



FCoE: Fabric Crosstalk

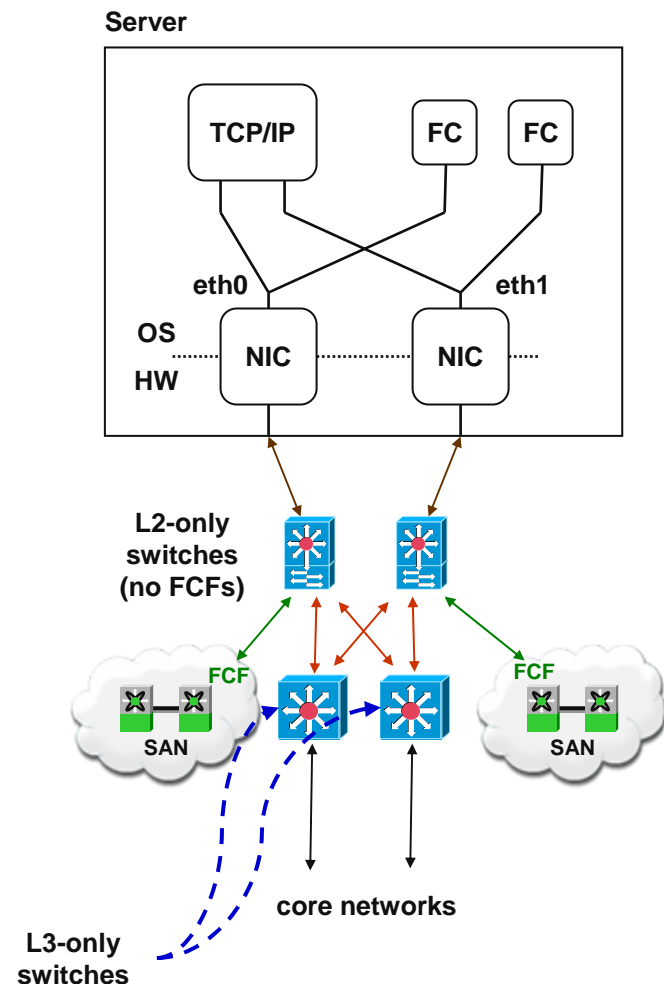
David L. Black, EMC
October, 2007

Background

- Action item from July meeting (Colorado Springs)
 - AI-2: EMC/HP/IBM group to provide concrete forwarding loop configuration examples
- FCoE and Layer 2 (L2) Ethernet-only switches (aka bridges)
 - Native FC links are always private (2 ends)
 - Ethernet-only L2 switches: FCoE links can be shared (3+ ends)
- This presentation: Shared link issues (three scenarios)
 - What goes wrong? (includes a forwarding loop scenario)
 - How can these problems be prevented?
- Focus: Layer 2 (Ethernet-only) switches
 - Necessary to create these problem scenarios
 - Necessary to support in the FCoE standard

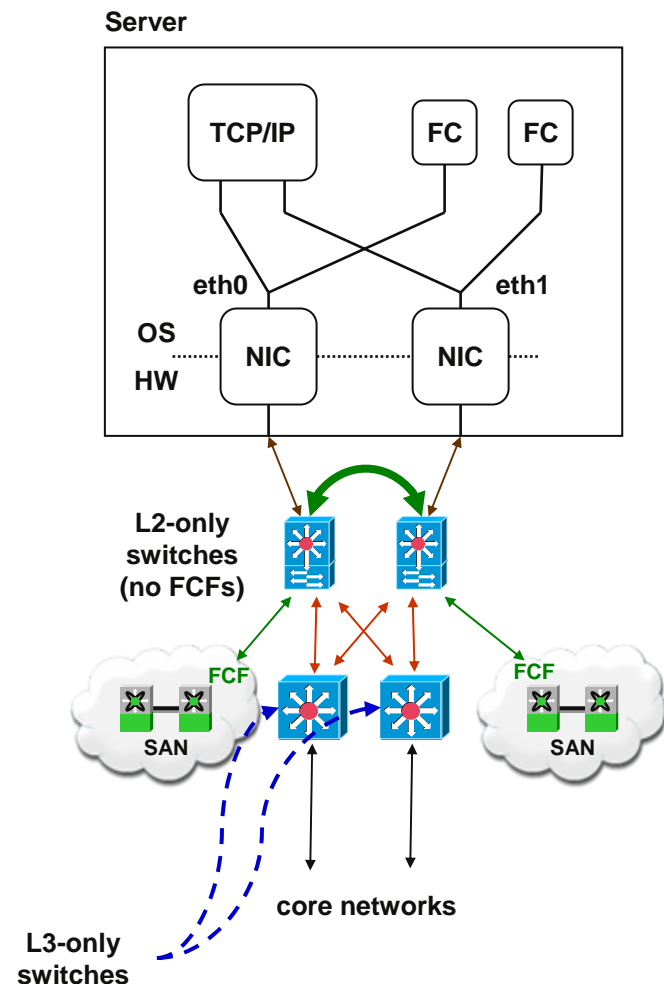
Fabric Crosstalk Problem Setup – Scenarios 1 & 2

- Edge switches are Ethernet-only (L2)
 - No IP (L3), no FCF (FC-awareness)
 - VF_Ports (FCFs): Edges of FC clouds
- **VLAN 1 (left and right instances) for FCoE: Dynamic or Virtual MACs**
 - Left and right FC fabrics
- **VLAN 2 (left and right instances) for TCP/IP: Hardware MACs**
- Each NIC – FCoE MAC in VLAN 1 plus HW MAC in VLAN 2
 - Assume same FCoE MAC assigned to each NIC (e.g., same FCID assigned by each fabric)
 - Software FCoE in virtualized guest OS's may also duplicate MACs across servers
 - **Hardware MACs should be different**



Fabric Crosstalk Problem Scenario 1

- **Mistake: Cross connect edge switches with VLAN 1 link**
 - Result: 2 instances of same FCoE MAC on new VLAN 1 [FC-MAC-1]
 - Spanning tree: new link stays active
- Nothing breaks immediately
 - Both switches learned not to forward FC-MAC-1 traffic to other switch.
- What if one of them forgets?
 - Or is “helped” to forget by an administrator or injected frame?
- **Read data corruption is possible!**
 - Assume same FC Exchange IDs used with both VN_Ports at same time

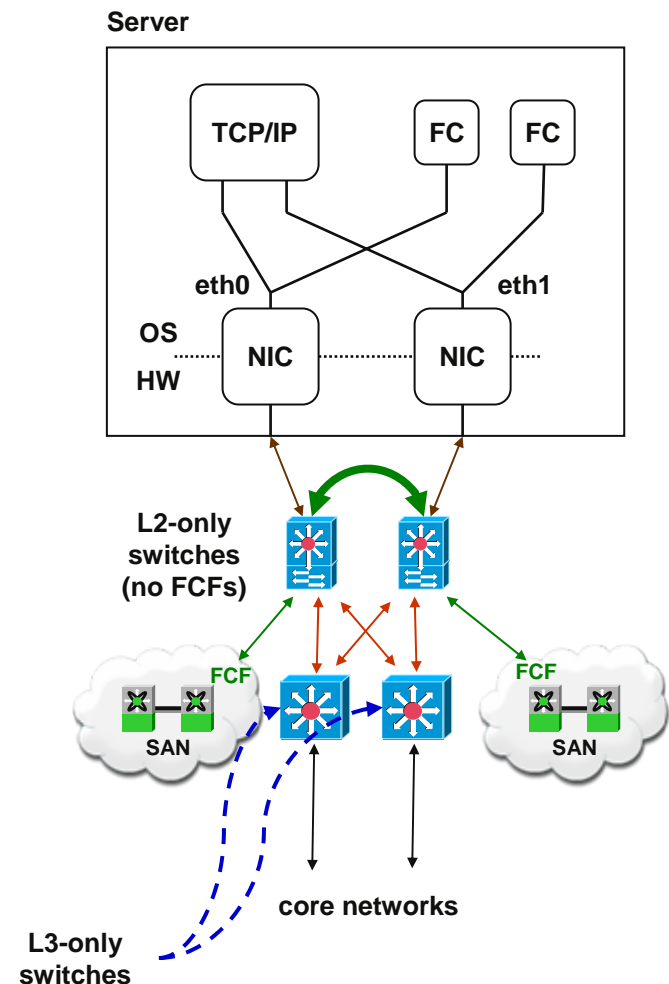


Problem Scenario 1: Diagnosis and Solution Approach

- Switched Fabric FC: N_Port to F_Port link is private
 - Q: Where will this frame arrive?
 - A: At the other end of the link, stupid!
- FCoE: VN_Port to VF_Port link is shared (> 2 ends)
 - Q: Where will this frame arrive?
 - A: At its destination MAC address.
 - Q: And where is that? (More than one instance of MAC address)
 - A: It depends on the L2 switches. Oops ...
- Solution Approach: Restrict traffic to correct destination
 - Ethernet switch ACLs: Restriction enforcement mechanism

Fabric Crosstalk Problem Scenario 2

- **Same Mistake: Cross connect edge switches with VLAN 1 link**
 - Same Result: 2 instances of same FC MAC on new VLAN 1 [FC-MAC-1]
- What happens if one of the top VN_Ports resets and sends FLOGI?
 - FLOGI multicast could go to both FCFs
 - Both VN_Ports may wind up in the same fabric with different FCIDs
- **Loss of multipath redundancy!**
- VRRP doesn't solve this problem
 - Disabling an FCF is catastrophic for other servers using that FCF.
 - Second FC fabric unlikely to handle first FC fabric's traffic correctly.

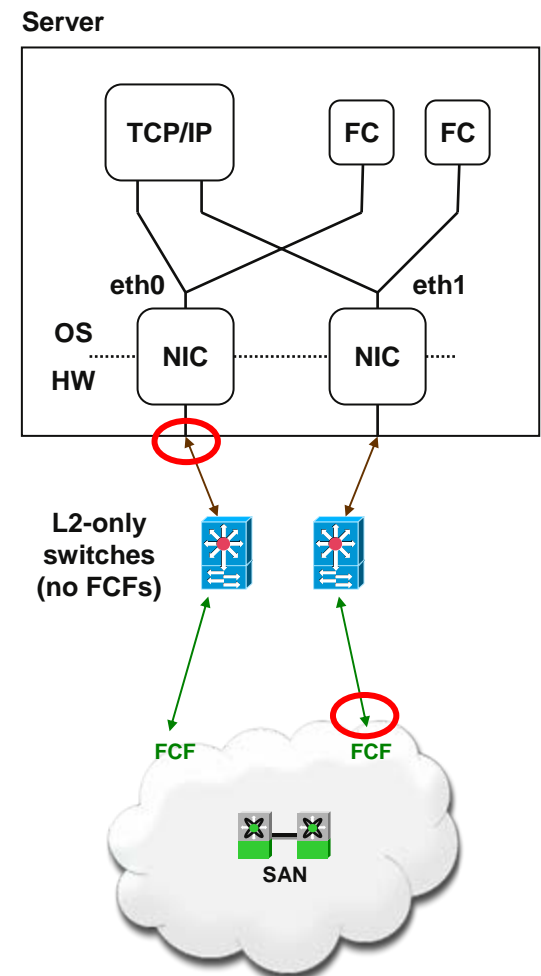


Problem Scenario 2: Diagnosis and Solution Approach

- Switched Fabric FC: N_Port to F_Port link is private
 - Q: What's the F_Port for this N_Port?
 - A: It's the other end of the link, stupid!
- FCoE: VN_Port to VF_Port link is shared (L2 switches)
 - Q: What's the VF_Port for this VN_Port?
 - A: It depends, e.g., on who answers first. Oops ...
- Solution Approach: VN_Port has FC fabric awareness
 - E.g., Don't switch fabrics unless administratively instructed to do so.
 - Baseline: Remember fabric from last login, don't talk to strangers.
- Also consider controls on where FLOGI is sent
 - Limit which fabrics a VN_Port attempts to log into
 - FLOGI has fabric side effects even if LOGO immediately follows

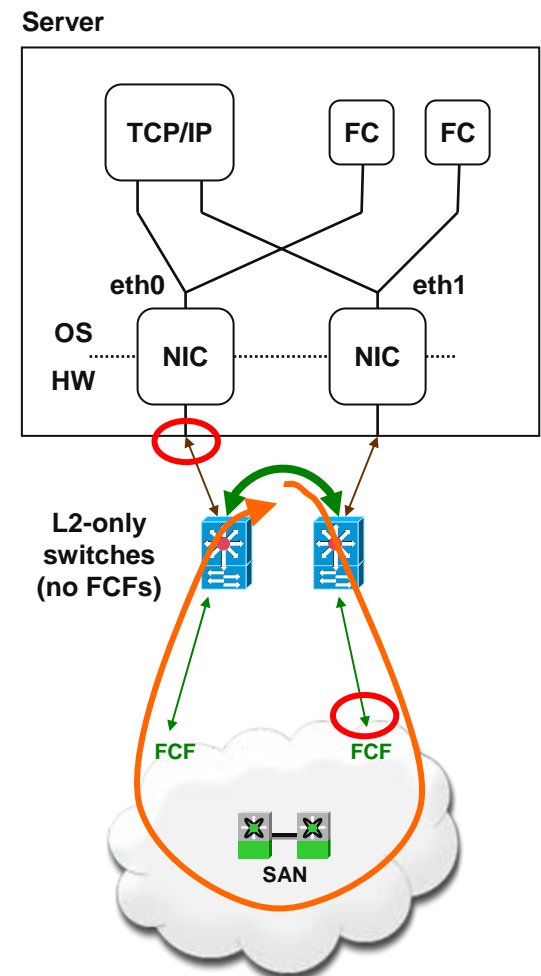
Fabric Crosstalk Problem Setup – Scenario 3

- Edge switches: L2, Ethernet-only
 - No L3 (IP), no FCF
 - FCFs are at edges of FC cloud
- One FC fabric (cloud) this time
- VLAN 1 (left and right instances) for FC: FCID-based MACs
- Get the FCFs involved
 - Suppose left NIC and right FCF have same FCoE MAC
 - Left FCF believes right FCF's MAC is a VN_Port



Fabric Crosstalk Problem – Scenario 3

- **Similar Mistake: Cross connect edge switches with VLAN 1 link**
- Send FC frames to left NIC's FCID
 - Left FCF thinks the right FCF's MAC is for a VN_Port
- Nothing breaks immediately, but
 - Suppose left switch forgets where that MAC is located
 - Right FCF may send frame that “helps” that switch forget
- Left NIC frames now loop forever:
 - Sent to left FCF for NIC VN_Port
 - Left FCF uses VN_Port's MAC
 - Frame arrives at right FCF
 - Destination FCID hasn't changed
 - Fabric forwards to left FCF
 - Lather, rinse, repeat ...



Scenario 3: Diagnosis and Solution Approach

- Shared Ethernet links (L2 switches) strike again:
 - FCoE frame transmission path that's impossible in native FC
 - FCoE can't isolate the logical FC link as native FC would
 - VE_Ports may make this worse
- Goal: Robust solution to make FCoE behave like native FC
 - Drop frames received from wrong type of port, and complain (e.g., log)
- Solution Approach: FCoE frame indicates FC port type of FCoE port that is intended to receive the frame
 - Example: VN_Port frame sent to fabric is received by VF_Port
 - Could use combination of FCoE header and native FC header
 - Example: Class F traffic must be received by a VE_Port
 - Also identifies VN_Port to VN_Port direct traffic (not via VF_Port)
- Alternative (poor): Forbid use of dynamic MACs for VF_Ports
 - Hard to enforce - which OUIs are banned and why?
 - Prohibits software FCFs in hypervisors and virtualized OS guests
 - FCF configuration mistake (left FCF) can still cause forwarding loop

Solution Approach Summary

- Scenario 1: Misdirected FCoE frames
 - Solution Approach: Restrict traffic to correct destination
 - Ethernet switch ACLs appear promising
- Scenario 2: Misdirected FCoE Fabric Login
 - Solution Approach: VN_Port has FC fabric awareness
 - Baseline: Remember fabric from last login, don't talk to strangers
 - Also consider controlling where FLOGI is sent
 - FLOGI has fabric side effects even if LOGO immediately follows
- Scenario 3: FCoE-created Forwarding Loop
 - Solution Approach: FCoE frame indicates FC port type of FCoE port that is supposed to receive the frame
 - Example: VN_Port frame sent to fabric is received by VF_Port
 - Could use combination of FCoE header and native FC header

EMC²[®]

where information lives[®]