



Education

# **Storage Protocol Analyzers: Not Just for R & D Anymore**

Brandy Barton, Medusa Labs - Finisar

# SNIA Legal Notice

- 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.

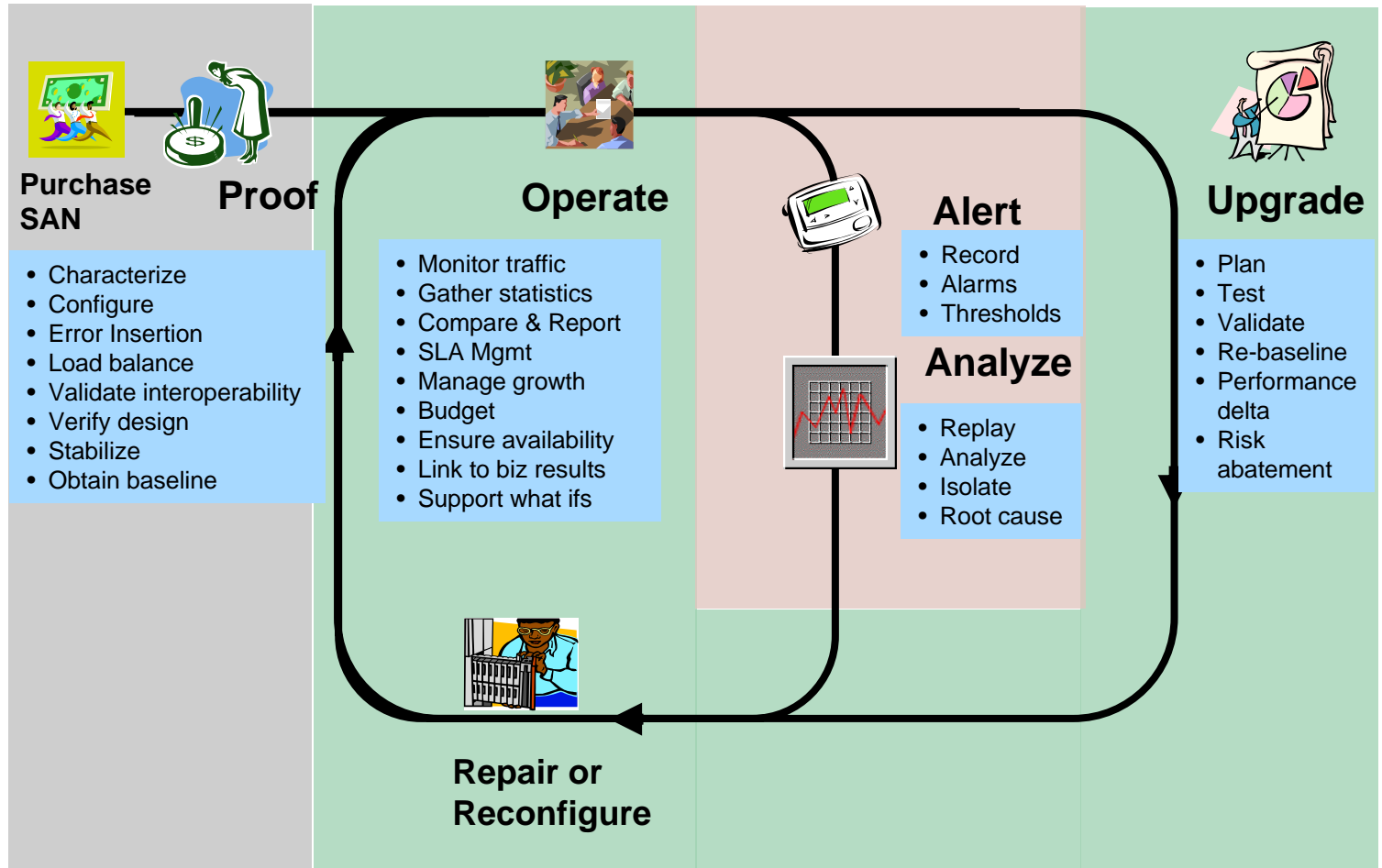
## Storage Protocol Analyzers – Not Just For R&D Anymore

In recent years, a growing number of data centers have begun to employ the use of traditional R&D analysis tools, such as protocol analyzers, into their production SAN environments. Reasons for this adoption have been cited as protecting the ROI of the SAN, decreasing downtime, reducing the risk of outages and preventing lost or corrupt data. This session will provide an overview of protocol analyzer tools and how to use them.

# Agenda Topics

- General SAN Tool Comparison
- How Protocol Analyzers Work
- Analyzer Placement
- Key Fibre Channel Concepts for Troubleshooting
- Key SAN Issues Found by Protocol Analyzers

# SAN Lifecycle



## Lifecycle Tool Matrix

	Proof	Operate	Alert	Analyze
Load Generators	<b>X</b>			
SAN Testers	<b>X</b>			
B.E.R.T.	<b>X</b>			
Switch Monitoring		<b>X</b>		
Storage Monitoring		<b>X</b>	<b>X</b>	
HBA Monitoring		<b>X</b>		
Network Monitoring		<b>X</b>	<b>X</b>	
Emulators	<b>X</b>	<b>X</b>		
Protocol Analyzers	<b>X</b>		<b>X</b>	<b>X</b>

# Protocol Analyzer Tools

## ➤ Proprietary implementations

- ◆ IO Traces Tools & Analysis group within SNIA working on standardization

## ➤ Capture I/O Traces

- ◆ Detailed protocol-level data
- ◆ No issue with heterogeneous configurations
- ◆ Do not see what is “inside” a server bus (e.g. HBA to OS, Storage Controller to Drive).

Bookmark	mm:ss.ms_us_ns (Rel)	DT/Port	Port	Count - OS - RCTI	Co
	01:17.980_991_497	0.038	FC_HBA 0	2 - Idle -	
Error	01:17.980_991_572	0.075	FC_HBA 0	1 - SOFn3 - ABTS	
	01:17.980_991_911	0.339	FC_HBA 0	3 - Idle -	
	01:17.980_992_005	0.094	FC_HBA 0	1 - R-Rdy -	
	01:17.980_992_043	0.038	FC_HBA 0	2 - Idle -	

General

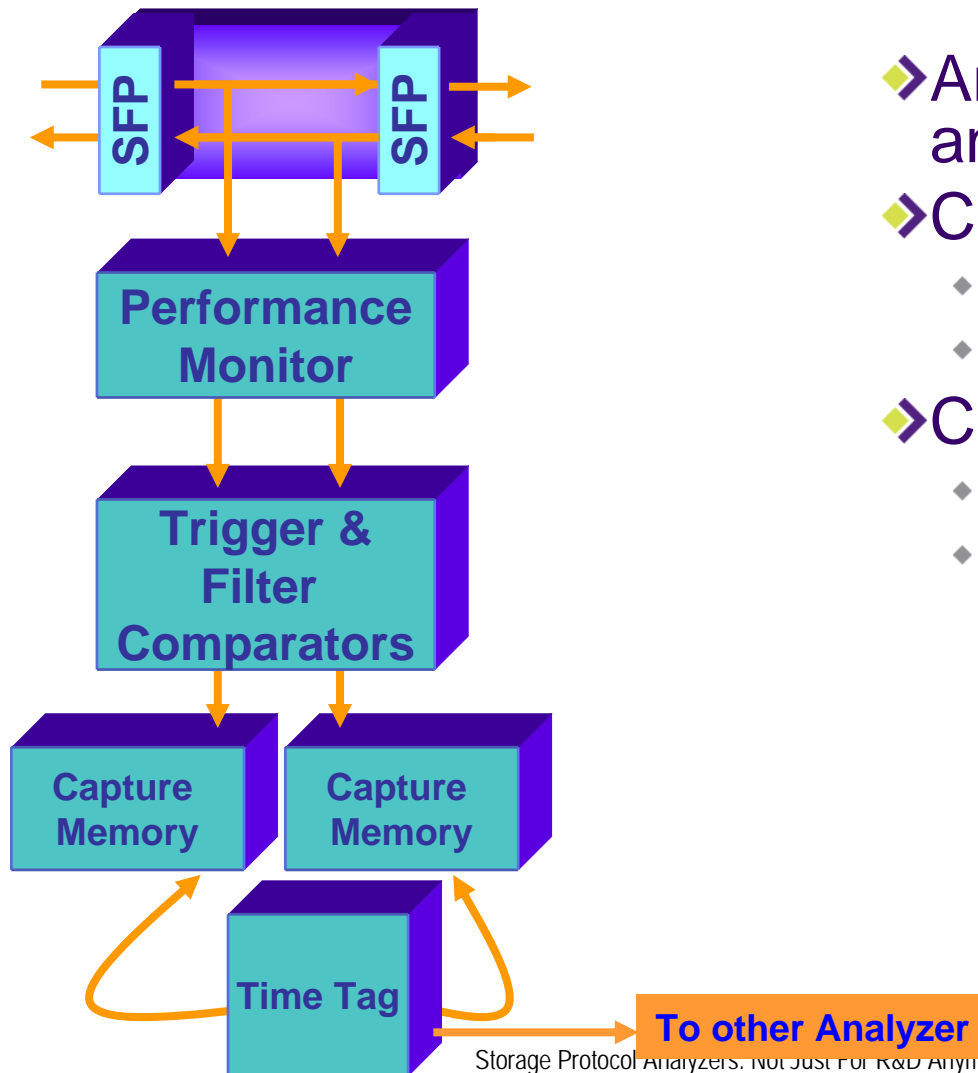
Tree 10 Bit

- [-] Fibre Channel Frame
  - SOE = 0xBCB53636 SOFn3
  - [-] Fibre Channel Header
    - RCTI = 0x81 ABTS
    - D\_id = 0x00000F
    - CS\_CTL = 0x00
    - S\_id = 0x000001
    - Type = 0x00
    - [+] F\_Ctl = **0x090008**
      - SEQ\_Id = 0x2A
      - DF\_Ctl = 0x00 No Opt Header
      - SEQ\_Cnt = 0x000C
      - OX\_Id = 0x0095
      - RX\_Id = 0xFFFF
      - PARA = 0x00007000
  - [-] End Of Frame
    - CRC = A256C220
    - EOF = EOF(-)

# Why Analyzers Are Making Their Way to the Data Center

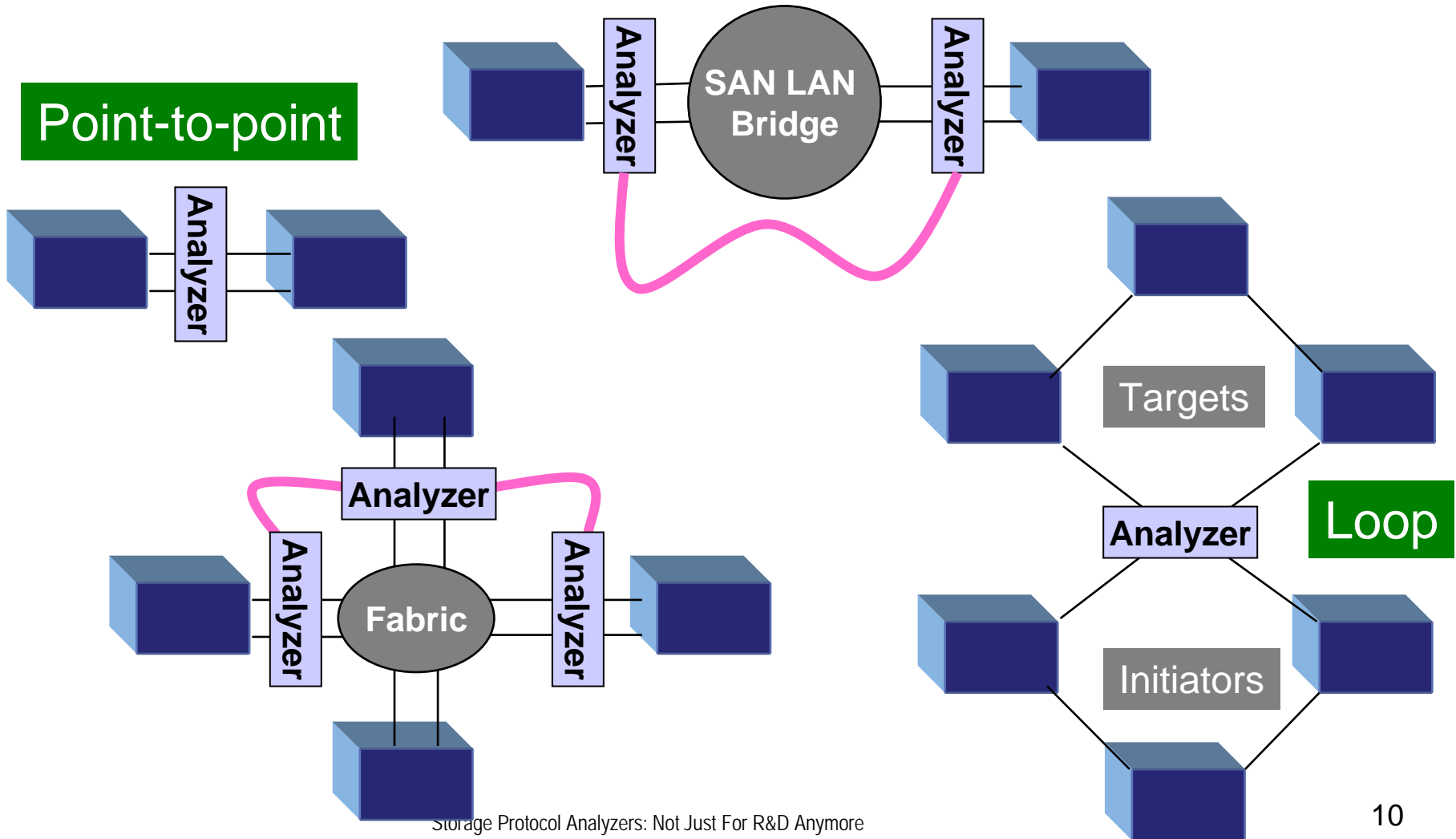
- Problem hardware and software must be pinpointed before changes are approved
  - ◆ SAN admins must prove they *know* where the problem is coming from.
- SAN error collection on intermediary hardware and software does not always show where the source of the problem resides.
- Switch trunking makes analyzing delivery issues more difficult.

# Analyzer Architecture



- An example of how an analyzer might work
- Choose position
  - ◆ In-line
  - ◆ Tap/sniff
- Choose probing method
  - ◆ Analog bypass (Repeating)
  - ◆ Retiming

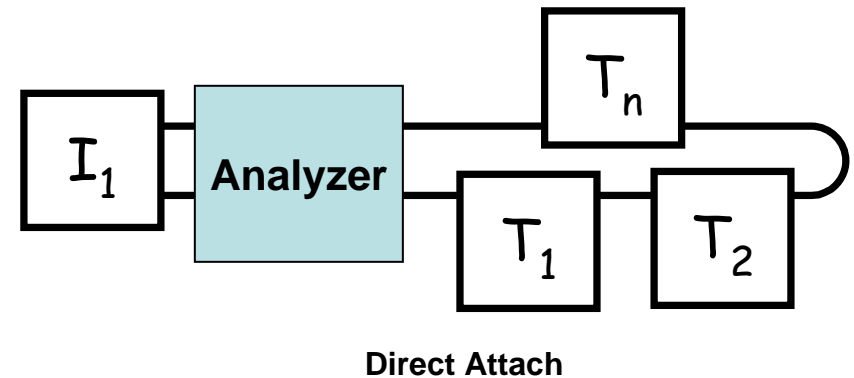
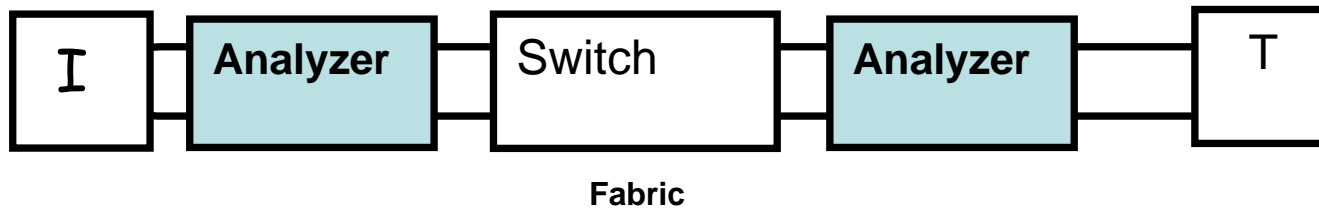
# Connect to any Topology (FC & iSCSI)



# Ways Into the Link

- ◆ In-line or in between the devices from which you want to capture the traffic or
- ◆ Placed outside the direct link with the use of an optical splitter (“sniffing”) or hub of optical splitters.

# In-Line Analyzer Placement

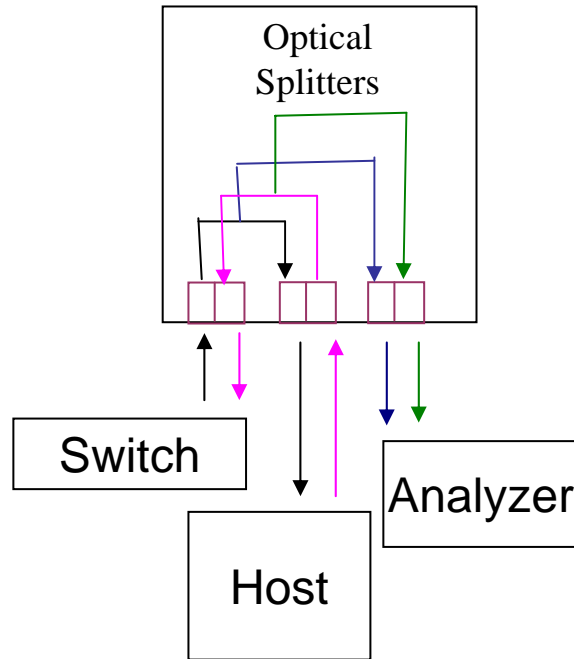


Key:

I – Initiator (Host)

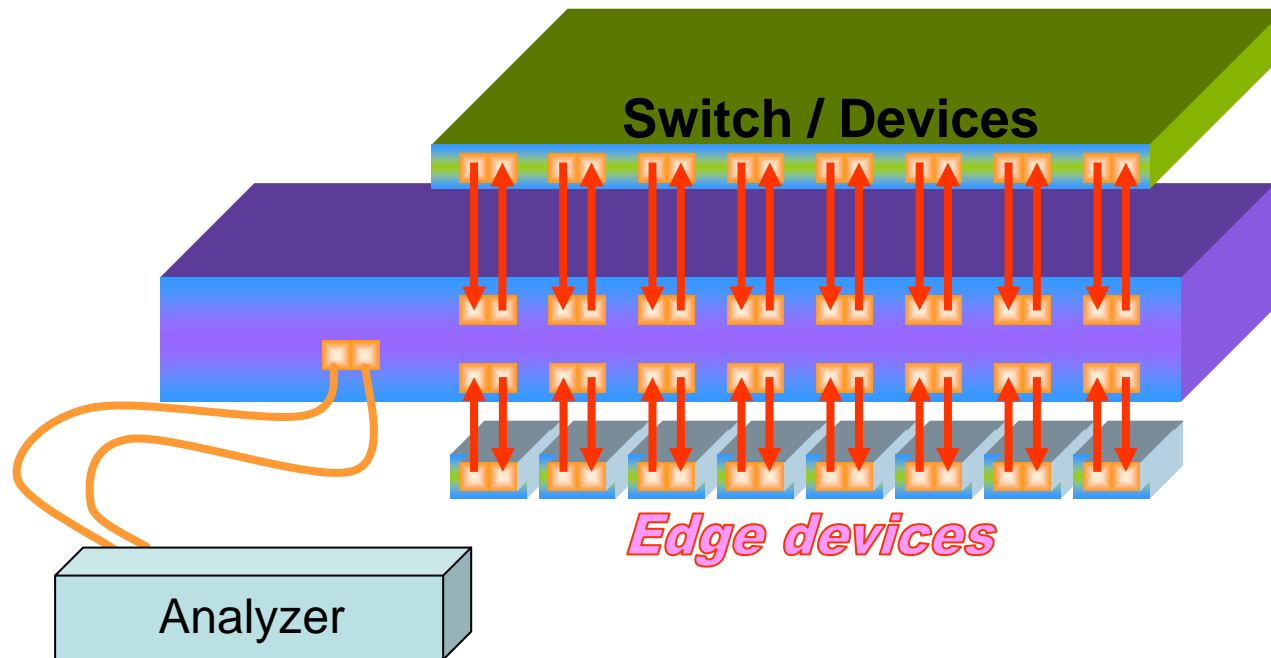
T – Target (Storage)

# Using Optical Splitters



- This analyzer is placed to capture traffic between the Host and the Switch by “sniffing” the link through a set of optical splitters.
- Optical splitters are installed while the network is down but once they are in you can add and remove analysis tools as needed

# Hub of Optical Splitters

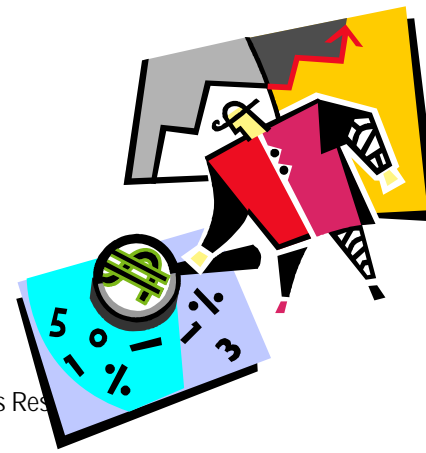


# Which Link Do I Look At?

- ▶ If you have only 2-ports you are challenged to place the analyzer in a position to capture the right data since you can only capture from one link at a time
  - ◆ Will see frames for SCSI I/O.
  - ◆ Will not see frames and credits between host and switch (or switch and storage) link if that's not the one you are plugged into.
  - ◆ May take moving the analyzer and taking more than one trace to accurately troubleshoot some issues.

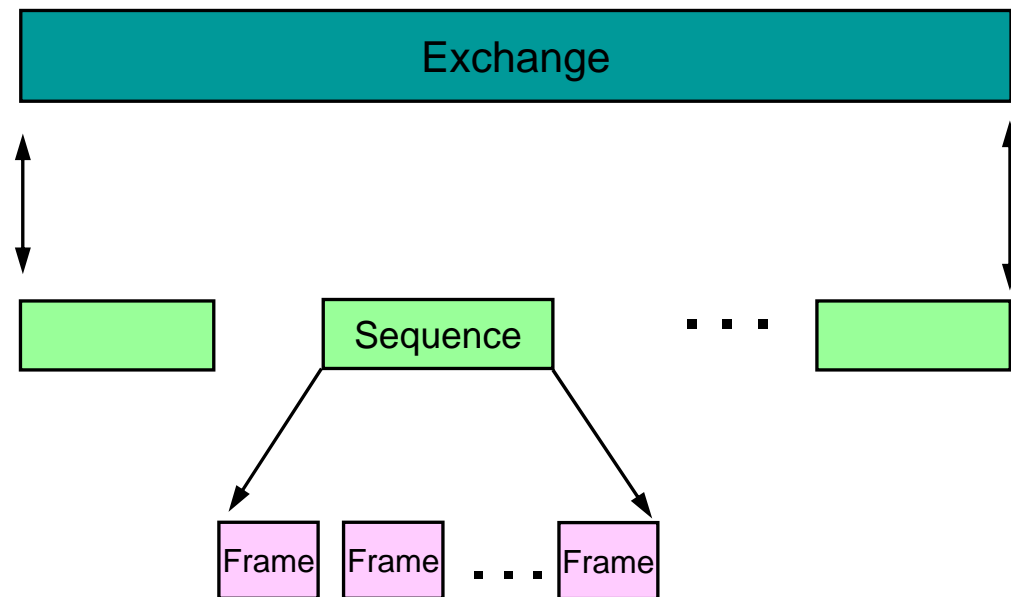
# Task: Troubleshooting I/O

- ▶ Sometimes finding the root cause of an issue requires a more focused look at the I/O flowing on the SAN.
  - ◆ This type of troubleshooting will require some familiarity with protocol specification standards.
- ▶ Data needed to provide analysis:
  - ◆ I/O capture (to “see” the flow of commands and their fulfillment).
  - ◆ Error



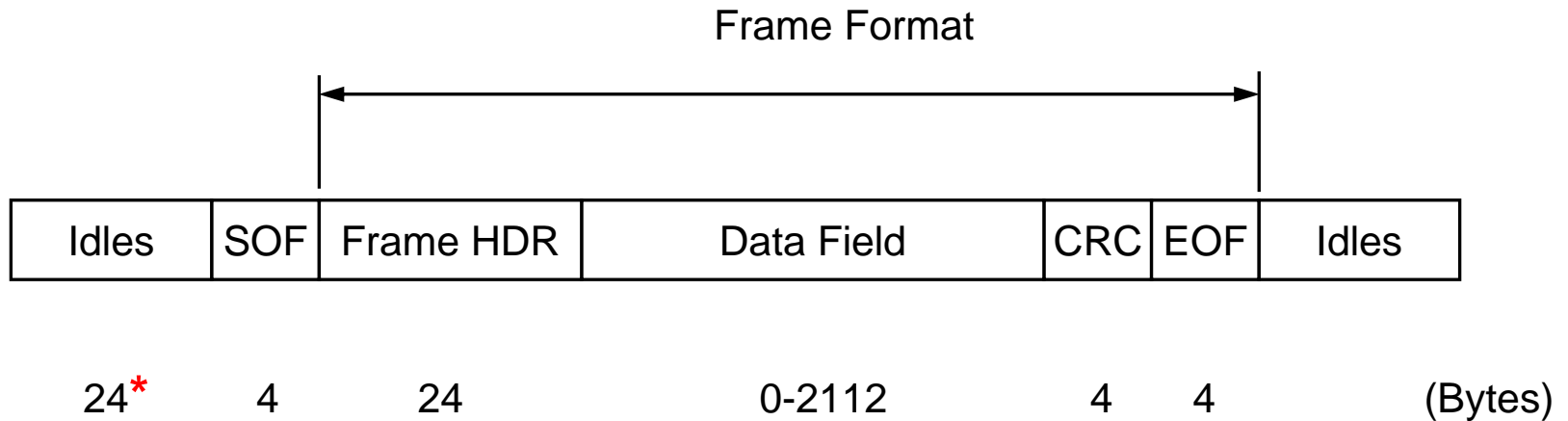
## ➤ The fundamental Protocol Constructs in Fibre Channel are:

- ◆ Exchanges
- ◆ Sequences
- ◆ Frames



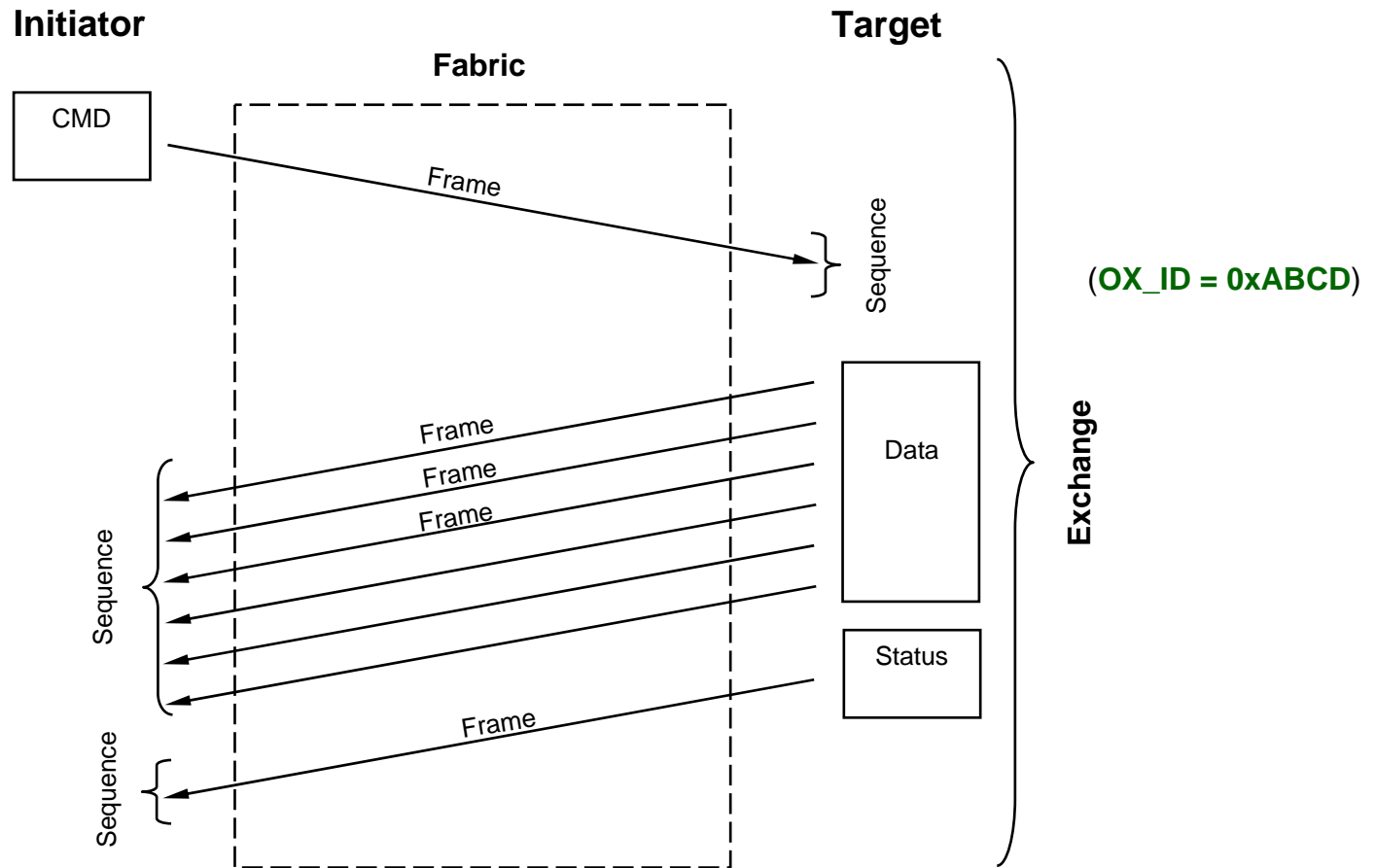
- The Upper Level Storage Protocol defines a “conversation” in order to exchange data
  - ◆ Command / Data / Status phases (SCSI)
- In Fibre Channel this conversation is called an Exchange
  - ◆ An Exchange consists of a set of related Sequences
  - ◆ Exchanges tend to be bi-directional
  - ◆ The Exchange is identified by the OX\_ID in Fibre Channel

# Frame Structure

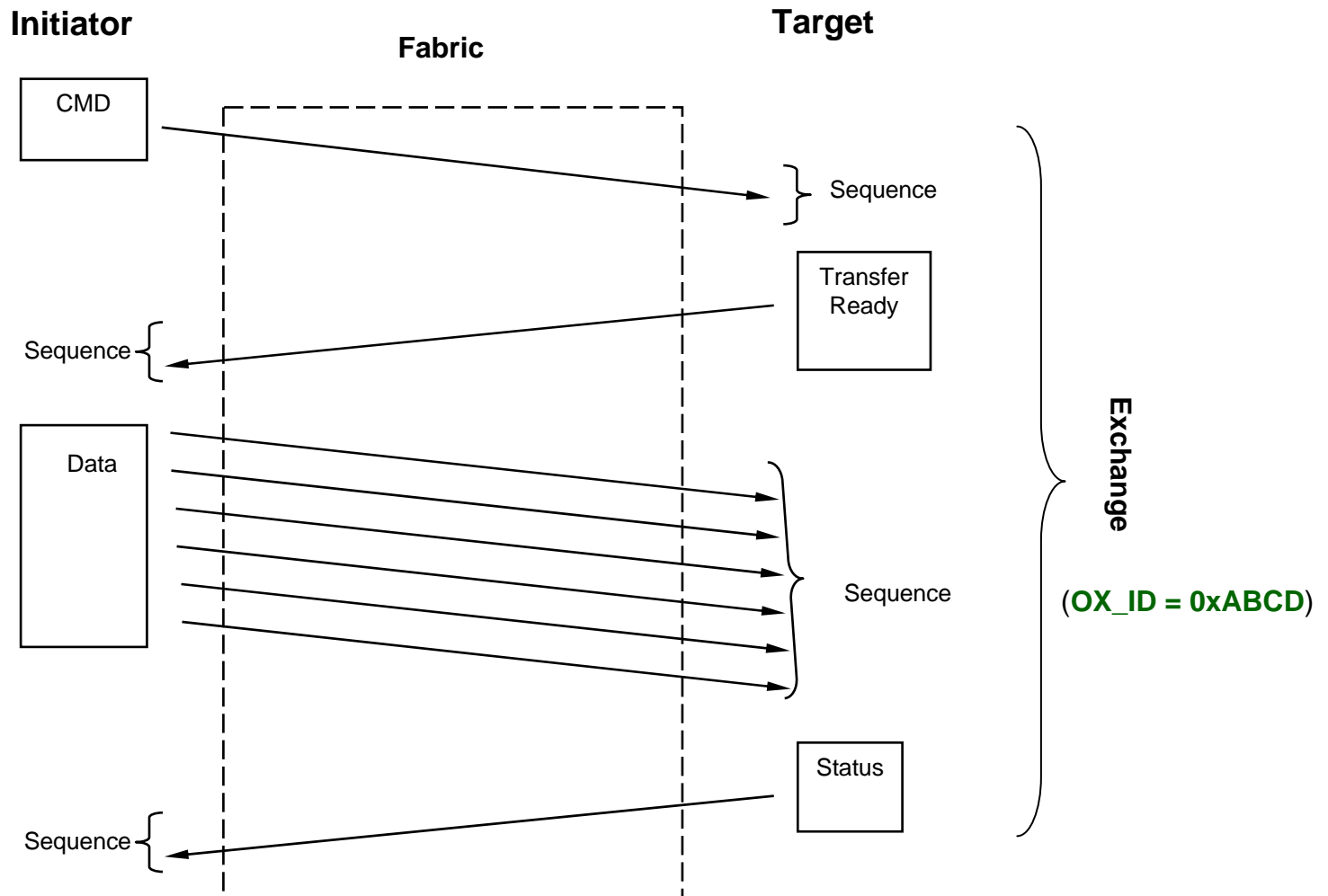


- \* 6 Idle word (24 bytes – 4 bytes each) required by transmitter  
2 Idle words (8 bytes) guaranteed to receiver

# SCSI Read Exchange



# SCSI Write Exchange/Task



# Upper Layer Protocol = SCSI

- 90+% of Fibre Channel implementations are transporting SCSI as the Upper Layer Protocol (ULP)
- A complete operation consists of the successful delivery of the SCSI Command (1 FC frame), if applicable to the command the SCSI Data (1-many FC frames), and the SCSI Status (1 FC frame).
- Failure to deliver any part of the SCSI operation will result in a SCSI Timeout situation.

- ▶ Most FC implementations are running Class 3 configurations:
  - ◆ No FC means of recovery for dropped or corrupted frames
  - ◆ Recovery is handled by SCSI
    - › Timeouts take from 60-120 seconds to occur depending on the OS on the host system.
  - ◆ Protocol Analyzers are the only tool available to pinpoint the actual failed I/O.



Education

## The Issue...SCSI Timeouts in FC

# What Happens When SCSI Fails & Times Out?

- Host (initiator) fails to deliver to the storage (target)
  - ◆ Target may return a Check Condition.
  - ◆ Initiator will normally retry the command (operation).
- Target fails to deliver to the host:
  - ◆ Initiator may return an ABTS (Abort Sequence) which terminates the operation.
  - ◆ Initiator will normally retry the operation.
  - ◆ (Note: some hosts do not send ABTS, they simply move to the retry process).

- In order to have enough data to analyze, all traces should adhere to the following:
  - ◆ Capture both frame and non-frame events
  - ◆ Capture Header plus a minimum of 64 bytes of payload
  - ◆ Valid and relevant data from both directions of the link must be captured (i.e. full duplex)

# Locating An Error

mm:ss.ms_us_ns (R)	Port	Count - Type	Count - Type	Summary	Destination [FC]	Source [FC]	OX_Id	RX_Id
00:08.112_589_550	FC Port(1,3,4)		1 - FC4Status	Good Status;	610000	6105DC	028D	FFFF
00:08.112_594_680	FC Port(1,1,2)		1 - FC4Status	Good Status;	610000	6105DC	028D	FFFF
00:08.112_658_935	FC Port(1,3,4)		1 - FC4SData	FC4SData; Offset = 0x00000000; Len = 0x800	610000	6105DC	026D	FFFF

The status came out of the Target ok.  
But it came out of the switch with a  
CRC error.

Destination & Source  
Device Addresses

Index	Hex	Interpretation	Error
FC 0000001	BC B5 56 56	SOF = SOFi3;	
FCH 0000001	07 61 00 00	Rctl = FC4Status; D_Id = 0x610000;	
FCH 0000002	00 61 05 DC	CS_CTL = 0x00; S_Id = 0x6105DC;	
FCH 0000003	08 98 00 00	Type = SCSI FCP; RX/OX = RX; S_C = S_C_; L_S; E_S;	
FCH 0000004	FF 00 00 00	SEQ_Id = 0xFF; DF_Ctl = No Opt Header; SEQ_Cnt = 0x0000;	
FCH 0000005	02 8D FF FF	OX_Id = 0x028D; RX_Id = 0xFFFF;	
FCH 0000006	00 00 00 00	PARA = 0x00000000;	
FCP 0000001	00 00 00 00		
FCP 0000002	00 00 00 00		
FCP 0000003	00 00 00 00	SCSI Status = Good Status;	
FCP 0000004	00 00 00 00	R_Count = 0x00000000;	
FCP 0000005	00 00 00 00	FCP_SNS_LEN = 0x00000000;	
FCP 0000006	00 00 00 00	FCP_RSP_LEN = 0x00000000;	
End 0000001	DE AD BE EF	CRC = 0xDEADBEEF (Computed: 0xD548C893);	CRC Error
End 0000002	BC 95 75 75	EOF = EOF(-);	

Affected Exchange  
OX\_ID = 0x028D

- Indicates corruption in the FC frame.
- Class 3 FC will drop this frame
  - ◆ The Status of this SCSI IO (the contents of this frame) will not be delivered.
- SCSI Initiator times out ~ 60 seconds.

# Determining the Fallout of the CRC Error – SCSI Timeout

A little over 57 seconds  
later...ABTS

mm:ss.ms_us_ns (R)	Port	Count - Type	Count - Type	Summary	Destination [FC]	Source [FC]	OX_Id	RX_Id
01:05.223_890_420	FC Port(1,1,1)	1 - FC4SData		FC4SData; Offset = 0x0000F800; Len = 0x800	6105E0	610000	015F	FFFF
01:05.223_898_900	FC Port(1,3,3)	1 - FC4SData		FC4SData; Offset = 0x0000F800; Len = 0x800	6105E0	610000	015F	FFFF
01:05.228_225_180	FC Port(1,1,1)	1 - ABTS		ABTS; Type = 0x00;	6105DC	610000	028D	FFFF
01:05.228_233_700	FC Port(1,3,3)	1 - ABTS		ABTS; Type = 0x00;	6105DC	610000	028D	FFFF
01:05.228_309_210	FC Port(1,1,1)	1 - FC4Cmd		Read(10); LUN = 0x0000; LBA = 0x00000A00; F	6105DC	610000	0145	FFFF
01:05.228_317_720	FC Port(1,3,3)	1 - FC4Cmd		Read(10); LUN = 0x0000; LBA = 0x00000A00; F	6105DC	610000	0145	FFFF
01:05.228_349_638	FC Port(1,3,4)		1 - BA_ACC	BA_ACC; Type = 0x00; OX_ID Aborted = 0x028D	610000	6105DC	028D	FFFF
01:05.228_354_750	FC Port(1,1,2)		1 - BA_ACC	BA_ACC; Type = 0x00; OX_ID Aborted = 0x028D	610000	6105DC	028D	FFFF

Note: the association is  
made from the OX\_ID,  
0x028D

# Check Condition Cause of SCSI Timeout Condition

mm:ss.ms_us_ns (R)	Port	Count - Type	Count - Type	Summary	Destination	Source	OX_Id	RX_Id	SEQ_Cnt	PARA
01:17.946_113_444	FC_HBA 0	1 - FC4Cmd		Write(10); LUN = 0x0000; LBA = 0x00001F00; FCP_DL = 0x00010000;	00000F	000001	0095	FFFF	0000	00004000
01:17.980_488_555	FC_HBA 1		1 - FC4XRdy	DATA_RO = 0x00000000; BURST_LEN = 0x00010000;	000001	00000F	0095	FFFF	0000	00000000
01:17.980_620_670	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00001000; Len = 0x800	00000F	000001	0095	FFFF	0000	00001000
01:17.980_640_537	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00000800; Len = 0x800	00000F	000001	0095	FFFF	0001	00000800
01:17.980_660_397	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00001000; Len = 0x800	00000F	000001	0095	FFFF	0002	00001000
01:17.980_680_227	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00001800; Len = 0x800	00000F	000001	0095	FFFF	0003	00001800
01:17.980_700_050	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00002000; Len = 0x800	00000F	000001	0095	FFFF	0004	00002000
01:17.980_719_917	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00002800; Len = 0x800	00000F	000001	0095	FFFF	0005	00002800
01:17.980_768_285	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00003000; Len = 0x800	00000F	000001	0095	FFFF	0006	00003000
01:17.980_788_116	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00003800; Len = 0x800	00000F	000001	0095	FFFF	0007	00003800
01:17.980_847_834	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00004000; Len = 0x800	00000F	000001	0095	FFFF	0008	00004000
01:17.980_867_684	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00004800; Len = 0x800	00000F	000001	0095	FFFF	0009	00004800
01:17.980_885_406	FC_HBA 1		1 - FC4Status	Check Condition; RO Mismatch; ABORTED COMMAND; DATA PHASE ERROR;	000001	00000F	0095	FFFF	0000	00000000
01:17.980_887_514	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00005000; Len = 0x800	00000F	000001	0095	FFFF	000A	00005000
01:17.980_907_345	FC_HBA 0	1 - FC4SData		FC4SData; Offset = 0x00005800; Len = 0x800	00000F	000001	0095	FFFF	000B	00005800
01:17.980_991_572	FC_HBA 0	1 - ABTS		ABTS; Type = 0x00;	00000F	000001	0095	FFFF	000C	00007000
01:17.983_720_022	FC_HBA 1		1 - BA_ACC	BA_ACC; Type = 0x00; OX_ID Aborted = 0x0095; RX_ID Aborted = 0xFFFF;	000001	00000F	0095	FFFF	000C	00007000

1. Data offsets overlap – Parameter Field

2. SCSI Target cites Check Condition for Relative Offset Mismatch. Aborts Command.

3. Initiator sends ABTS. Target accepts it.

# What Do I Do Now?

- Congratulations, you have just identified the source of a single IO failure!!
- Take additional captures to watch for increased errors on the link upon which the CRC error occurred
- Plan and justify a course of action (e.g replace a cable, replace an SFP, call your vendor)



- Please send any questions or comments on this presentation to SNIA: [trackstorage@snia.org](mailto:trackstorage@snia.org)

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

*SNIA Education Committee*

**Steve Klotz  
Jim Marrone**

**Elaine Silber  
Rob Peglar**