



# Storage Area Network Extension Solutions and Their Performance Assessment

Radha Telikepalli, Tadeusz Drwiega, and James Yan,  
Nortel Networks

Presented by: Xiangnan Xu



# Outline

- Introduction
- Protocols used for Storage Area Networks
- Performance Analysis
- Conclusions



# Introduction

- Origin
- With the emergence of client-server architecture, the information is now distributed across a network.
- Definition
- A SAN is a high-speed network that contains different physical connections between different storage elements, computer systems, and a management layer to organize the connections.
- Usage
- SAN is mainly used to provide secure and robust data transfer between computer systems and storage elements, and among storage elements.
- Storage Area Networking
- The practice of creating, installing, and administering SANs



# Introduction

- Storage Area Network Extension Solutions
- SONET-based networks: Fibre-Channel or end-to-end Fibre-Channel (FC)
- IP-based networks: Internet SCSI (iSCSI), Internet Fibre Channel Protocol (iFCP), and Fibre Channel over TCP/IP (FCIP)



# Introduction

- Evaluation Measurements
- Ease of implementation
- Costs involved
- Complexity of network management
- Interfaces
- hardware- or software-based implementation
- Technical capabilities( e.g., throughput and latency)



# Introduction

- Synchronous data replication has a limitation on the distance between SAN islands because of its low tolerance to delays. So asynchronous replications are more suitable for SAN extensions over long distance.
- SAN extensions require transfer capability adequate to transport huge amounts of data within the given time limit, so throughput is a key characteristic.



# Introduction

- Contribution
- Provide a quantified assessment of the impact of the main network characteristics on the application throughput of asynchronous replication:
  - ① distance between SAN islands
  - ② available bandwidth
  - ③ packet loss



## Protocol Used For Storage Area Networks

- Fibre Channel
- Fibre Channel over TCP/IP
- Internet SCSI
- Internet Fibre Channel Protocol

# Fibre Channel

- The protocol has been defined in five layers, FC-0 to FC-4 and can be mapped onto the open systems interconnection (OSI) stack as following:
- FC-0 and part of FC-1 -> physical layer
- FC-1 and FC-2 -> data link layer
- FC-3 -> network layer
- FC-4 -> transport layer



# Fibre Channel

- Fibre Channel supports five different classes of service (1-6, with class 5 undefined) to suit the requirements of different network topologies and services.
- The performance of Fibre-Channel-based SAN depends on the number of buffer credits, class of service at the link layer, and physical layer error correction mechanisms.



# Fibre Channel over TCP/IP

- FCIP is a tunneling protocol that interconnects Fibre-Channel-based SAN islands over IP-based networks and offers an alternate solution for SAN extension.
- Two sessions maintained:
  - end-to-end session between Fibre Channel devices that is based on Fibre Channel
  - TCP/IP session between the gateways in the IP network



# Fibre Channel over TCP/IP

- All Fibre Channel classes of service are supported by FCIP except class 1, which requires a mandatory dedicated connection.



# Internet SCSI

- iSCSI utilizes a TCP/IP network for the transport of SCSI commands and responses.
- iSCSI depends on a client-server architecture with the client the SCSI initiator and the server the SCSI target.
- Error recovery mechanisms have been defined by three classes ranging from data retransmissions to complete teardown and restarting of the iSCSI session.

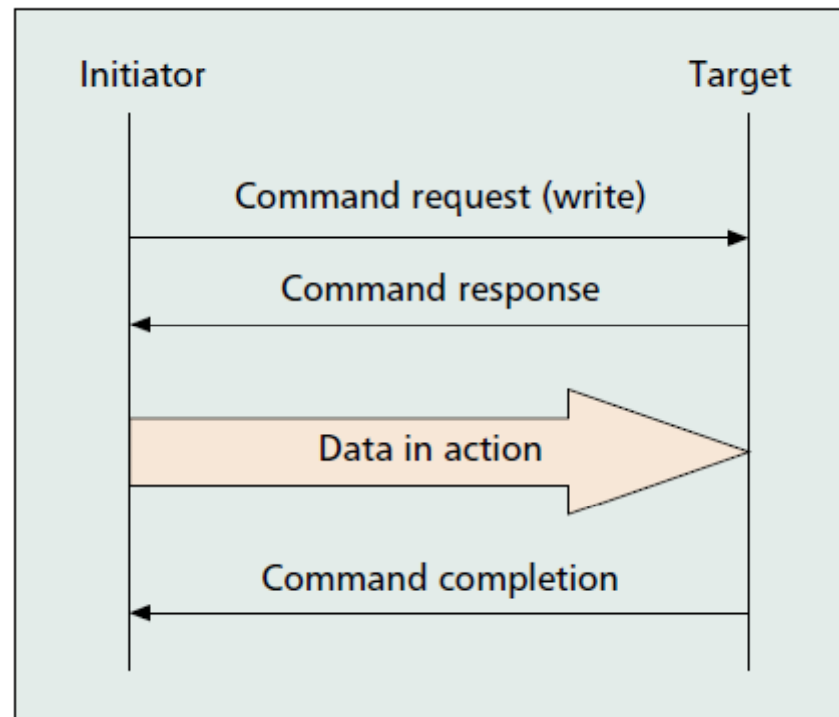


# Internet Fibre Channel Protocol

- iFCP is a gateway-to-gateway protocol to transport Fibre Channel frames over TCP/IP switching and routing elements.
- The protocol enables attachment of existing Fibre Channel storage products to an IP network.
- iFCP only supports Class 2 and 3 fiber channel transport services.

# Performance Analysis

- Analytical models developed for asynchronous replication are based on the SCSI command sequence for a data write.



■ Figure 1. *SCSI command sequence for data write.*

# Performance Analysis

- Analytical models for asynchronous replication use two types of variables: network-based and protocol-specific.

Parameter	Fibre-Channel-based	iSCSI-based	FCIP- and iFCP-based
Data window	Maximum number of buffer credits available: 125	Single TCP connection with a window of 256 kbytes	Single TCP connection with a window of 256 kbytes
Frame size (bytes)	2104 payload + 44 overhead	1404 payload + 114 overhead (GFP) or 112B overhead (POS)	1384 payload + 134 overhead (GFP) or 112B overhead (POS)
Zero copy	Confirmed to have zero copy architecture	Zero copy architecture with TCP processing overhead	Processing delays in gateways are of microsecond order
TCP processing delays	Not applicable	TCP processing overhead: protocol-related (packet-based) + checksum (byte-based)	TCP processing occurs in gateways
I/O writes	Parallel writes are available in the disk	Parallel writes are available in the disk	Parallel writes are available in the disk
I/O block	64 kbytes	64 kbytes	64 kbytes
Command sequence	Fibre Channel Protocol	iSCSI draft	Combination of FC and IP
<b><u>Network-based</u></b>			
SAN separation distance	0 to 5000 km; all the equipment collocated in the same office for 0 km; optical equipment/IP routers are placed every 50 km		
Available bandwidth	1 Gb/s		
Packet loss	Point-to-point over optical: negligible < 1 e-06; IP network: 1e-06 to 1e-03 ; typical loss range in commercial IP networks: 1e-04 to 1e-03		
Framing	GFP for point-to-point optical		

■ Table 1. Parameters used in modeling.



# Performance Analysis

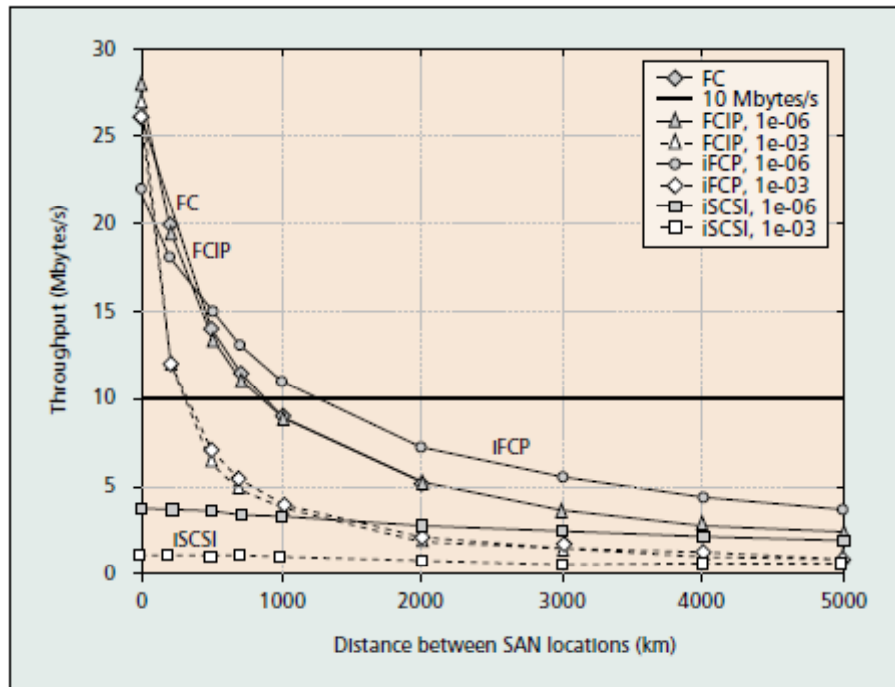
- Application throughput is modeled as a function of SAN separation distance, packet loss, available bandwidth, and so on.
- The comparison of performance of different SAN extension solutions is for a single TCP session with a maximum window of 256 kbytes, constant availability of a 1 Gb/s pipe in the access, and a SAN separation distance of 1500 km.



# The Impact of Network Parameters on the Performance of SAN Extensions

- Distance
- Packet Loss
- Available Bandwidth
- Advances in TCP Implementations
- Increased Maximum TCP Window
- Parallel TCP Sessions

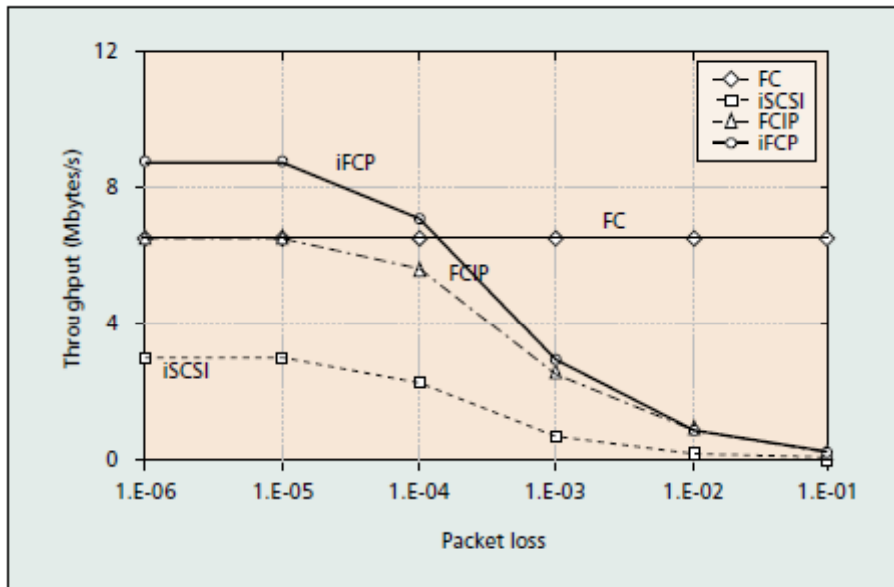
# Distance



■ Figure 2. Application throughput performance with distance (solid lines: packet losses < 1e-06, dotted lines: packet losses < 1e-03).

- For Fibre Channel (SONET-based), FCIP, and iFCP, throughput decreases with distance.
- For iSCSI, it has lower throughput than other solutions, with little dependence on distance. (Reason?: TCP processing latencies and disk latencies are so high that they are comparable to the propagation delays over the distance range.)

# Packet Loss



■ Figure 3. Application throughput performance at 1500 km with packet loss.

- Packet loss results in packet retransmissions and reduced throughput when reliable data delivery is involved.
- Packet loss is minimal in SONET-based networks with inherent error correction mechanisms. In IP networks, it is a few orders of magnitude higher.
- Packet loss  $< 1.e-5$  &  $> 1.e-4$

# Available Bandwidth

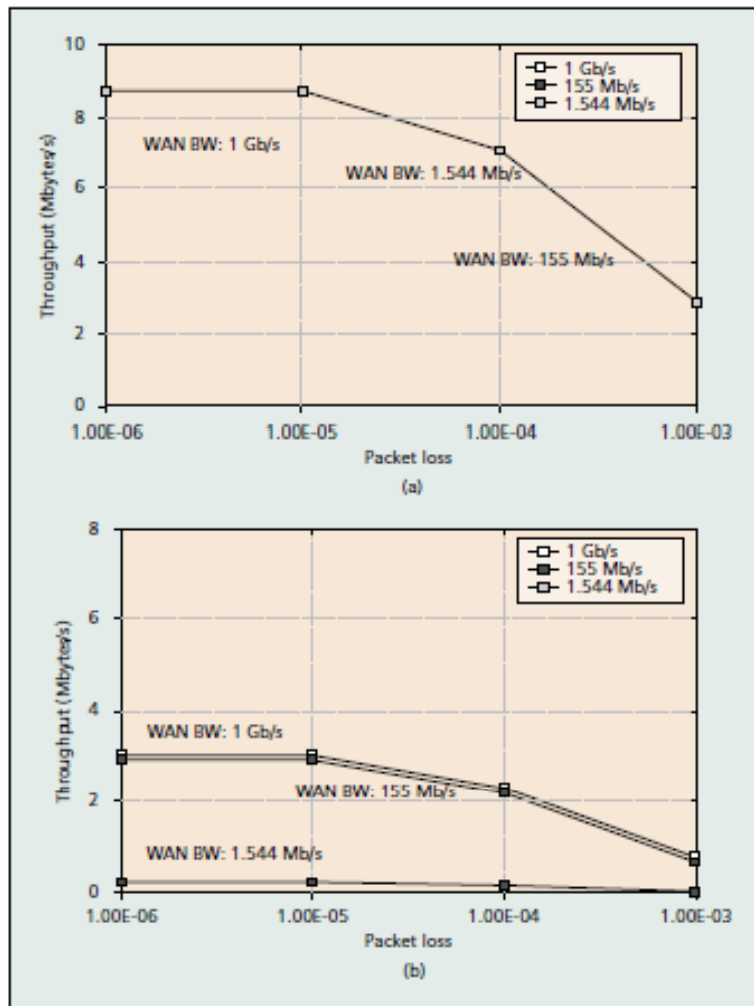
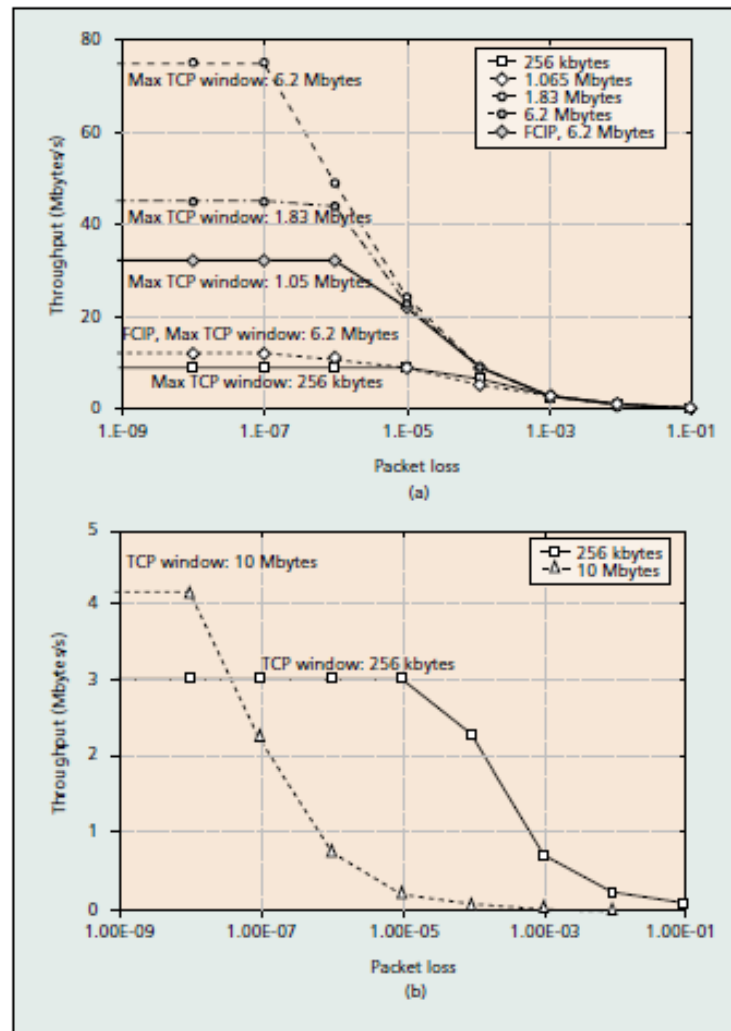


Figure 4. Application throughput performance of a) iFCP and b) iSCSI with variable bandwidth (BW).

- For iFCP, the effect of bandwidth is negligible due to the data sending and acknowledgement receiving involved in any TCP process.
- For iSCSI, the effect is similar but not exactly the same as TCP processing delays are dominant.

# Advances in TCP Implementations – Increased Maximum TCP Window



■ Figure 5. Application throughput performance of a) iFCP and b) iSCSI with increased maximum TCP window.

- The gain due to increased maximum TCP window is on the order of 10 when the packet loss is less than  $1e-06$ , the throughput falls and approaches that of 256 kbytes.

# Advances in TCP Implementations – Parallel TCP Sessions

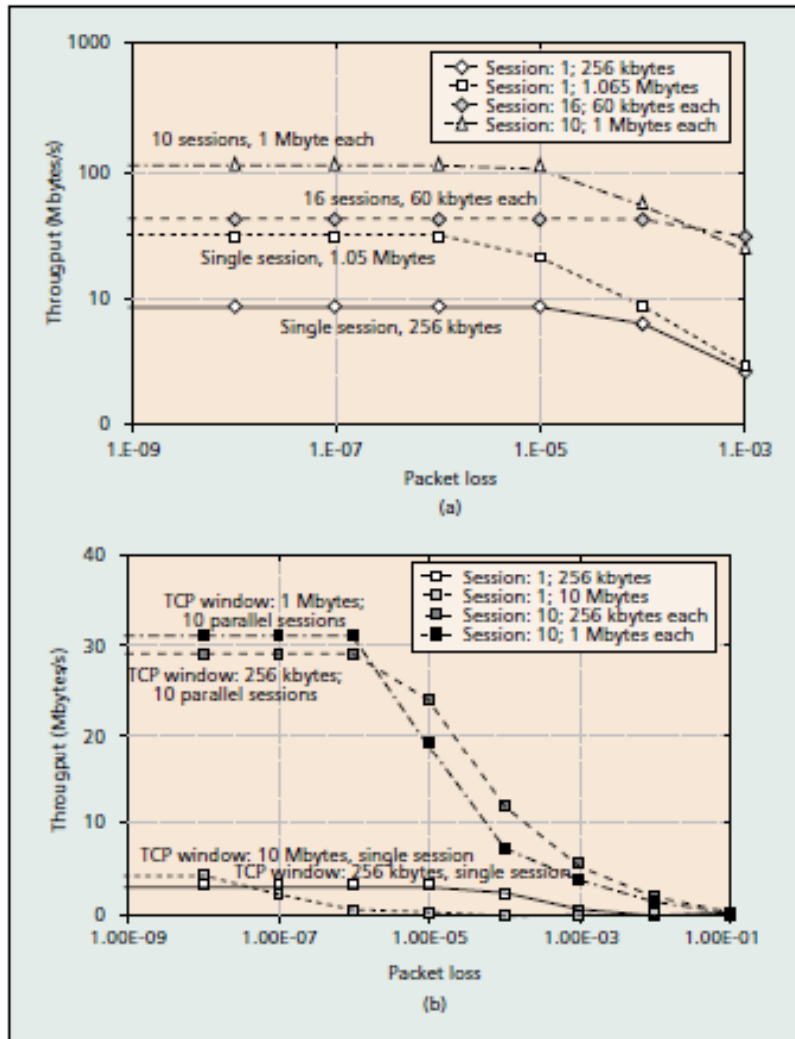


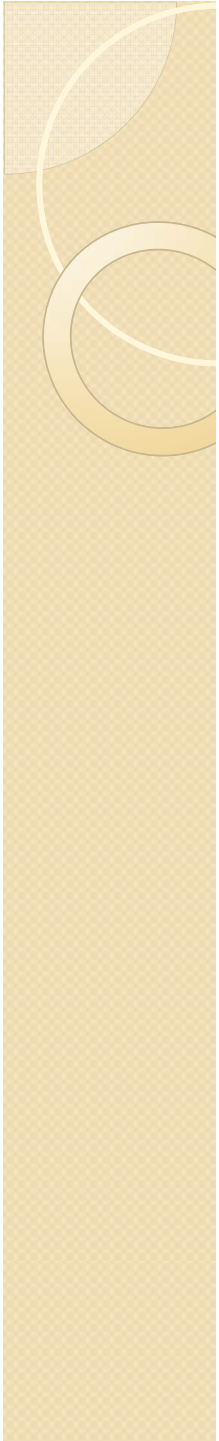
Figure 6. Application throughput performance of a) iFCP and b) iSCSI with parallel TCP sessions.

- TCP throughput can be increased by increasing the number of parallel TCP sessions between a sender and a receiver pair, the maximum number of TCP sessions a system can support with a given bandwidth needs to be calculated in advance.



# Conclusions

- IP-based solutions are sensitive to packet loss. When implemented with default TCP features, they have high throughputs (> 10 Mbytes/s) up to a distance of 300 km, with realistic packet losses around  $1e-03$ . When implemented with advanced TCP features, they may exhibit higher throughputs with realistic packet losses than SONET-based extension solutions.
- The selection of any SAN extension solution, either SONET- or IP-based, should be based on application throughput requirements, availability of buffer credits for Fibre-Channel-based solutions, and the complexity of implementation.



**Thank you!**