Customer Deployment Examples for FICON Technologies

Mike Blair – mblair@cisco.com
Howard Johnson - hjohnson@Brocade.COM
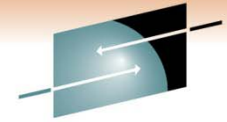Lou Ricci - lricci@us.ibm.com

2 March 2011 (4:30pm – 5:30pm)
Session Number 8990
Room 211B

Abstract

- Extending a cascaded storage area network over long distances requires specific technologies suited for the purpose. Fibre Channel over Internet Protocol (FCIP) is the fundamental technology used to drive these solutions. This technology allows FICON to be extended in order to meet the disaster recovery and business continuance needs of today's enterprise environments. In this session, we explore several customer deployments using FCIP and FCIP based products. Please join industry experts from Brocade, Cisco, and IBM as we delve into the nuances of this critical technology.
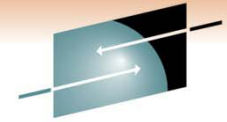
Agenda

- **FCIP Overview**
- FCIP Topologies
- Cisco FCIP – explanations
  - Customer 1 – Prioritized workload with QOS ~ 1000 miles
  - Customer 2 – STK extension over 1000 miles
  - Customer 3 – Overcome DWDM switchover issue – 80km
- Brocade FCIP – explanations
  - Customer 1 – Extension for asynchronous replication
  - Customer 2 – XRC emulation for remote vaulting
  - Customer 3 – Extension solution differences
- IBM comments
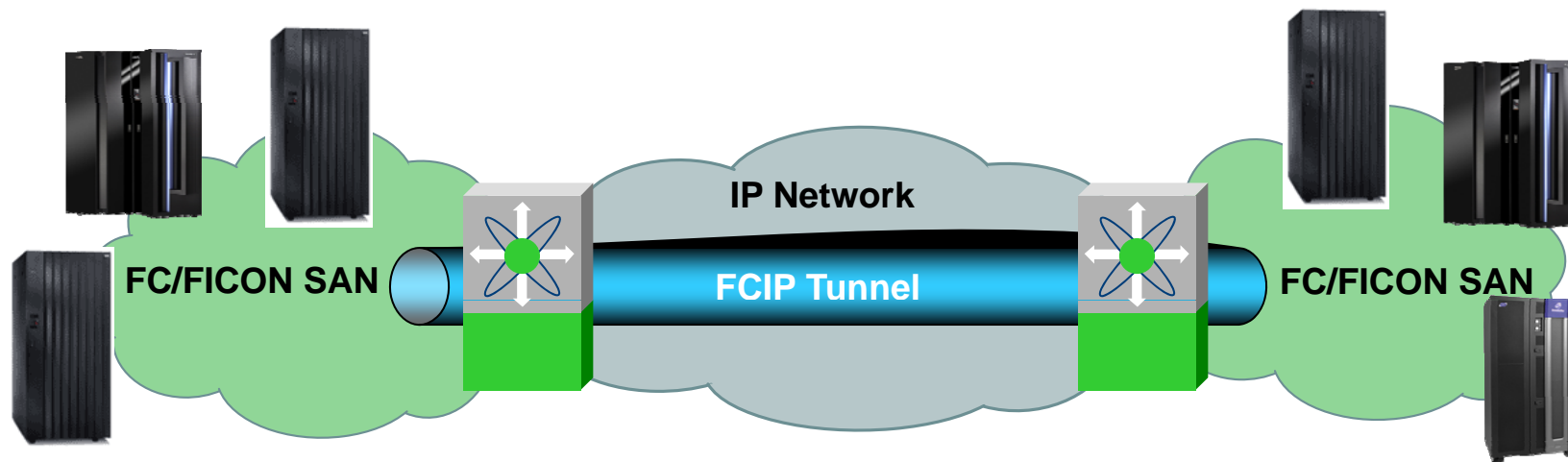 - System z Requirements
- Q & A

FCIP Overview
Fibre Channel over IP

IP Network

FCIP Tunnel

FC/FICON SAN

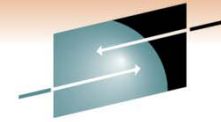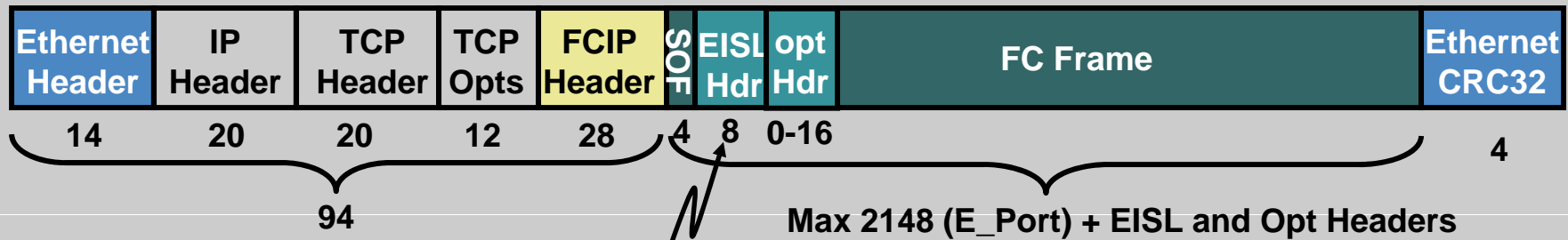FC/FICON SAN

FCIP Is a Standard from the IETF IP Storage WG for Linking FibreChannel SANs over IP (RFCs 3821 and 3643)

- Point-to-point tunnel between FCIP link end-points
- Appears as one logical FC fabric with single FSPF routing domain

FCIP Overview
FCIP Frame Detail

Ethernet Header	IP Header	TCP Header	TCP Opts	FCIP Header	SOF	EISL Hdr	opt Hdr	FC Frame	Ethernet CRC32
14	20	20	12	28	4	8	0-16		4

94

Max 2148 (E_Port) + EISL and Opt Headers

FCIP Overhead for Ethernet Frames:

94 Byte Header + 4 Byte CRC = 98 Bytes

EISL and Optional Headers

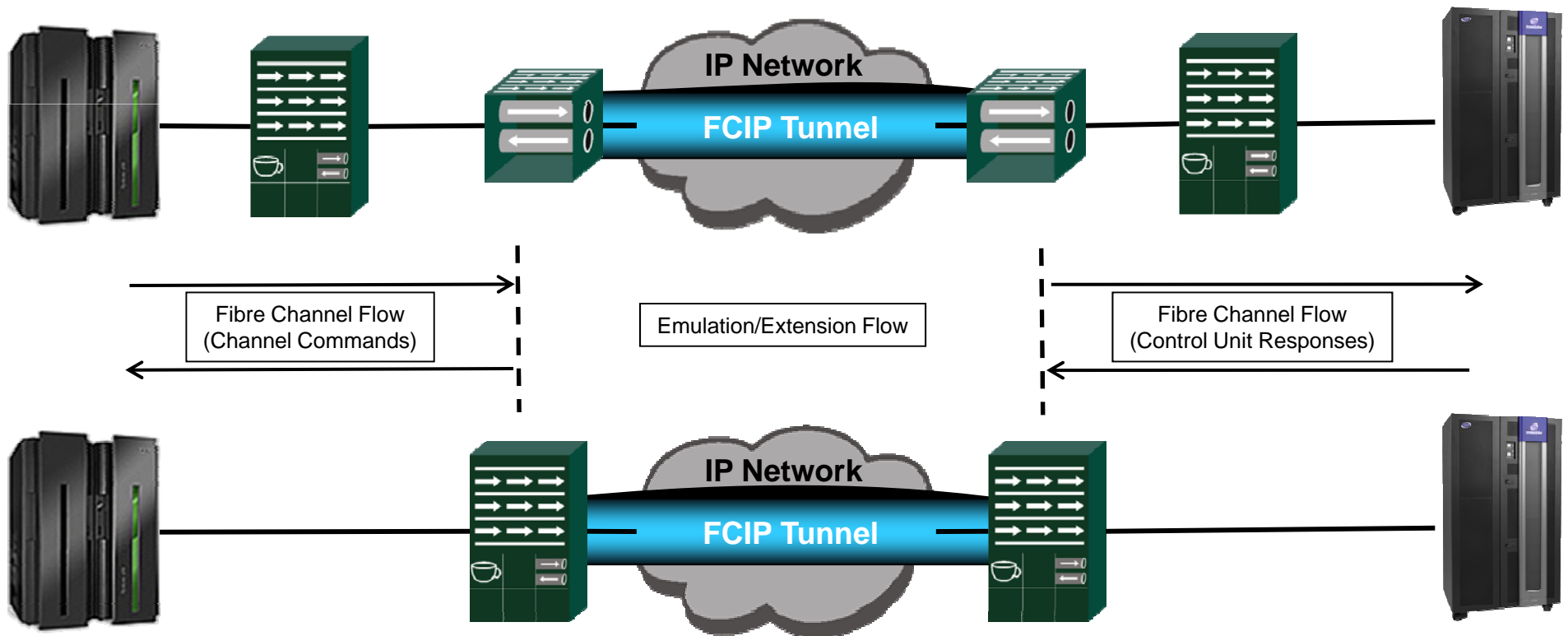**If TE_Port, then 8 Bytes Added to FC Frame (after SOF) for VSAN Routing**

- Max FibreChannel frame is 2148 bytes plus optional extras
- FC frames are segmented and reassembled if MTU too small (TCP payload on second or subsequent packets)
- Jumbo frames may increase performance
  - IP MTU of 2300 avoids splitting of TCP frames

5

Agenda

- FCIP Overview
- **FCIP Topologies**
- Cisco FCIP – explanations
  - Customer 1 – Prioritized workload with QOS ~ 1000 miles
  - Customer 2 – STK extension over 1000 miles
  - Customer 3 – Overcome DWDM switchover issue – 80km
- Brocade FCIP – explanations
  - Customer 1 – Extension for asynchronous replication
  - Customer 2 – XRC emulation for remote vaulting
  - Customer 3 – Extension solution differences
- IBM comments
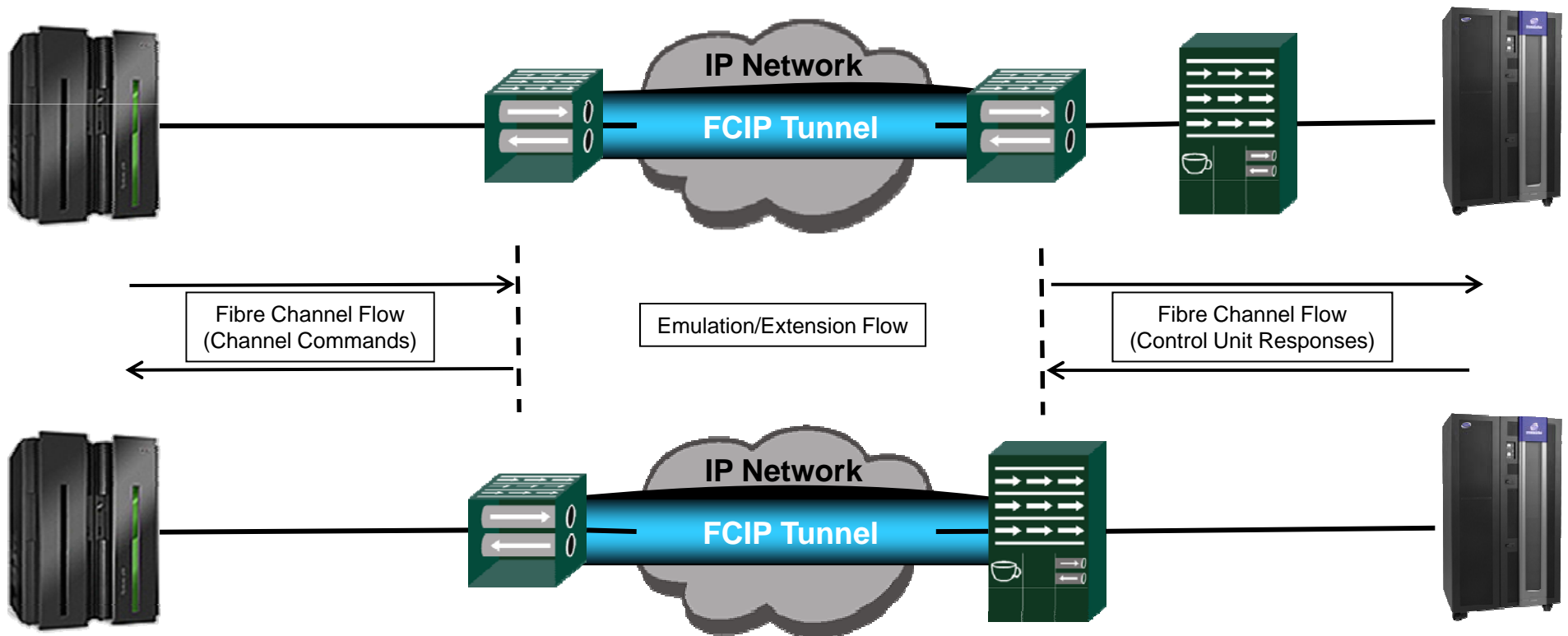 - System z Requirements
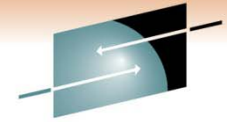- Q & A

FCIP Topologies
ISL Link Extension

IP Network

FCIP Tunnel

Fibre Channel Flow
(Channel Commands)

Emulation/Extension Flow

Fibre Channel Flow
(Control Unit Responses)

IP Network

FCIP Tunnel

FCIP Topologies
Channel Link Extension

IP Network

FCIP Tunnel

Fibre Channel Flow
(Channel Commands)

Emulation/Extension Flow

Fibre Channel Flow
(Control Unit Responses)

IP Network

FCIP Tunnel

FCIP Topologies
Control Unit Link Extension

Fibre Channel Flow
(Channel Commands)

Emulation/Extension Flow

Fibre Channel Flow
(Control Unit Responses)

IP Network

FCIP Tunnel

IP Network

FCIP Tunnel
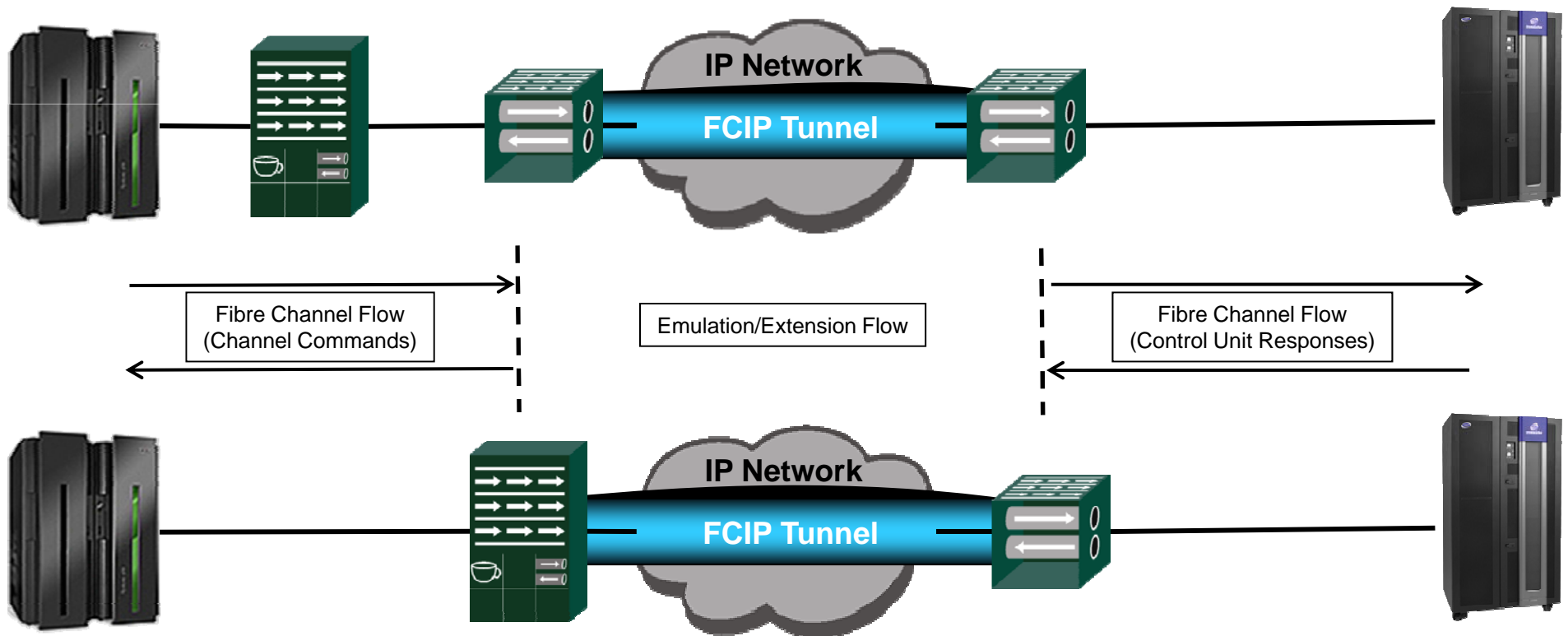
FCIP Topologies
Point to Point Link Extension
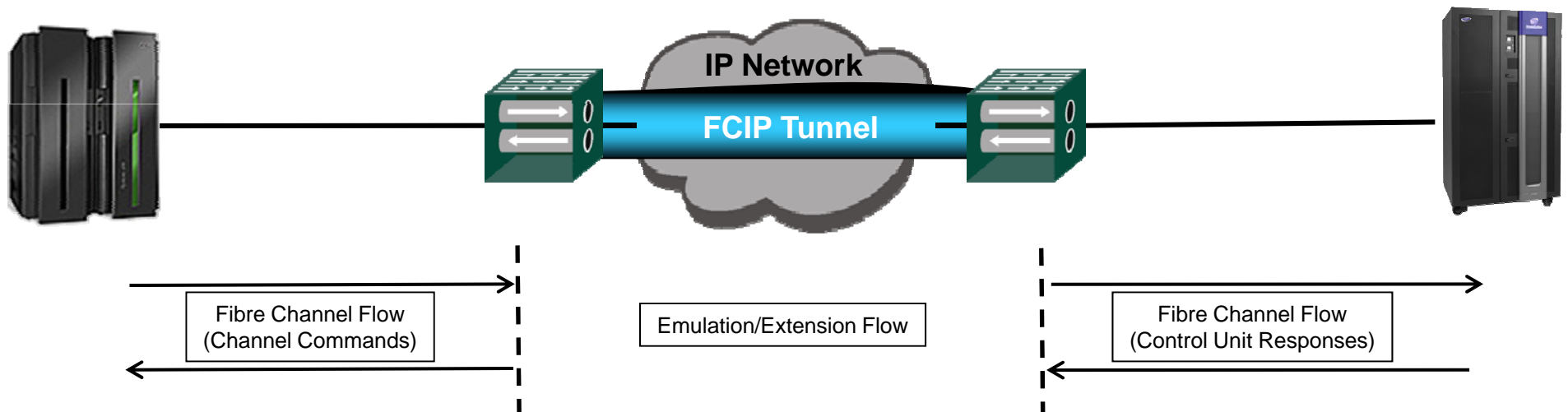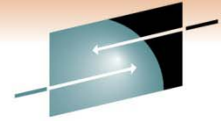
IP Network

FCIP Tunnel

Fibre Channel Flow
(Channel Commands)

Emulation/Extension Flow

Fibre Channel Flow
(Control Unit Responses)

FCIP Topologies
ISL Link Extension with Storage Replication

Fibre Channel Flow
(Channel Commands)

Emulation/Extension Flow

Fibre Channel Flow
(Control Unit Responses)

FICON Traffic

IP Network

FCIP Tunnel

FCP Traffic
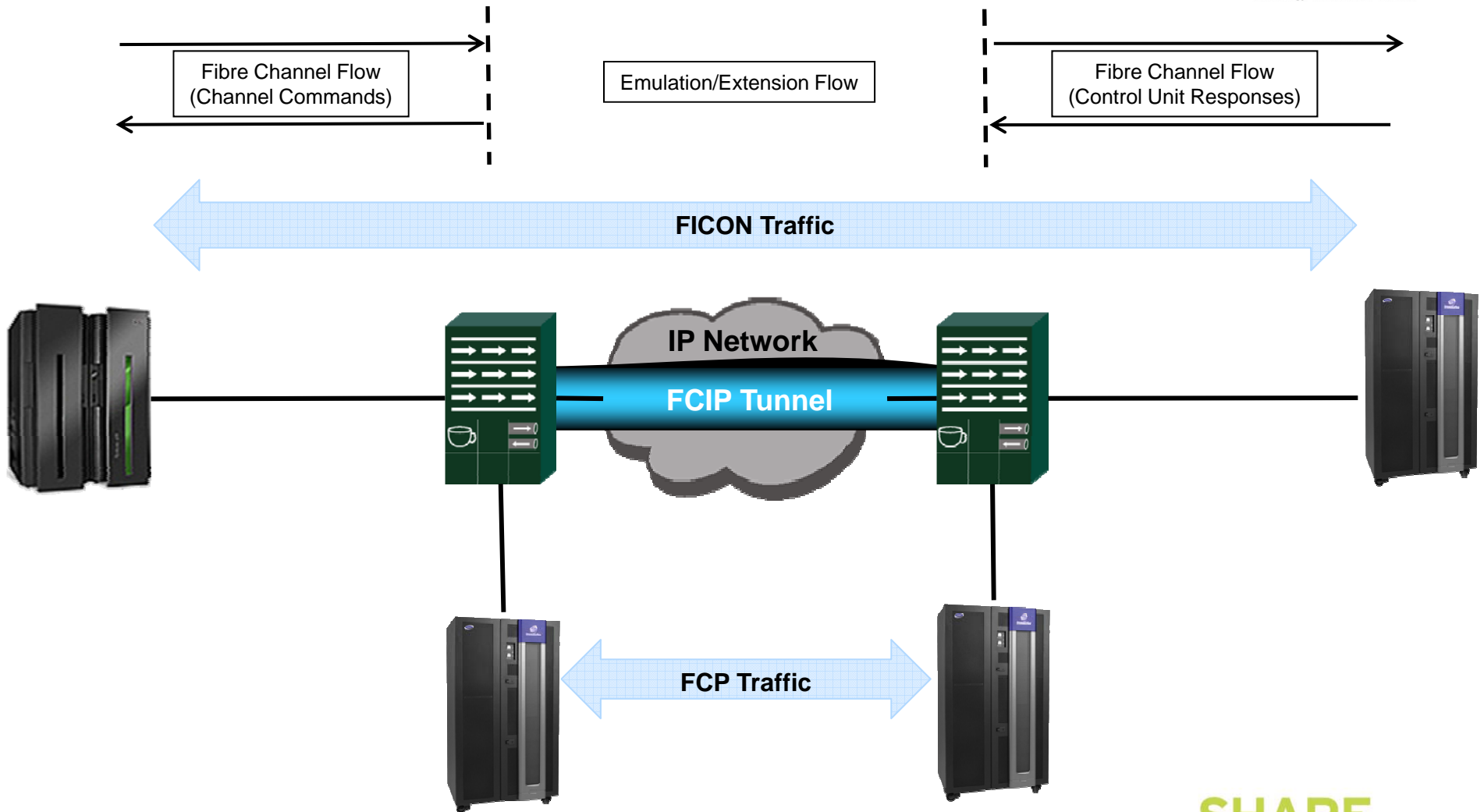
FCIP Topologies
Channel / Control Unit Link Extension with Storage Replication
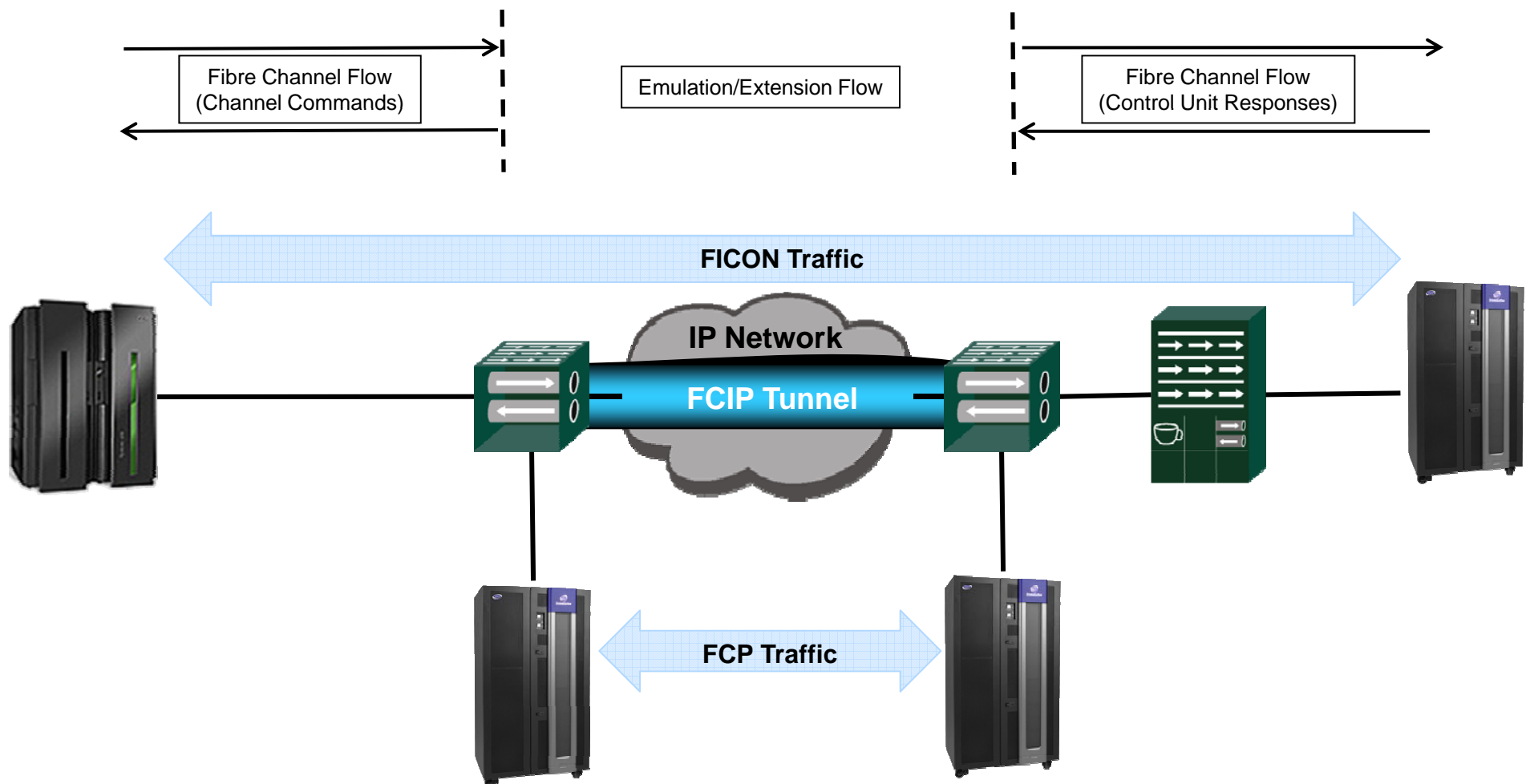
Fibre Channel Flow (Channel Commands)

Emulation/Extension Flow

Fibre Channel Flow (Control Unit Responses)

FICON Traffic

IP Network

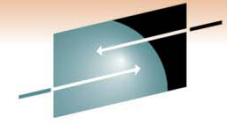FCIP Tunnel

FCP Traffic

Agenda

- FCIP Overview
- FCIP Topologies
- **Cisco FCIP – explanations**
  - Customer 1 – Prioritized workload with QOS ~ 1000 miles
  - Customer 2 – STK extension over 1000 miles
  - Customer 3 – Overcome DWDM switchover issue – 80km
- Brocade FCIP – explanations
  - Customer 1 – Extension for asynchronous replication
  - Customer 2 – XRC emulation for remote vaulting
  - Customer 3 – Extension solution differences
- IBM comments
  - System z Requirements
- Q & A

Cisco MDS FCIP TCP Behavior

- Reduce probability of drops
  - Bursts controlled through per flow shaping and congestion window control → less likely to overrun routers

- Increased resilience to drops
  - Uses SACK, fast retransmit and shaping

- Aggressive slow start q
  - Initial rate controlled by "min-available-bandwidth"
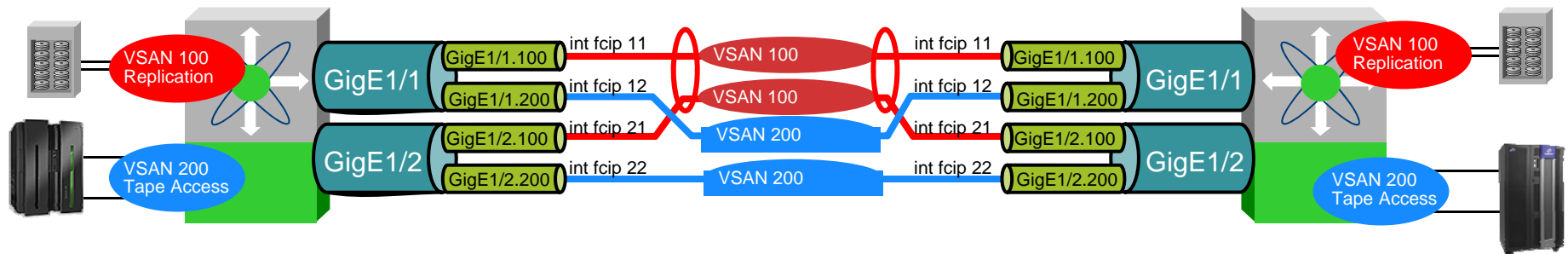 - Max rate controlled by "max-bandwidth"

Differences with Normal TCP:

- When congestion occurs with other conventional TCP traffic, FCIP is more aggressive during recovery ("bullying" the other traffic)
  Aggression is proportional to the min-available-bandwidth configuration

# Cisco FCIP – Multiple FCIP Tunnels

Now, Configure QOS based on business priorities of data

- VSAN 100 – high priority – disk mirroring
- VSAN 200 – med priority – Tape backups
- VSAN 300 (not shown) – low priority (open systems SAN stuff)

Making the assumption that this is a dedicated SAN WAN infrastructure – but within that, prioritization is needed.

Note: Routers and Switches MUST be QOS aware.
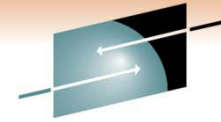
Cisco FCIP QoS Markings

Customer networks can have several types of business-critical traffic, including voice over IP (VoIP), video, FCIP, business applications, etc…

Traffic is normally classified as it enters the network, where it is marked for appropriate treatment.

Application	L3 Classification			L2 CoS
	IPP	PHB	DSCP	
Routing/FCIP control	6	CS6	48	6
Voice	5	EF	46	5
Video Conferencing	4	AF41	34	4
Streaming Video	4	CS4	32	4
FCIP SYNC	3	AF31	26	3
Call Signaling	3	CS3	24	3
FCIP ASYNC	2	AF21	18	2
Network Management	2	CS2	16	2
FCIP backup/FCIP ASYNC	1	AF11	10	1
Scavenger	1	CS1	8	1
Best Effort	0	0	0	0

High →

Low →

Med ←

Cisco's FICON Tape Acceleration

- Accelerates Writes by means of local acknowledgement
  - Command Response
  - Status
  - Write data never fully owned by FTA
    - Sync command is not emulated – insures data integrity
- Accelerated Reads by means of pre-fetching of data
  - Watches first read(s) to learn
  - Pre-reads to help fill the data pipe
  - Re-position logic to handle if we pre-read too far
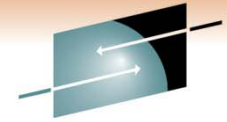- Tape control, label processing, etc are not accelerated

Backup protocol without acceleration

Mainframe   Cisco MDS   FCIP   Cisco MDS   VSM/Tape Library

ScratchVol mount, Write VolHdrs etc.

Write Chain

Backup protocol without acceleration …

Mainframe   Cisco MDS   FCIP   Cisco MDS   VSM / Tape Library

ScratchVol mount, Write VolHdrs etc.

Write Chain

Status

TAPE IDLE

Write Chain

Backup protocol without acceleration …

Mainframe   **Cisco MDS**   FCIP   **Cisco MDS**   **VSM / Tape Library**

ScratchVol mount, Write VolHdrs etc.

Write Chain

Status

TAPE IDLE

Write Chain

Status

SYNC

All data on Media

Status

Rewind Unload

Backup protocol with acceleration

Mainframe

Host-side FCIP

Tape-Side FCIP

VSM / Tape Library

ScratchVol mount, Write VolHdrs etc.

Write Chain 1

Write Chain 1

New OXID

Backup protocol with acceleration …

Backup protocol with acceleration …
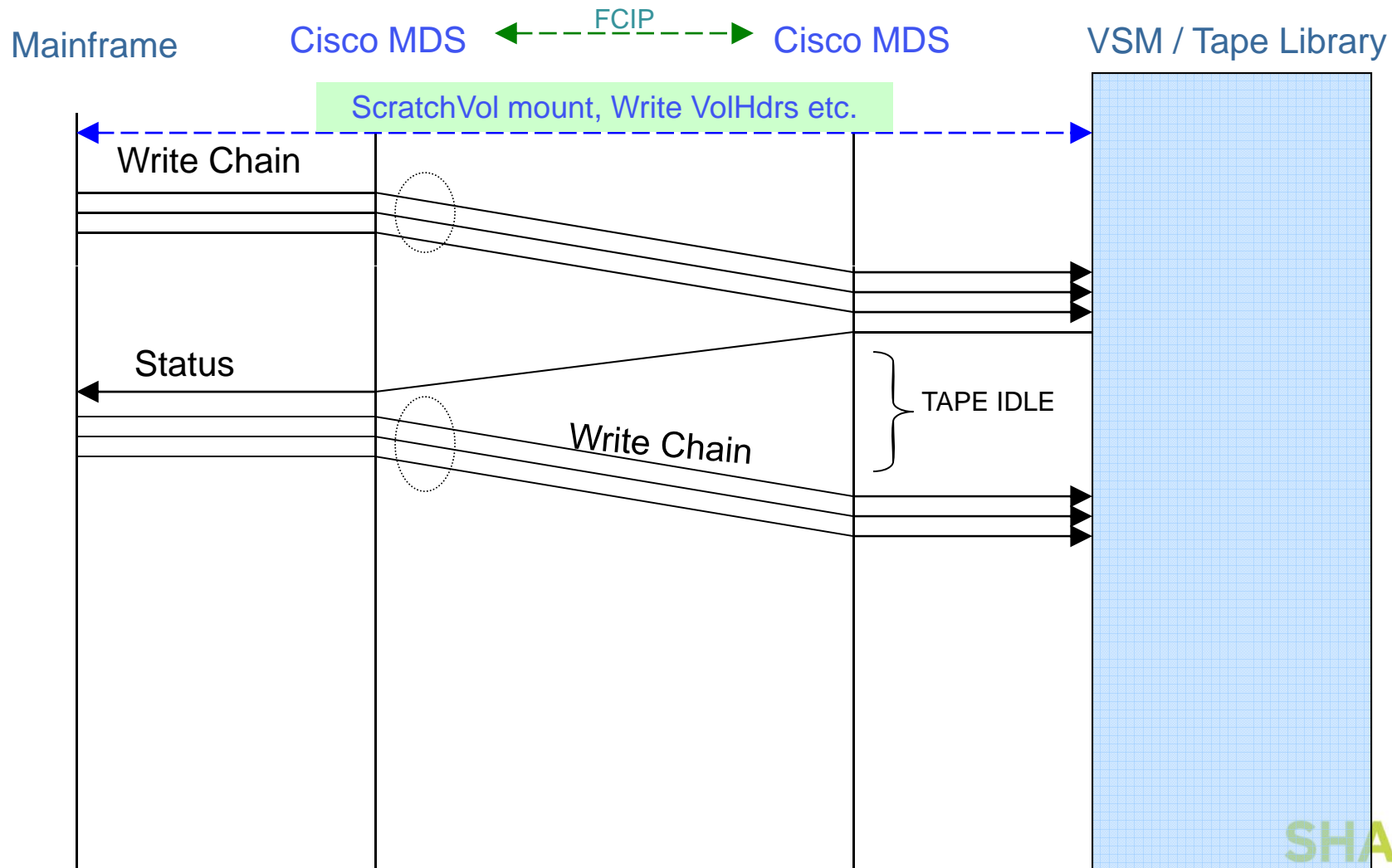
Mainframe | Host-side FCIP | Tape-Side FCIP | VSM / Tape Library

ScratchVol mount, Write VolHdrs etc.

Write Chain 1

Write Chain 1

New OXID

SYNC

Write Chain 2

All data on Media

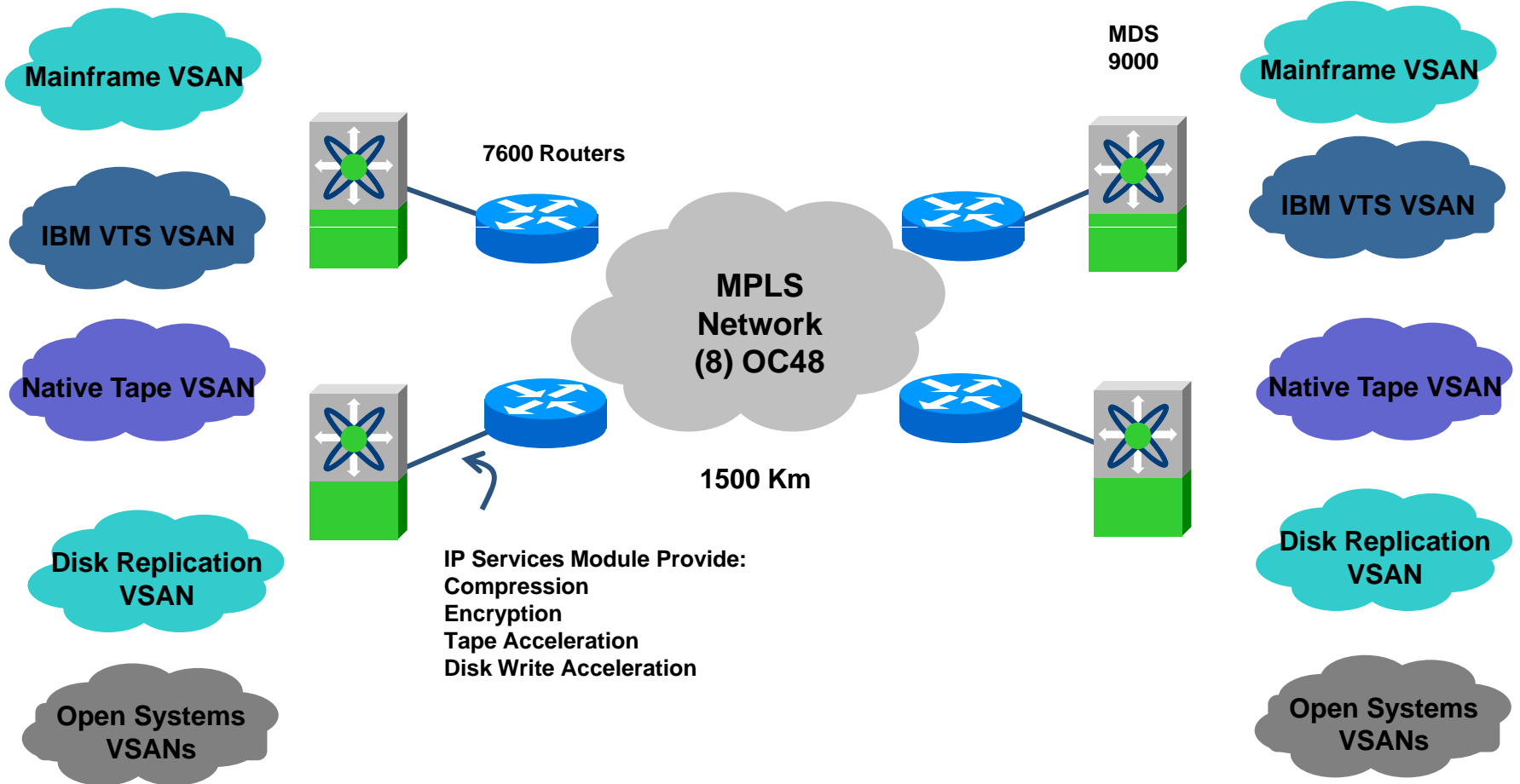Rewind Unload
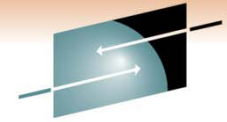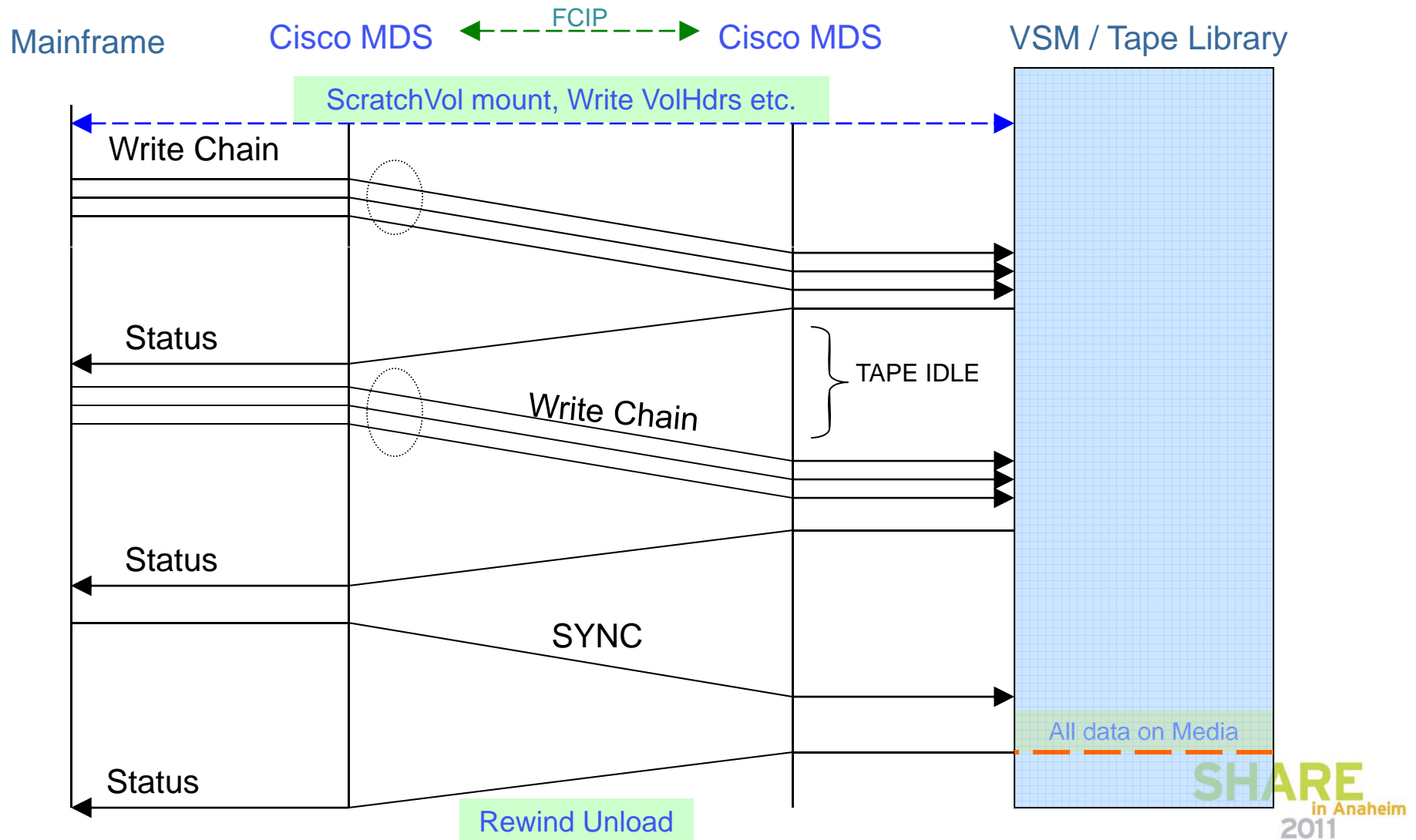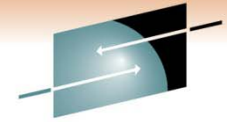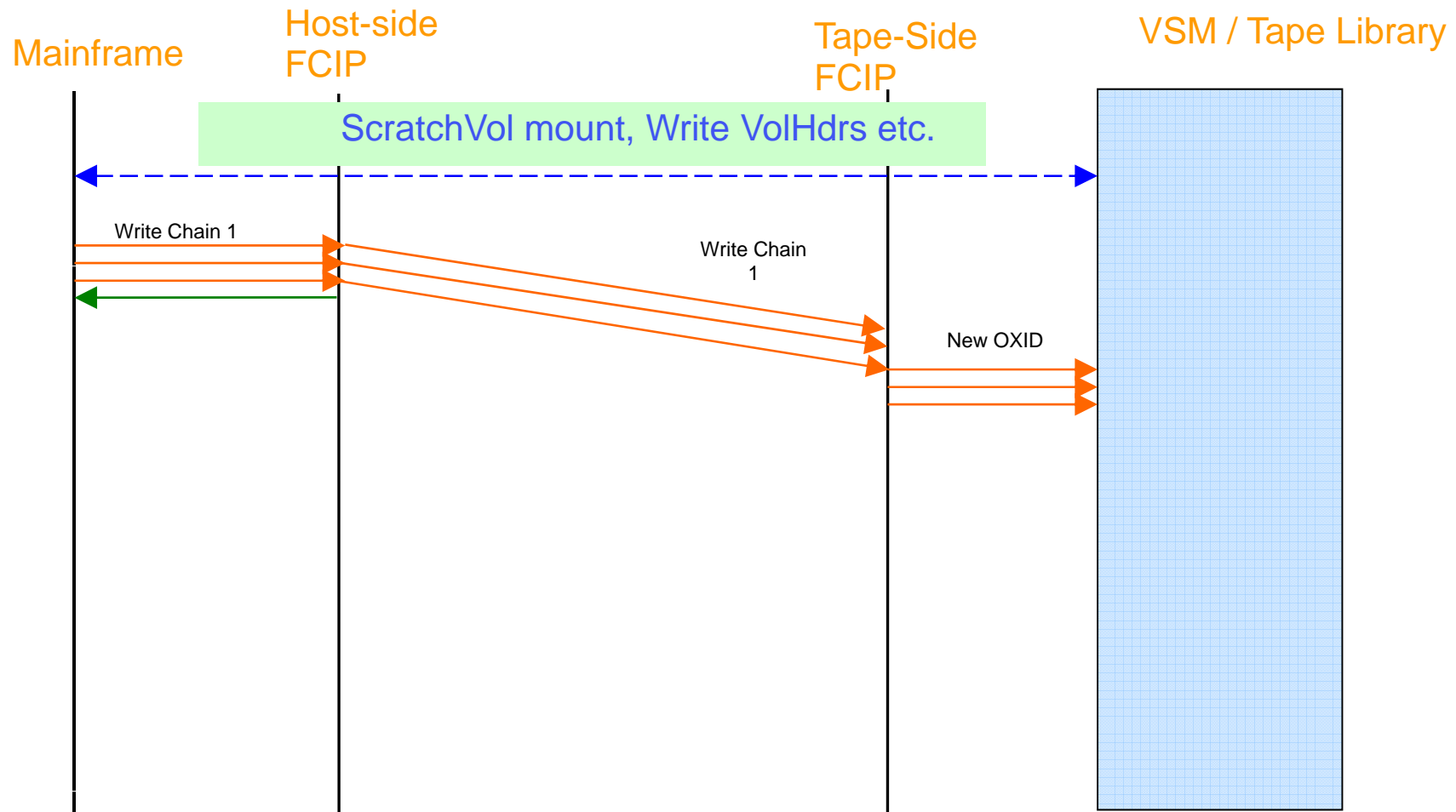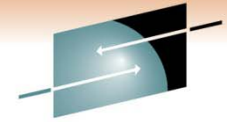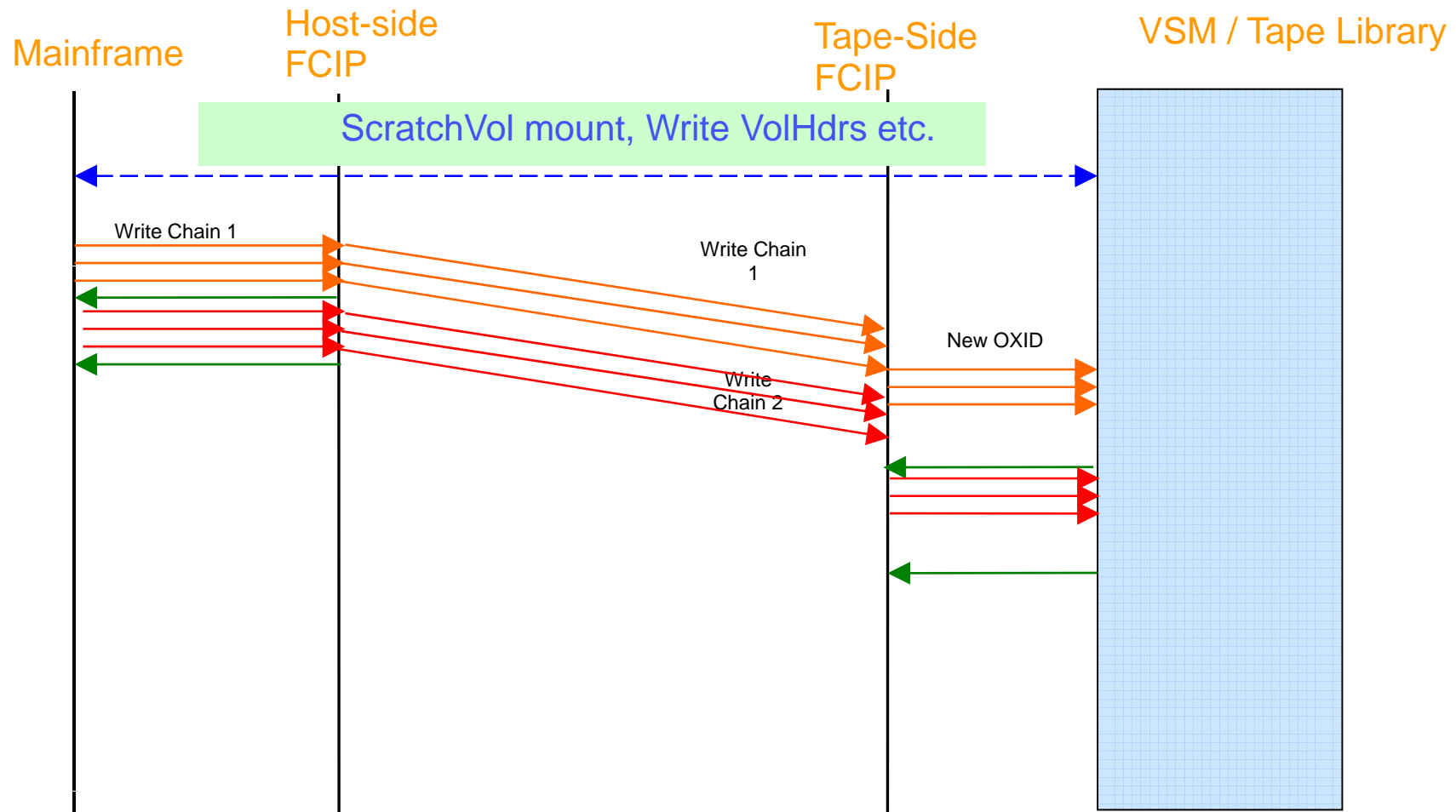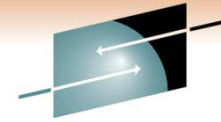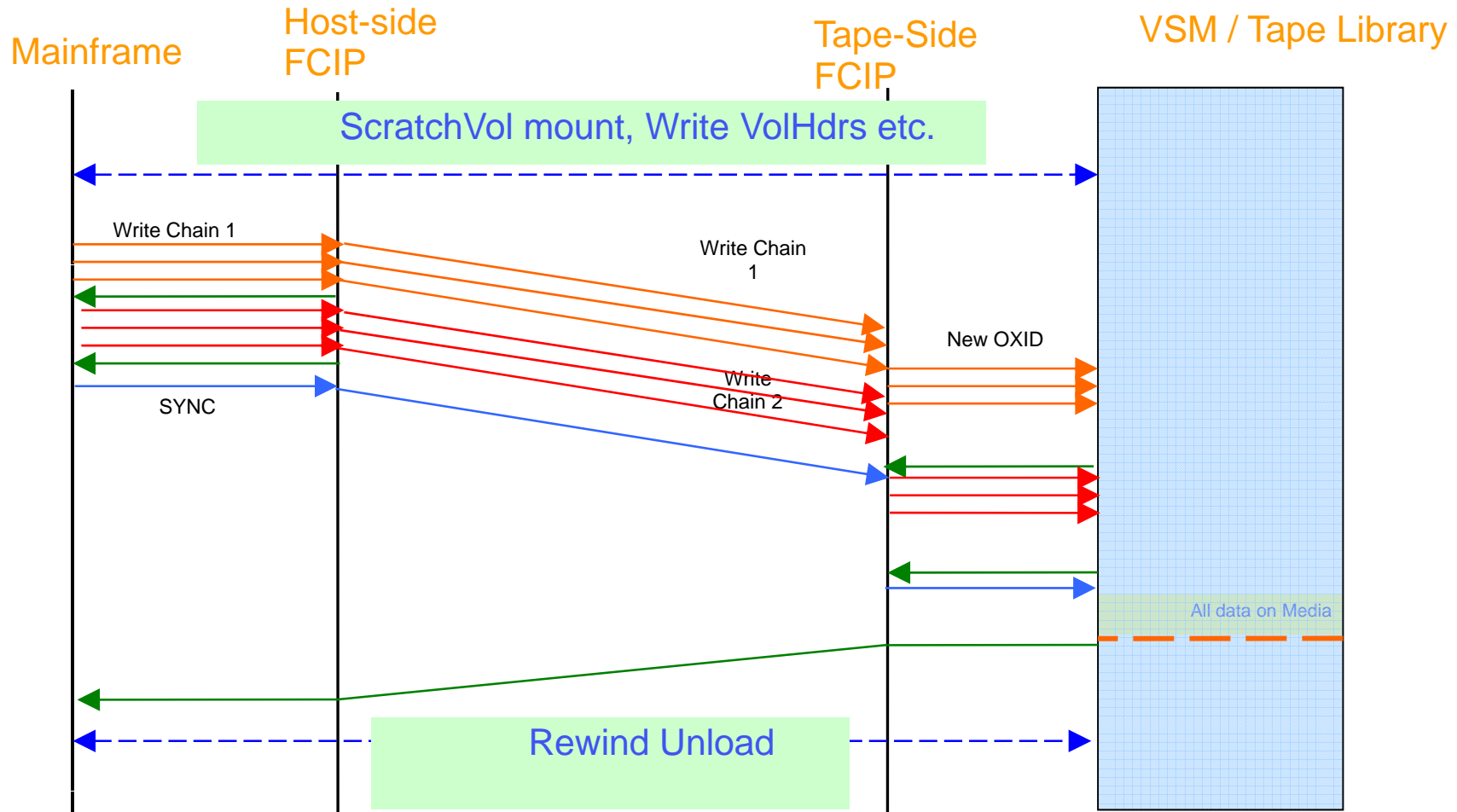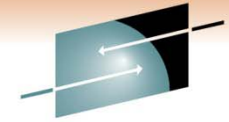
Sample Results: Write Throughput
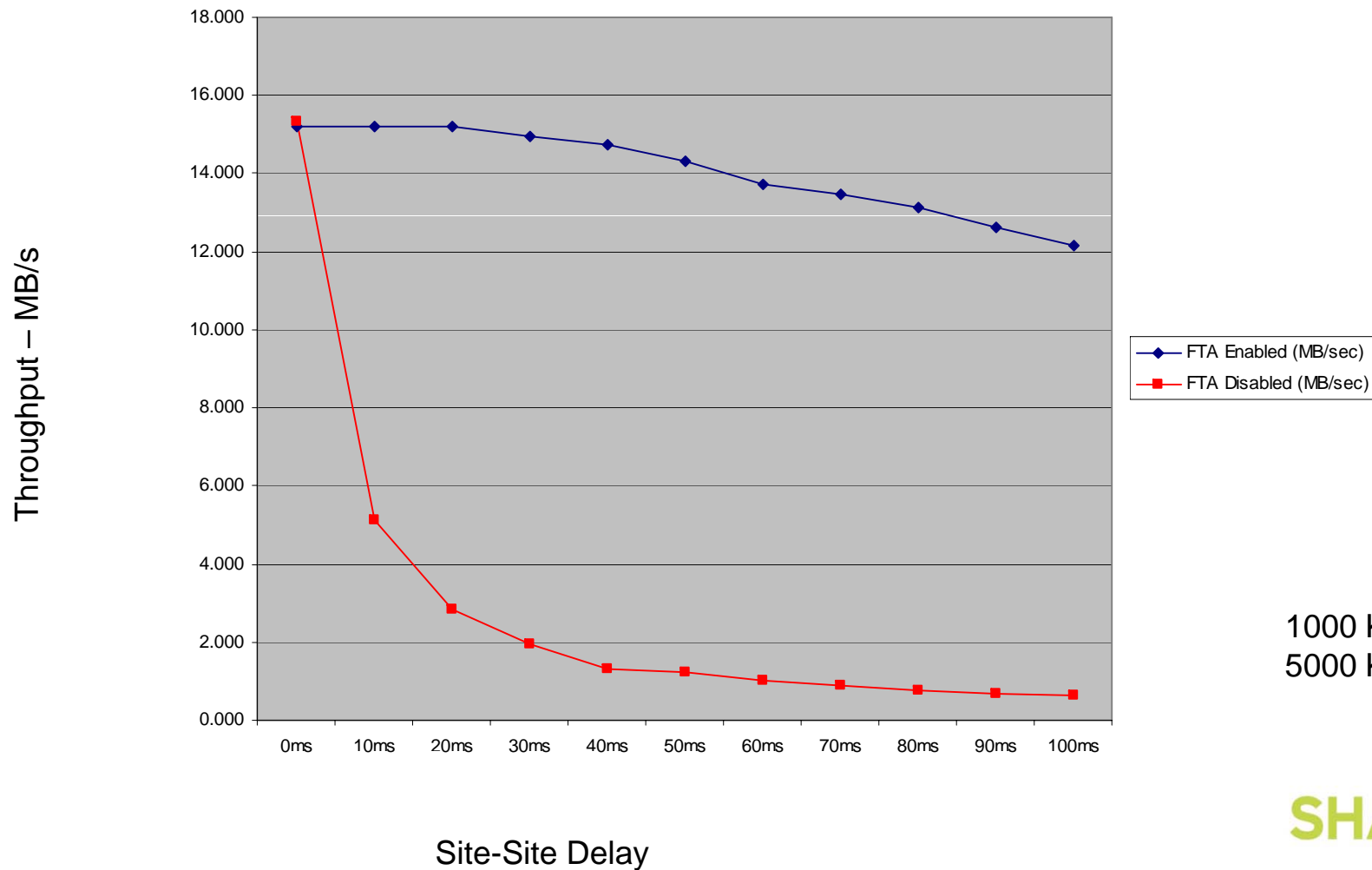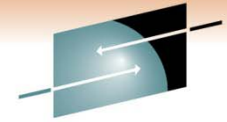

1000 Km – 10ms
5000 Km – 50ms

Cisco's XRC Acceleration

- Flow based acceleration to accelerate RRS commands
  - SDM indicates how many RRS commands in a chain
  - Remote Cisco MDS pre-issues these RRS commands
  - Helps to fill the pipe – working around IU pacing
- Fully supports Cisco Port channels (bundled ISLs)
  - Less disruption when WAN errors occur
- Works with all models of z Systems
  - Integrates smoothly with z10 Extended distance feature
- Can utilize integrated hardware compression / encryption
- Supports all three major vendors DASD systems

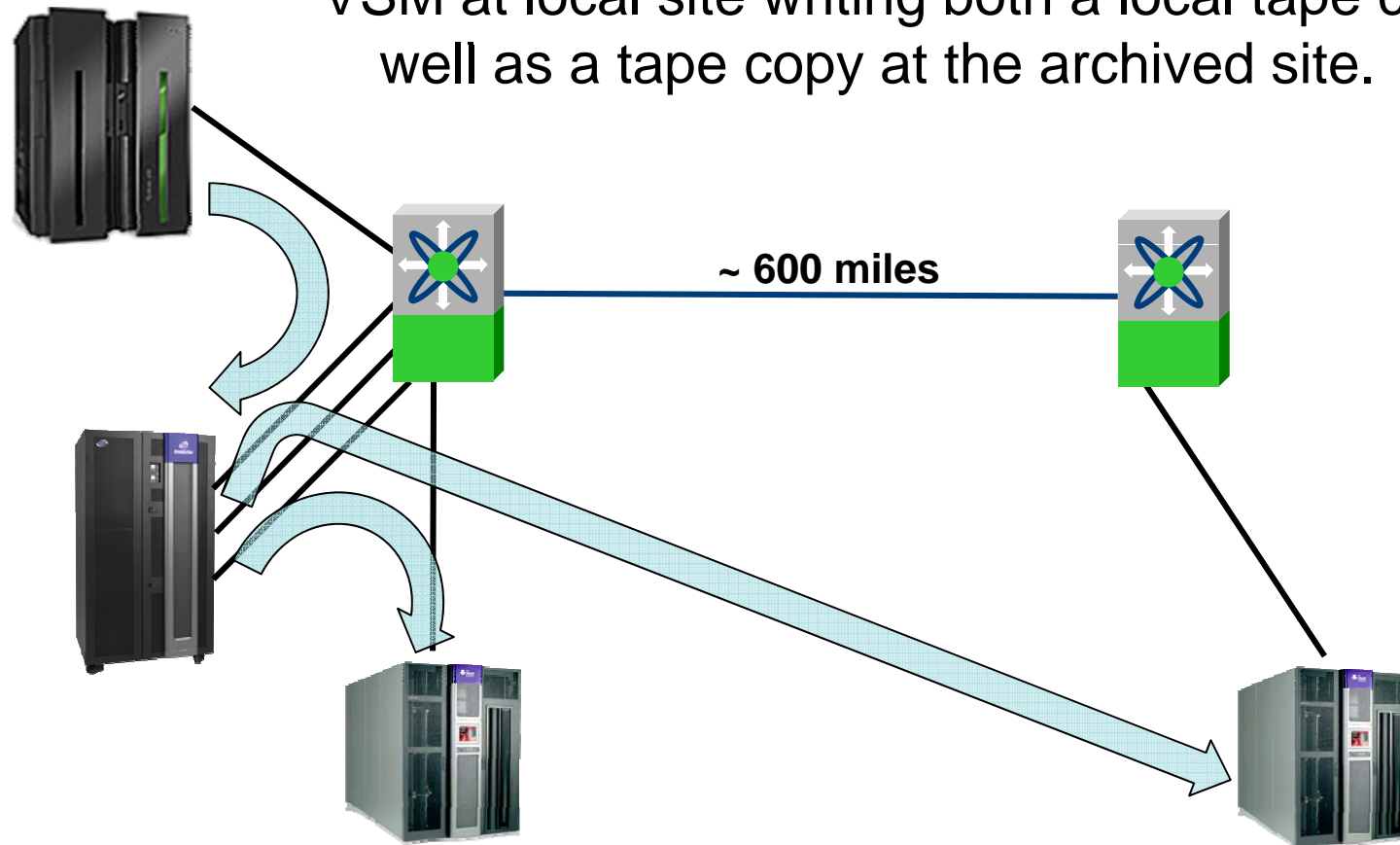# Customer 1 - Large Provider of Business Outsourcing Services

Mainframe VSAN

IBM VTS VSAN

Native Tape VSAN

Disk Replication VSAN

Open Systems VSANs

MDS 9000

7600 Routers

MPLS Network (8) OC48

1500 Km

IP Services Module Provide:
Compression
Encryption
Tape Acceleration
Disk Write Acceleration

Mainframe VSAN

IBM VTS VSAN

Native Tape VSAN

Disk Replication VSAN

Open Systems VSANs

Customer 2 – Large European Bank

VSM at local site writing both a local tape copy as well as a tape copy at the archived site.

~ **600 miles**
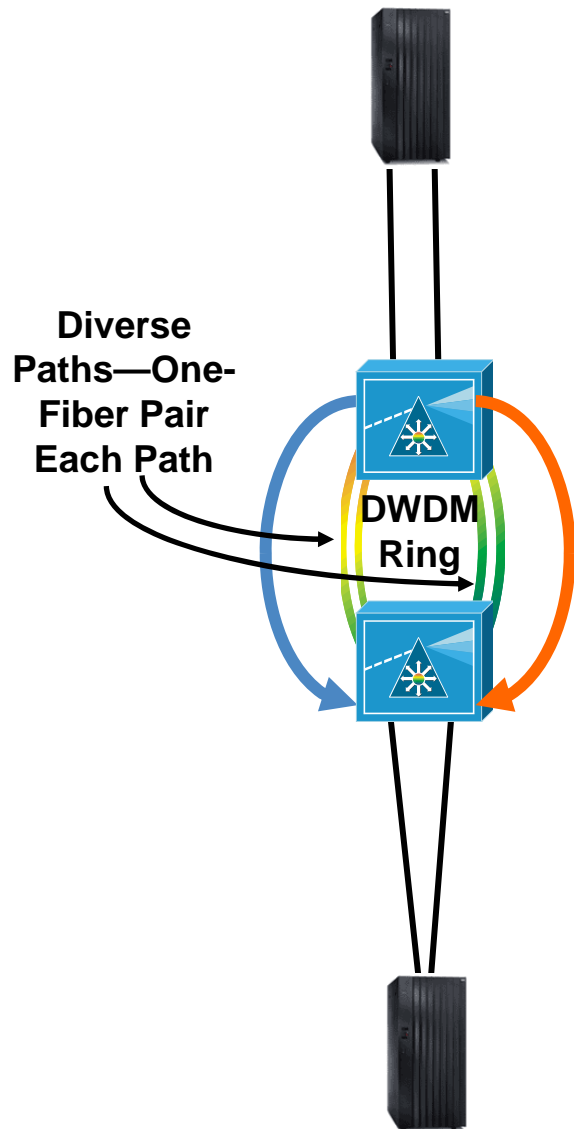
Diverse Paths—One-Fiber Pair Each Path

DWDM Ring

Problem to solve:

Disk mirroring links hooked directly to DWDM

50ms DWDM switchover causing mirror to drop

Customer 3 – Large Financial Institution

Diverse Paths—One-Fiber Pair Each Path

DWDM Ring
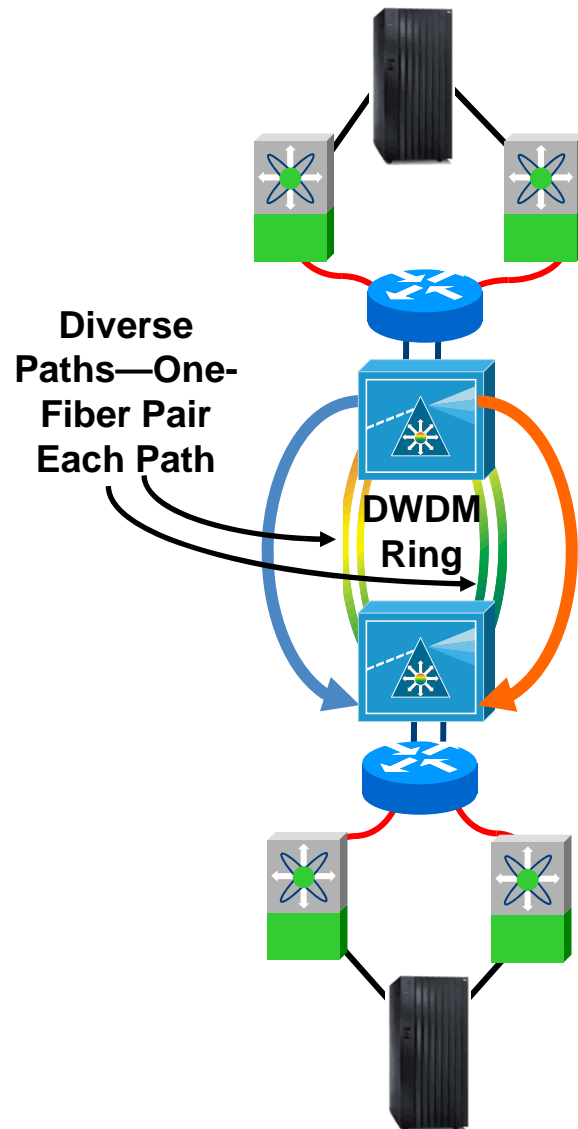
Advantages:

Consolidate to 10G Extended infrastructure

FCIP retry logic masks optical switchover

Agenda

- FCIP Overview
- FCIP Topologies
- Cisco FCIP – explanations
  - Customer 1 – Prioritized workload with QOS ~ 1000 miles
  - Customer 2 – STK extension over 1000 miles
  - Customer 3 – Overcome DWDM switchover issue – 80km
- **Brocade FCIP – explanations**
  - Customer 1 – Extension for asynchronous replication
  - Customer 2 – XRC emulation for remote vaulting
  - Customer 3 – Extension solution differences
- IBM comments
  - System z Requirements
- Q & A

Brocade Extension Characteristics Asynchronous Replication

- Virtual Fabric Isolation
  - Stabilizes local and remote environments
  - Isolates environments from each other
- Fabric Stability
  - Network behaviors reduced to device access
- Device Access
 - Limited to defined devices
  - Local devices isolated from remote

Customer 1 – Large North American Financial Institution

Asynchronous Replication

Virtual Inter-switch Links

Synchronous Replication

Customer 1 – Large North American Financial Institution

Asynchronous Replication

Virtual Fabrics

Virtual Inter-switch Links

Synchronous Replication

Customer 1 – Large North American Financial Institution

Asynchronous Replication

Logical Access

Virtual Inter-switch Links
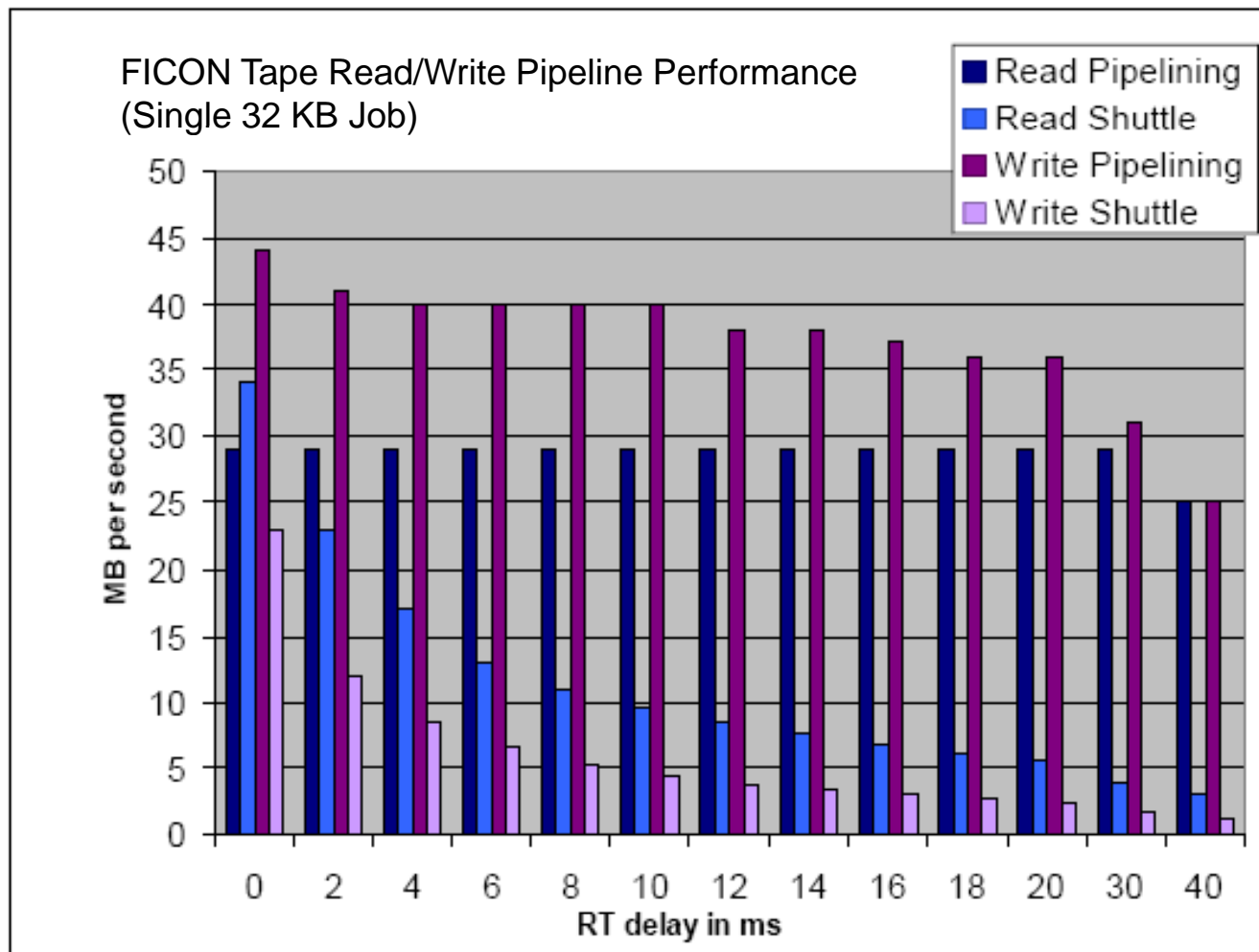
Synchronous Replication
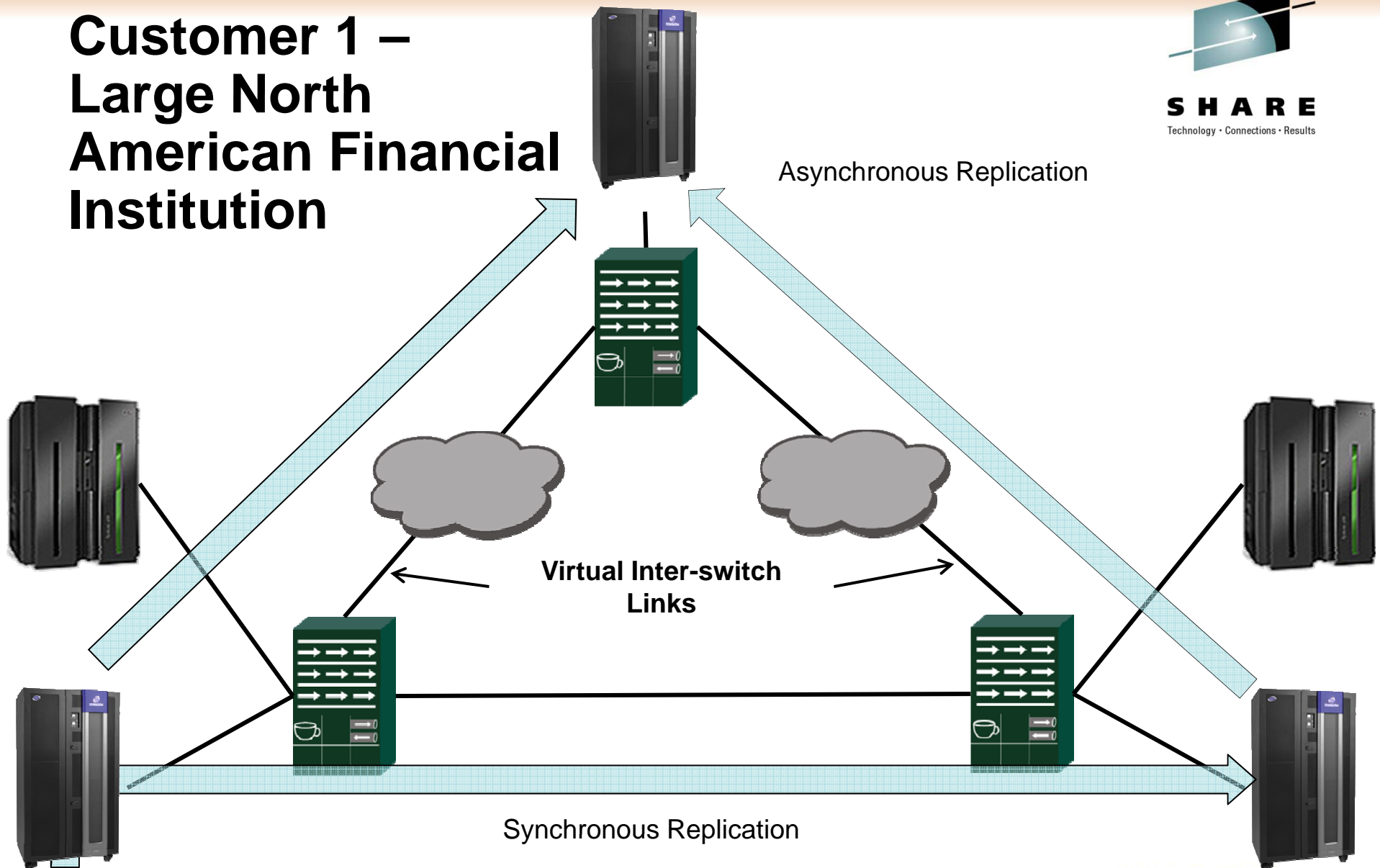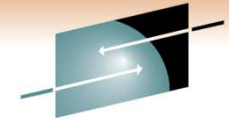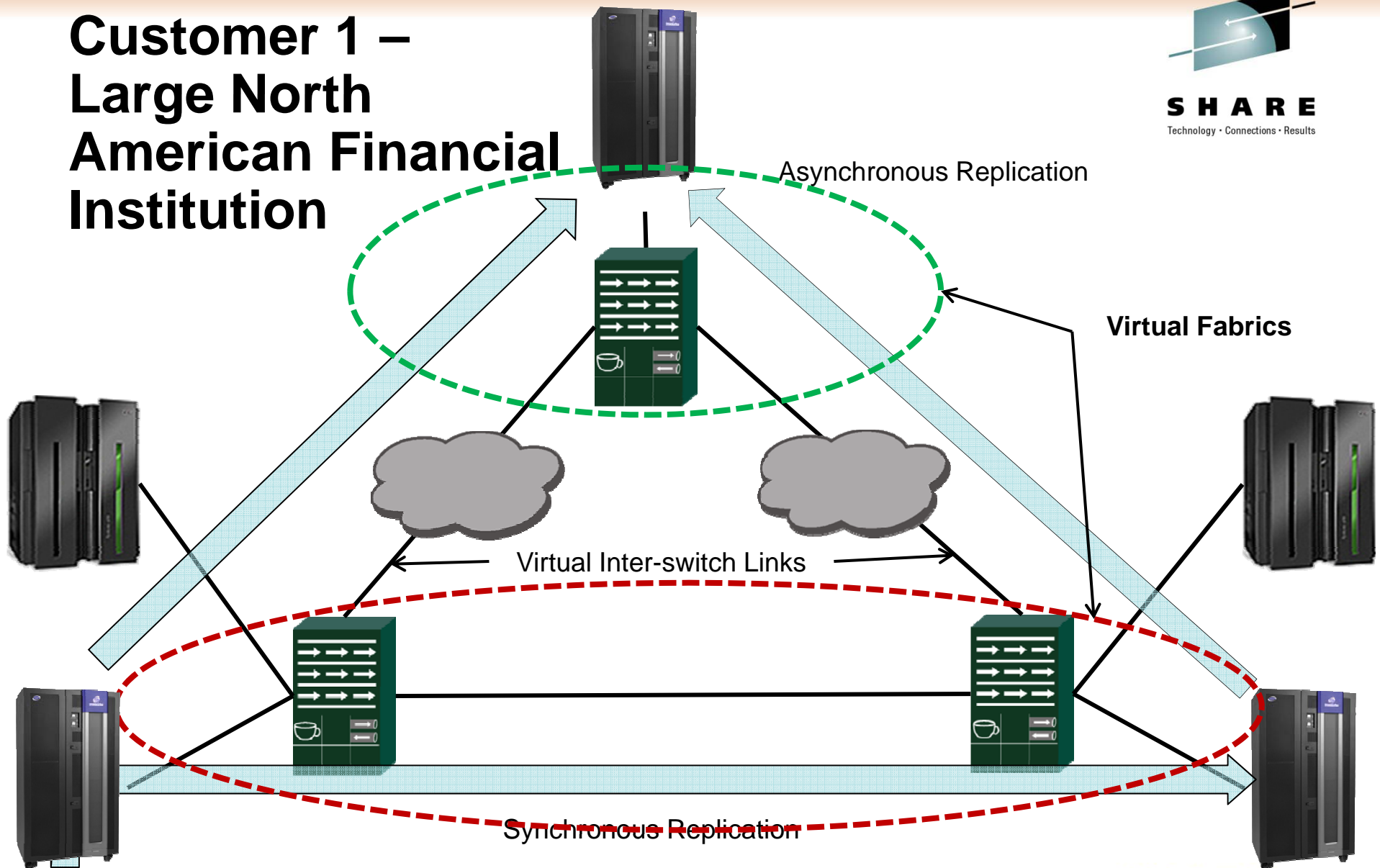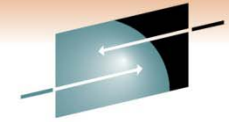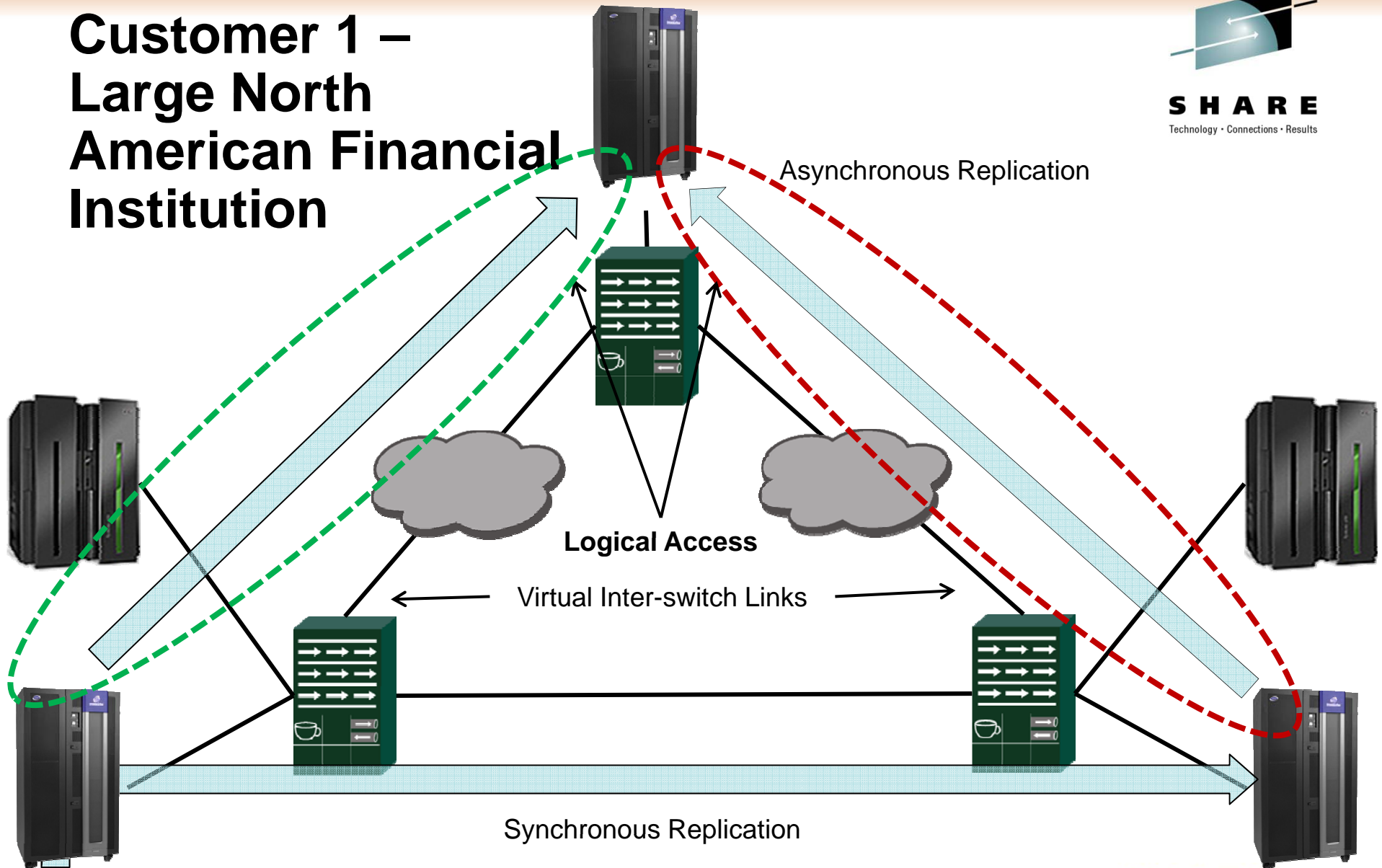
Brocade Emulation Functions
XRC Environment

- Serves uniquely formatted channel program
  - Identified by a uniquely prefixed DSO
    - Defined Subsystem Operation command Comand-Data IU
  - All other channel programs are shuttled across WAN
    - No additional processing
- Seeks to alleviate dormancy
  - Buffering delays
  - Signal propagation delays
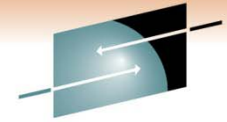 - Bandwidth restrictions of WAN links
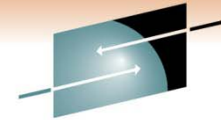
Brocade Performance Gains
Device Emulation

FICON Tape Read/Write Pipeline Performance
(Single 32 KB Job)

Legend: Read Pipelining, Read Shuttle, Write Pipelining, Write Shuttle

Y-axis: MB per second
X-axis: RT delay in ms

Customer 2 – American Financial Institution
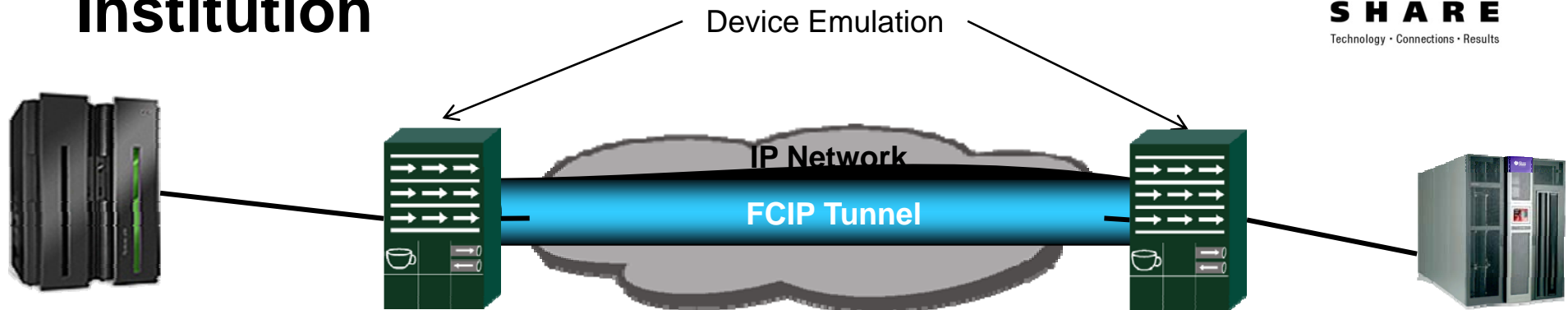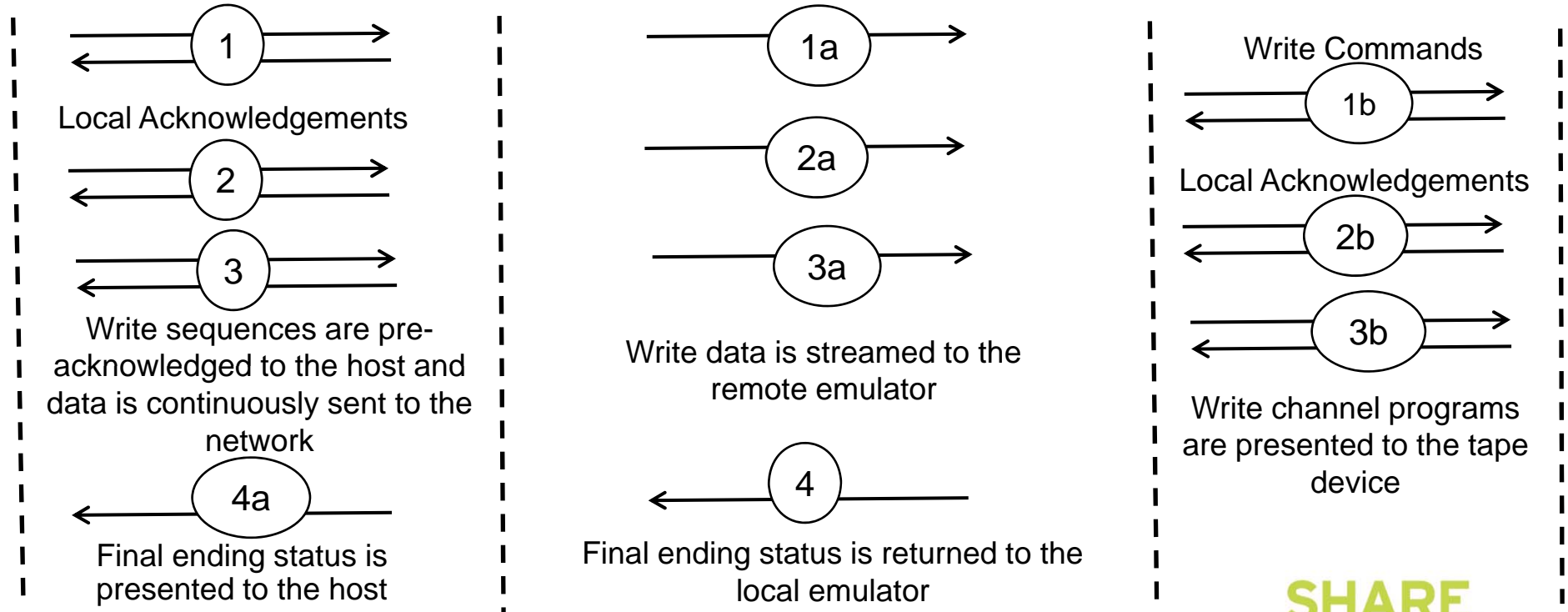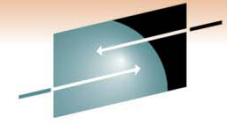
# Brocade Extension and Emulation Solution Considerations

- Dedicated Emulation Products
  - These solutions appear as a "bump in the wire"
- Switched Emulation Products
  - These solutions integrate Fibre Channel switching with emulation services
    - Can be standalone switches or Directors with emulation blades
    - When emulating only they are a "hop of no concern"
- Gotcha's
  - Mixing solutions requires an understanding of Fibre Channel routing rules

Customer 3 – Australian Solution Provider
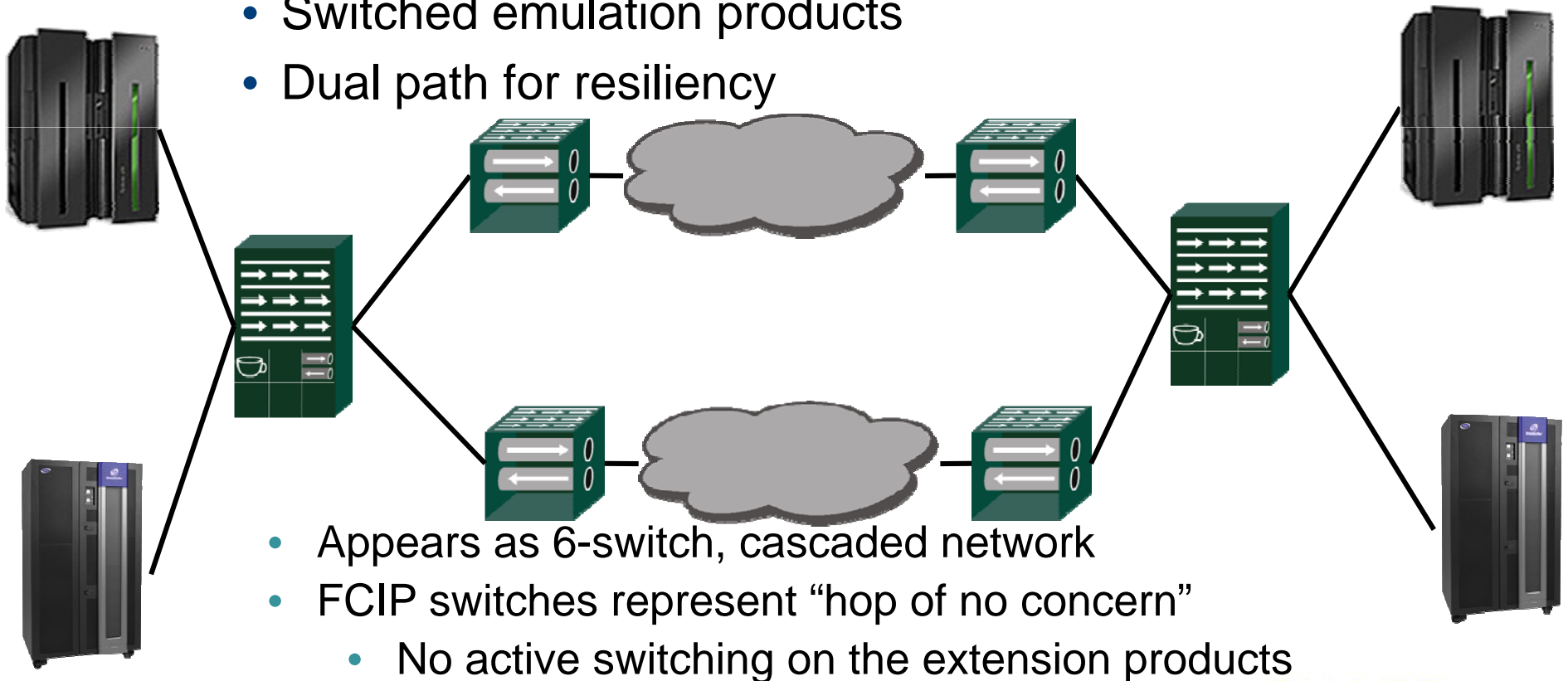
- Existing solution
  - Dedicated emulation products
  - Dual path for resiliency
  - Appears as 2-hop cascaded network with two ISLs
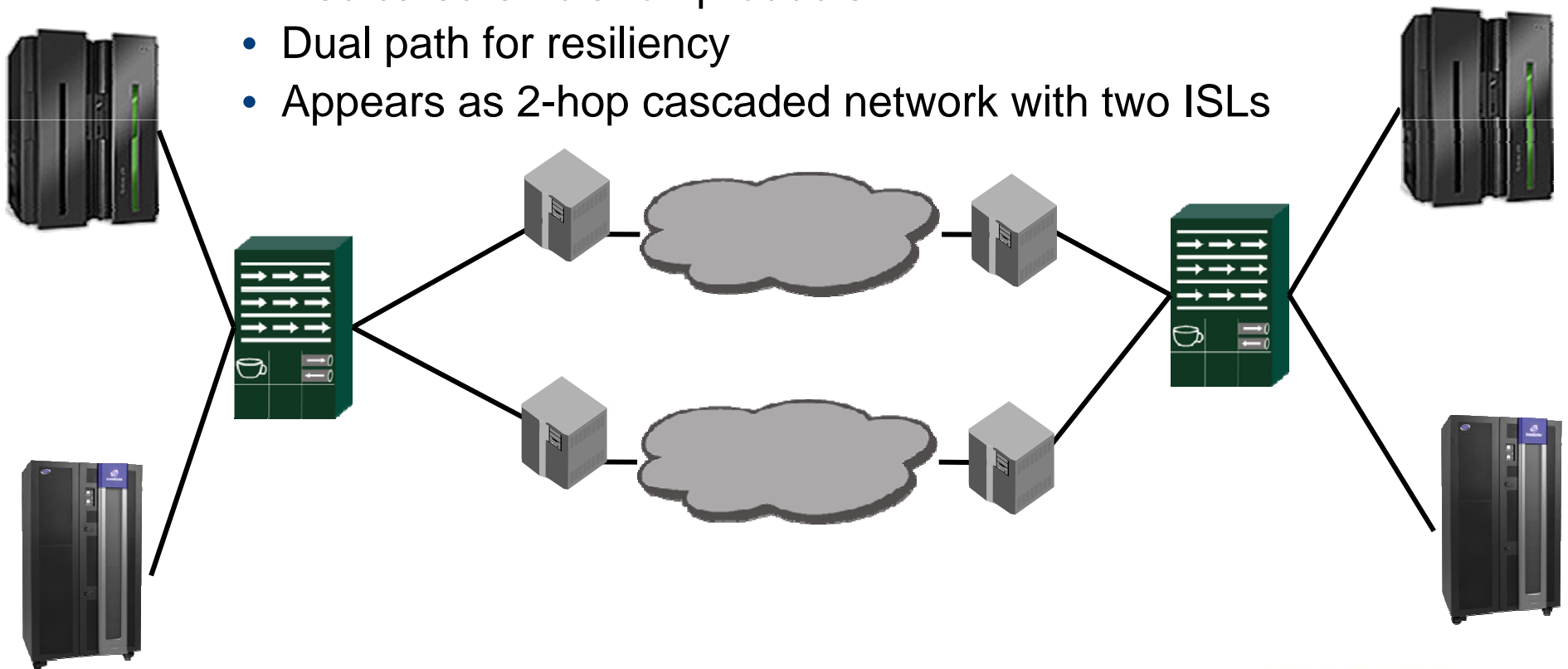
Customer 3 – Australian Solution Provider

- Updated solution
  - Switched emulation products
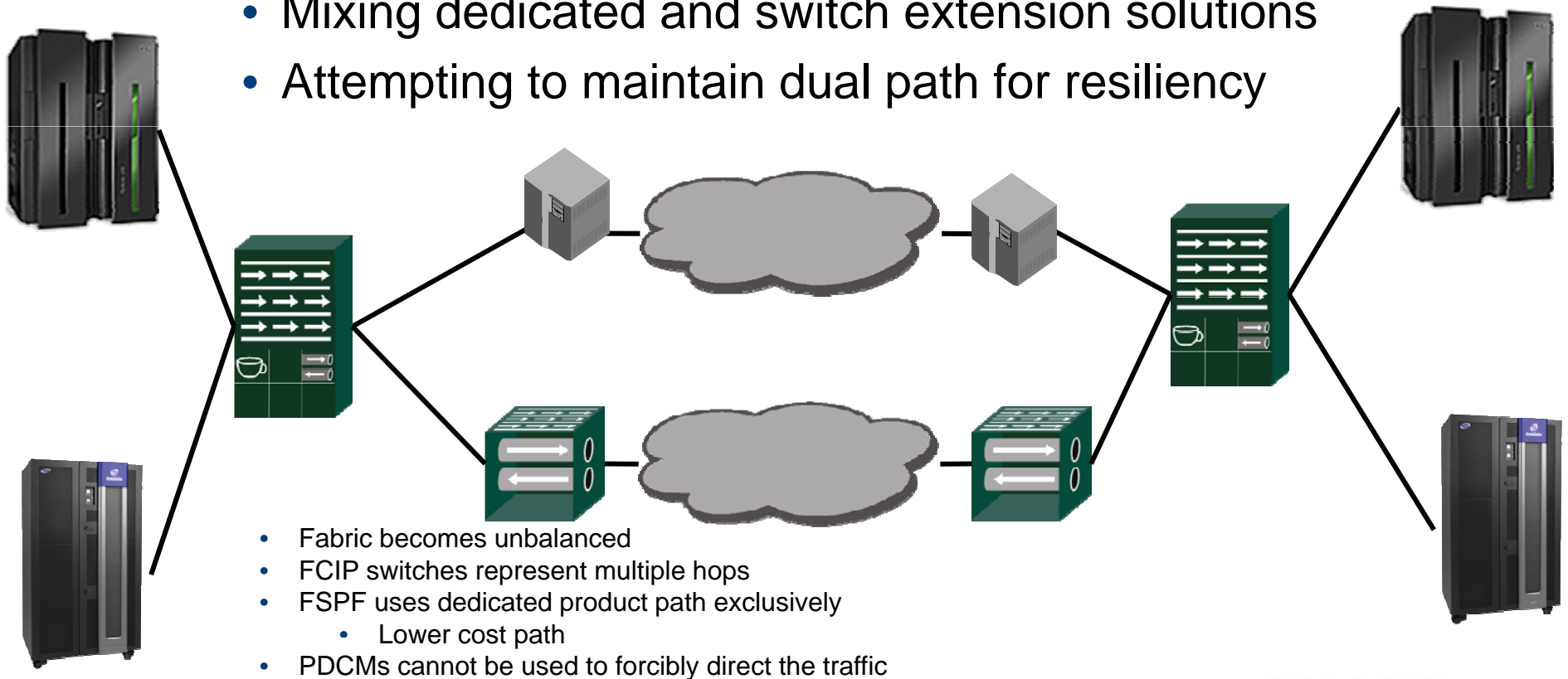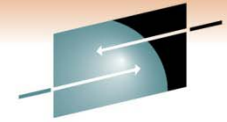 - Dual path for resiliency

  - Appears as 6-switch, cascaded network
  - FCIP switches represent "hop of no concern"
    - No active switching on the extension products

Customer 3 – Australian Solution Provider

- Transitional conundrum
  - Mixing dedicated and switch extension solutions
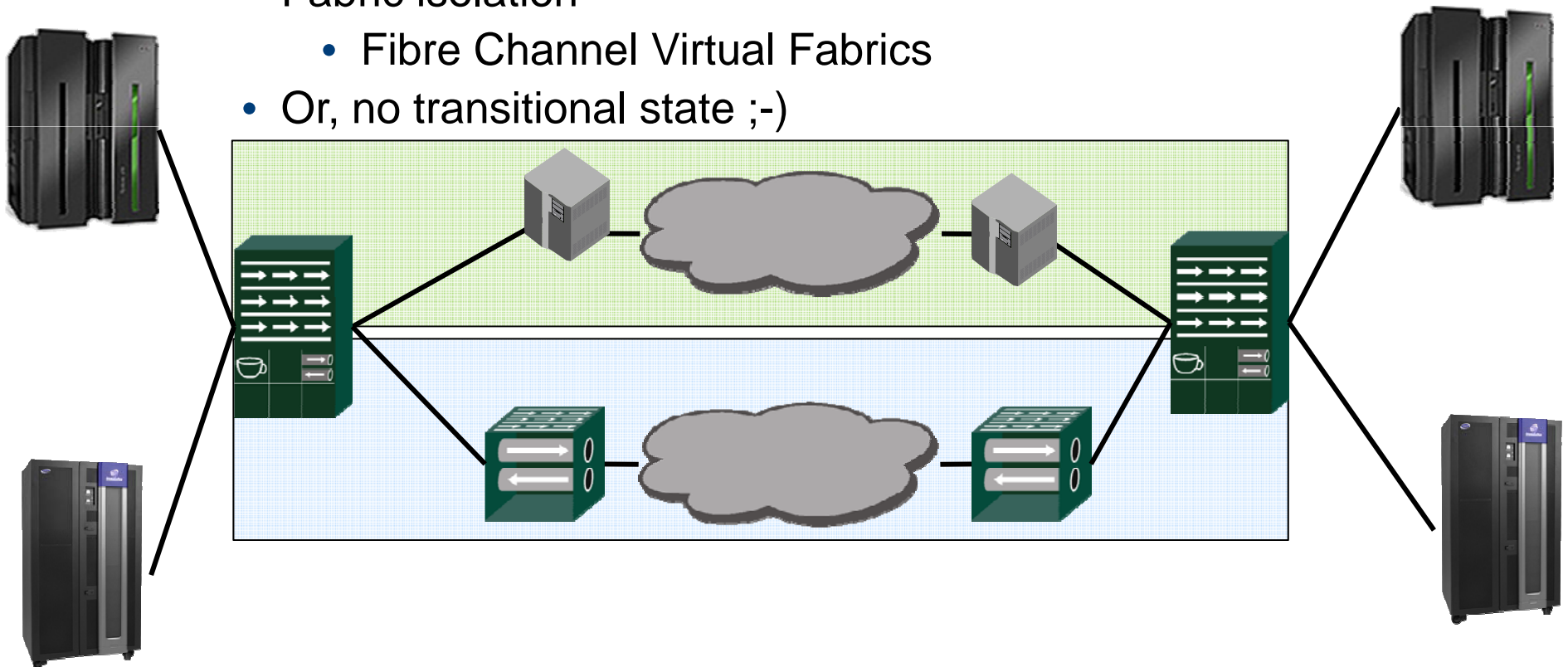 - Attempting to maintain dual path for resiliency

- Fabric becomes unbalanced
- FCIP switches represent multiple hops
- FSPF uses dedicated product path exclusively
  - Lower cost path
- PDCMs cannot be used to forcibly direct the traffic

Customer 3 – Australian Solution Provider

- Resolution
  - Fabric isolation
    - Fibre Channel Virtual Fabrics
  - Or, no transitional state ;-)

Agenda

- FCIP Overview
- FCIP Topologies
- Cisco FCIP – explanations
  - Customer 1 – Prioritized workload with QOS ~ 1000 miles
  - Customer 2 – STK extension over 1000 miles
  - Customer 3 – Overcome DWDM switchover issue – 80km
- Brocade FCIP – explanations
  - Customer 1 – Extension for asynchronous replication
  - Customer 2 – XRC emulation for remote vaulting
  - Customer 3 – Extension solution differences
- **IBM comments**
  - System z Requirements
- Q & A

IBM Qualification

- Moving data is easy

- Handling the unusual or error cases is HARD

- Testing is 80% Science and 20% Art

- Test to Architecture

- Plethora of error inject cases

IBM Qualification

- IBM develops emulated control units and devices for each of the SAN traffic types (ESCON, FICON, FCP) supported on System z
- IBM develops a proprietary test tool to test the S/390 architecture
  - Performs limit testing by stressing each capability of the channel to eliminate windows of failure
  - Can run to real devices as well as emulated devices
    - Emulated devices used for error injection
 - Over 30 years of experience built in
  - Updated as architecture evolves

IBM Qualification

- Switch Vendor Qualification Test content includes:
  - Architectural stress
  - z/OS stress
  - Serviceability
  - Fabric Security and Event Notification
  - CUP testing
  - ISL extension and balancing
  - GUI – function/human factors/ useability
  - Basic performance testing for data droop
  - Power cycling, faults, and redundant power
  - Link "up/downs" – IMLs, LPAR activate/deactivate, channel path varies
  - Improvised testing

IBM Qualification

- Collaborative effort
- Exit Criteria:
  - 'Clean' run without un-expected errors
  - Minor exceptions documented
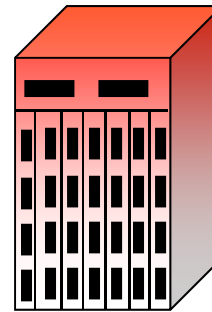
Duration of test can range from several weeks for a 'minor' release to several months for a 'major' release
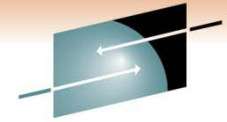
IBM Qualification

Fibre Channel Switch

IBM Ficon Test Vehicle

System z
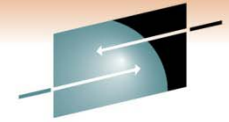
SHARE

Anaheim

February 2011

THANK YOU!
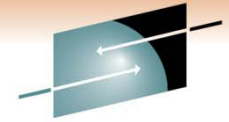
THIS SLIDE INTENTIONALLY LEFT BLANK