



*Interoperability
through
Simplicity*

Controlling Switch Redundancy Protocol

Claudio DeSanti
T11/11-227v0

June 2011

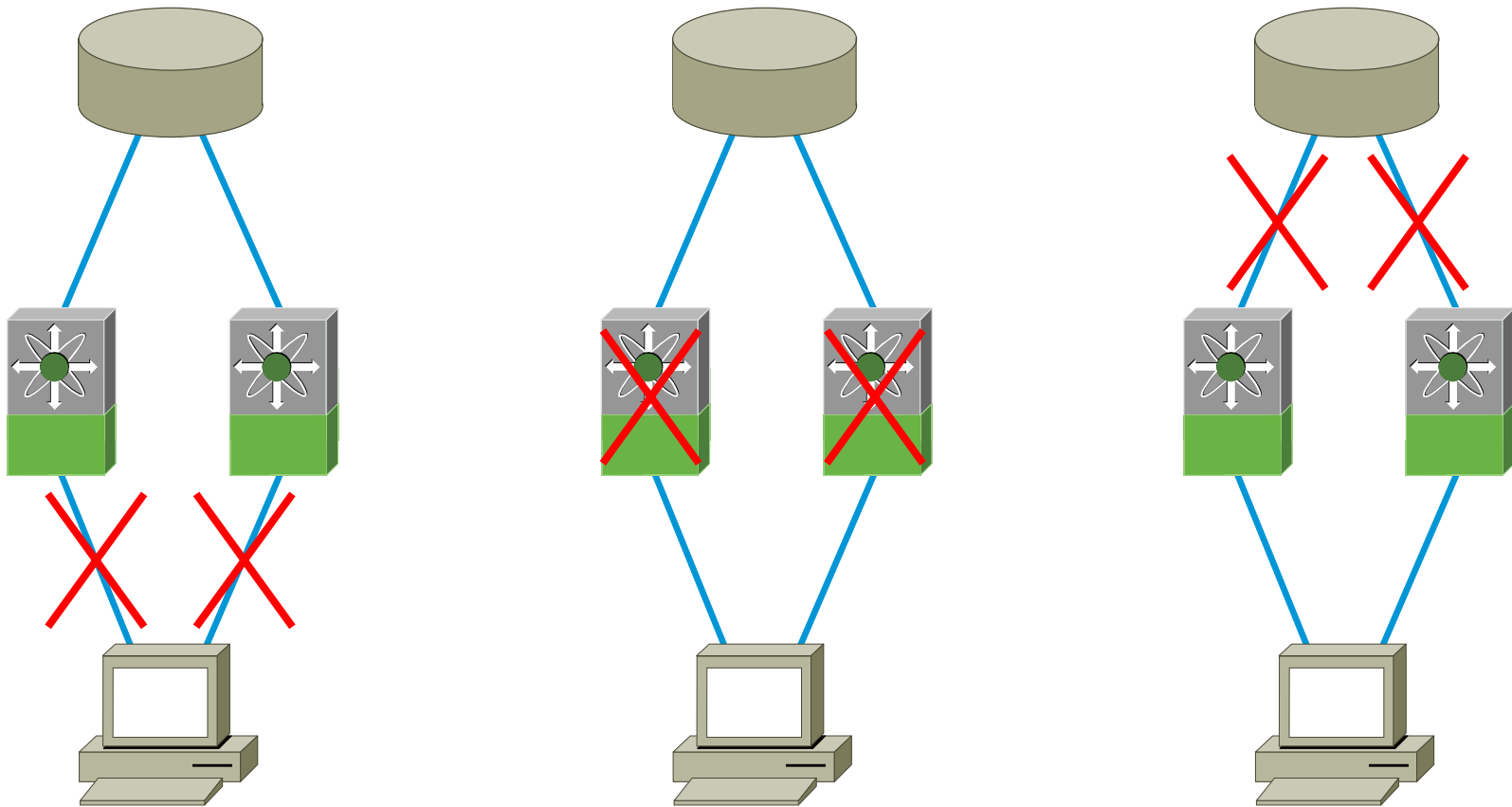
Agenda

- Overview
- State Machine
- Redundancy Cases

Redundancy Protocol

- Inspired by a successful protocol: VRRP
Virtual Router Redundancy Protocol, RFC 5798
- Goals:
 - Avoid any single point of failure in a Distributed Switch
 - Be simple
- Non goal:
 - Avoid double points of failure

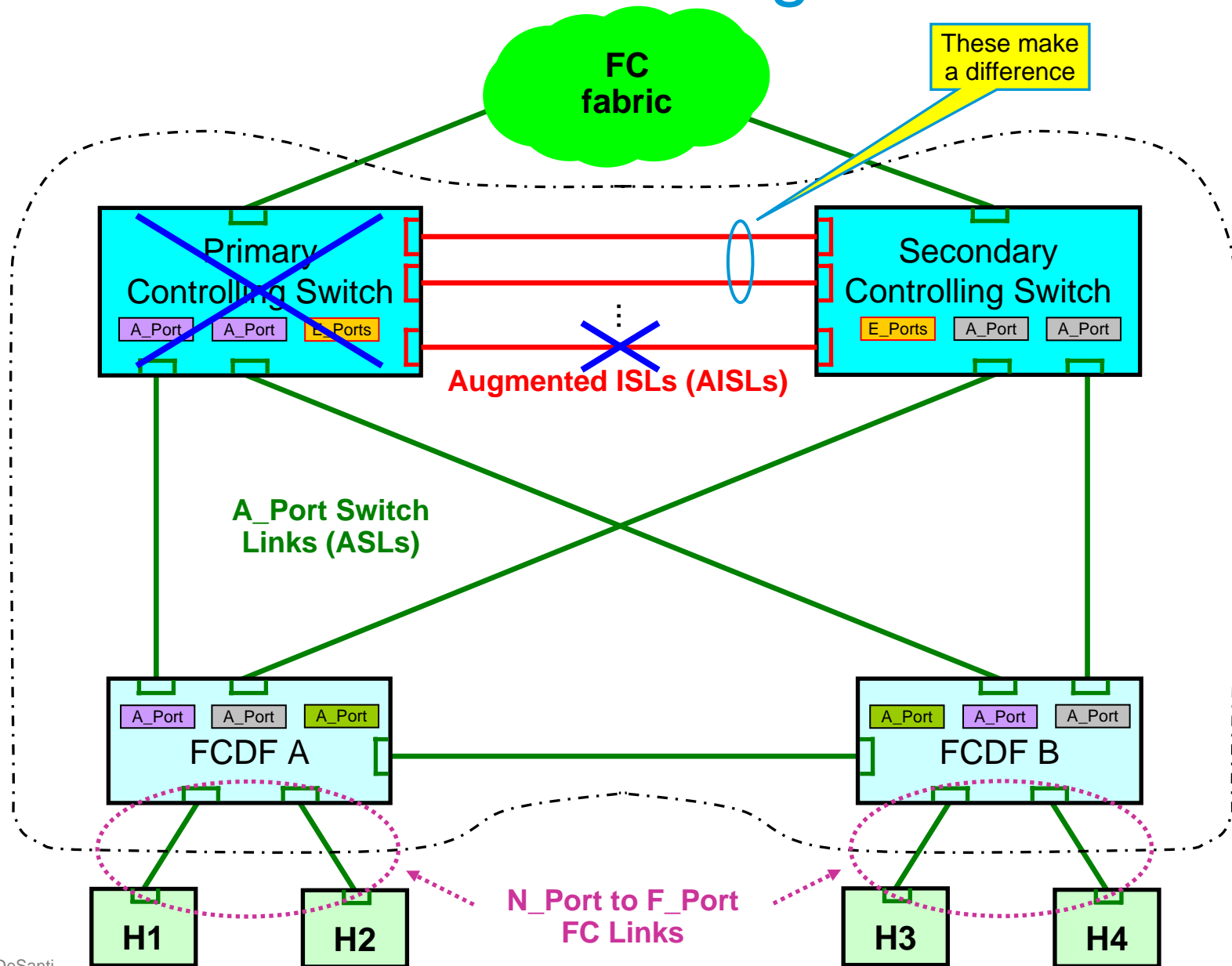
Double Failures with Current FC



Fundamental Issue

- How to distinguish a Controlling Switch failure from an Augmented ISL (AISL) failure?
- Solution: use at least two AISLs

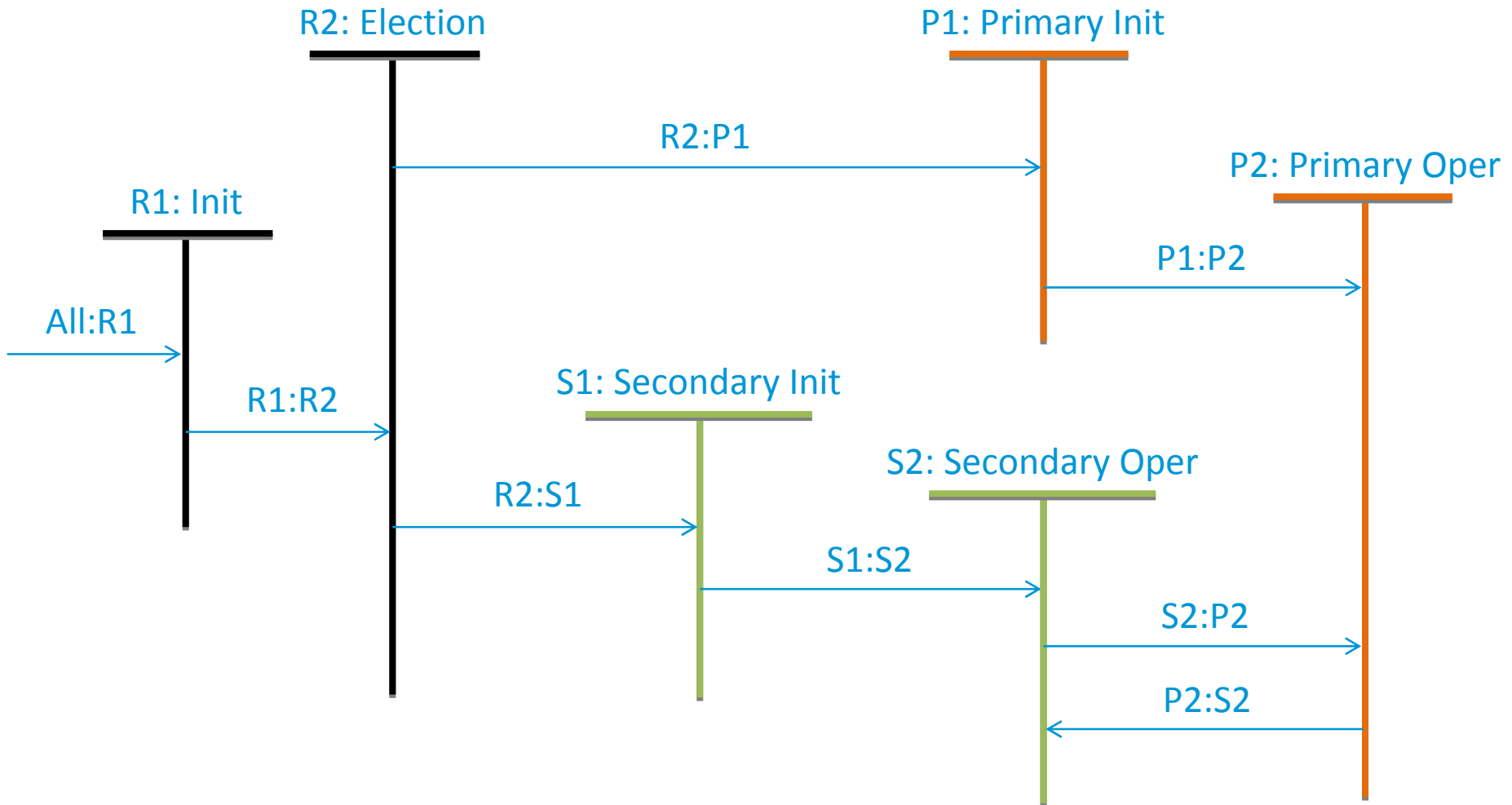
Link Failure vs. Controlling Switch failure



Agenda

- Overview
- State Machine
- Redundancy Cases

Redundancy Protocol State Machine



Controlling Switch Priority

- Used in the Primary/Secondary election process

Value	Description
00h	Reserved
01h	Highest Priority value. This value is administratively configured to force the election of a Controlling Switch to Primary.
02h ^a	Primary Controlling Switch priority. This value is used by the Redundancy protocol to identify a Controlling Switch as Primary.
03 .. FEh	Higher to lower Priority values. The default value is 128.
FFh ^a	This value indicates that a Controlling Switch is not willing to operate as Primary. This is used by the Primary Controlling Switch to trigger a transition of the Secondary Controlling Switch to Primary without having to wait for the current Primary to timeout, if appropriate.
^a These values are used by the Redundancy protocol and not available to an administrator.	

- A table that should look familiar to FC-SW-x people... 😊

Primary/Secondary Election

- The Controlling Switch sends unidirectional RHellos every RHello_Interval over each of its AISLs and waits until:
 - Down_Timer expires; or
 - An RHello is received
- If Down_Timer expires without having received RHellos, the Controlling Switch becomes the Primary
- If an RHello is received:
 - If local Priority < remote Priority, the Controlling Switch becomes the Primary
 - If local Priority > remote Priority, the Controlling Switch becomes the Secondary
 - If local Priority = remote Priority, the Switch_Name determines who is what

Primary Behavior

- State P1:Initialization
 - Sets its Priority to 02h
 - Obtain a Virtual Domain_ID from the Principal Switch
- State P2:Operational
 - Sets its Priority to 02h
 - Sends a DFMD SW_ILS to all reachable FCDFs declaring itself as Primary
 - Establishes VA_Port to VA_Port (Virtual) link
 - On both FC and FCoE interfaces, if any
 - Performs the VA_Port Protocols
 - Sends RHellos every RHello_Interval over each of its AISLs
 - Resets the Down_Timer to Down_Interval everytime an RHello is received
 - When an SSA SW_ILS is received (i.e., the Secondary completed its synchronization), sends a DFMD SW_ILS to all reachable FCDFs declaring itself as Primary and the Secondary as Secondary
 - When Down_Timer expires (i.e., when the Secondary is not anymore available) sends a DFMD SW_ILS to all reachable FCDFs declaring itself as Primary

Secondary Behavior

- State S1:Initialization

Synchronizes its state with the one of the Primary

1. Requests to the Primary the Virtual Domain_ID and the FCDF topology through the GFSS (Get FCDF Set Status) SW_ILS
2. Requests to the Primary the N_Port_ID Allocation state in the Distributed Switch through the GNAS (Get N_Port_ID Allocation State) SW_ILS
3. Obtains the information associated with each N_Port_ID in the Name Server through the GE_ID CT Request
4. Communicates the achieved state synchronization to the Primary through the SSA (Secondary Synchronization Achieved) SW_ILS

- State S2:Operational

Sets its Priority to its configured value

Establishes VA_Port to VA_Port (Virtual) link

On both FC and FCoE interfaces, if any

Performs the VA_Port Protocols

Sends RHellos every RHello_Interval over each of its AISLs

Resets the Down_Timer to Down_Interval everytime an RHello is received

Operational Transitions

- Transition S2:P2. The Secondary becomes Primary

The Down_Timer expires (i.e., no RHellos have been received over any AISL for Down_Interval). This indicates the failure of the Primary; or

The Priority field in the received RHellos has a value of FFh. This is an indication that the Primary determined to become Secondary.

- Transition P2:S2. The Primary determines to become Secondary

By setting its Priority to FFh

This may happen when the Primary determines that the number of FCDFs reachable from the Secondary is greater than the number of FCDFs reachable from the Primary (e.g., after the failure of an ASL with the Primary)

SW_ILS Payloads (1)

Table 2 – RHello Request Payload

Item	Size (bytes)
SW_ILS Code	4
Originating Controlling Switch Switch_Name	8
RHello_Interval	4
Reserved	3
Originating Controlling Switch Priority	1

Table 3 – GFSS Request Payload

Item	Size (bytes)
SW_ILS Code	4
Originating Controlling Switch Switch_Name	8

Table 4 – GFSS SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Originating Controlling Switch Switch_Name	8
Reserved	3
Virtual Domain_ID Value	1
Distributed Switch Switch_Name	8
Number of FCDF Connectivity Records (n)	4
FCDF Connectivity Record #1	
FCDF Connectivity Record #2	
...	
FCDF Connectivity Record #n	

Table 5 – FCDF Connectivity Record Format

Item	Size (bytes)
FCDF Switch_Name	8
Number of ASL Neighbors (m)	4
Switch_Name of Neighbor #1	8
ASL cost to Neighbor #1	4
Switch_Name of Neighbor #2	8
ASL cost to Neighbor #2	4
...	
Switch_Name of Neighbor #m	8
ASL cost to Neighbor #m	4

SW_ILS Payloads (2)

Table 6 – GNAS Request Payload

Item	Size (bytes)
SW_ILS Code	4
Originating Controlling Switch Switch_Name	8
FCDF Switch_Name	8

Table 7 – GNAS SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Originating Controlling Switch Switch_Name	8
FCDF Switch_Name	8
Number of Allocated N_Port_IDs (q)	4
Allocated N_Port_ID #1	4
Allocated N_Port_ID #2	4
...	
Allocated N_Port_ID #q	4

Table 8 – SSA Request Payload

Item	Size (bytes)
SW_ILS Code	4
Originating Controlling Switch Switch_Name	8

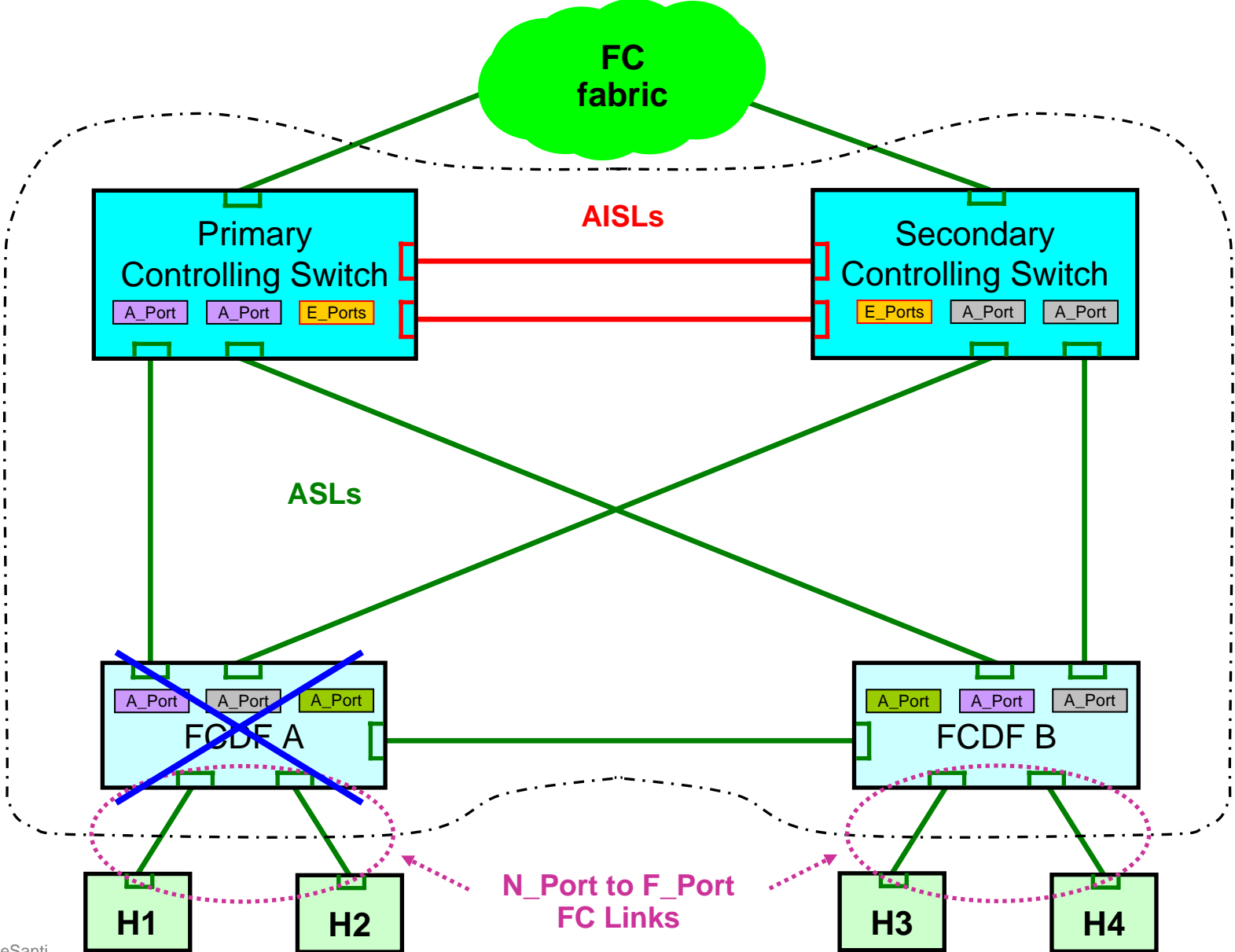
Table 9 – SSA SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Originating Controlling Switch Switch_Name	8

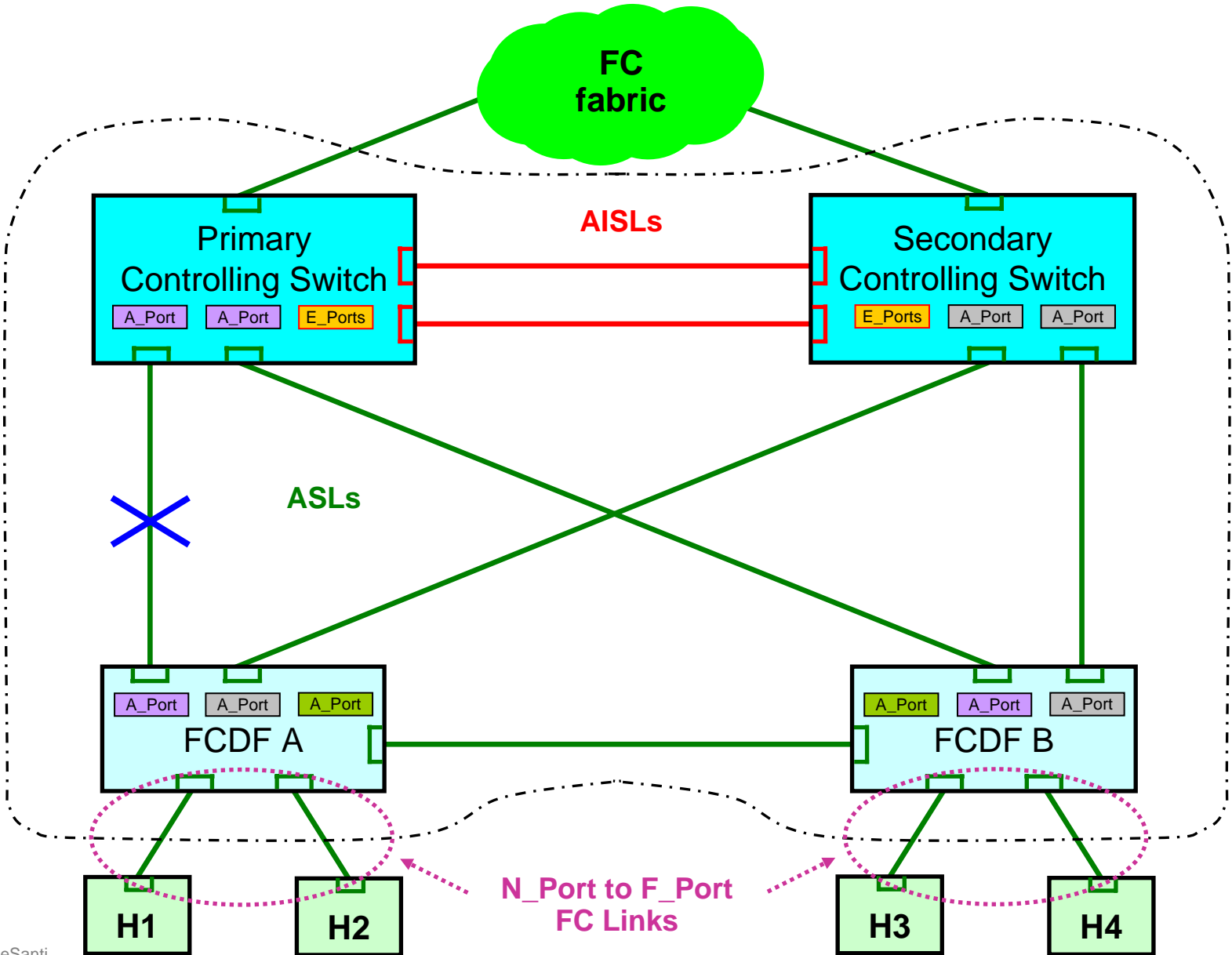
Agenda

- Overview
- State Machine
- Redundancy Cases

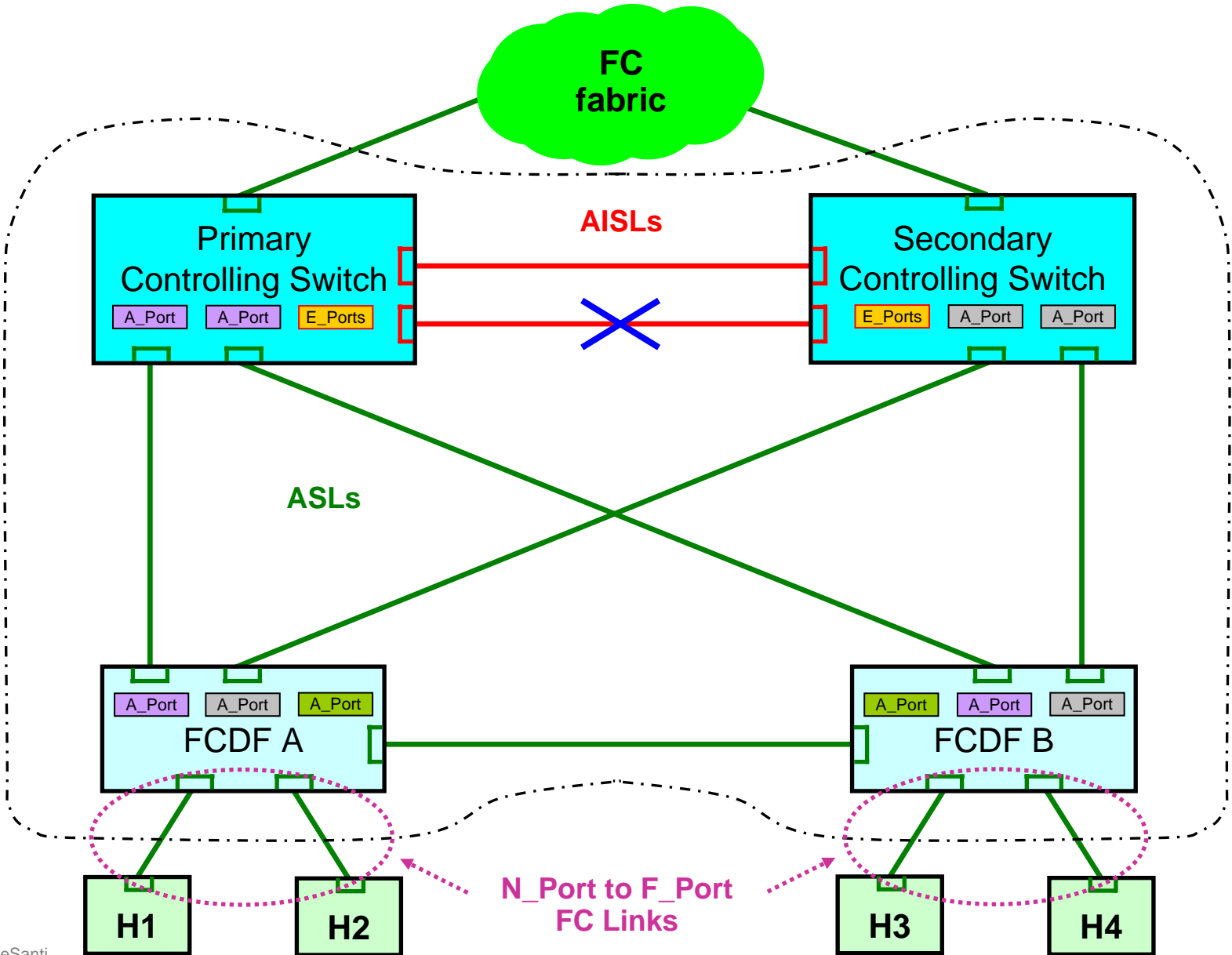
Distributed Switch Reliability (1)



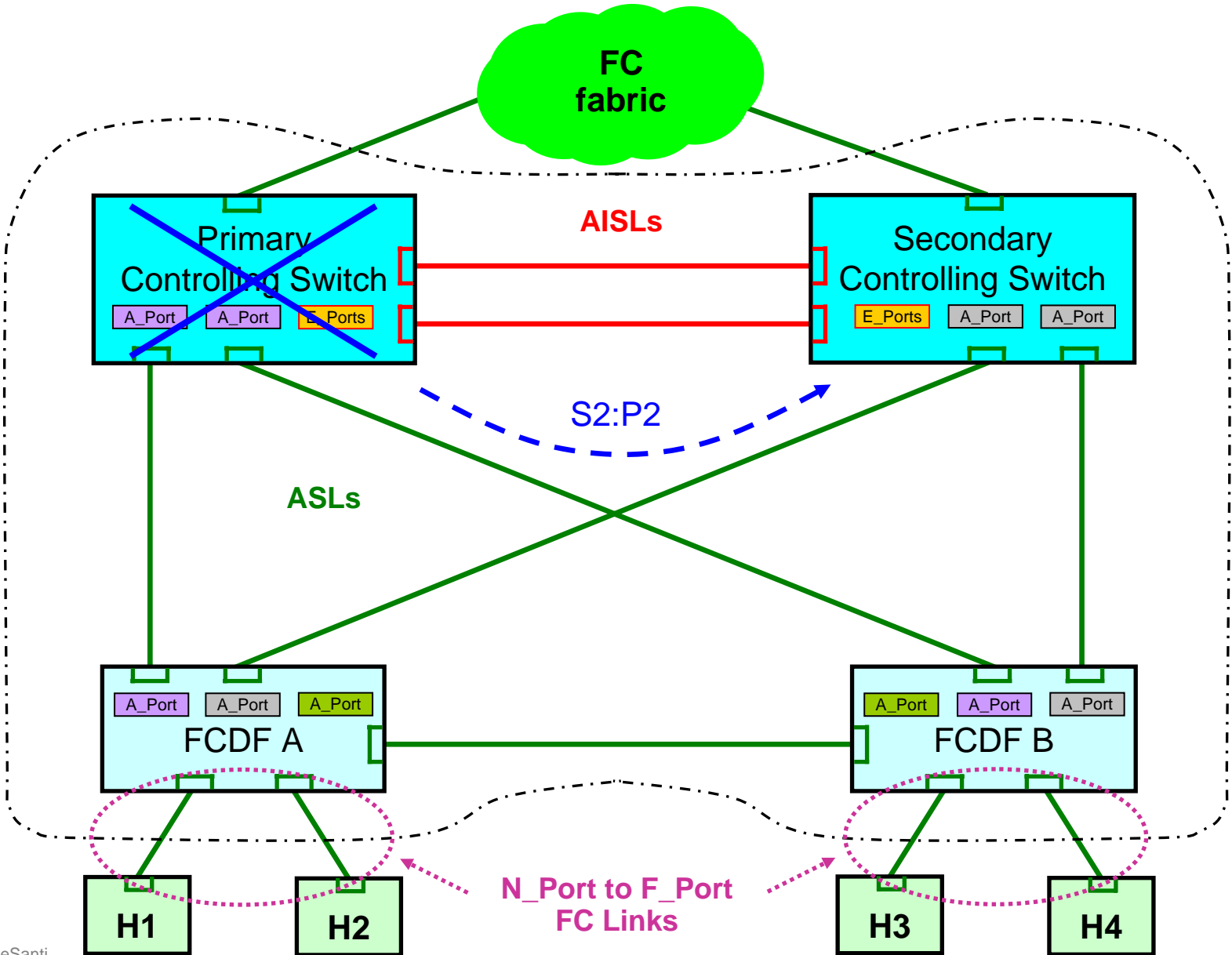
Distributed Switch Reliability (2)



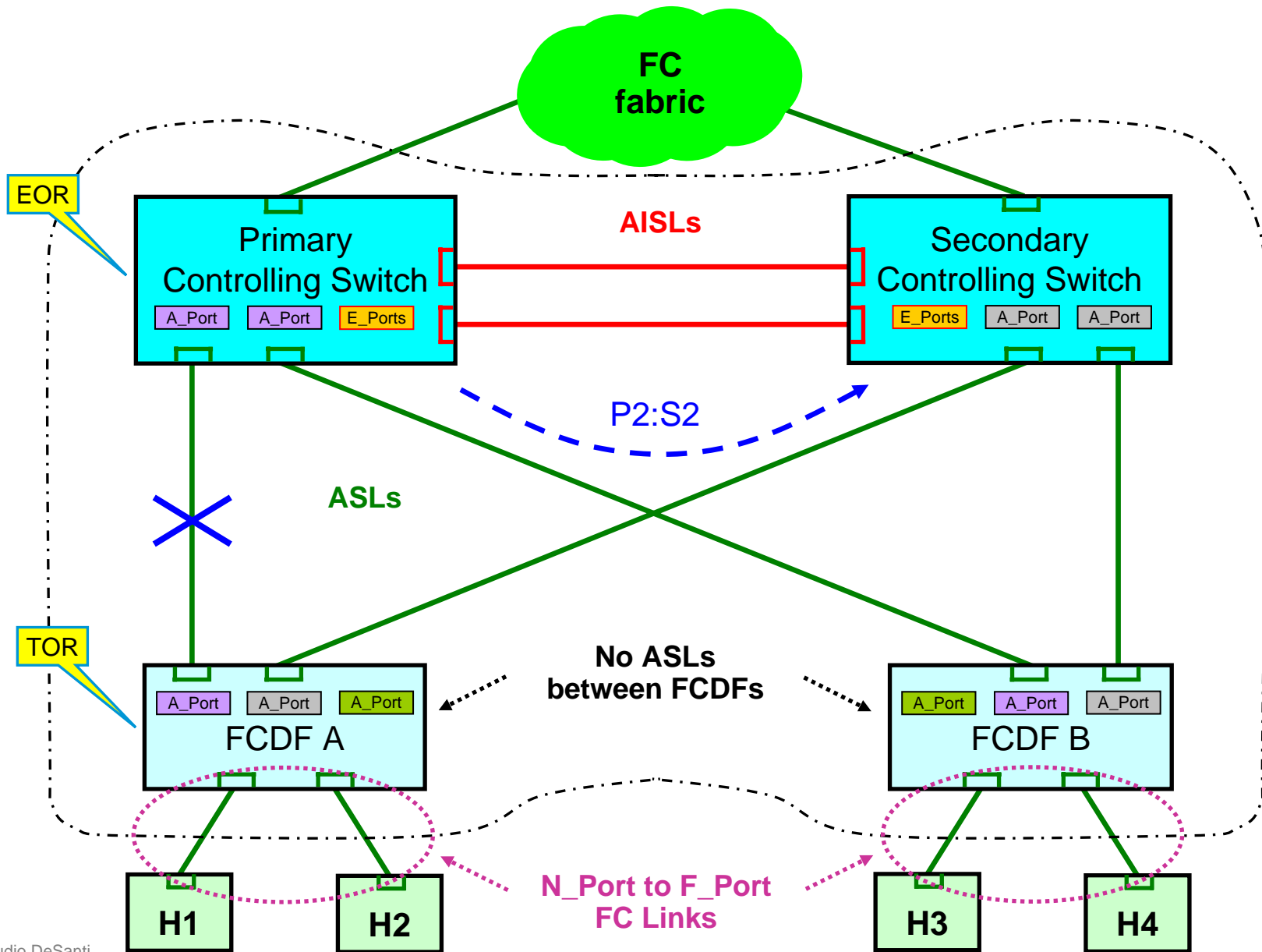
Distributed Switch Reliability (3)



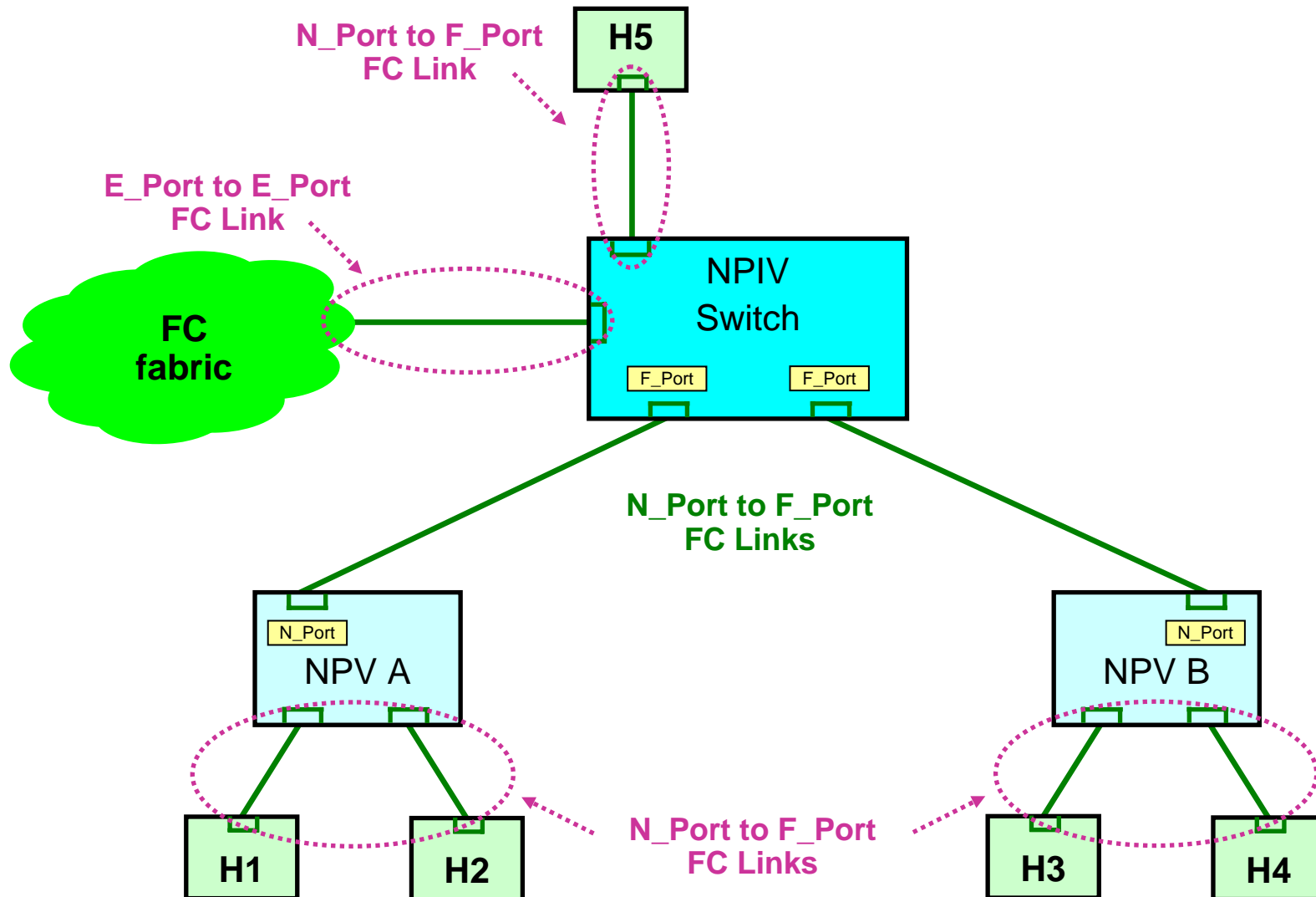
Distributed Switch Reliability (4)



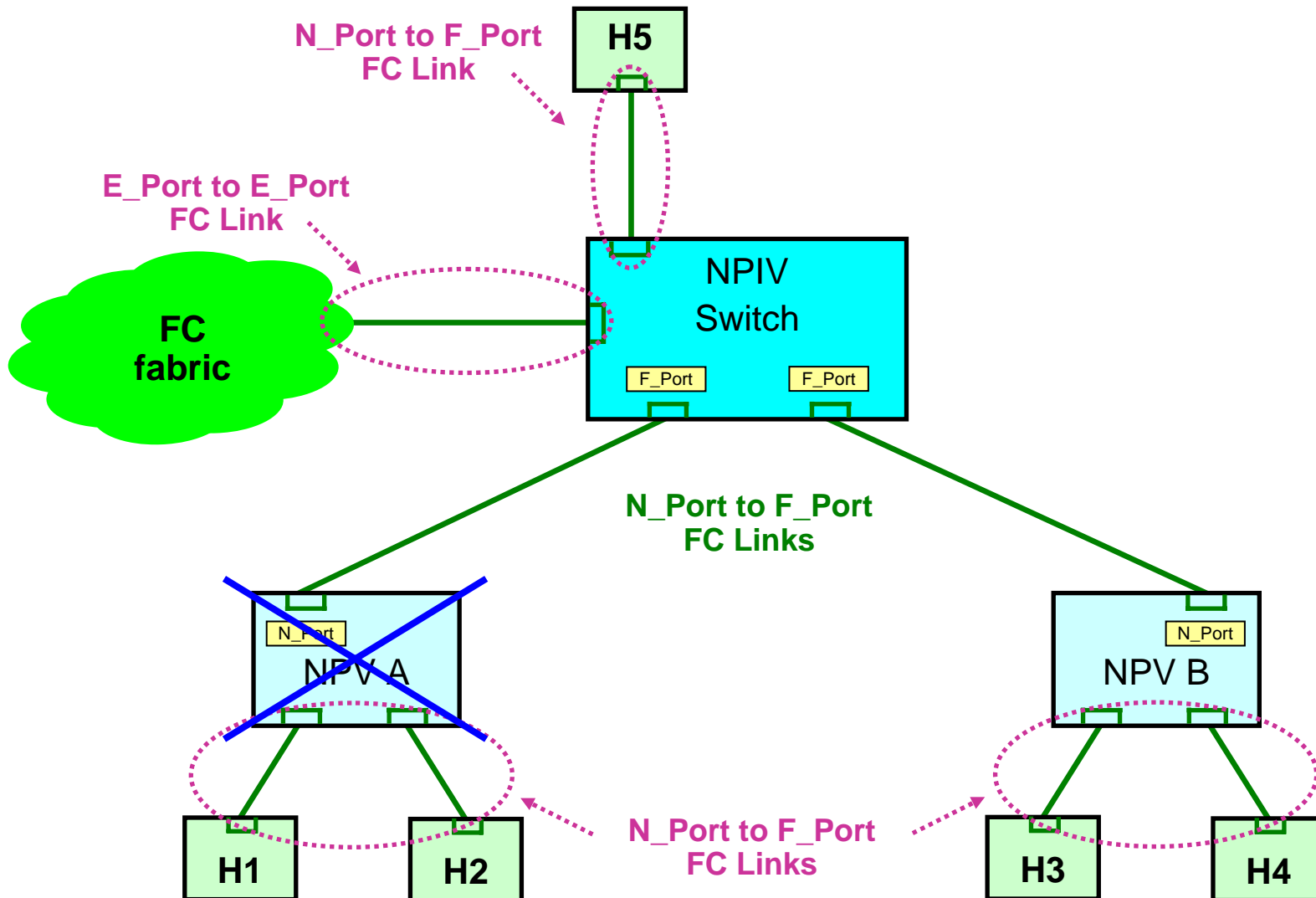
A Relevant Case



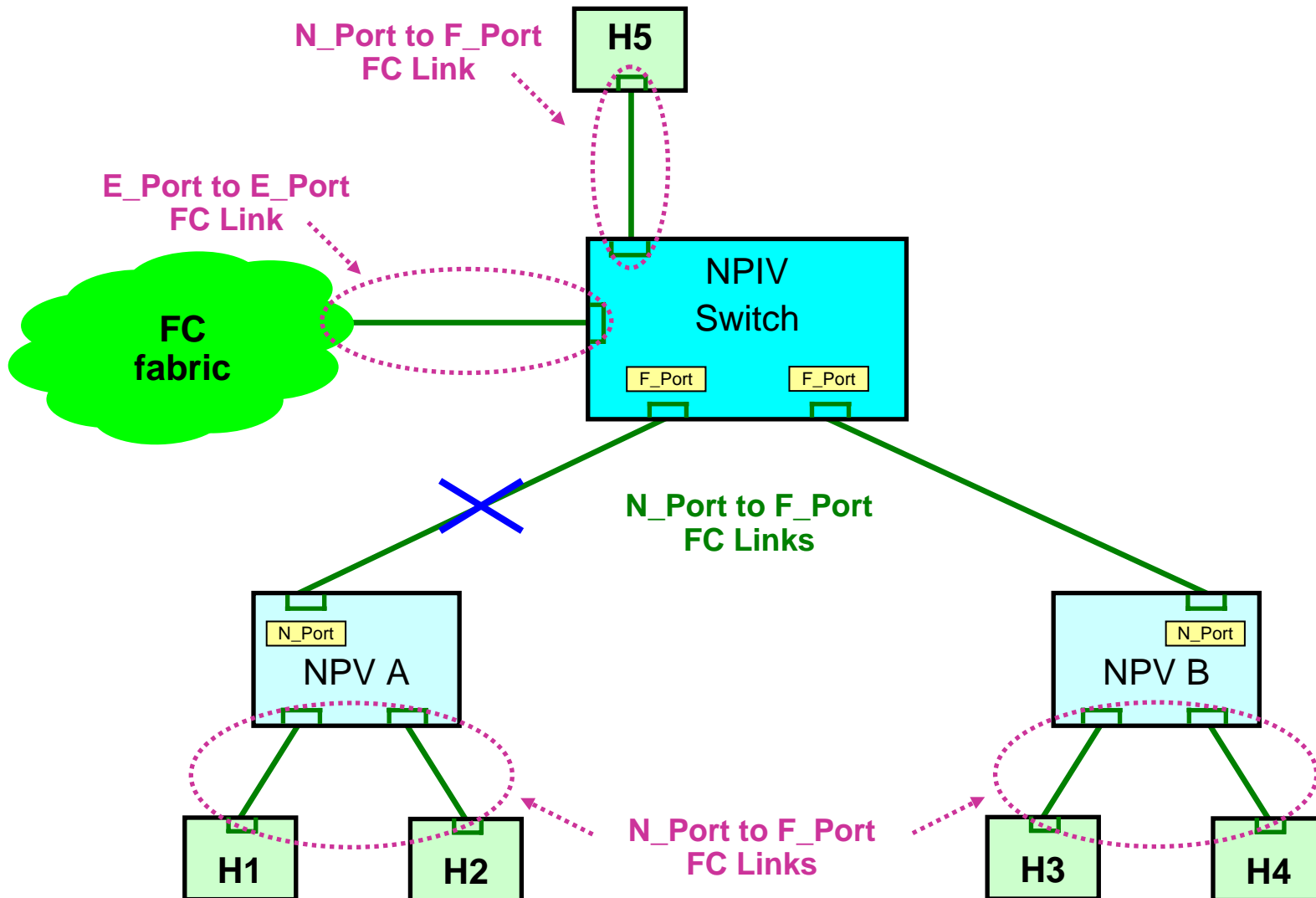
Existing Devices: N_Port Virtualizers



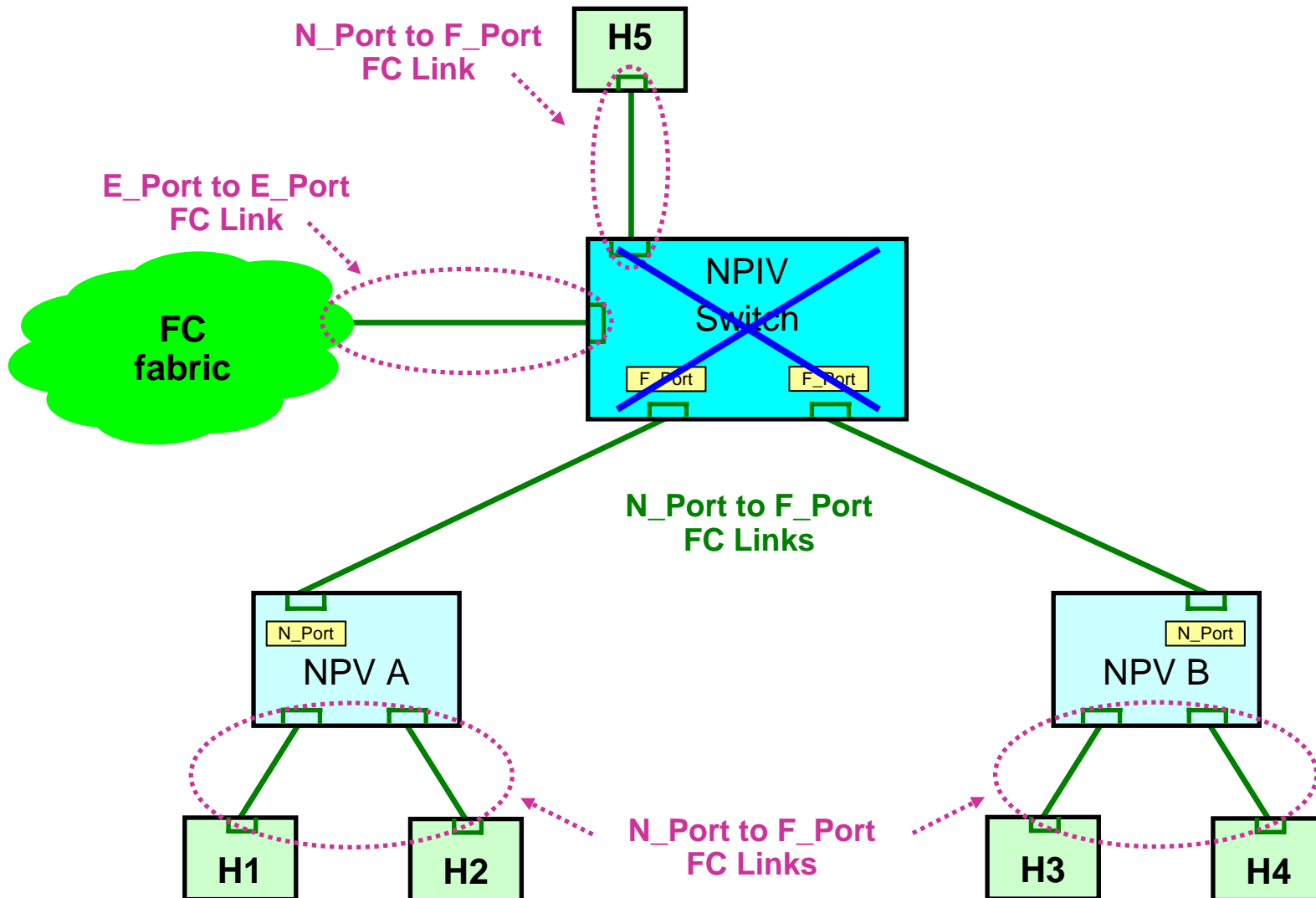
N_Port Virtualizer Reliability (1)



N_Port Virtualizer Reliability (2)



N_Port Virtualizer Reliability (3)



Summary

- A simple redundancy protocol between Primary and Secondary Controlling Switches/FCFs

It is a Fibre Channel protocol

As such, it applies to both FC and FCoE

Just replace the physical FC links with FCoE Virtual Links

- A Distributed Switch is more reliable and scalable than an N_Port Virtualizer

More reliable because of the Redundancy protocol

More scalable because it supports:

Cascaded configurations

Distribution of Zoning enforcement

*Interoperability
through
Simplicity*

Thank you!

