



Education

Ethernet Enhancements for Storage

Deploying FCoE

Sunil Ahluwalia, Intel Corporation
Errol Roberts, Cisco Systems Inc.

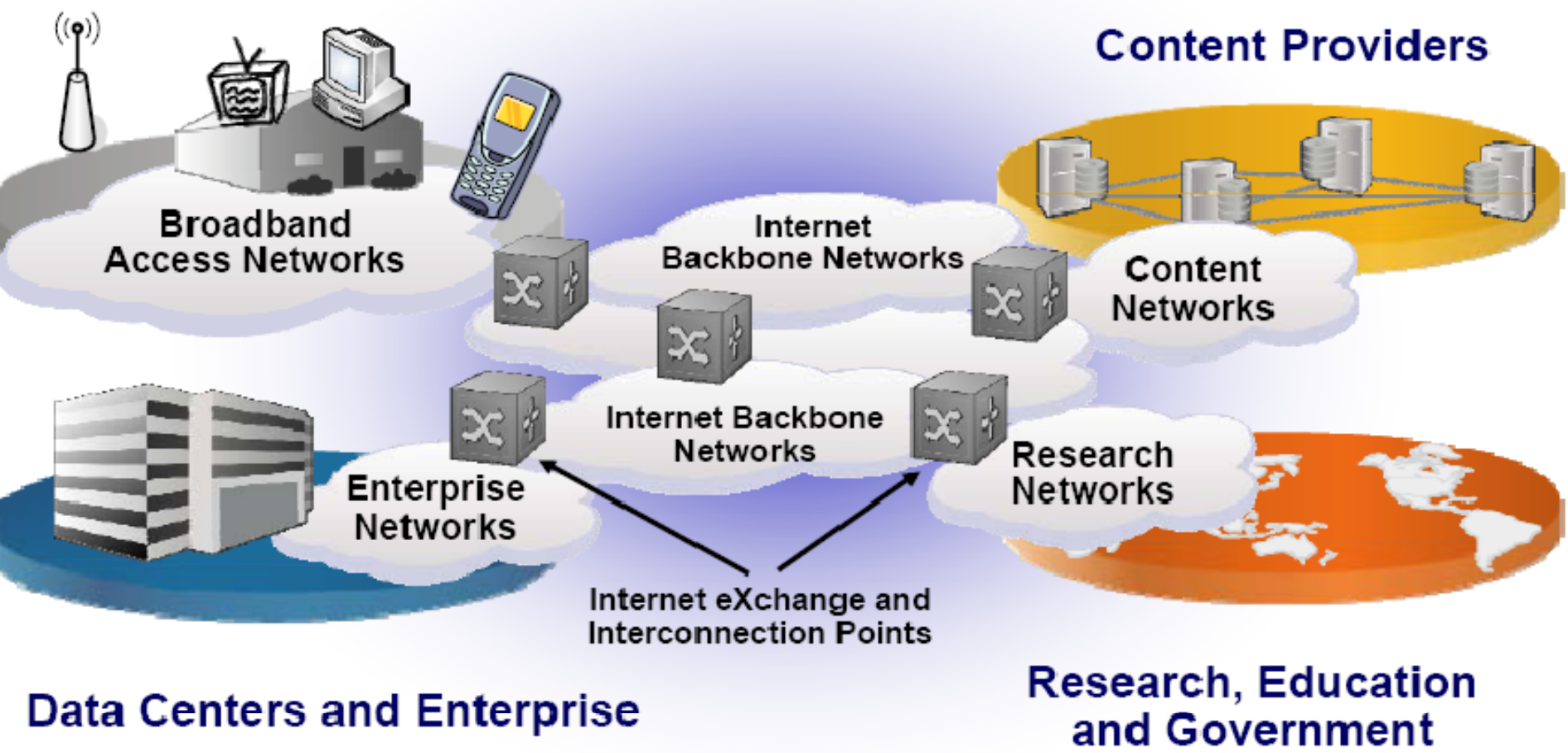
- The material contained in this tutorial is copyrighted by the SNIA.
 - Member companies and individuals may use this material in presentations and literature under the following conditions:
 - ◆ Any slide or slides used must be reproduced without modification
 - ◆ The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
 - This presentation is a project of the SNIA Education Committee.
 - Neither the Author nor the Presenter is an attorney and nothing in this presentation is intended to be nor should be construed as legal advice or opinion. If you need legal advice or legal opinion please contact an attorney.
 - The information presented herein represents the Author's personal opinion and current understanding of the issues involved. The Author, the Presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.
- NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.**

- **Ethernet Enhancements for Storage: Deploying FCoE**
This session discusses the Ethernet enhancements required for deploying FCoE. It reviews an end-to-end view to evaluate FCoE benefits from a host and switch perspective. The session also provides results from real life implementation of FCoE

- Market Requirements
- Ethernet Enhancements
- What is FCoE?
- FCoE Deployment
- Case Study: Deploying FCoE

Ethernet Everywhere!

Broadband Access



Nearly all of the traffic on the Internet either originates or terminates with an Ethernet connection

Server Migration from 1GbE to 10GbE



Multi-Core CPU architectures allowing bigger and multiple workloads on the same machine



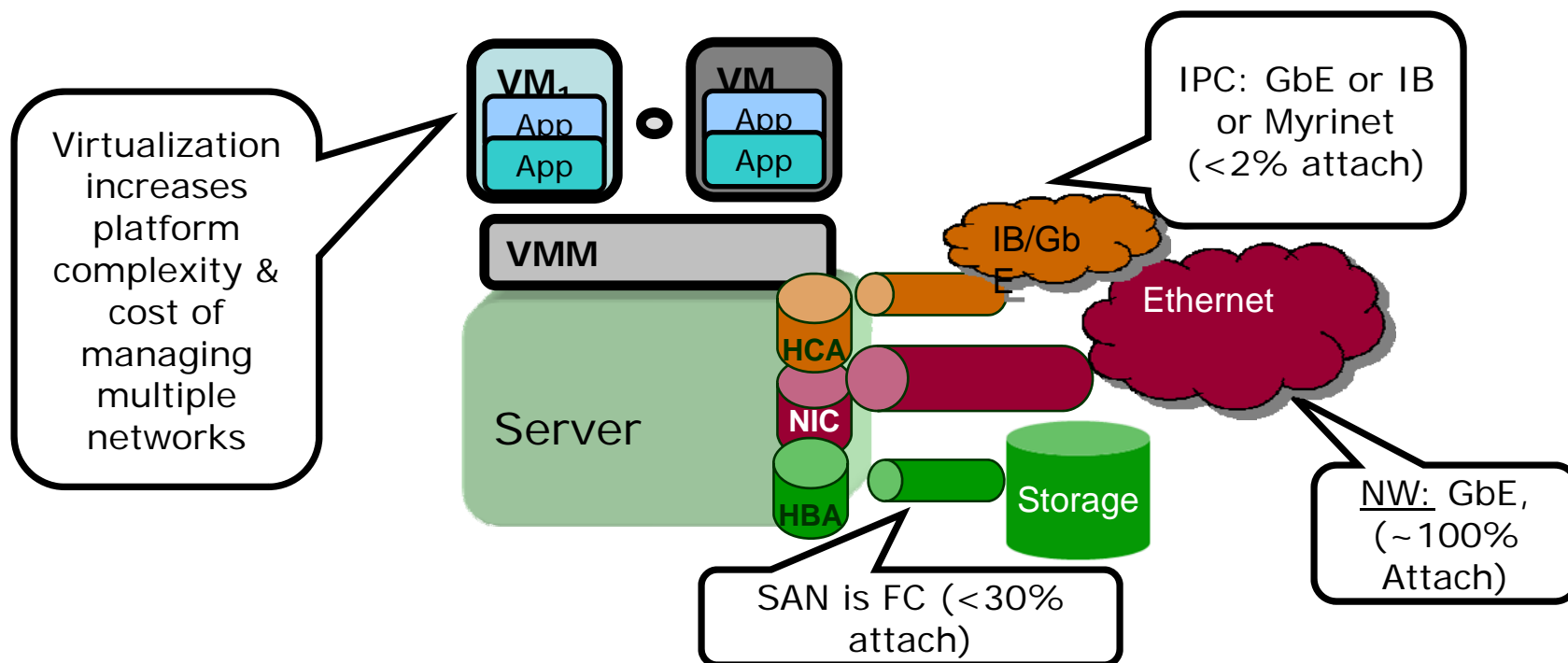
Server virtualization driving the need for more bandwidth per server due to server consolidation



Growing need for network storage driving the demand for higher network bandwidth to the server

Multi-Core CPUs, Server Virtualization and Storage, driving the adoption of 10GE network connections

Ethernet Enhancements [Data Center Bridging]



➤ Multiple networks, one per traffic class

- ◆ IP and other LAN protocols over an Ethernet network
- ◆ SAN over a Fibre Channel network
- ◆ IPC over an InfiniBand network

Merging the Requirements

LAN/IP

- **Must be Ethernet!**
 - Too much investment
 - Too many applications that assume Ethernet

Storage

- **Must follow the Fibre Channel model**
- **Losing frames is not an option**

IPC (Inter-Process Communication)

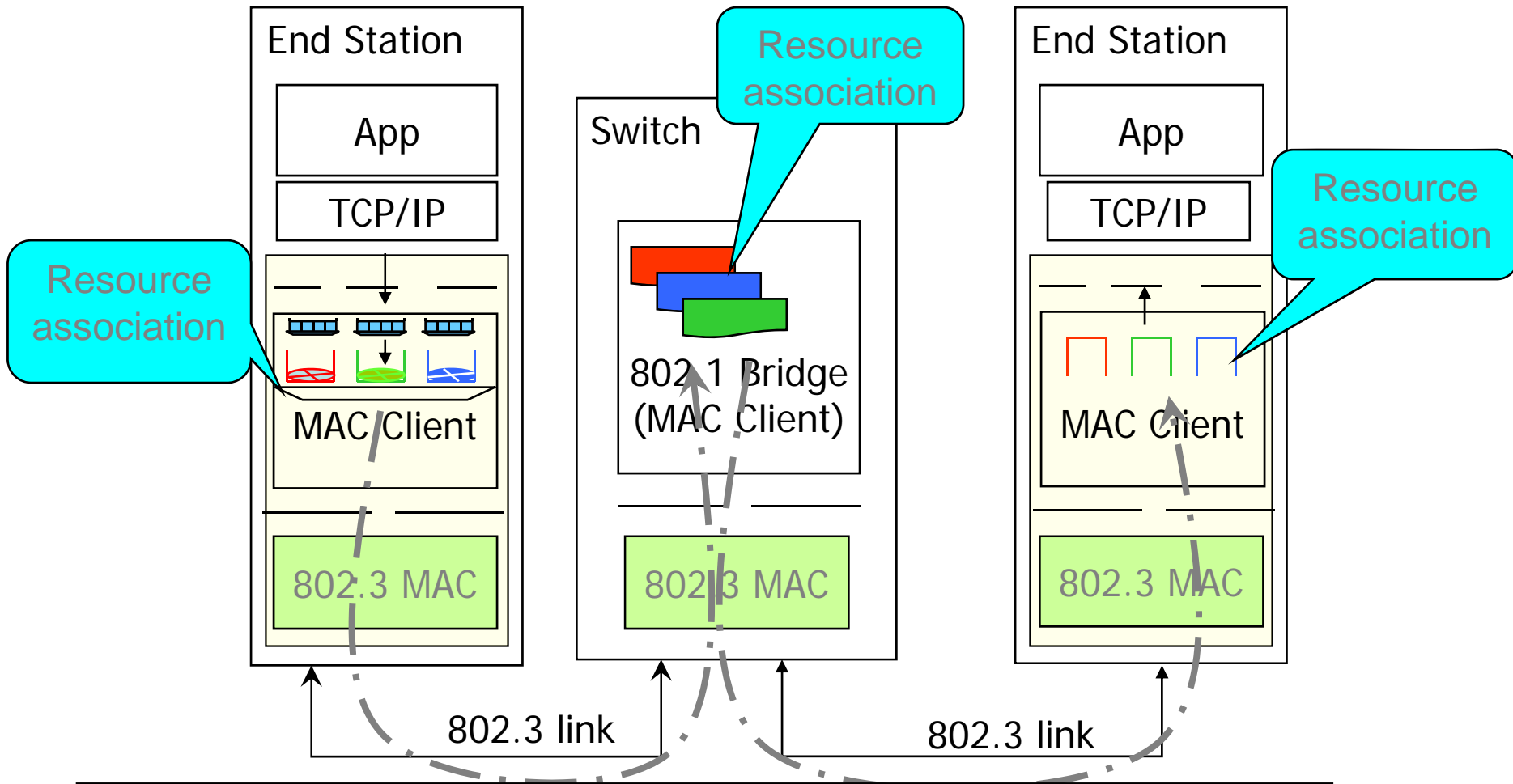
- **Doesn't care about the underlying network, provided that**
 - It is cheap
 - It is low latency
 - It supports APIs like OFED, MPI, sockets

- Traffic Differentiation
 - ◆ Provides end-to-end traffic differentiation for LAN, SAN and IPC traffic

- “Lossless” Fabric: Reliable Transport in Ethernet
 - ◆ Transient congestion - Priority Based Flow Control
 - ◆ Persistent congestion - Backward Congestion Notification

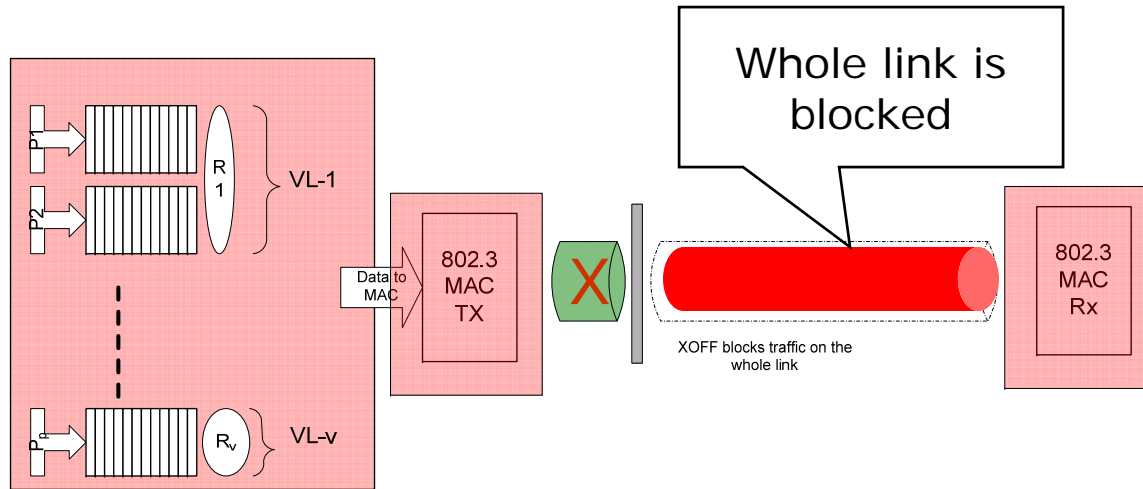
- Bi-sectional Bandwidth: Shortest-Path Bridging
 - ◆ Allow L2-Multipathing within Data Center

Enhanced Transmission Selection (IEEE 802.1Qaz)

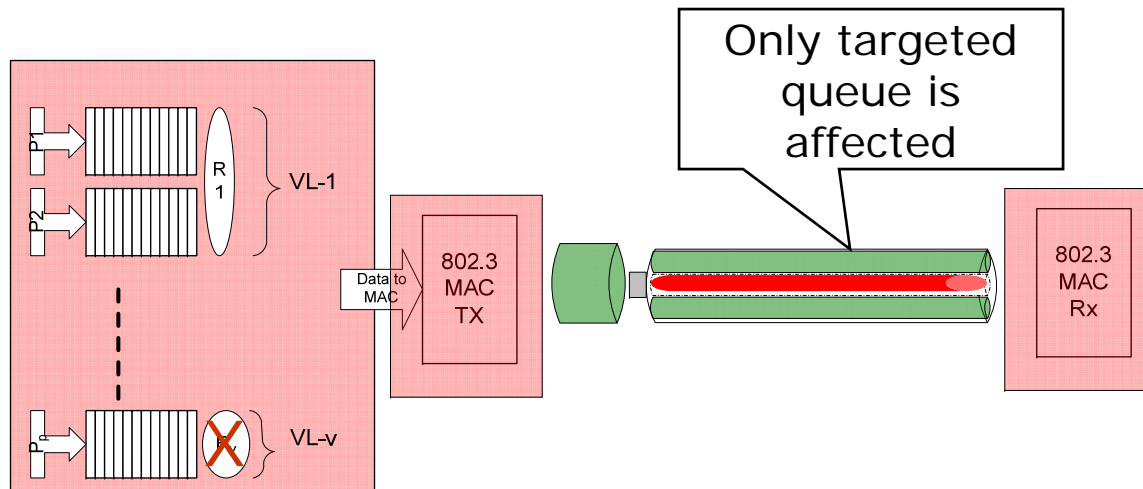


PG's allow latency optimization for one application while allowing throughput optimization for other application

Priority-based Flow Control (IEEE 802.1Qbb)

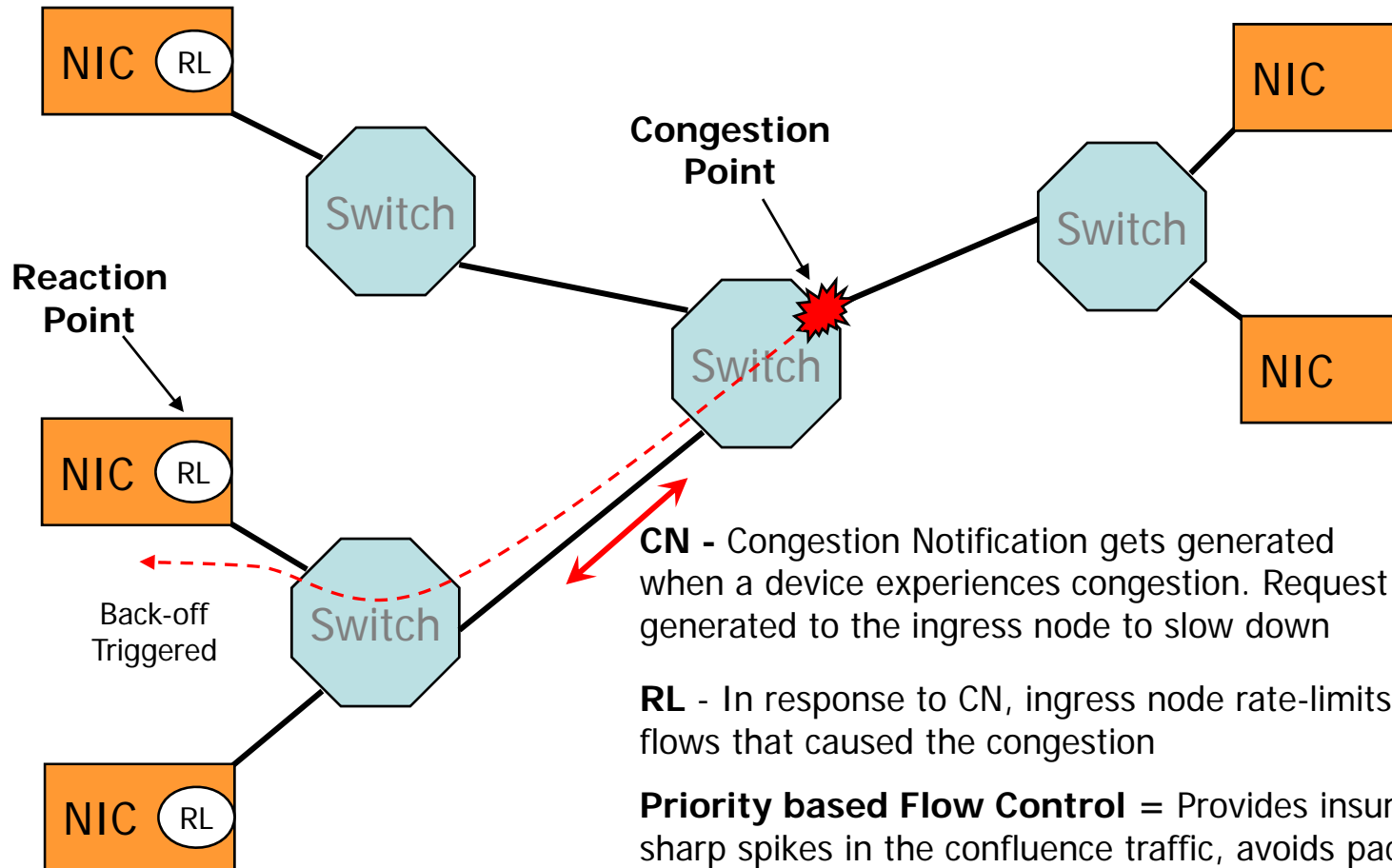


Link
Pause



Granular
Pause

Congestion Notification (IEEE 802.1Qau)

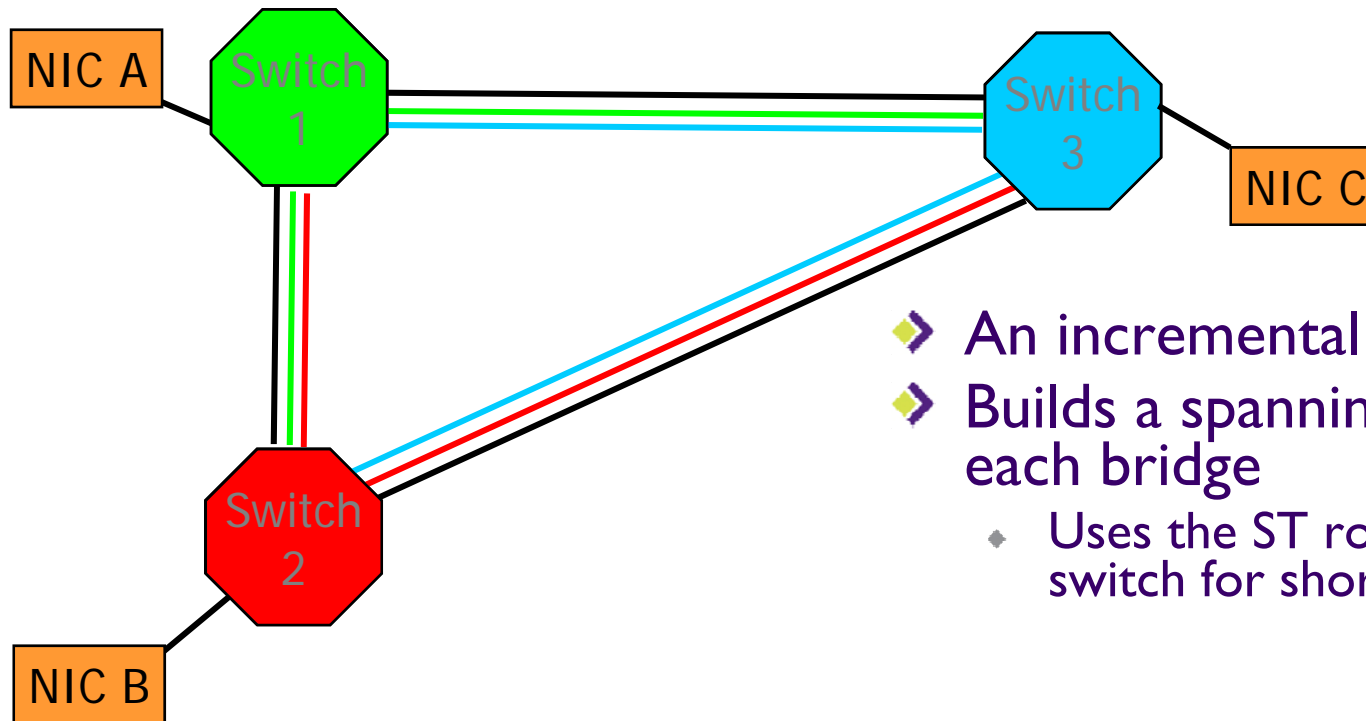


CN - Congestion Notification gets generated when a device experiences congestion. Request is generated to the ingress node to slow down

RL - In response to CN, ingress node rate-limits the flows that caused the congestion

Priority based Flow Control = Provides insurance against sharp spikes in the confluence traffic, avoids packet drops

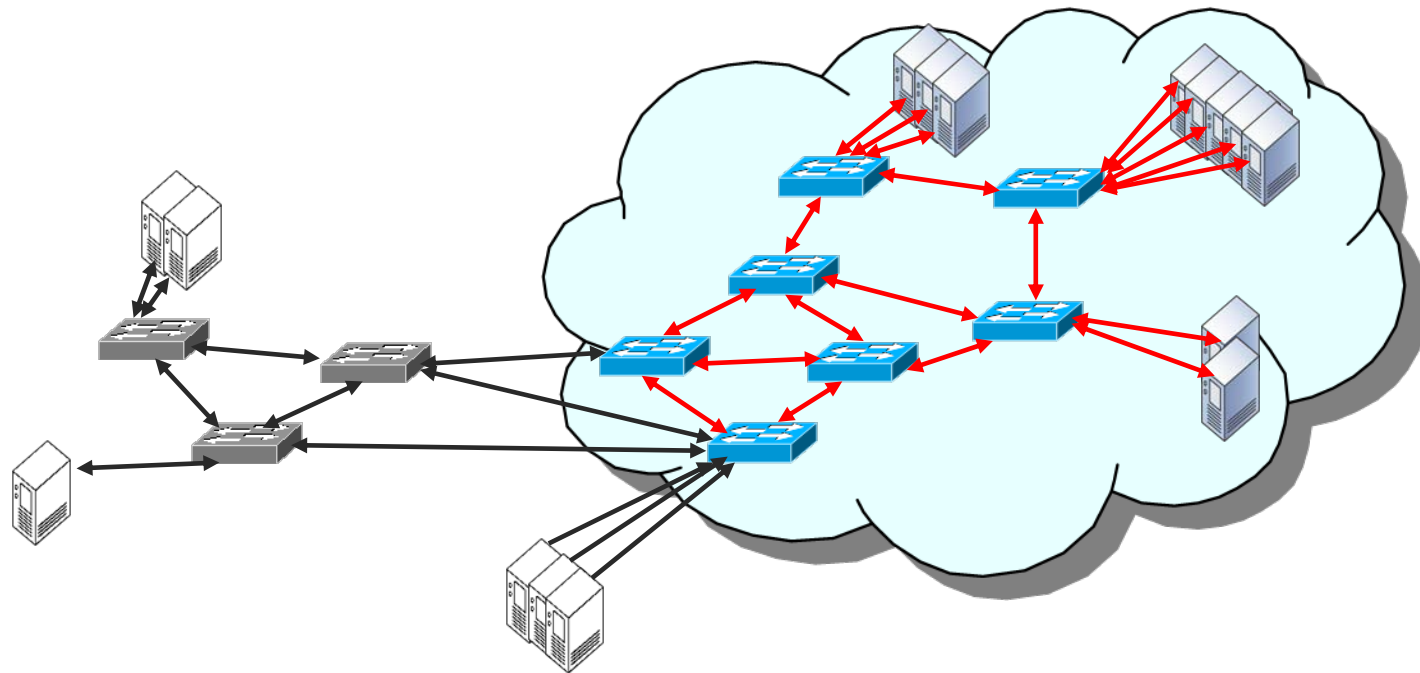
Shortest Path Bridging (IEEE 802.1aq)



- An incremental advance to MSTP
- Builds a spanning tree (ST) for each bridge
 - ◆ Uses the ST rooted at the source switch for shortest path bridging

- Ensures forward and reverse paths are aligned
 - ◆ Reflection Vector and other ideas being investigated to “align” spanning trees

Gluing it all together



- Capability Exchange Protocol allows discovery of compliant devices, capabilities
- Allows formation of cloud of compliant devices
- Allows incremental deployment – rack at a time

IEEE Enhancements for Data Center

- Effort underway to provide DC enhancements in IEEE
 - ◆ 25+ companies actively championing in IEEE
 - ◆ Work is called Data Center Bridging (DCB)
- IEEE projects necessary for I/O Consolidation in Data Center
 - ◆ Congestion Notification: Approved project IEEE 802.1Qau
 - ◆ Shortest Path Bridging: Approved project IEEE 802.1aq
 - ◆ Enhanced Transmission Selection: Approved project IEEE 802.1Qaz
 - ◆ Priority based Flow Control: Approved project in IEEE 802.1Qbb
 - ◆ DCB Capability Exchange Protocol: Part of various projects above
- DCB Standards trending for ratification in ~2009

10GbE build-out continues



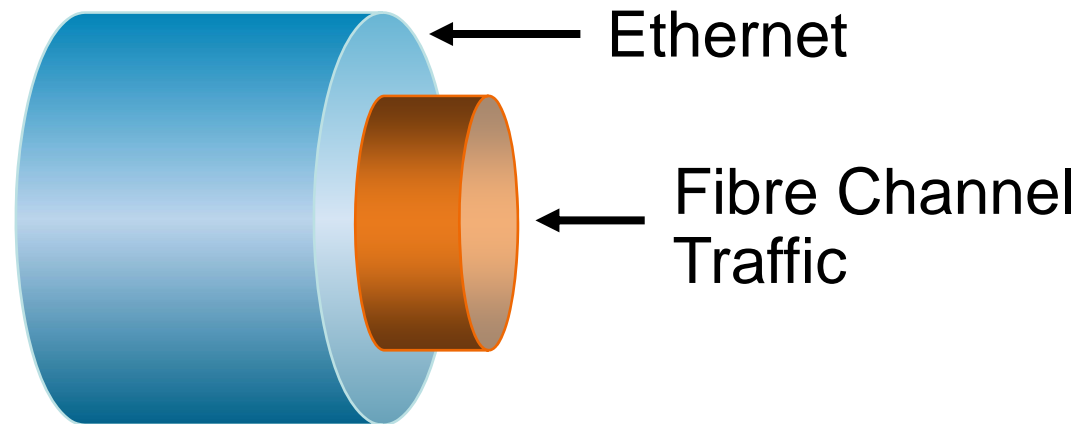
- Dense 10GbE Core switches
- Multiple top-of-rack switches announced
- Virtualization vendors support 10GbE
- New usage models drive lower cost and power

Lossless 10GbE is the fabric for I/O consolidation

FCoE: Fibre Channel over Ethernet

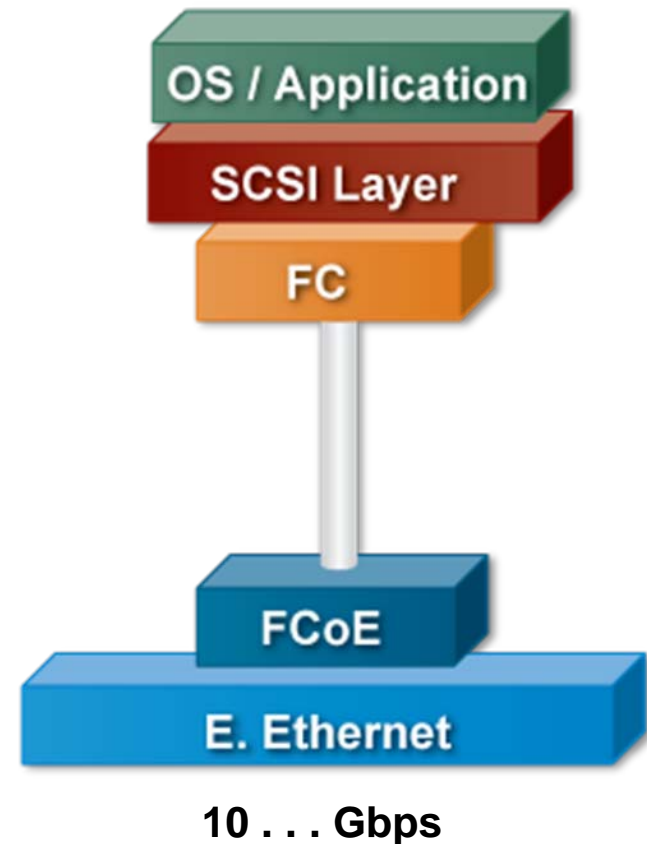
FCoE: FC over Ethernet

- FCoE is I/O consolidation of FC storage traffic over Ethernet
 - ◆ FC traffic shares Ethernet links with other traffics
 - ◆ Requires a lossless Ethernet fabric



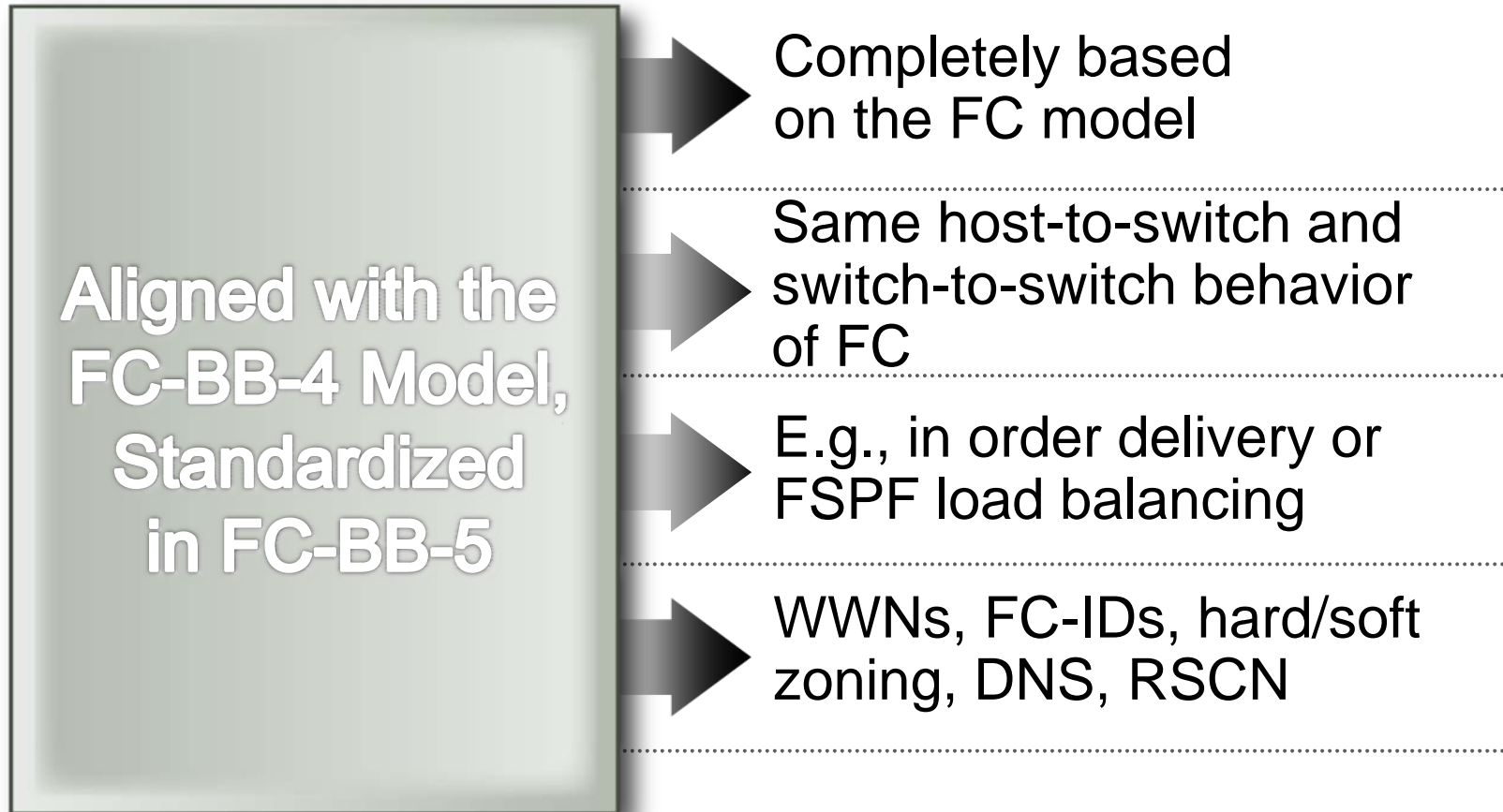
FCoE Protocol Stack

- From a Fibre Channel standpoint it's FC connectivity over a new type of cable called an Ethernet cloud
- From an Ethernet standpoint it's yet another ULP (Upper Layer Protocol) to be transported



FCoE is Fibre Channel

FCoE is Fibre Channel at the host and switch level



Protocol Organization

FCoE is really two different protocols:

FCoE itself

- Is the data plane protocol
- It is used to carry most of the FC frames and all the SCSI traffic

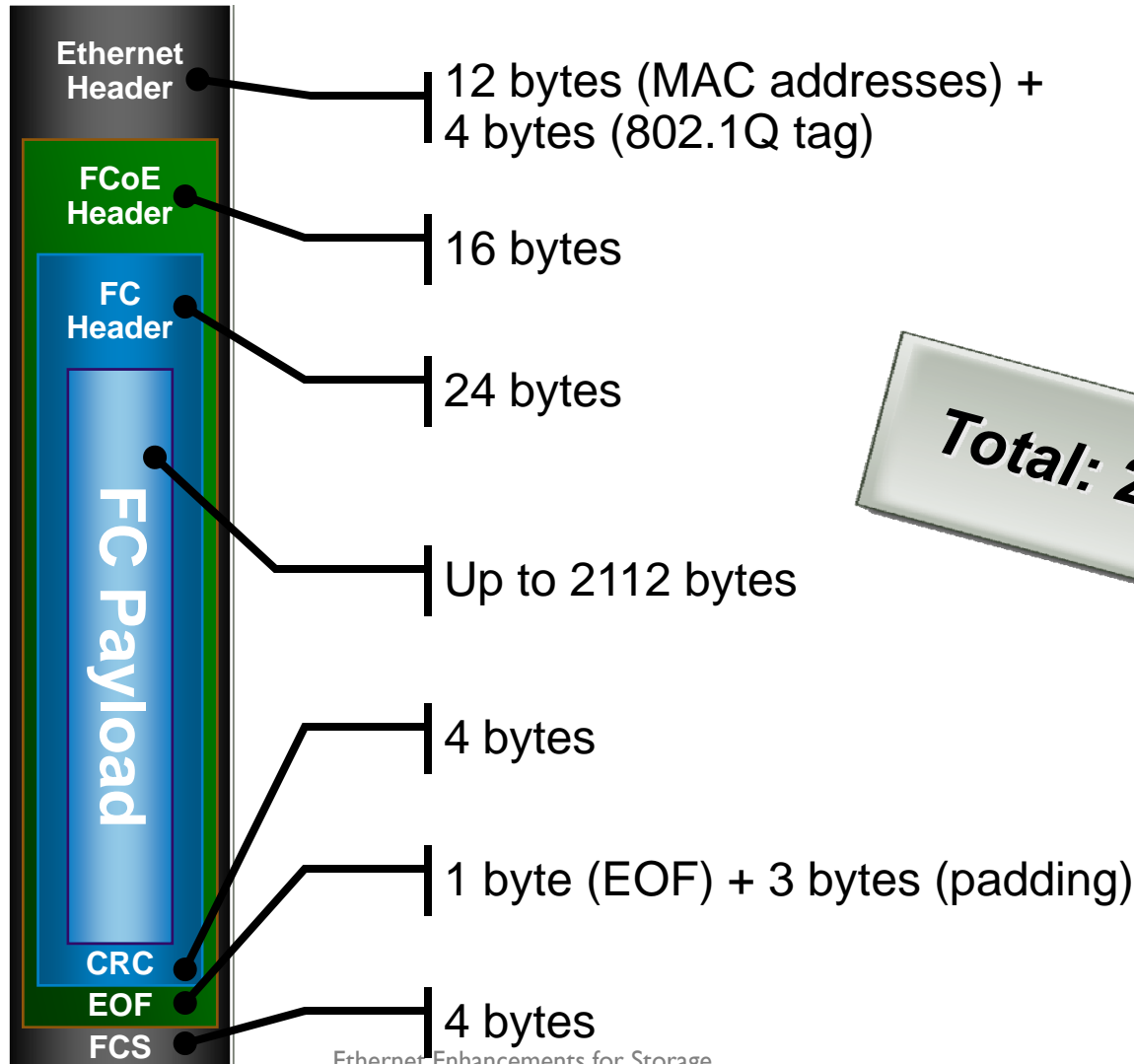
FIP (FCoE Initialization Protocol)

- Is the control plane protocol
- Is used to discover the FC entities connected to an Ethernet cloud
- Is also used to login to and logout from the FC fabric

The two protocols have:

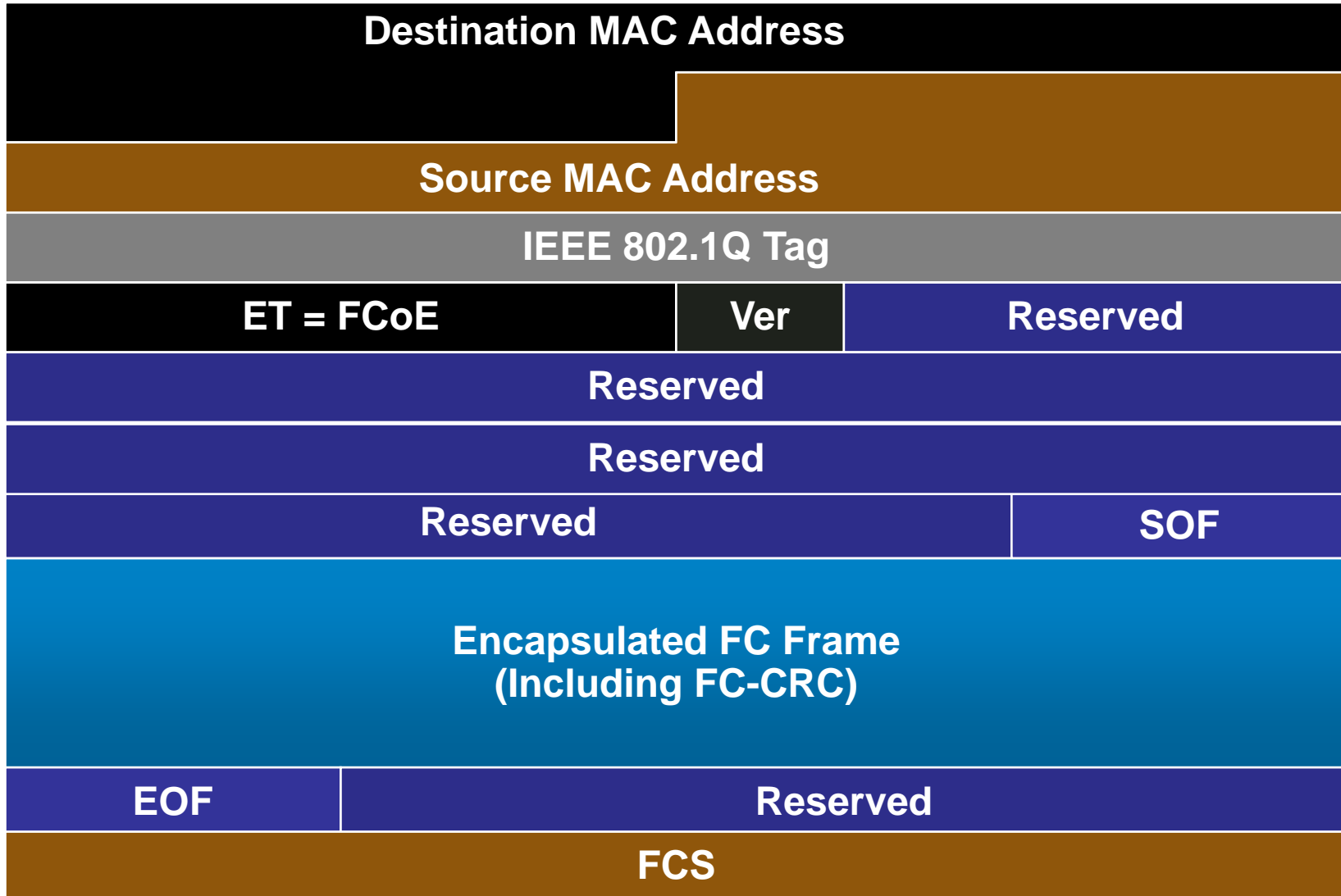
- Two different Ethertypes
- Two different frame formats

FCoE frame size



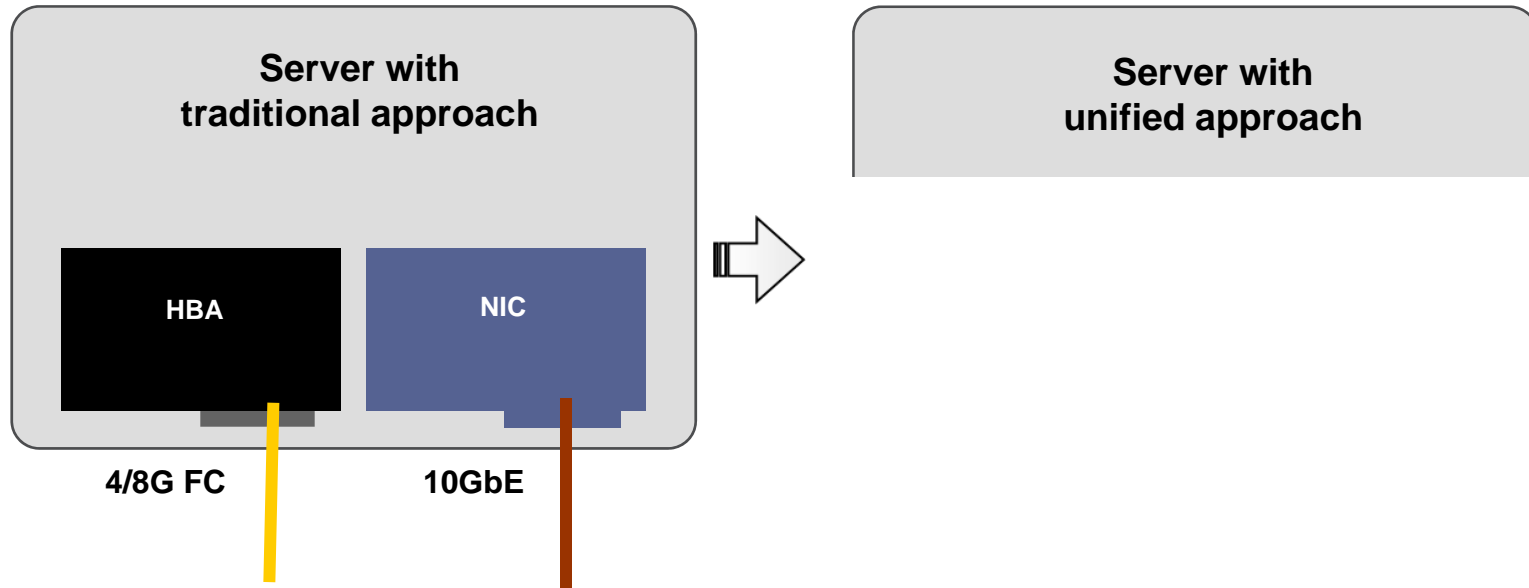
Total: 2180 bytes

FCoE Frame Format



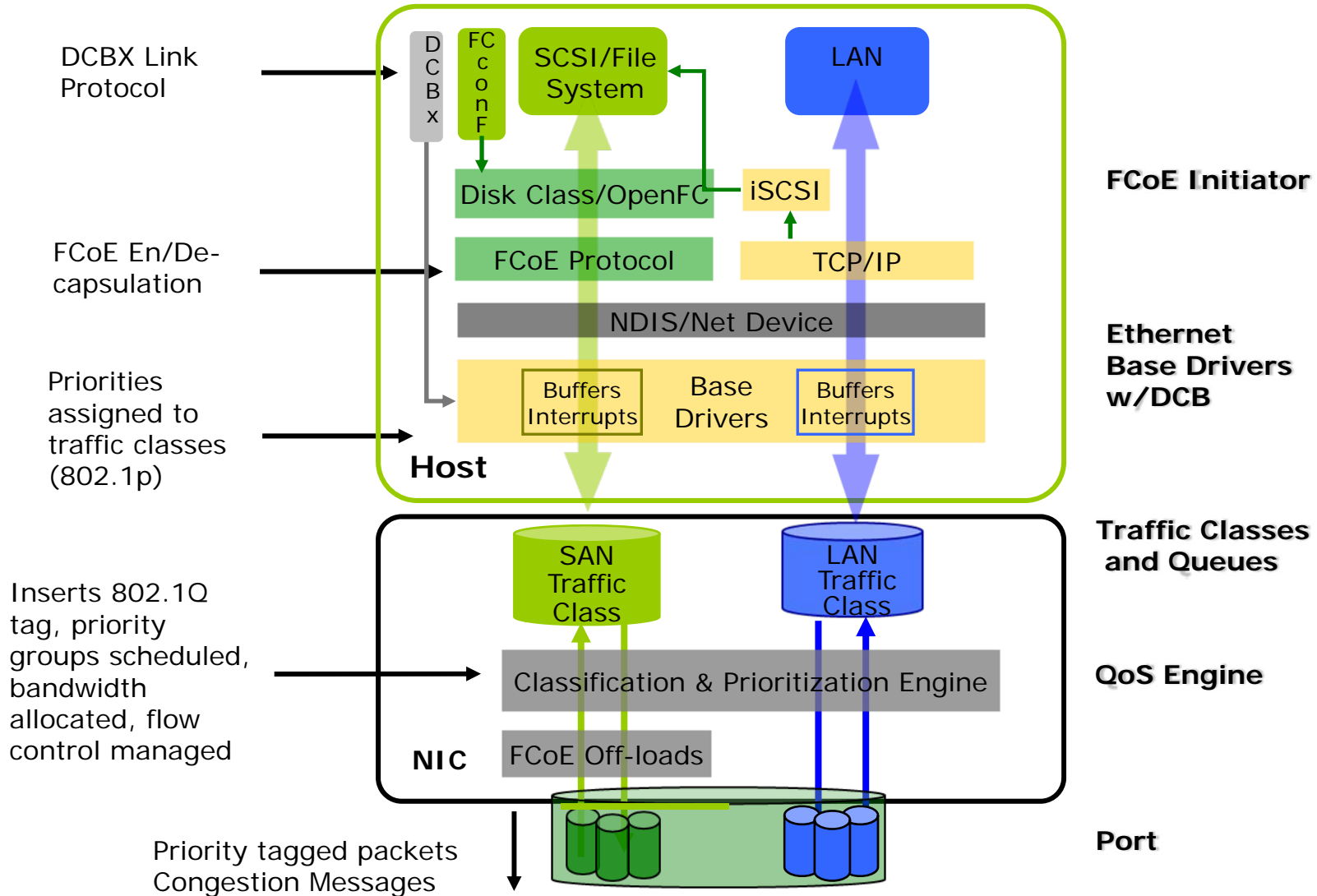
FCoE and I/O Consolidation [Server Perspective]

Unified Networking



- Enables cost effective SAN expansion in the data center
- Single adapter / single wire carries FC and IP traffic
- Continues to look like separate NIC & HBA to host OS

FCoE Stack



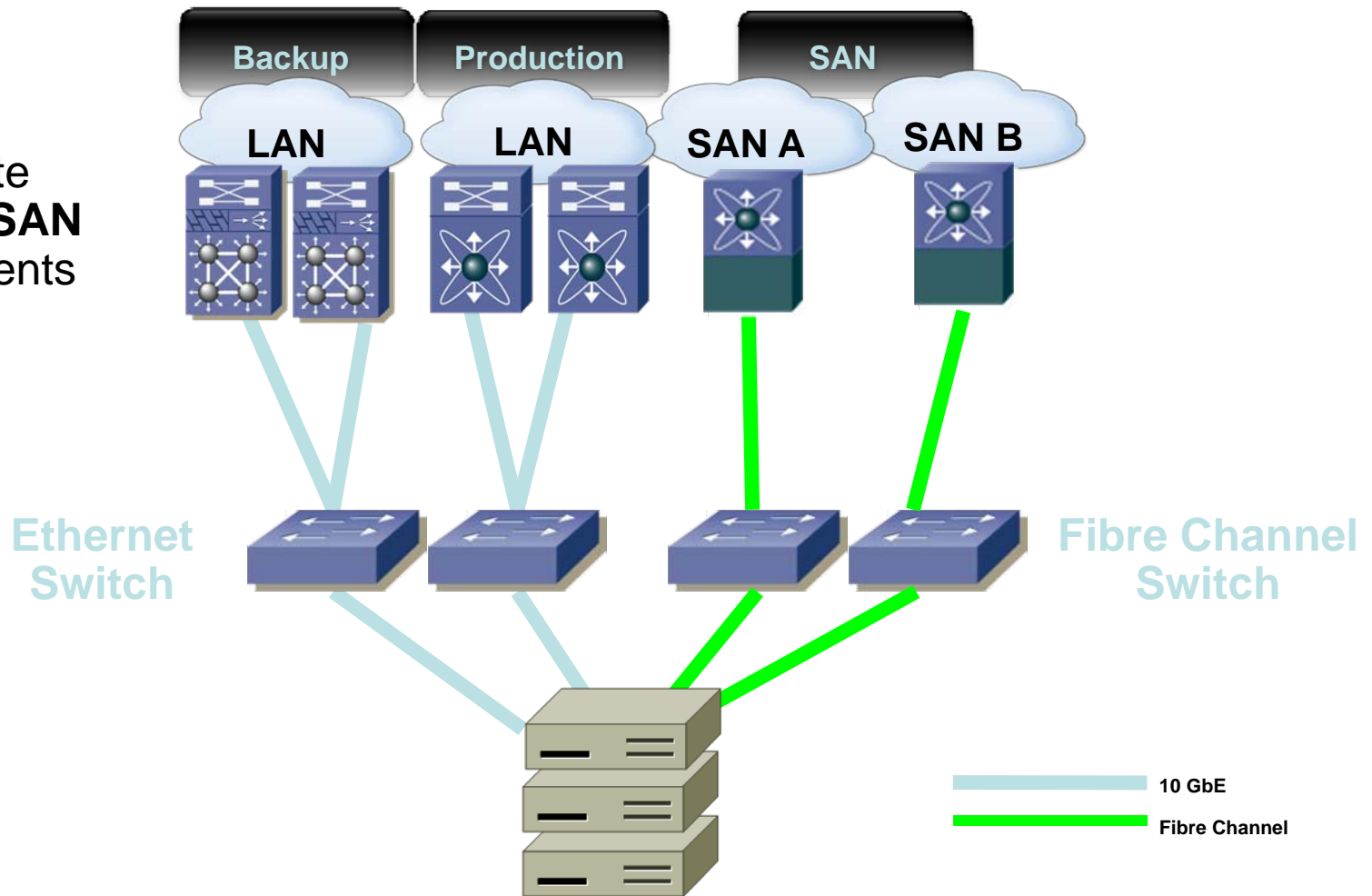
Open-FCoE

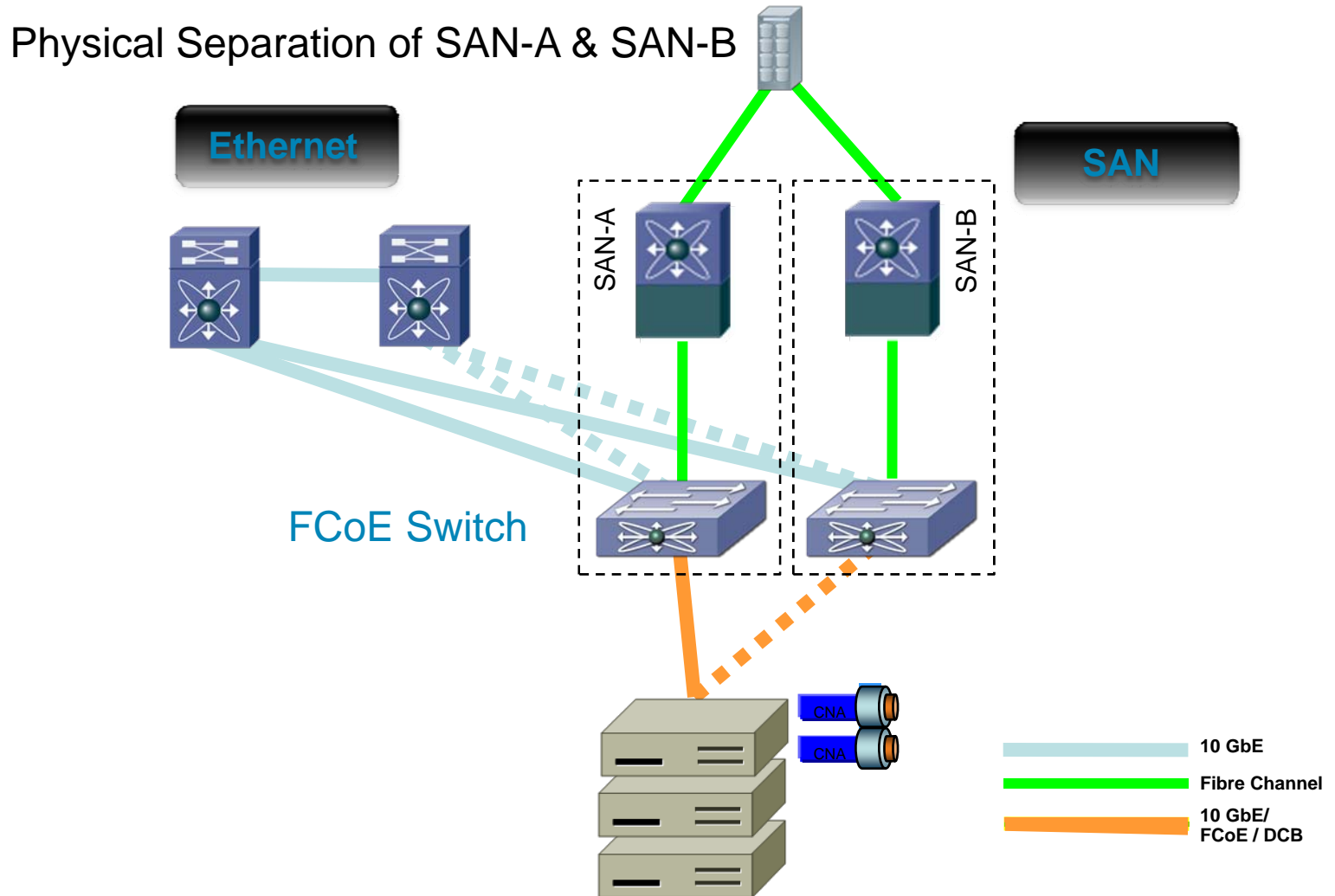
- The Open-FCoE project is an implementation of a FCoE initiator for the Linux OS targeting in-kernel acceptance and then distribution inclusion. The project enhances the Linux kernel by adding the following,
 - ◆ A modular and reusable FC library
 - ◆ A FCoE module that works for any Ethernet-capable adapter
- Open-FCoE is a LLD of the SCSI subsystem as is any existing FC HBA's driver
- Hosted at www.Open-FCoE.org
 - ◆ git repositories
 - ◆ Development mailing list
 - ◆ Bugzilla
 - ◆ Wiki
 - ◆ Announcements / Blog
- Building a FCoE community

FCoE Deployment

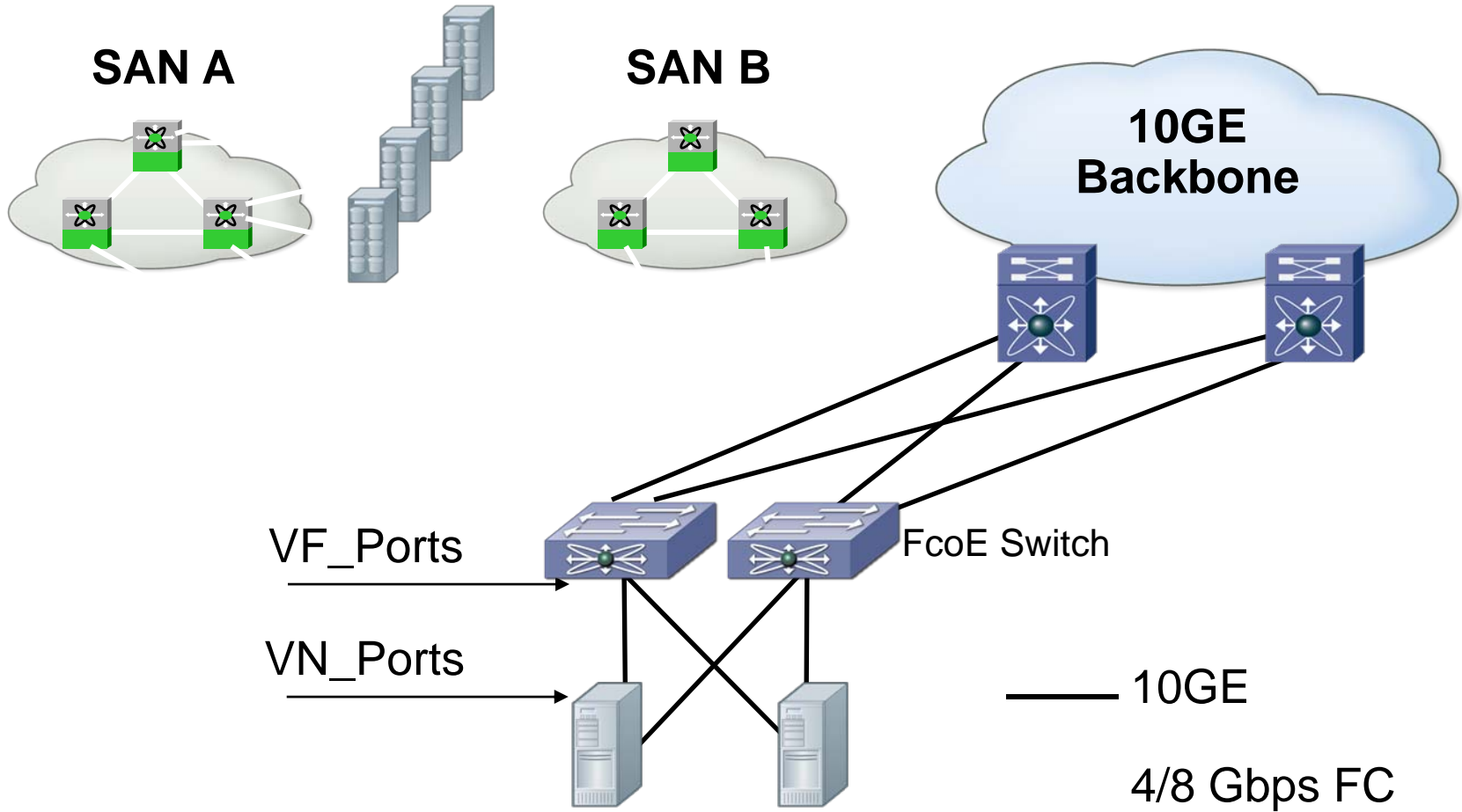
Existing Data Center Network Topology

Separate
LAN and **SAN**
Environments

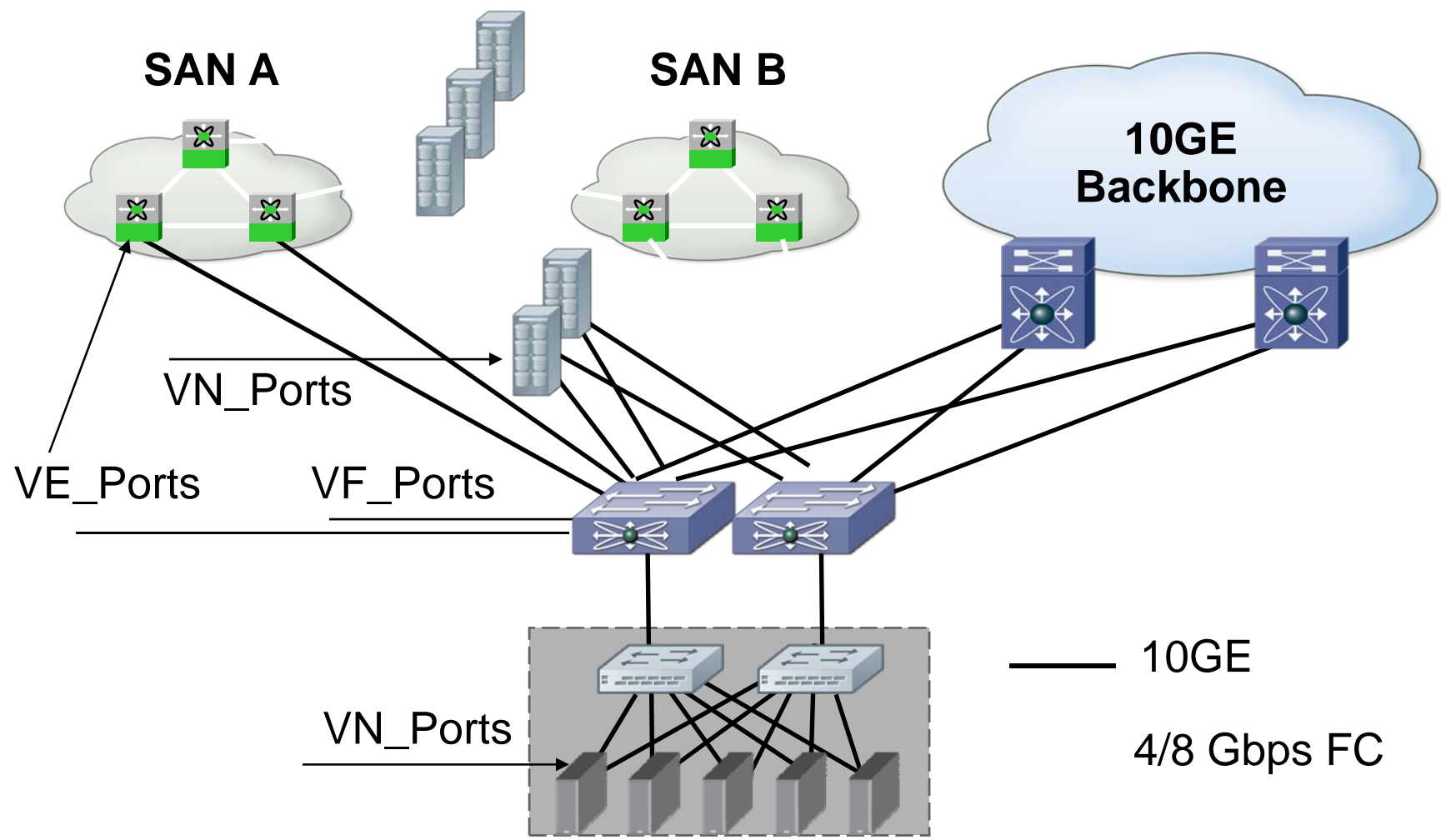




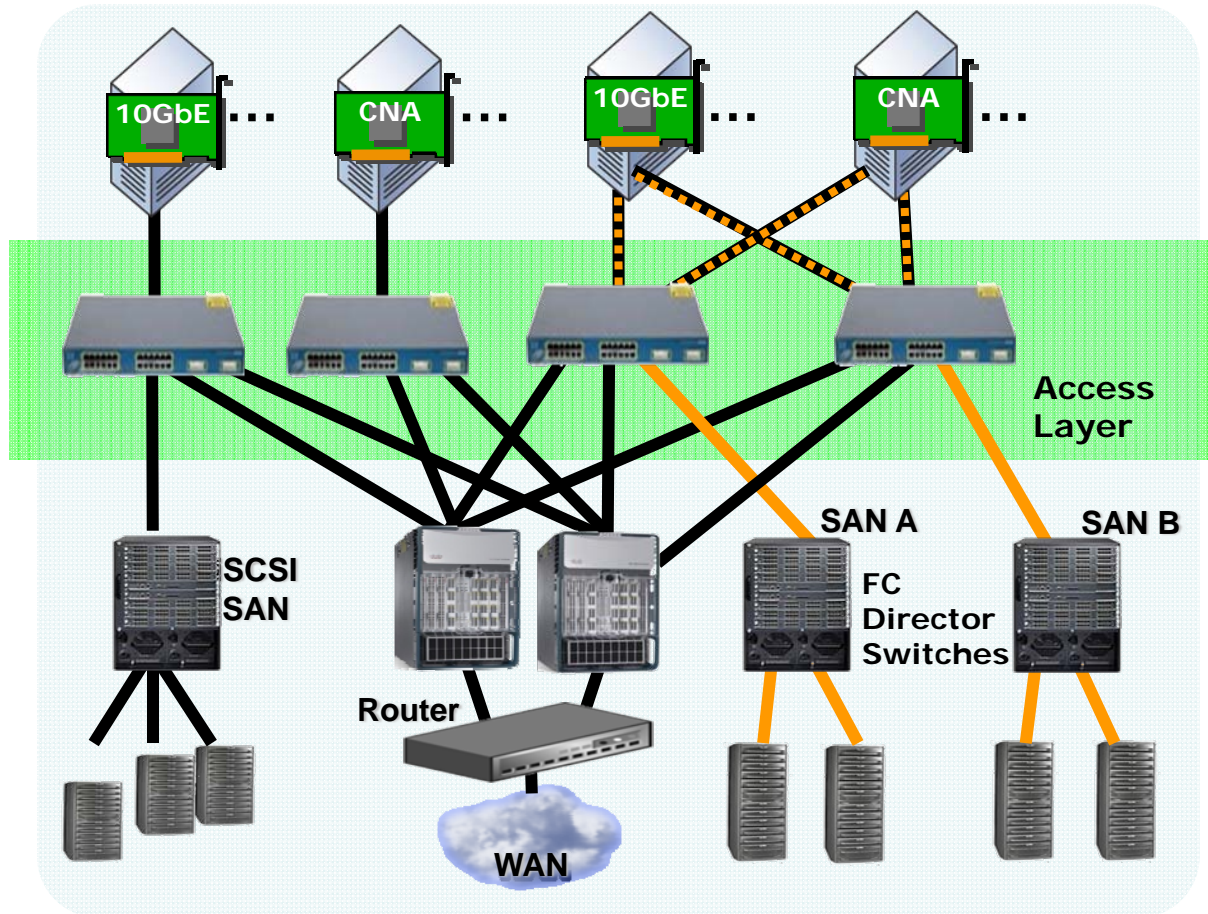
FCoE: Initial Deployment



FCoE: Adding Native FCoE Storage



FCoE Unified Access Deployment Model



Ethernet Enhancements for Storage

© 2008 Storage Networking Industry Association. All Rights Reserved.

..... FCoE — Ethernet — Fibre Channel

Case Study: Deploying FCoE



Goals

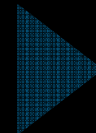


Assessment of 10GbE

Test FCoE in real life Environment



Objectives

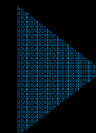


Characterize performance of 10GbE and FCoE

Cost analysis of 10G with FCoE and FC



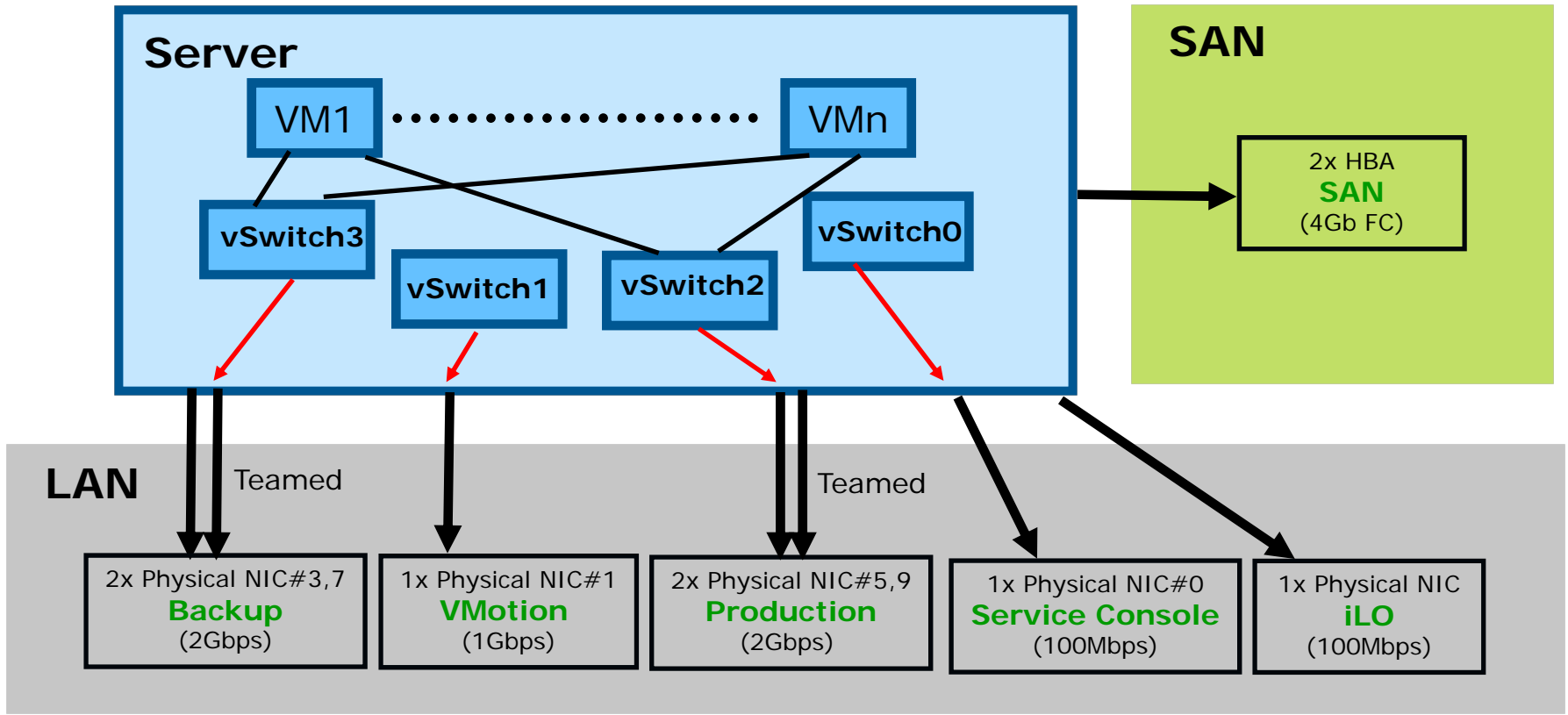
Motivation



Consolidate 9 ports (7 GbE + 2 FC) down to 2 10GbE

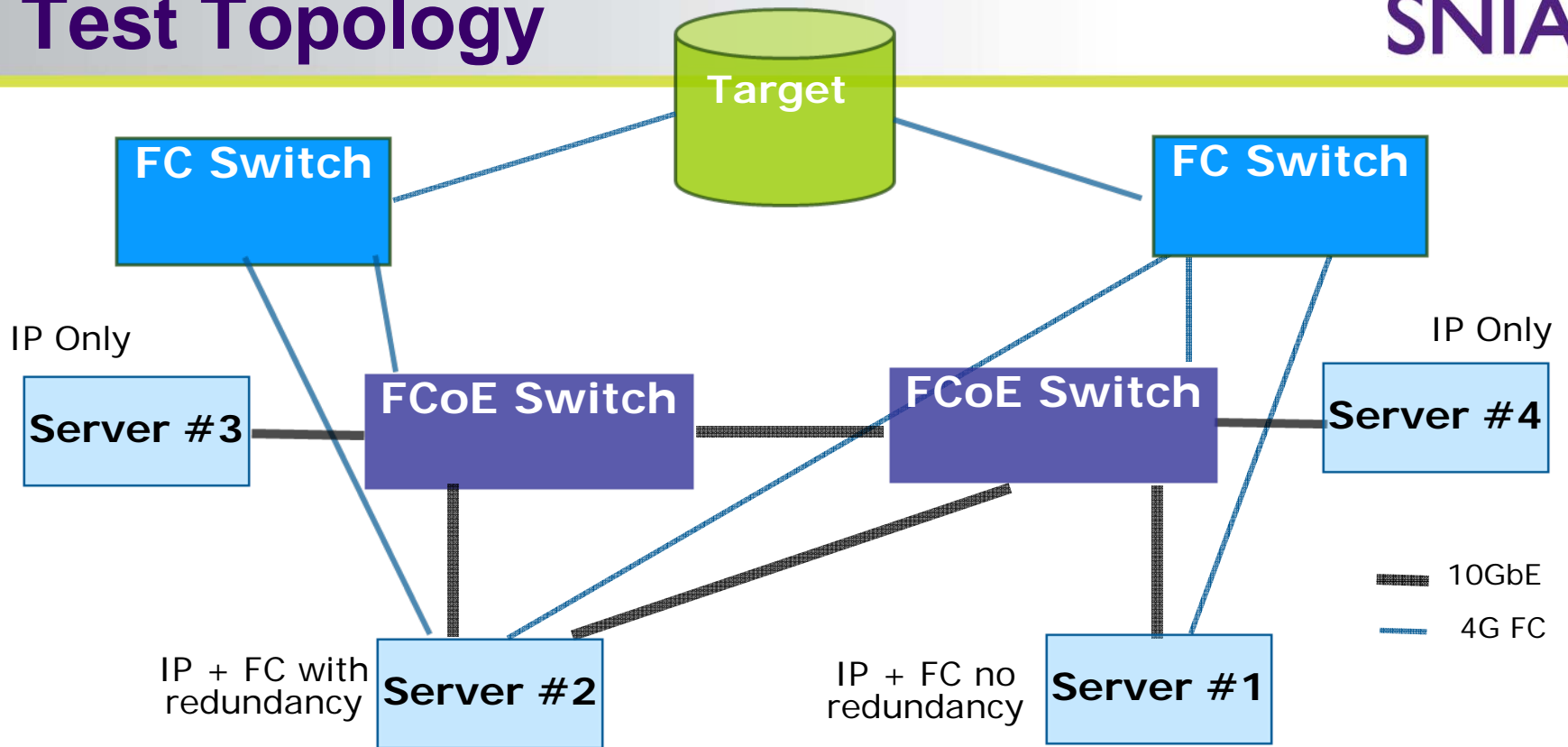
Reduce NAS ports with 10GbE

Typical Virtualized Server



Five Gig, Two Fast Ethernet and Two FC Connections

Test Topology



Source - Destination	Data Flow	Results
Server 1 – Server 2	Through Fibre Channel (baseline)	375 Mbps throughput
Server 3 – Server 4 (IP)	Through FCoE (new baseline)	375 Mbps throughput
Server 2 – Server 3	Mixed mode (FC and FCoE)	375 Mbps throughput

Test Results

- Performance characterization performed at 512B, 4KB, 8KB, 16KB, 32KB, 64KB, 128KB, 256KB, and 1MB block I/O sizes
- Max Network bandwidth observed through one port was 1.5 GBps
- No “penalty” for disk i/o when combining TCP/IP and FCoE traffic
- No network errors detected during any of the performance tests
- Induced failure of a port did not affect performance.

FCoE delivers same performance as FC and at 25% lower cost

Summary

- Data Center Bridging standards are driving Ethernet Enhancements for multiple traffic types
- Lossless 10GbE is the fabric for I/O consolidation
- Early adoption of FCoE is in the access layer
- Case Study results show that FCoE delivers same performance as FC and at 25% lower cost

- Please send any questions or comments on this presentation to SNIA: tracknetworking@snia.org

**Many thanks to the following individuals
for their contributions to this tutorial.**

- SNIA Education Committee

**Rob Peglar
Walter Dey
Steve Wilson
Joe White**