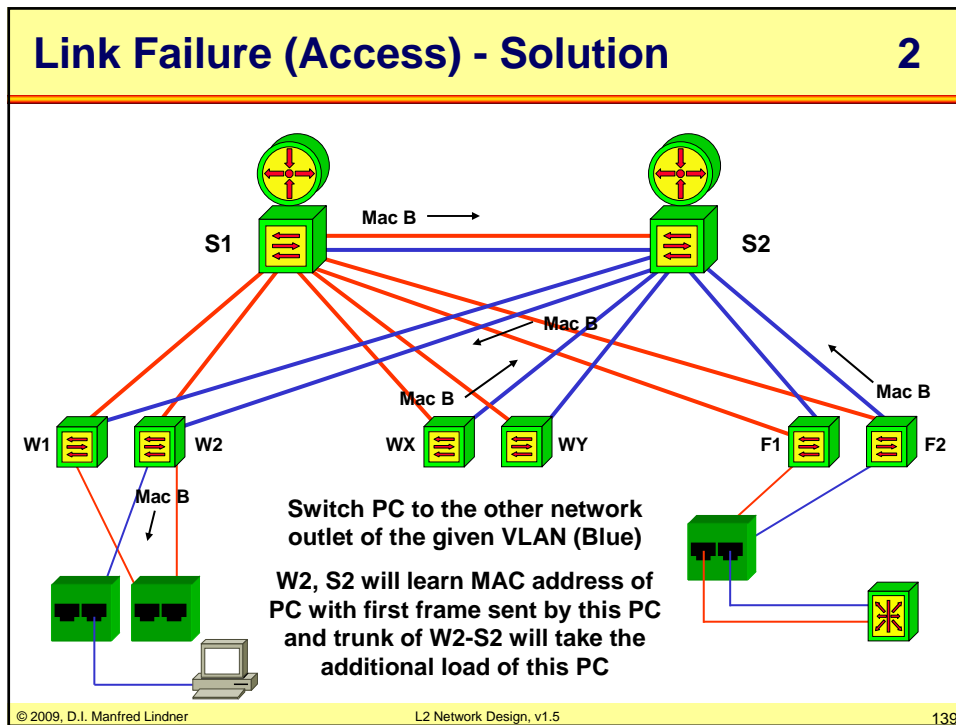
L101 - L2 Network Design



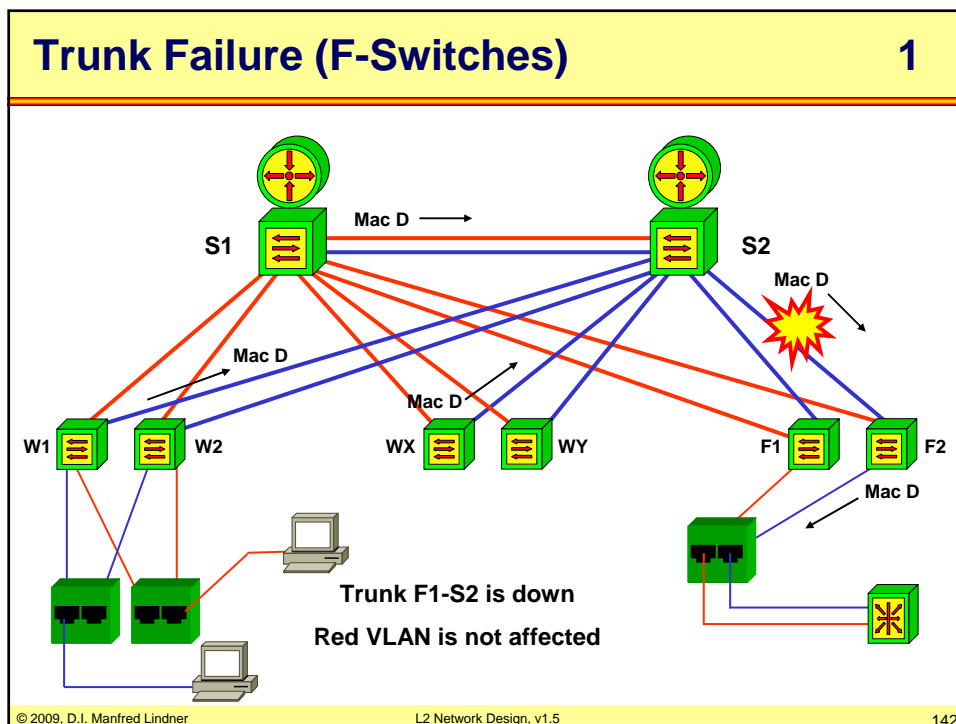
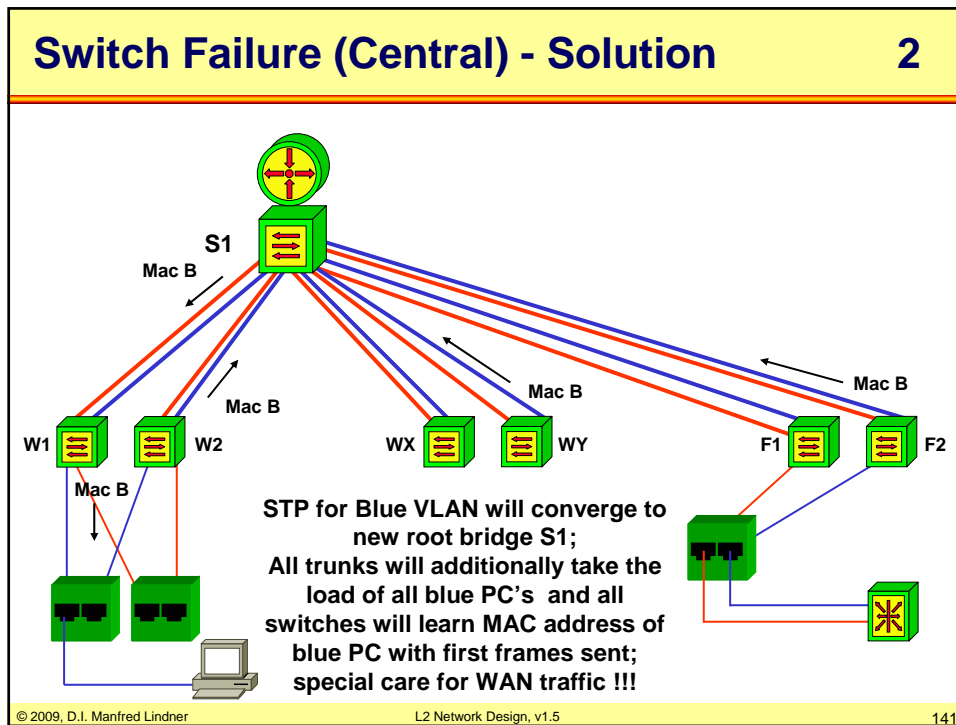
L101 - L2 Network Design



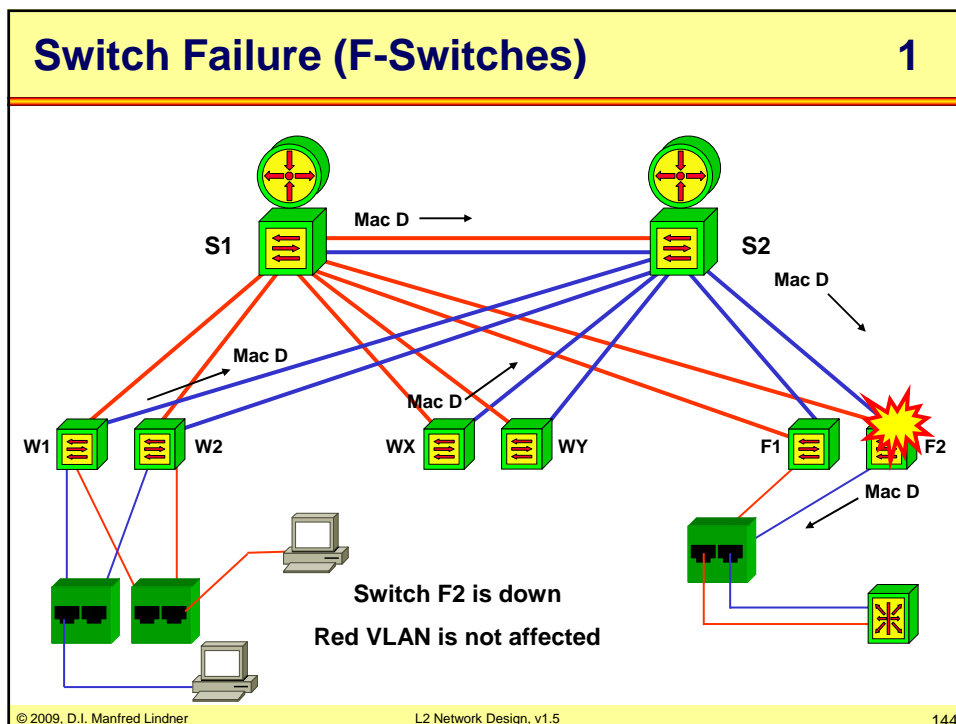
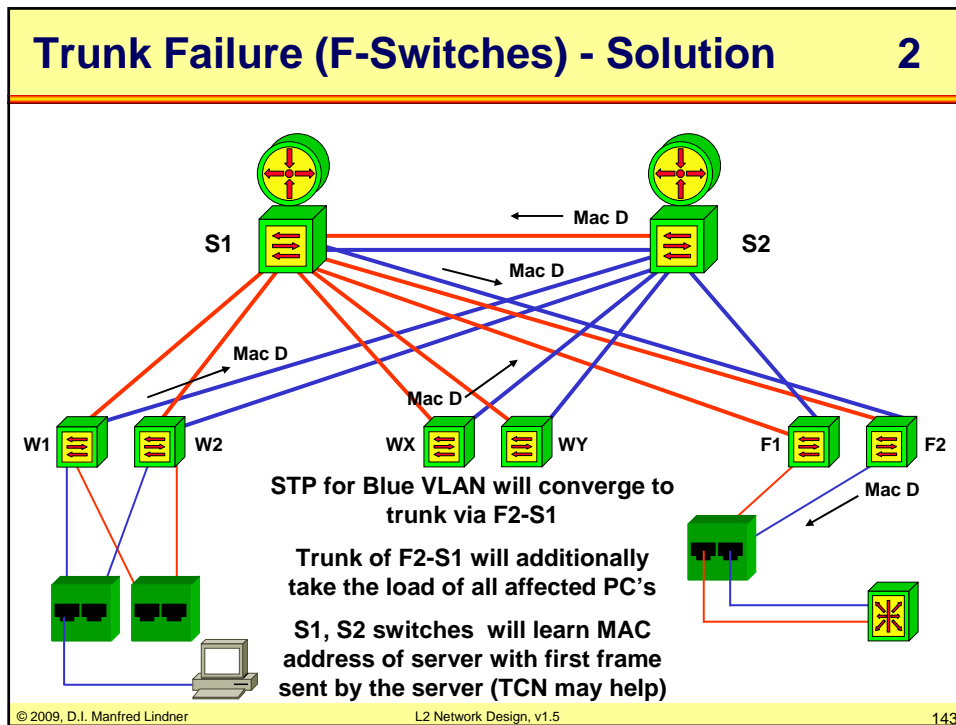
L101 - L2 Network Design



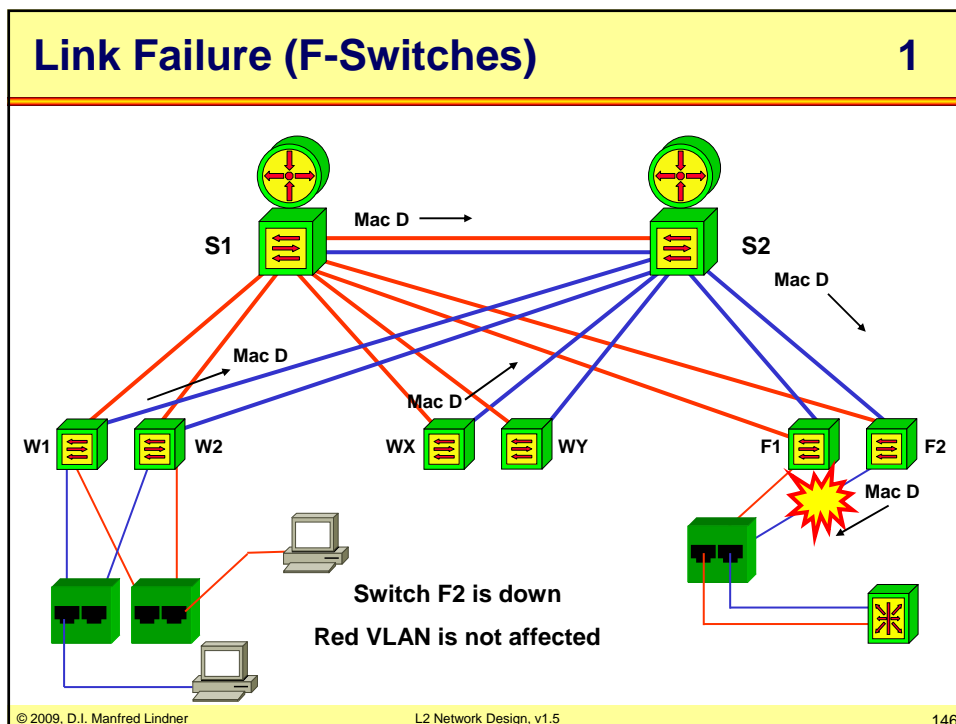
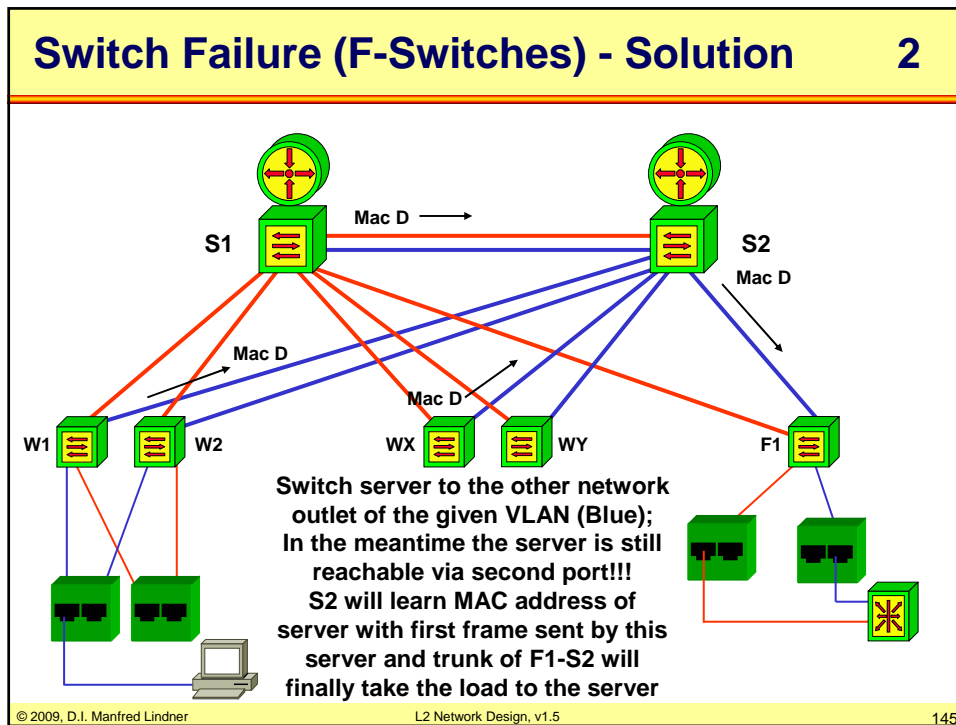
L101 - L2 Network Design



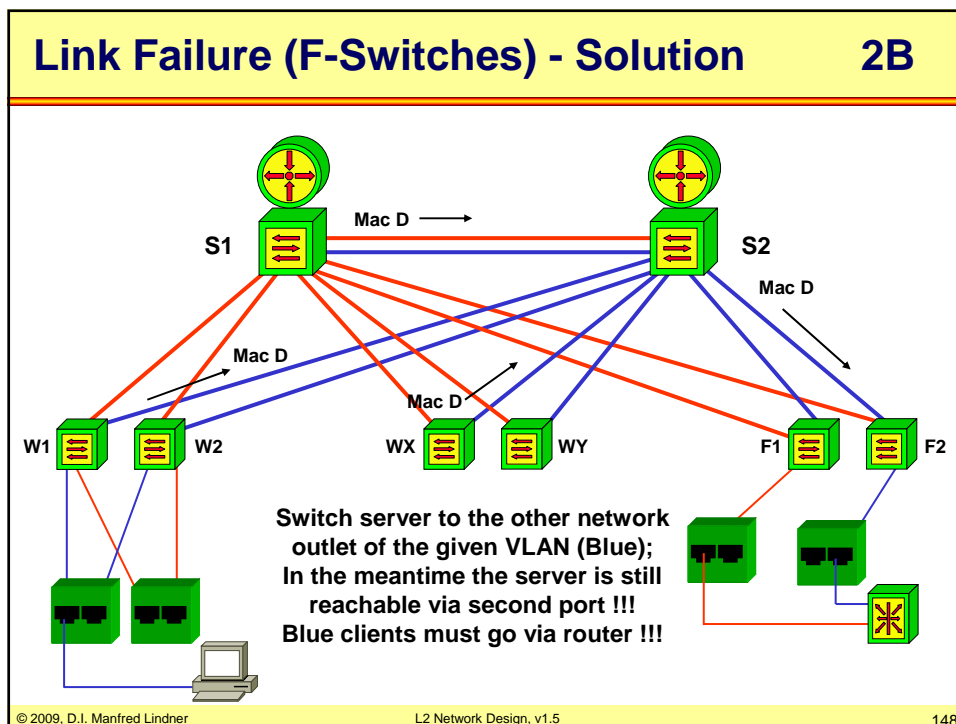
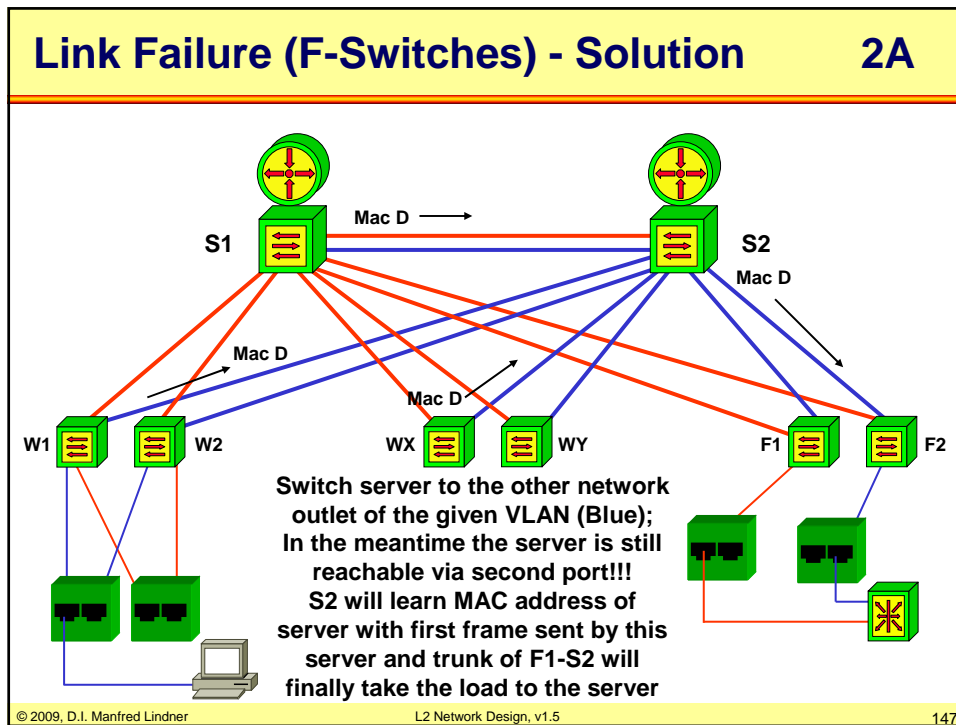
L101 - L2 Network Design



L101 - L2 Network Design



L101 - L2 Network Design



L101 - L2 Network Design

Agenda

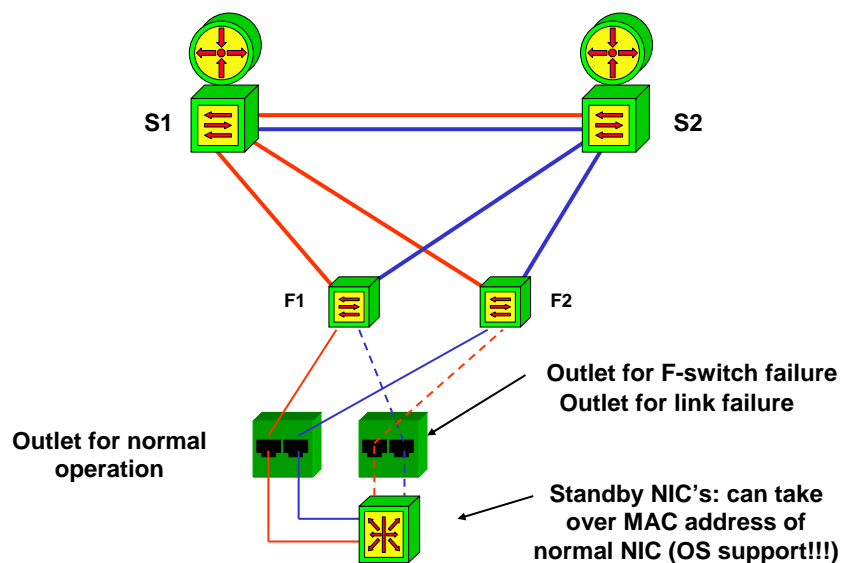
- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

149

Server Connections to F-Switches – Advanced Techniques

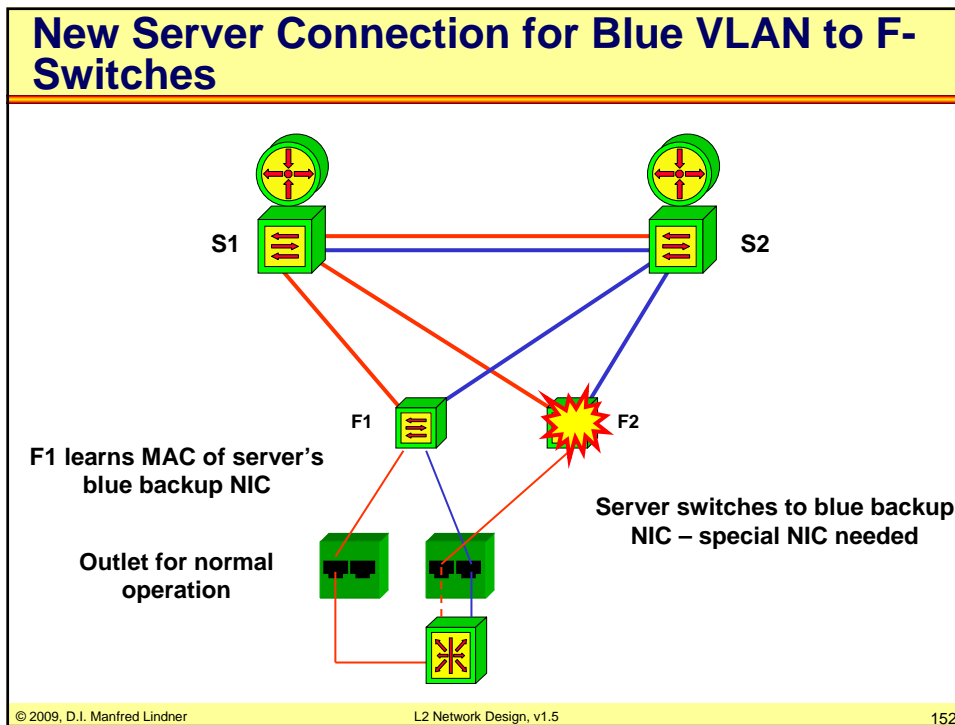
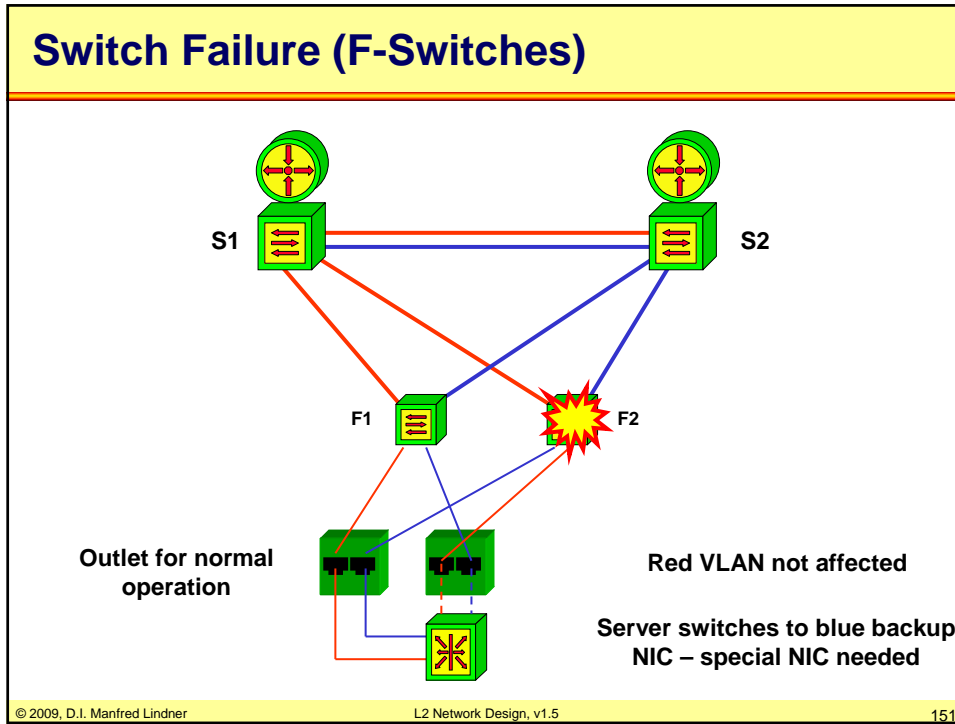


© 2009, D.I. Manfred Lindner

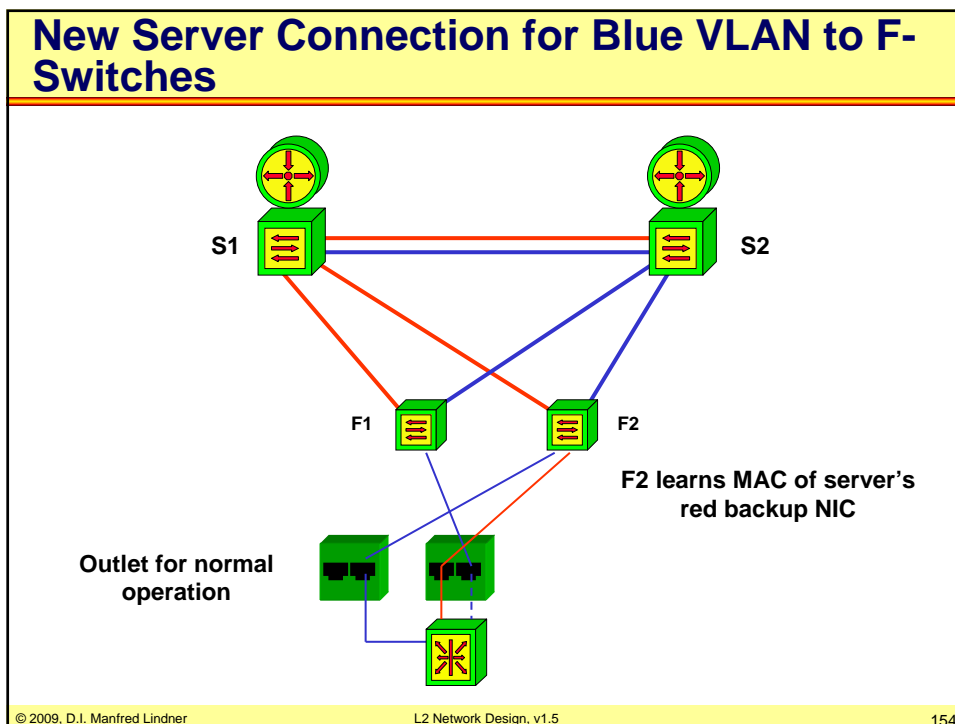
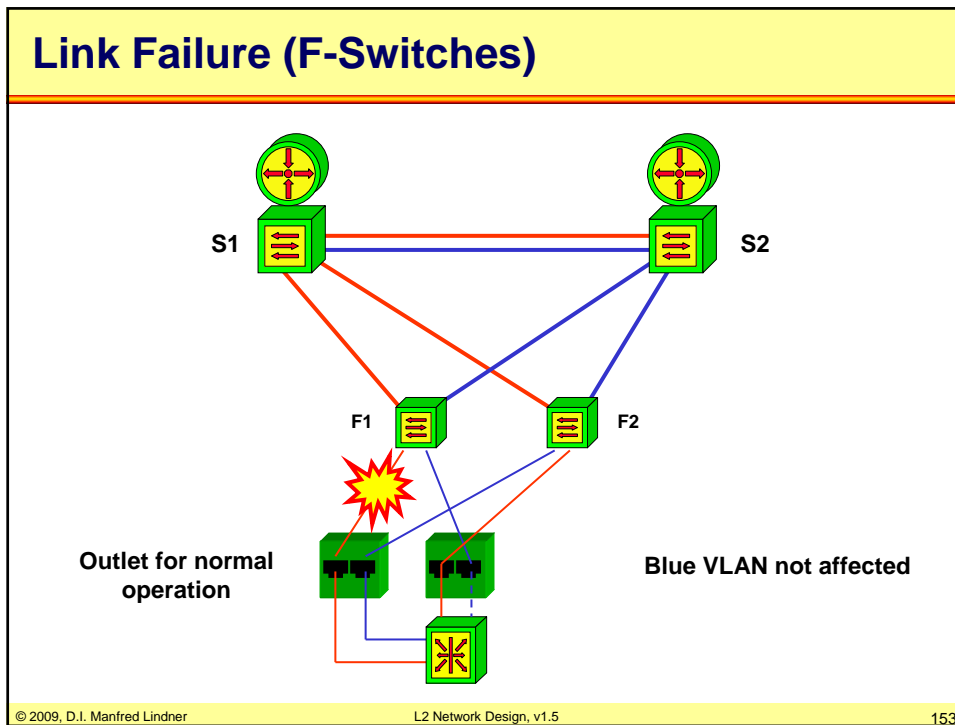
L2 Network Design, v1.5

150

L101 - L2 Network Design

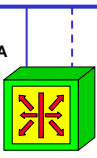
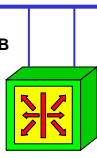

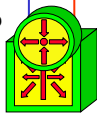


L101 - L2 Network Design



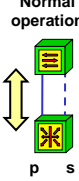
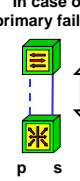
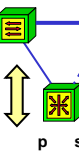
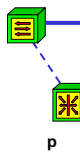
L101 - L2 Network Design

Configuration Options for Redundant NIC

<p>VLAN Blue / IP NET Blue</p> <p>Type-A</p>  <p>only one NIC active, the other NIC in standby (will take over in case of failure)</p> <p>1 server IP address and 1 server MAC address, 1 Default Gateway</p>	<p>VLAN Blue / IP NET Blue</p> <p>Type-B</p>  <p>both NIC's active, load balancing possible</p> <p>2 server IP addresses and 2 server MAC addresses, 1 Default Gateway, strategy which NIC to use is necessary</p>
<p>VLAN Red / IP NET Red</p> <p>VLAN Blue / IP NET Blue</p> <p>Type-C</p>  <p>both NIC's active, load balancing possible</p> <p>2 server IP addresses and 2 server MAC addresses, 2 Default Gateways, strategy which NIC to use is necessary</p>	<p>VLAN Red / IP NET Red</p> <p>VLAN Blue / IP NET Blue</p> <p>Type-D</p>  <p>both NIC's active and connected to internal router, load balancing via routing protocol is possible</p> <p>1 server IP addresses on the loopback interface, which connects the server to the internal router</p>

© 2009, D.I. Manfred Lindner L2 Network Design, v1.5 155

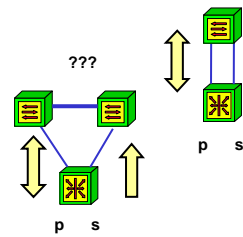
Redundant NIC Type-A, Intel Teaming 1

<p>Normal operation</p>  <p>In case of primary failure</p> 	<p>Intel Teaming Basic AFT (Adapter Fault Tolerance) Feature; one primary adapter, up to 7 secondary adapters primary for transmitting and receiving all traffic from and to the unique server MAC address; multicast/broadcast probes to ensure a secondary adapter is available, therefore all links are physically active and all secondary adapters have their own MAC address; secondary will take over in case primary fails that means it starts sending traffic with the server MAC address and receiving traffic for the server MAC address; STP must be switched off</p>
<p>Normal operation</p>  <p>In case of primary failure</p> 	<p>Intel Teaming SFT (Switch Fault Tolerance) Feature; one primary adapter, one secondary adapter primary for transmitting and receiving all traffic from and for the unique server MAC address; multicast/broadcast probes to ensure a secondary adapter is available, therefore all links are active and all secondary adapters have their own MAC address; secondary will take over in case primary fails that means it starts sending traffic with the server MAC address and receiving traffic for the server MAC address; STP must be switched on</p>

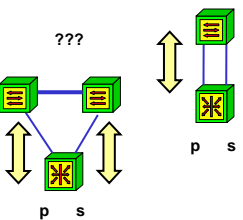
© 2009, D.I. Manfred Lindner L2 Network Design, v1.5 156

L101 - L2 Network Design

Redundant NIC Type-A, Intel Teaming



Intel Teaming ALB (Adaptive Load balancing) Feature;
 one primary adapter, up to 7 secondary adapters
 primary for receiving all traffic to the unique server MAC
 address and unique server IP address;
 secondary are used for balancing the load for transmit traffic;
 all links are active and all secondary adapters have their own
 MAC address;
 secondary send with their own MAC address and will not
 answer ARP Requests to the server IP address;
 thus the server MAC address will not be seen on switch ports
 leading to secondary adapters
 (? Doing so will not solve the ARP cache problem of the client-
 PCs because every received Ethernet frame at the client-PC will
 refresh/change the ARP cache ?)



Intel Teaming RLB (Receiver Tolerance) Feature;
 same as ALB, but now secondary answer ARP requests based
 on a internal scheduling decision hence populating the ARP
 cache of different client-PCs with different MAC addresses for
 the same unique server IP address;
 Tricky procedure in case the server itself sends an ARP request
 for a client with its unique server MAC address -> client-PC ARP
 caches would be refreshed and traffic would be directed to the
 primary -> hence appropriate ARP replies must be sent out to
 correct ARP cache again

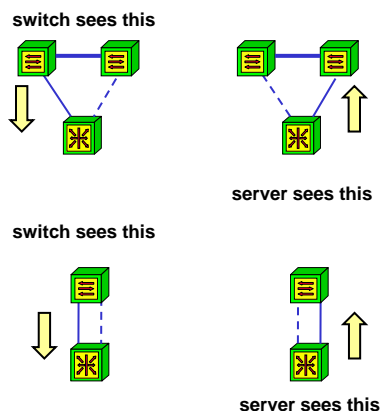
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

157

Redundant NIC Critical Aspect

In case of active / standby it is important that both sides (PC and the switch(es))
 have the same sight who is active and who is standby (symmetric view)



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

158

L101 - L2 Network Design

Intel Advanced Network Services Software (ANS)

- What is Intel ANS
- Teaming Features
- Teaming Modes
- Dependencies
- Details – How it works
 - Probe Packets
 - Server Load Balancing
 - Receive Load Balancing
 - Static Link Aggregation
 - 802.3ad Dynamic Mode

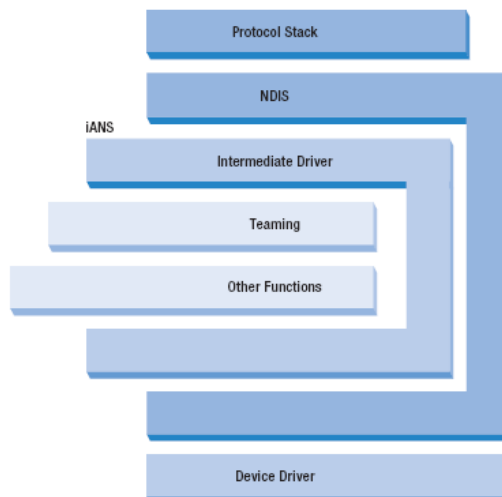
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

159

What is Intel ANS

- Implemented as an intermediate driver in the servers driver stack
- Windows and Linux supported
- Should Work also with „non-Intel Adapters“
 - Min. 1 Intel Pro Server Adapter need but we have bad experiences with such scenarios in FRQ -> **Avoid it**



© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

160

L101 - L2 Network Design

Teaming Features

- **Fault Tolerance**
 - 1 or more secondary adapter take over if primary fails
- **Link Aggregation**
 - Combine multiple adapters into a single channel
 - Bandwidth increase only available to multiple destination addresses
 - Must be supported by connected switch!
- **Load balancing**
 - Distribution of transmission and reception load among aggregates network adapters
 - Agent in ANS analyzes traffic and distributes the packets based on destination addresses

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

161

Teaming Modes

- **Adapter Fault Tolerance (AFT)**
 - 2-8 adapter supported
 - If primary fails -> secondary takes over
 - All adapters must be connected to same network
- **Switch Fault Tolerance (SFT)**
 - Failover relationship between 2 Adapters connected to different switches
 - STP must be enabled on the switches
 - STP must be disabled on connected Ports

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

162

L101 - L2 Network Design

Teaming Modes

- **Adaptive Load Balancing (ALB)**
 - Load balancing of transmit traffic
 - Receive Load Balancing (RLB) is advanced feature – enabled by default
- **Static Link Aggregation (SLA)**
 - IEEE 802.3ad static and dynamic mode
 - Needs compatible switch!
 - Intel Link Aggregation (LA), Cisco Fast EtherChannel (FEC), Gigabit EtherChannel (GEC) replaced by static link aggregation mode
 - 2-8 Adapters – all ports same speed
 - Incorporates AFT and ALB modes

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

163

Teaming Features and Modes

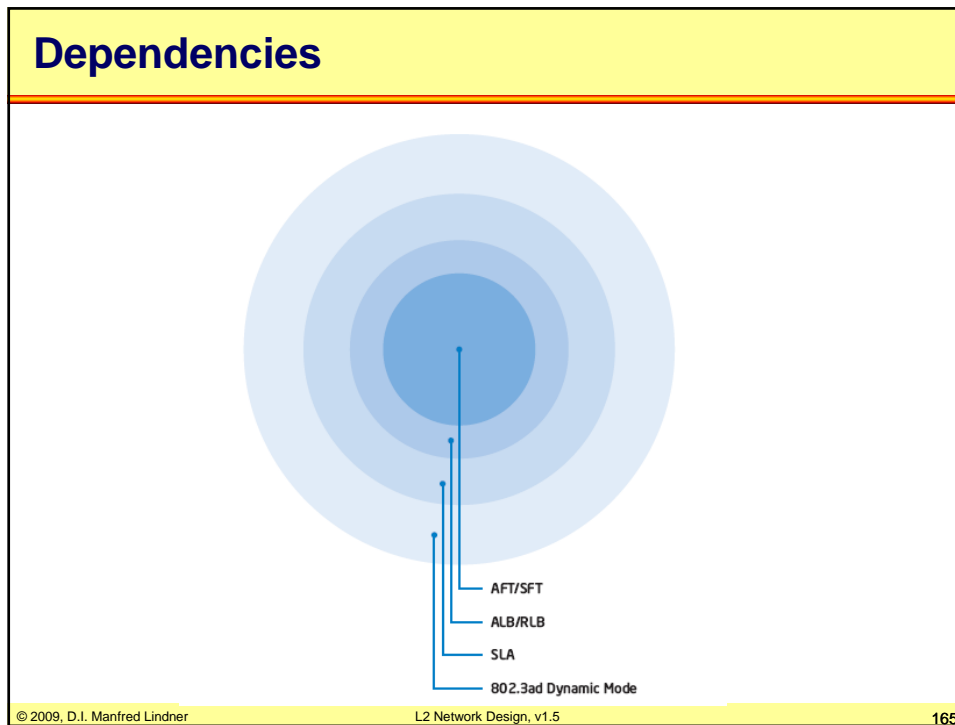
Features	Modes				
	AFT	ALB	RLB	SLA	Dynamic 802.3ad LACP
Fault Tolerance	X	X	X	X	X
Link Aggregation		X	X	X	X
Load Balancing		Tx	Tx/Rx	Tx/Rx	Tx/Rx
Layer 3 Address Aggregation		X	IP only	X	X
Layer 2 Address Aggregation				X	X
Mixed Speed Adapters	X	X			

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

164

L101 - L2 Network Design



Details – How does it work

- **How To Detect State And Health Of Adapters**
 - Probe Packets
 - Adapters send and receive them to determine presence and state of other adapters
 - Either broadcast or multicast – configurable in software
 - Activity Based Tolerance
 - If probe packets are not used or do not reach their destination -> sensing activity on the line
 - Link Based Tolerance
 - Used if neither probe packets nor activity based tolerance are available or successful

© 2009, D.I. Manfred Lindner L2 Network Design, v1.5 166

L101 - L2 Network Design

Probe Packets Details

- **2 different types of user configurable probes**
- **Each member uses 2 flags – Send and Receive – to track status**
- **When adapter sends probe sets both flags to Pending state**
- **When packet is received by a member of same team – it sets its receive flag to ReceiveComplete and sets sending Flag to SendComplete**
- **If Primary Adapter is set to disabled -> Secondary Adapter takes this role – new Secondary will be elected**

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

167

Server Load Balancing Methods

- **Adaptive Load Balancing (ALB)**
 - Receive Load Balancing (RLB) is a subset of ALB
 - Transmit Traffic balanced by Hash Table of last Octet of receivers IP address
 - New Dataflows are assigned to least loaded team member
 - After timeout of load bal. timer Dataflows are rebalanced
 - ALB without RLB uses Primary Team Members MAC in ARP Reply Packets
 - Send Packets include Team Members MAC as source
 - Failover: Secondary Adapter gets MAC of Primary
 - Do not Hotplug Primary and reuse somewhere else until Server Reboot

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

168

L101 - L2 Network Design

Receive Load Balancing (RLB)

- **When receiving ARP Request -> Intel ANS answers with MAC Address of the port which is chosen to service this client**
- **Clients are allocated in a “Round-Robin” manor**
- **RLB client table is refreshed after ReceiveBalancing Interval**
- **OS ARP requests are send through primary port**
 - Receive load collapses to primary
 - ANS sends gratuitous ARP to all clients in hash to restart RLB

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

169

Static Link Aggregation (SLA)

- **All Ports share same MAC Address**
- **For the switch this is a single link**
- **No designated primary port in the team**
- **Links must be same speed**
- **Switch handles receive load balancing**

© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

170

L101 - L2 Network Design

802.3ad Dynamic Mode

- All members share same MAC
- Switch ports must use LACP protocol
- Switch communicates with Intel ANS to add or remove members of team
- No designated primary – but first teamed port is initiator to switch
- Removal of Initiator could lead to packet loss
- To avoid this -> preconfigure the switch ports for added or removed members

Overview

Function	Intel (AFT, SFT, ALB)	Intel (SLA)	Intel (802.3ad)
Number of NICs per team	8 (2 for SFT)	8	8
NIC Fault Tolerance	Yes	Yes	Yes
Switch Fault Tolerance	2	Yes	Yes
Tx Load Balance	Yes	Yes	Yes
Rx Load Balance	Yes	Yes	Yes
Requires compatible switch	Yes	Yes	Yes
Heartbeats to check connectivity	Yes	No	No
NICs with different media	Yes	No	Yes
NICs with different speeds	Yes	No	Yes
Load balances TCP/IP	Yes	Yes	Yes
Load balances other protocols	Yes	Yes	Yes
Same MAC address for all team members	No	Yes	Yes
Same IP address for all team members	Yes	Yes	Yes
Load balancing by IP address	Yes	No	No
Load balancing by MAC address	Yes	Yes	Yes
802.1q tagged VLANs	Yes	Yes	Yes
Untagged VLANs	Yes	Yes	Yes

L101 - L2 Network Design

Information - Sources

- **Intel Advanced Network Service Software Whitepaper**
 - www.intel.com/network/connectivity/resources/doc_librarywhite_papers/254031.pdf
- **Ethertype 886 from the Intel Website**
 - <http://thetechfirm.com/packets/886/886.ppt>

Agenda

- **Scenarios**
- **Physical Layer**
 - Introduction
 - LAN
 - WAN
- **LAN (L2)**
 - Review Ethernet Technology
 - Design Considerations
 - Design Solution – Best Practices
 - Failover Handling
 - Advanced Techniques - Teaming
 - LAN – WAN Interconnection

L101 - L2 Network Design

LAN – WAN Interconnection VLAN Interconnection

- **Now let us look to Layer 3 (IP)**
- **We need routers**
 - For connecting the two VLAN's
 - For connecting the LAN infrastructure of a site to the WAN infrastructure
- **Be very careful to differentiate between**
 - L1 look of your network
 - L2 look of your network (VLAN, STP)
 - L3 look of your network (IP, ARP)

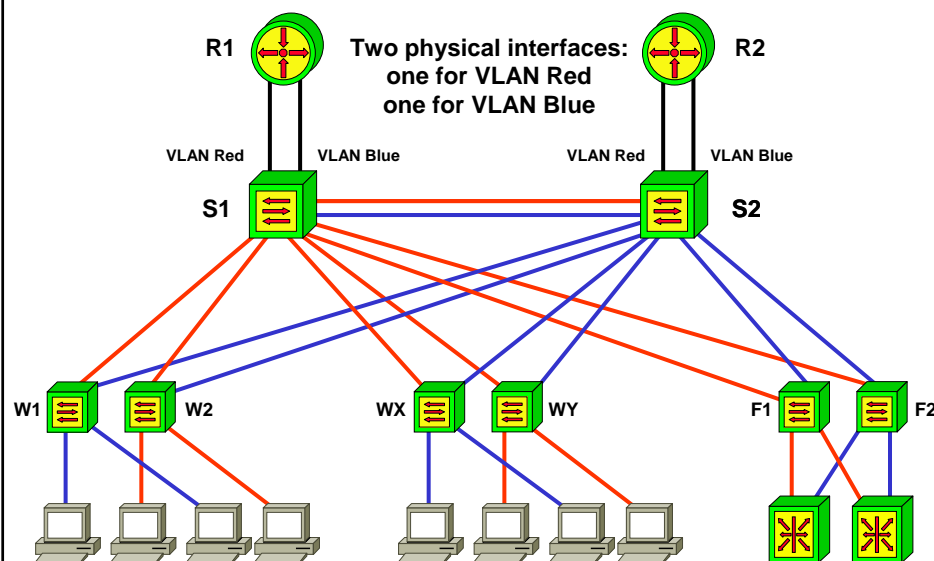
© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

175

Router Connections

1

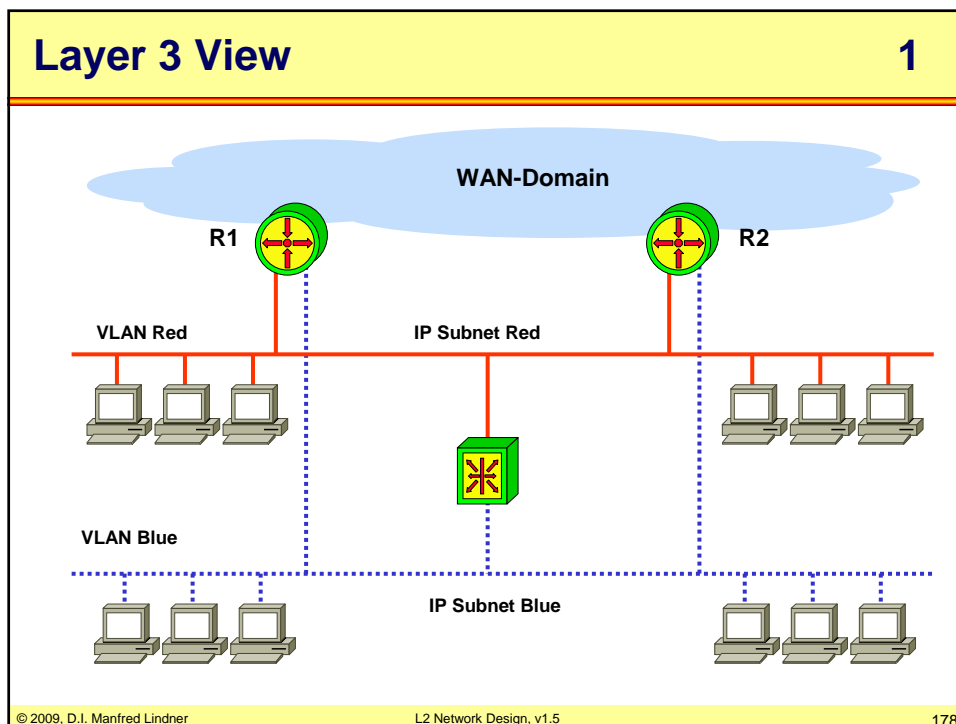
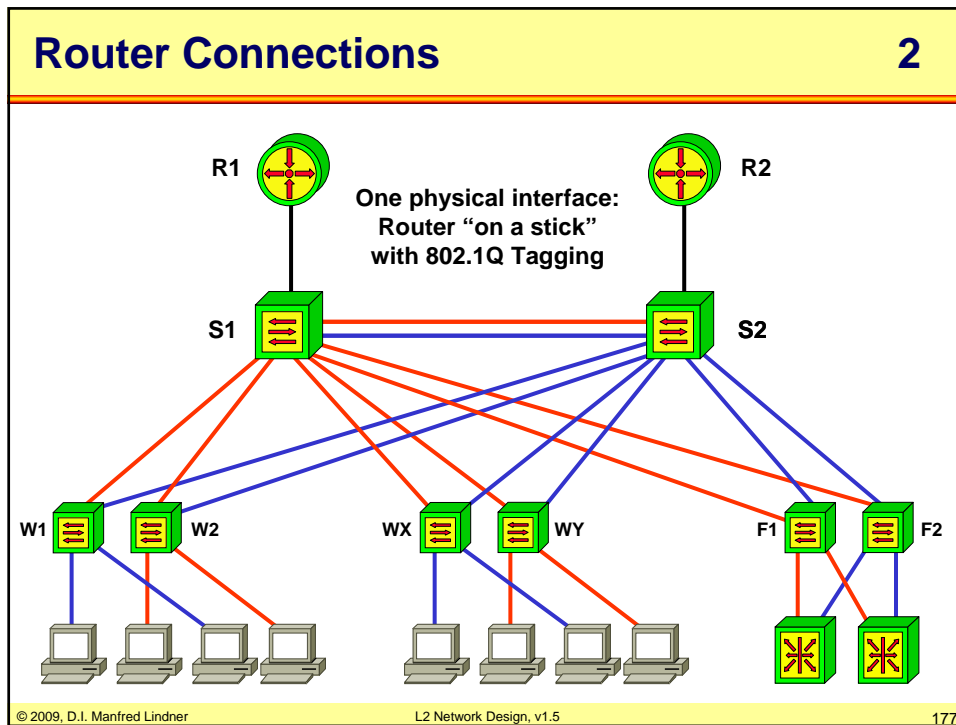


© 2009, D.I. Manfred Lindner

L2 Network Design, v1.5

176

L101 - L2 Network Design



L101 - L2 Network Design

