# Data Storage Institute

**SANSIM: A PLATFORM FOR SIMULATION AND DESIGN OF A STORAGE AREA NETWORK**

**Zhu Yaolong**

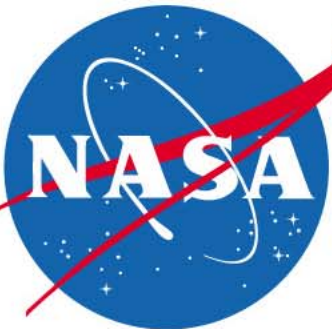**e_mail:zhu_yaolong@dsi.a-star.edu.sg**

NASA/IEEE MSST 2004
12th NASA Goddard/21st IEEE Conference on
Mass Storage Systems & Technologies
The Inn and Conference Center
University of Maryland University College
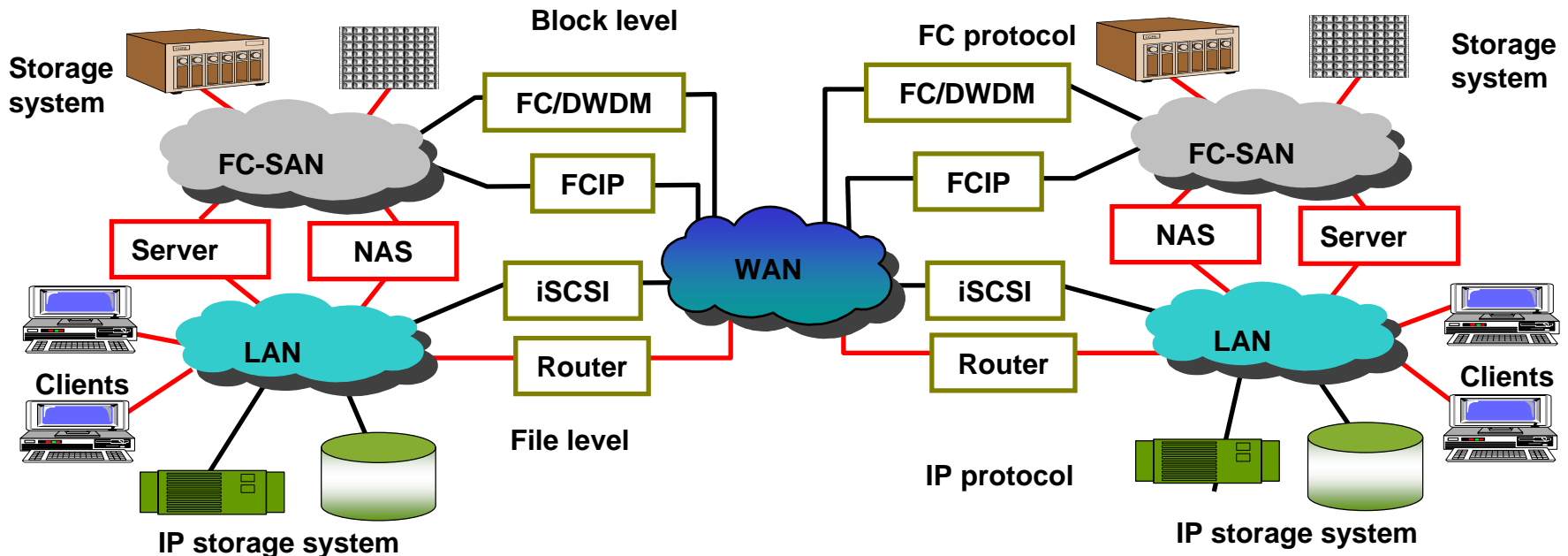Adelphi MD USA
April 13-16, 2004

# Outline

- Motivation
- Key Focuses
- Simulation Methodology
- SANSim Structure
  - Workload Module
  - Host Module
  - Network Module
  - Storage Module
- Calibration and Validation
- FC Network Analysis
- Summary
- Future Work

# Motivation of SANSim

- New idea is everything. New algorithm is the key.
- The motivation of this research is to develop modeling and simulation capability and to use this capability to explore and research on new ideas and algorithms.
- Modeling and simulation are fast and effective way to conduct in-depth algorithm researches.
- There is lack of the such kind of methodology and tool in network storage industry.
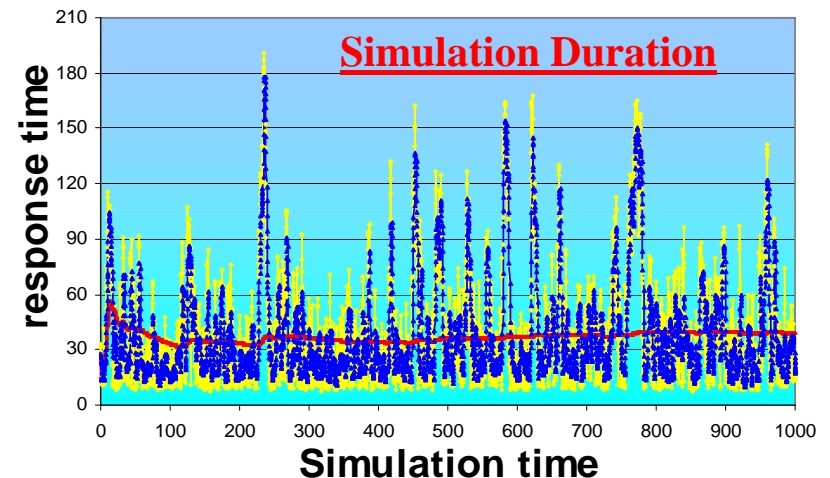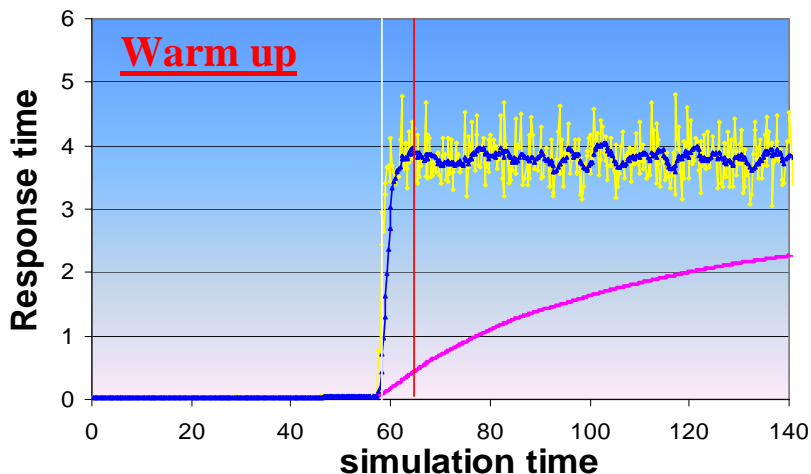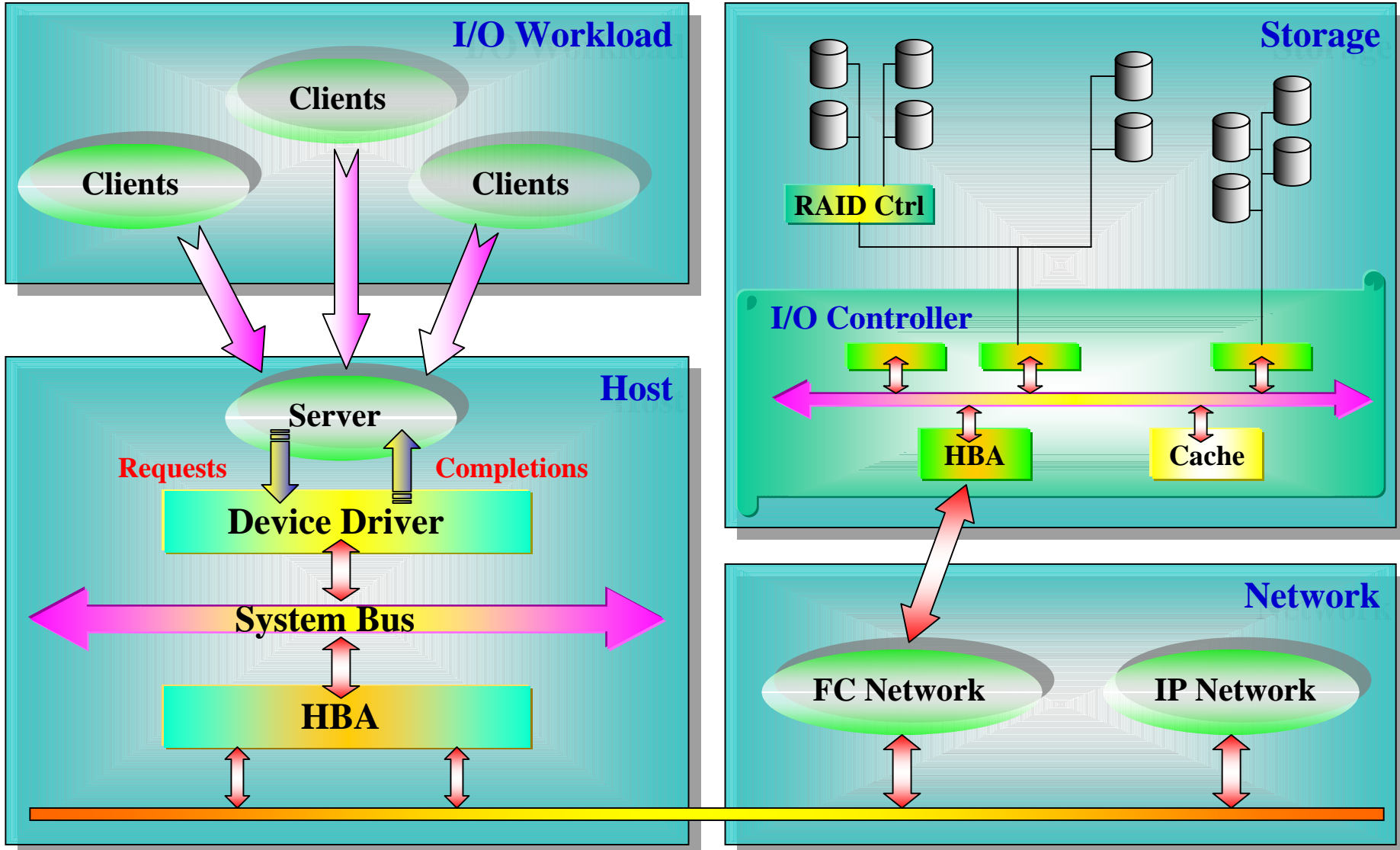
# SANSim Main Focuses

- Tools development
  - Build up the capability of modeling and simulation of network and storage
  - Simulation methodology
  - Modular structure

- Algorithms research
  - Modeling of storage architecture: DAS/NAS/SAN/OSD
  - File system
  - I/O workload analysis
  - FC networks and their schedule algorithms
  - Cache management
  - Disk array design & algorithms (normal, degradation and rebuild mode)
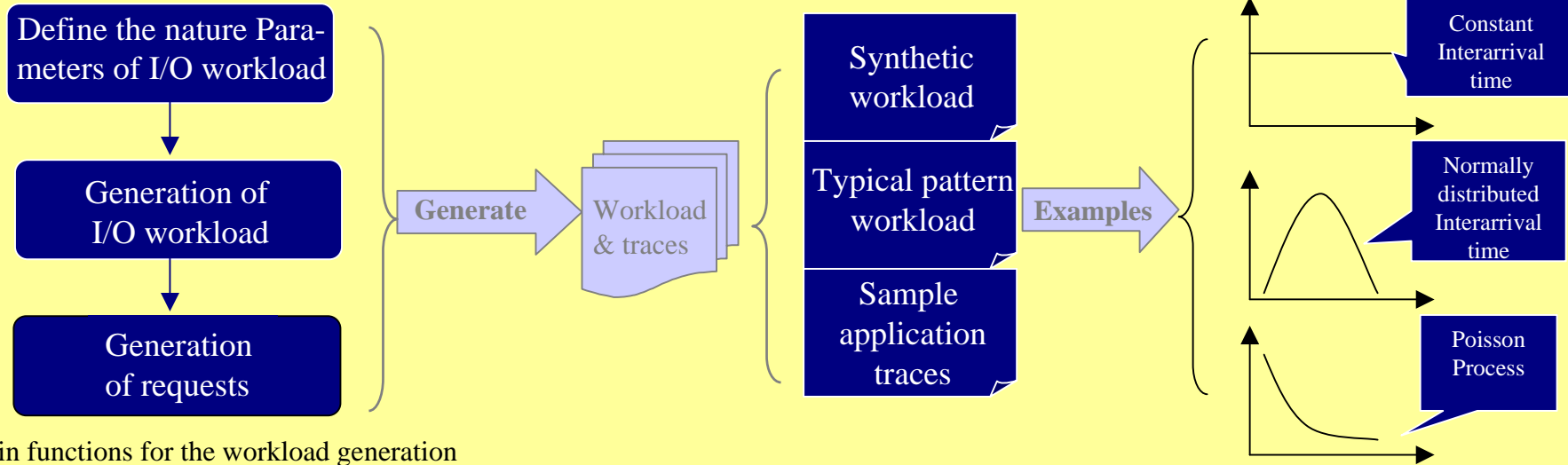
# Simulation Methodology

- ➢ **Event-Driven simulation**
  - ➢ **Trace-Driven System-Level Simulation**
  - ➢ **Synthetic I/O Workload**
- ➢ **Virtual Device**
- ➢ **High-Resolution Timestamp Mechanism**

- ➢ **Globe Scheduling and queuing monitoring**
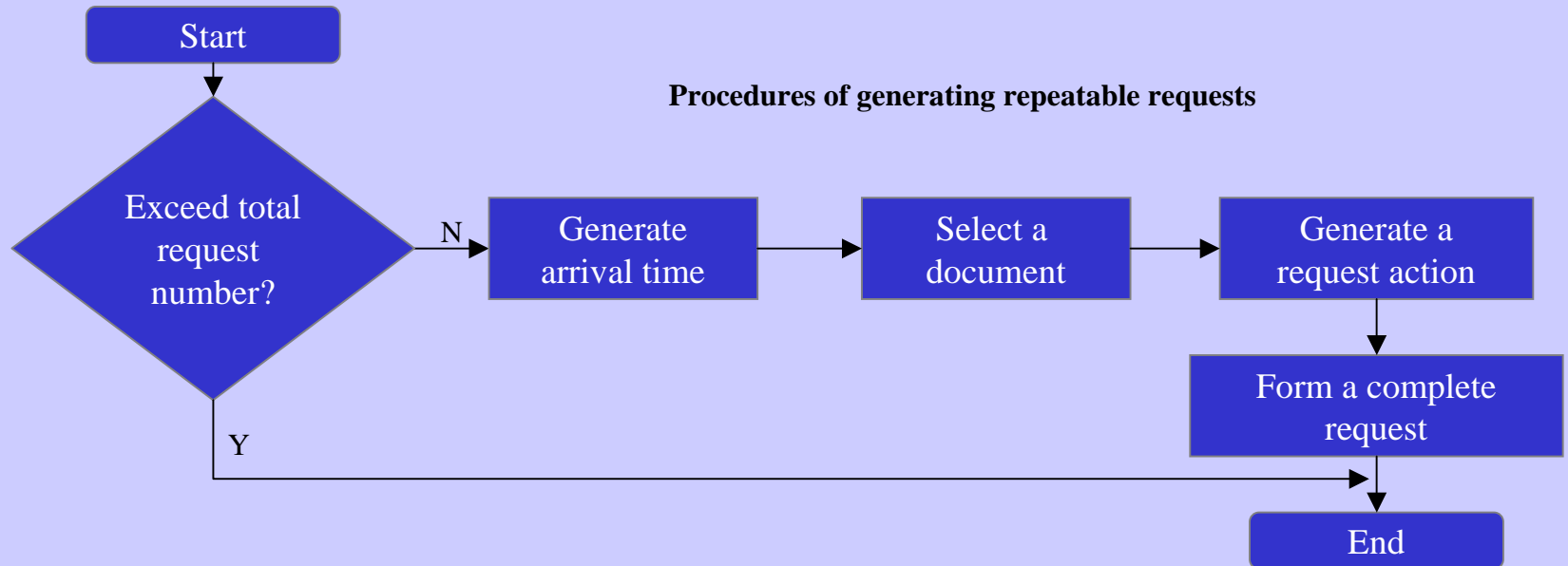- ➢ **Closed System/Open System Models**
- ➢ **Warn-up processing**
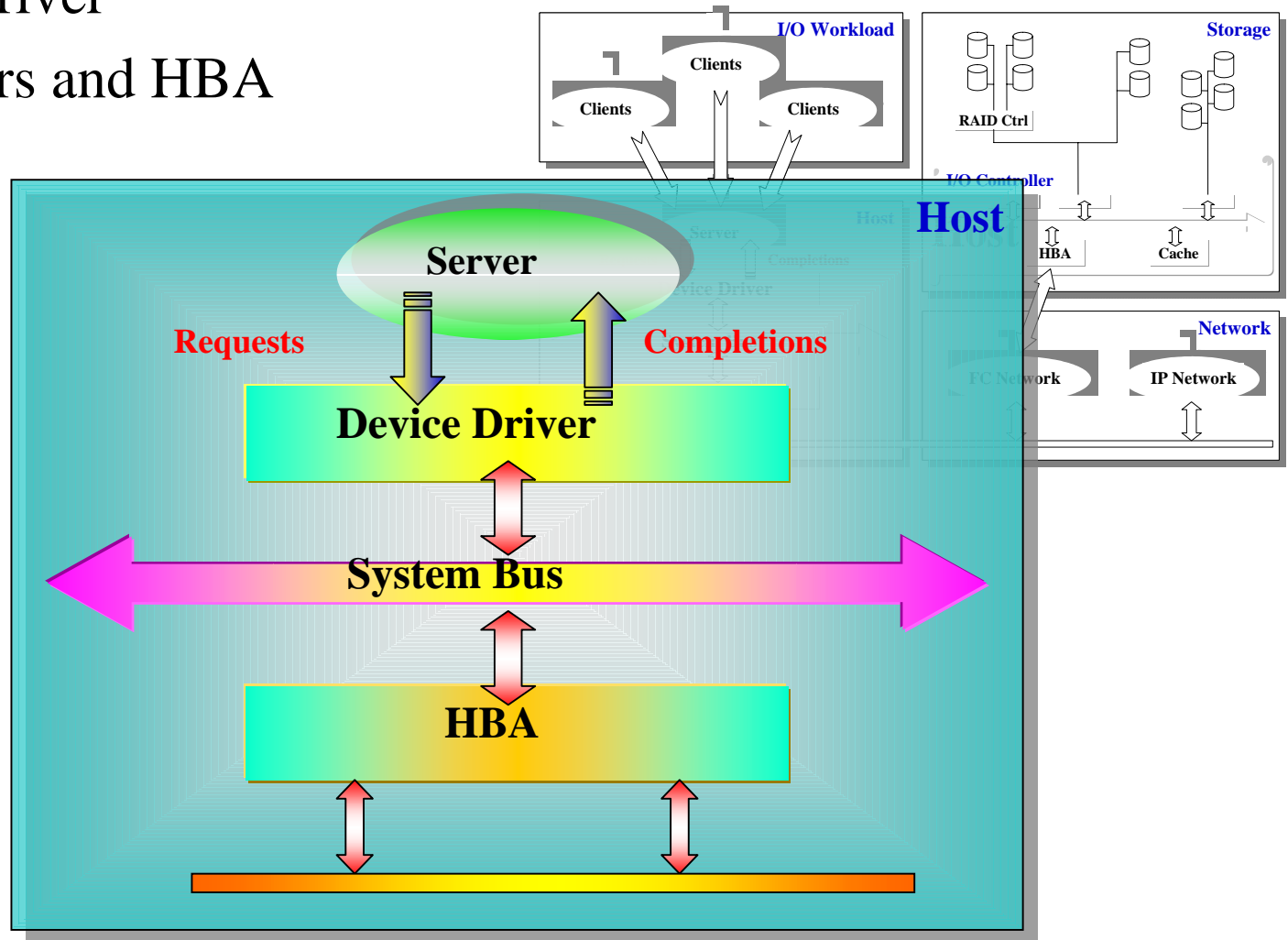
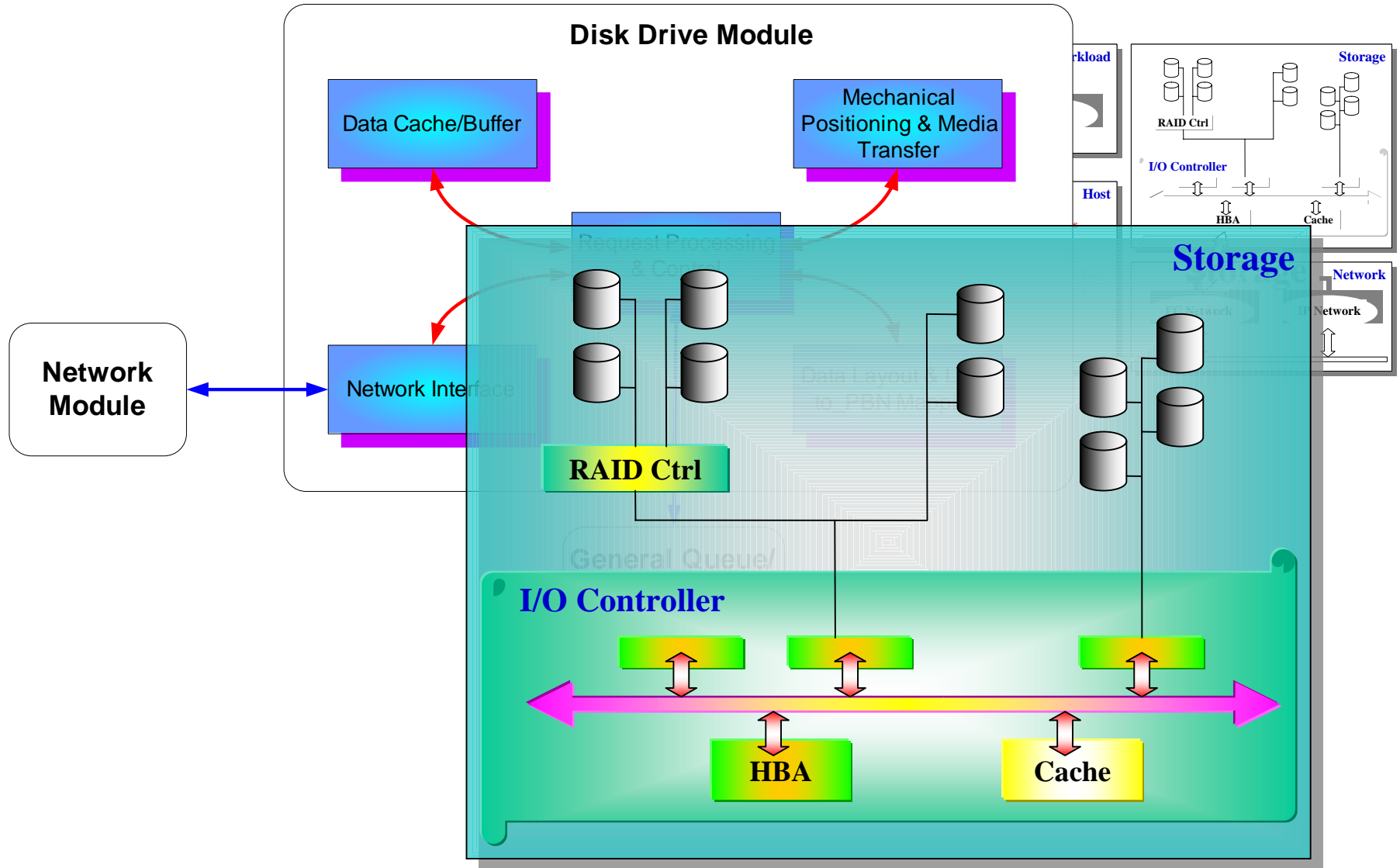# SANSim Module Structure

# SANSim Workload Generation

Define the nature Para-meters of I/O workload

↓

Generation of I/O workload

↓

Generation of requests

**Generate** → Workload & traces

Synthetic workload

Typical pattern workload

Sample application traces

**Examples** →

Constant Interarrival time

Normally distributed Interarrival time

Poisson Process

Main functions for the workload generation

**Procedures of generating repeatable requests**

Start

↓

Exceed total request number?

— N → Generate arrival time → Select a document → Generate a request action

↓

Form a complete request

↓

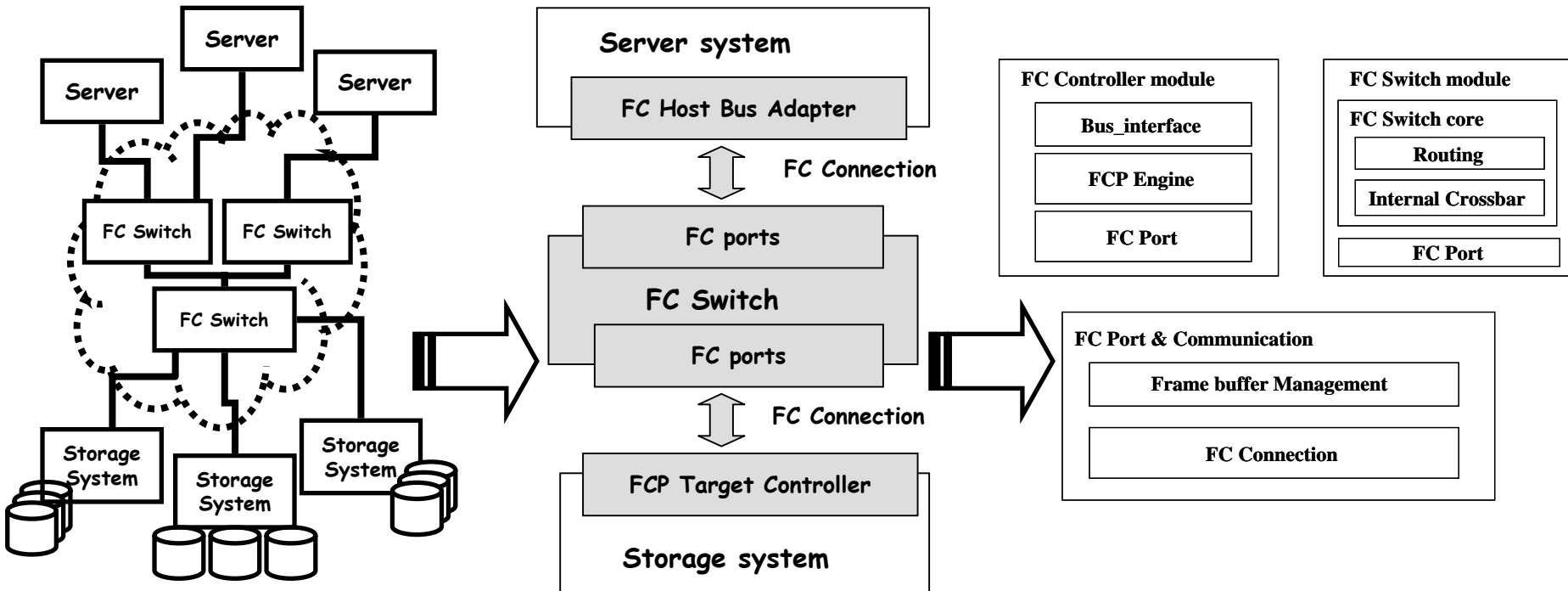Y ———————————————————→ End

# SANSim Host Modules

- Device Driver
- Controllers and HBA
- Buses
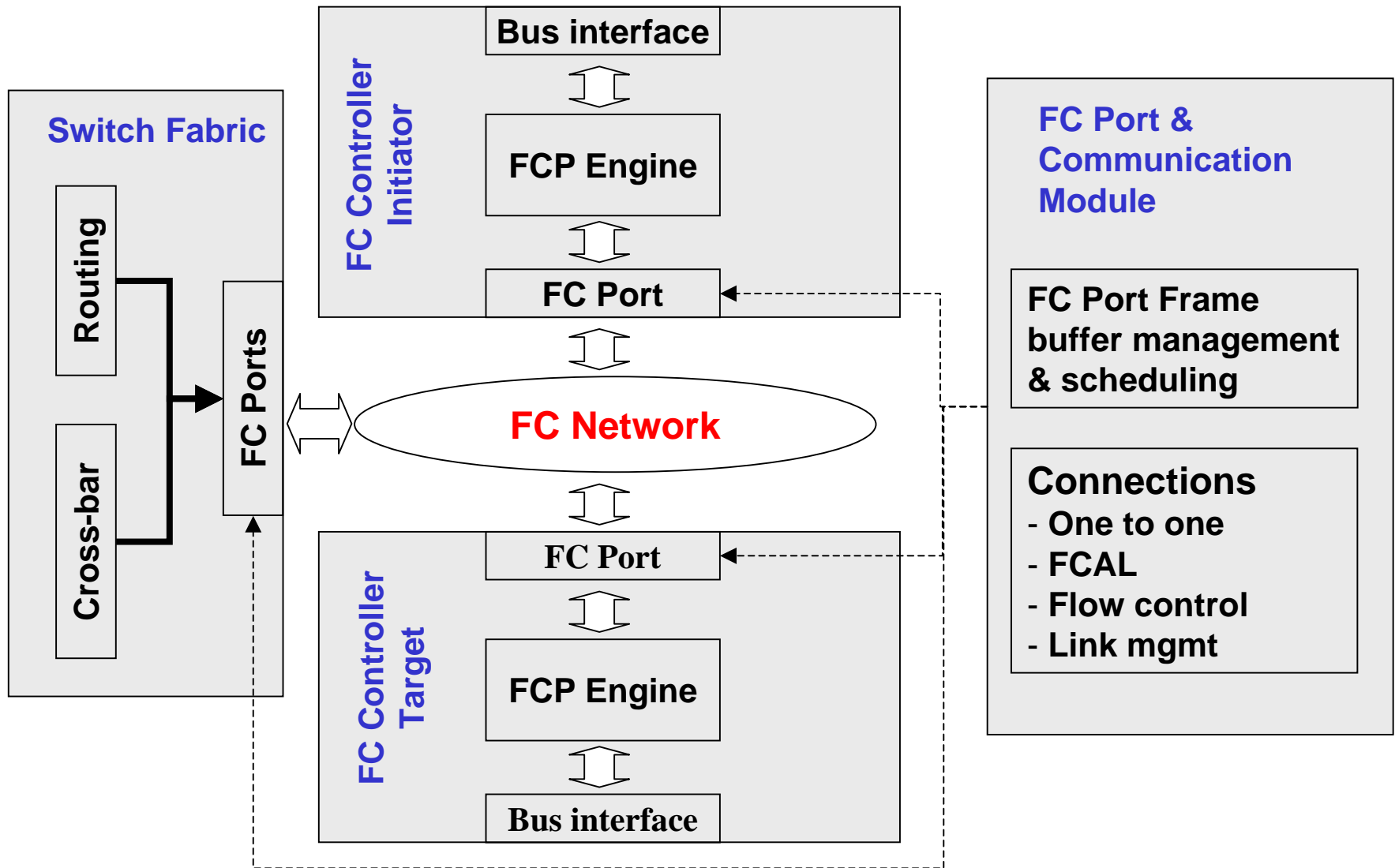
# SANSim Disk Module
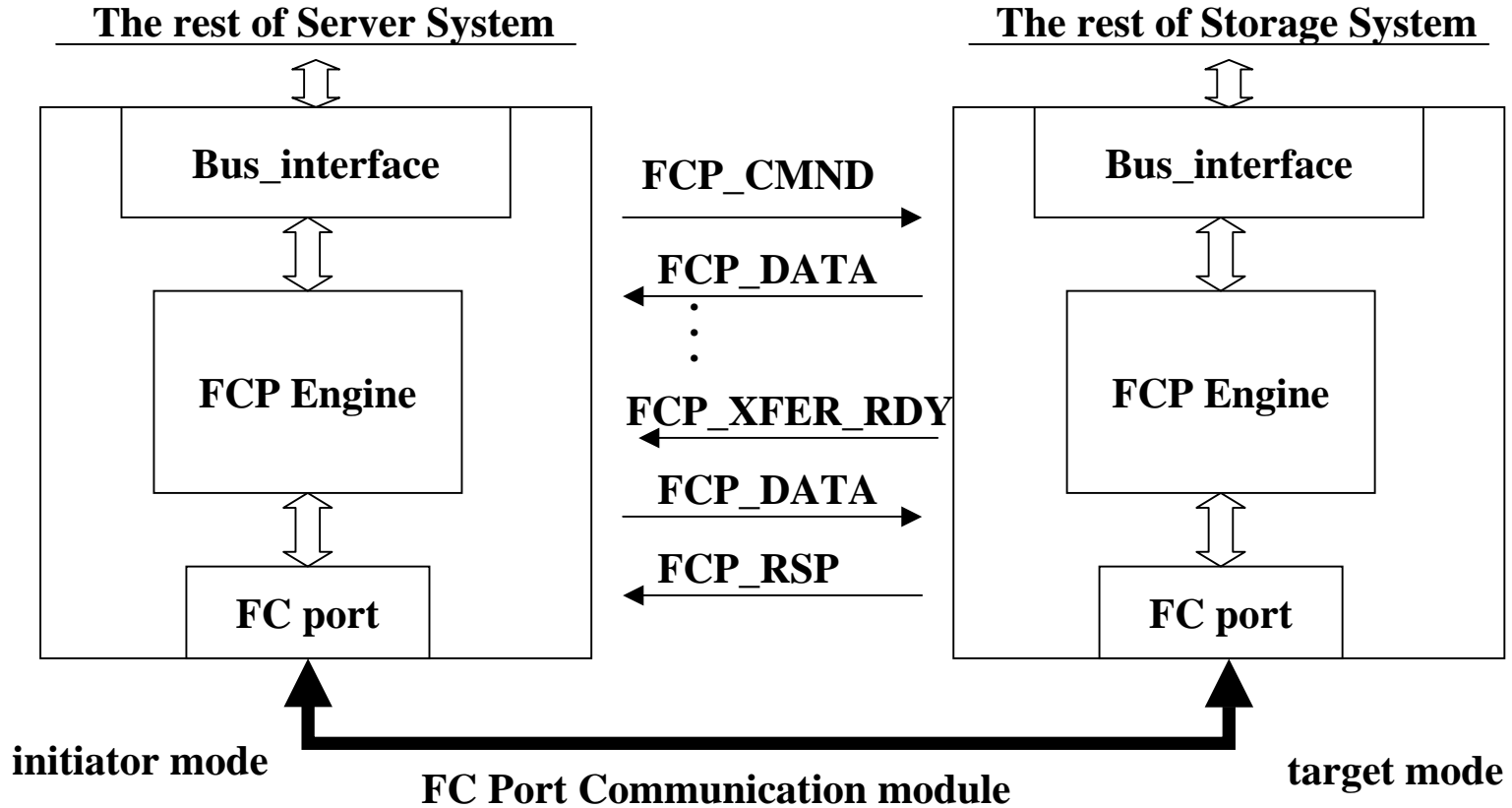
# SANSim FC Networking Module



(a)An example of FC SAN

(b) Abstracted view of participating component

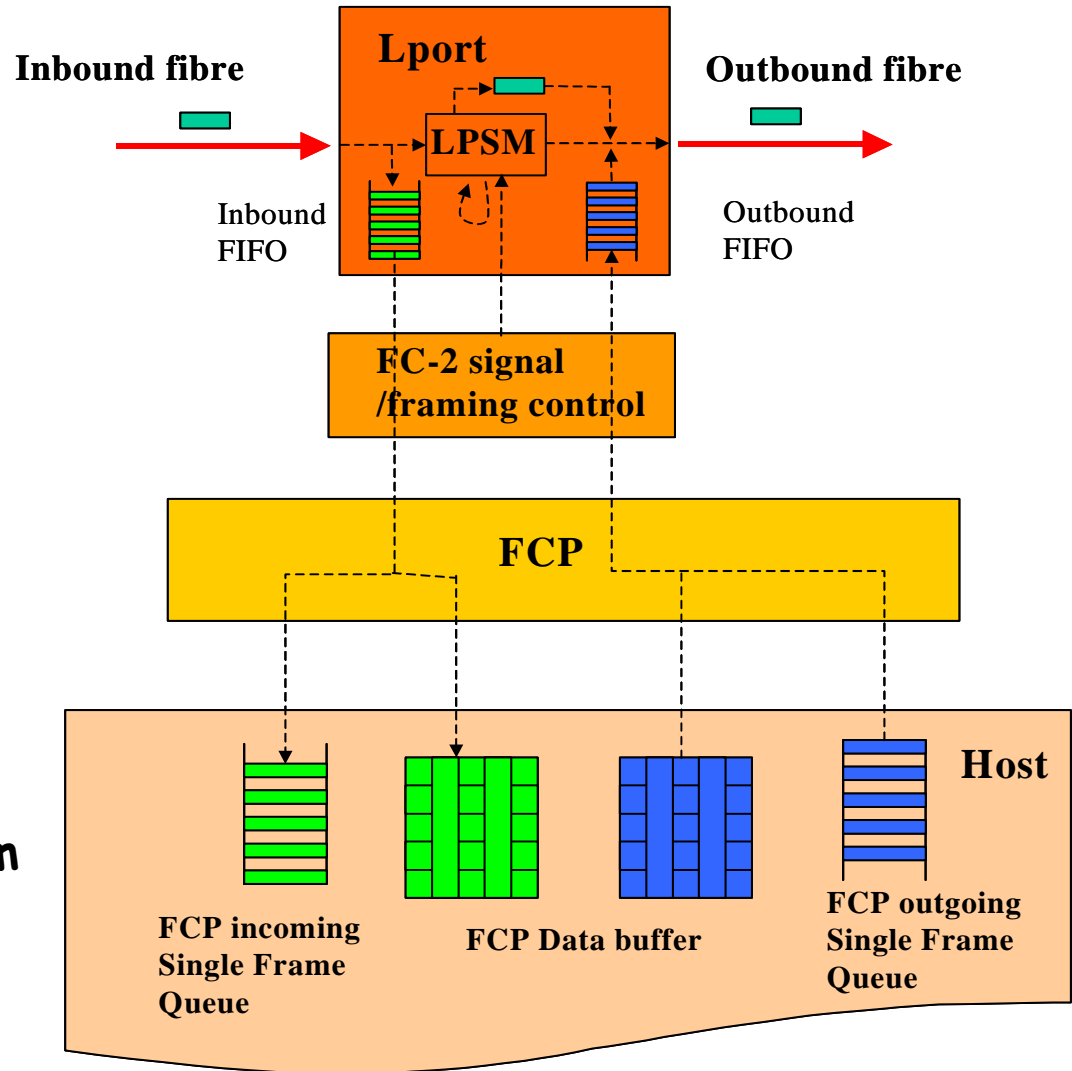(c) Simulation modules of the essential component s

# SANSim Network Modules
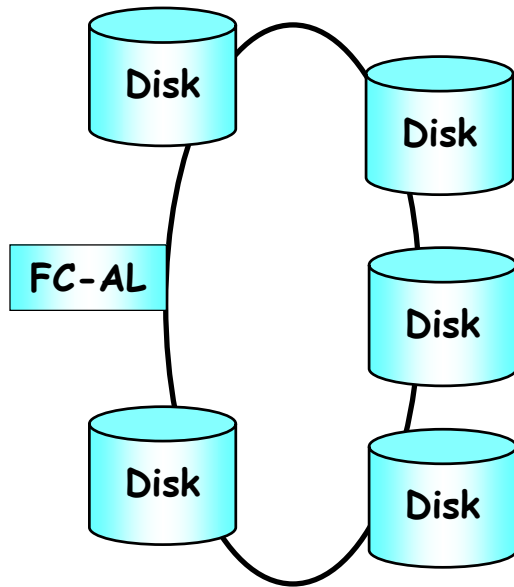
**Data Storage Institute**
**A*STAR**

**Switch Fabric**

Routing

Cross-bar

FC Ports

**FC Controller Initiator**

**Bus interface**

**FCP Engine**

**FC Port**

**FC Network**

**FC Controller Target**

**FC Port**

**FCP Engine**

**Bus interface**

**FC Port & Communication Module**

**FC Port Frame buffer management & scheduling**

**Connections**
- **One to one**
- **FCAL**
- **Flow control**
- **Link mgmt**

# FCP Operation

The rest of Server System

The rest of Storage System

| Bus_interface | | Bus_interface |

FCP_CMND →

FCP_DATA ←

⋮

FCP_XFER_RDY ←

FCP_DATA →

FCP_RSP ←

| FCP Engine | | FCP Engine |

| FC port | | FC port |

initiator mode

FC Port Communication module
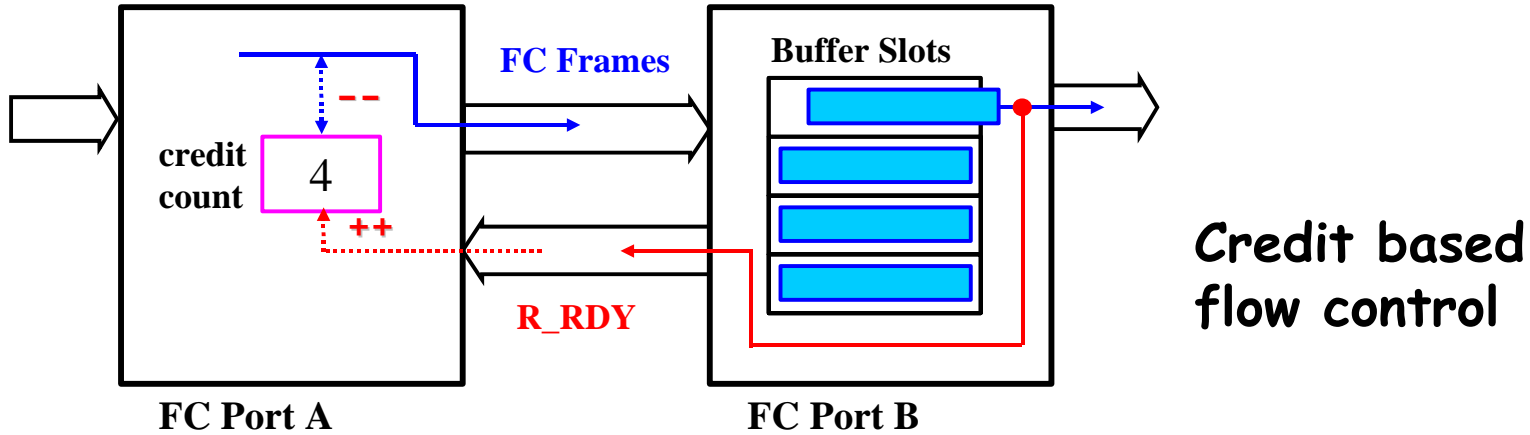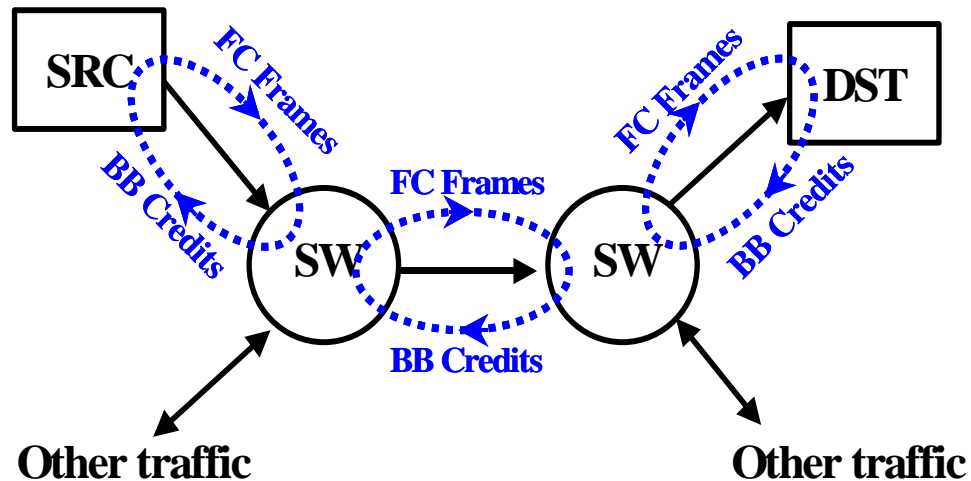
target mode

# FC_port Module



- ❖ **Frame/Order Set level**
- ❖ **Fairness Access Algorithm**
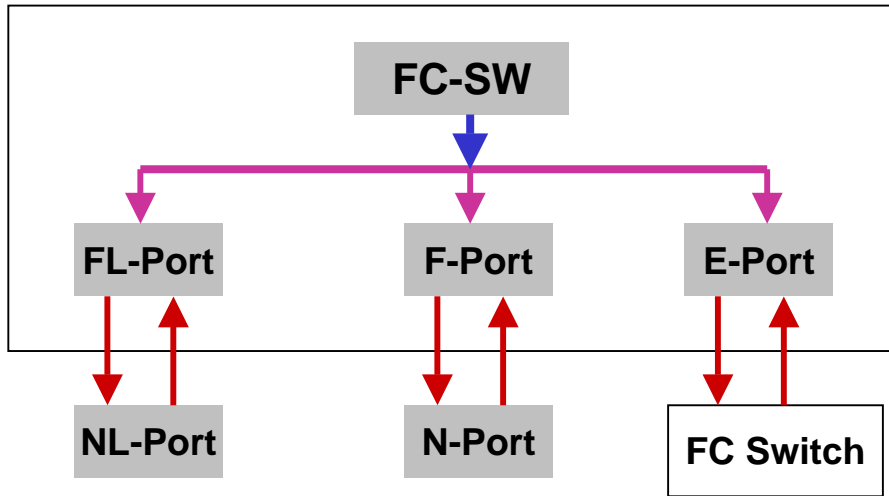- ❖ **Propagation delay due to physical distance**

# Flow Control Module
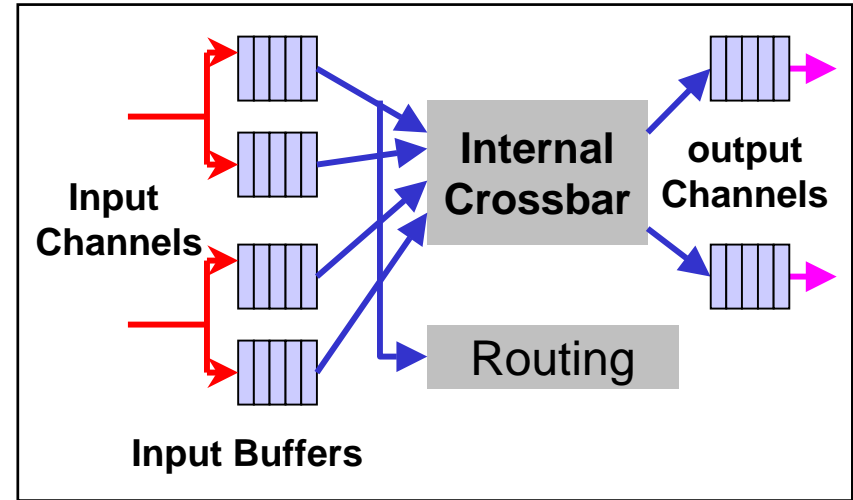


Credit based flow control

Multi-stage Buffer-to-Buffer
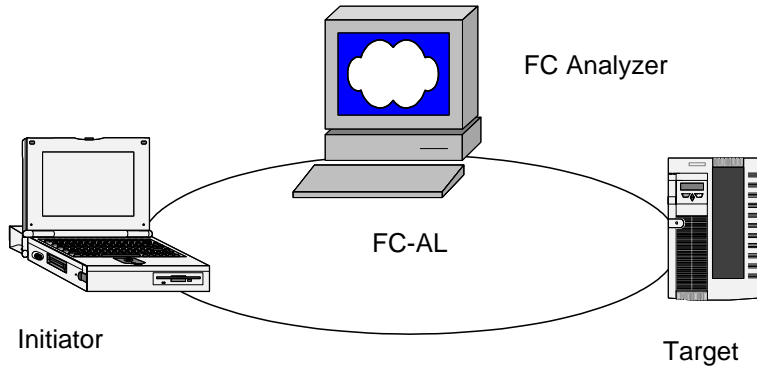
# FC_Switch Modeling


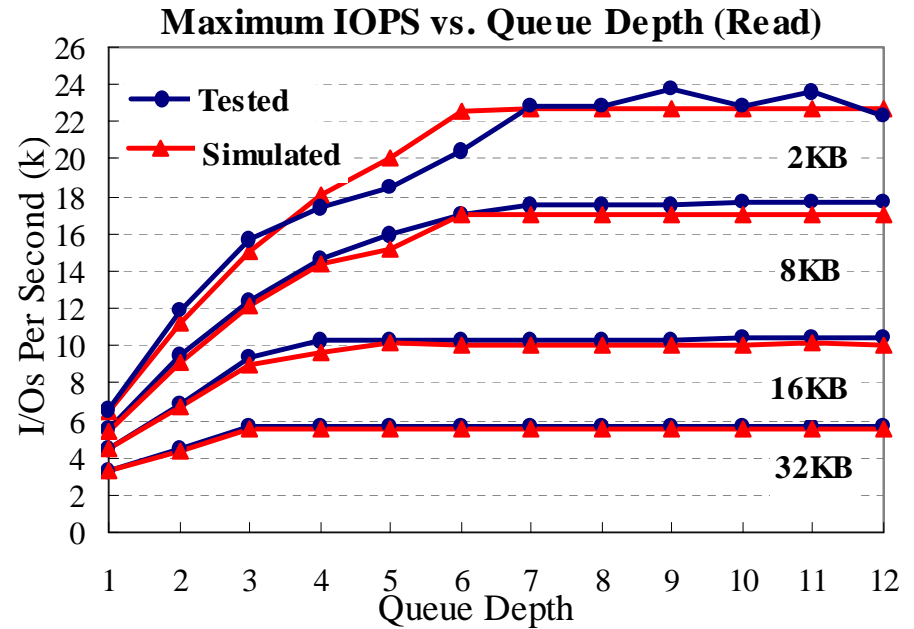
(a) FC_Switch module

(b) FC_SW sub-module

## Feature of SANSim FC Switch Module:

- ❖ Load balancing on multi-link inter FC Switch connection
- ❖ FSPF routing/path selection
- ❖ Domain_ID for each Fibre Channel switch
- ❖ Area_ID and Port_ID for each N/NL port

# Model Validation



Initiator

FC Analyzer

FC-AL

Target

| Initiator | |
|---|---|
| Hardware | CPU: AMD AthonMP 1600+<br>FC HBA: Qlogic 2300<br>RAM: 2x256MB DDR SDRAM<br>Mainboard: 64 bit PCI Tyan Tiger MP2466N |
| Software | OS: Windows XP Professional SP1<br>Driver: Qlogic Driver Version 8.1.5.12<br>Tool: Intel IOMeter Version2003.02.15 |
| Target | |
| Hardware | CPU: Intel PIII 1GHz<br>FC HBA: Qlogic 2300<br>RAM: 4 x 1GB Kingston ECC Reg. PC133<br>Mainboard: 64bit PCI,  Supermicro 370 |
| Software | OS: RedHat 8.0<br>Kernel: 2.4.18<br>Driver:  In-house 2300 target driver Ver 1.0,<br>In-house Linux **RAM Disk** Ver 2.0 |

**Maximum IOPS vs. Queue Depth (Read)**



I/Os Per Second (k) vs Queue Depth, with curves labeled Tested and Simulated for 2KB, 8KB, 16KB, 32KB

**Compare experimental data
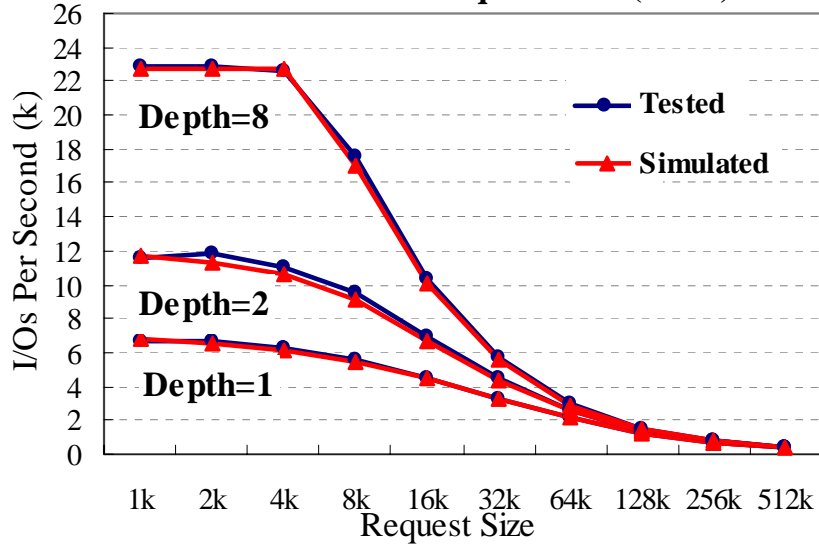with simulation result**

Error range < 10%.
Read operation <3%

**System configuration for testing  experiment**
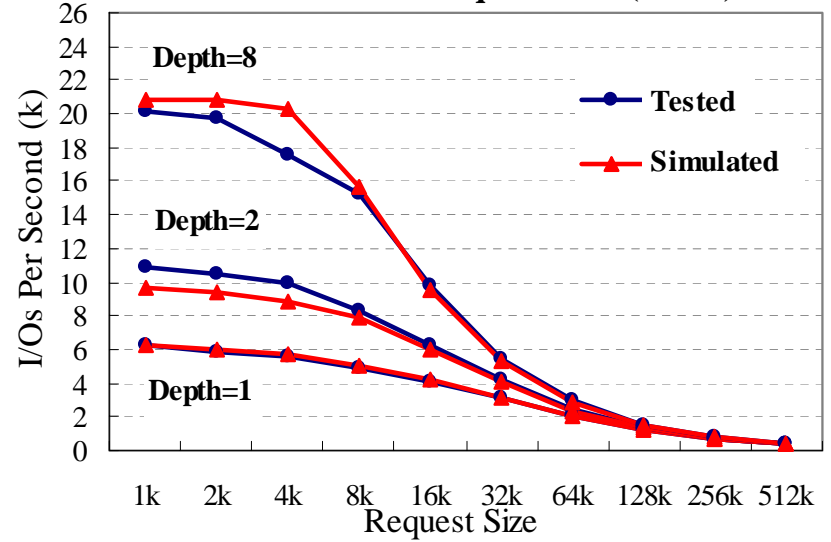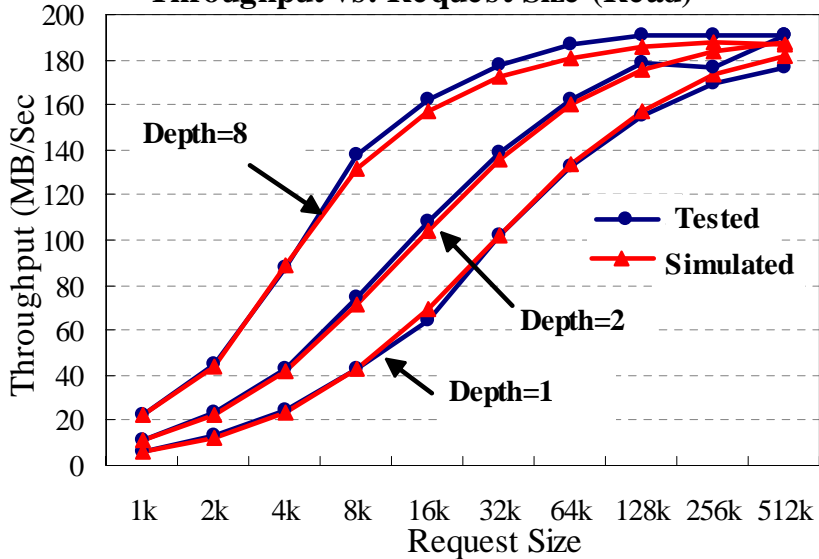
# Model Validation – cont'
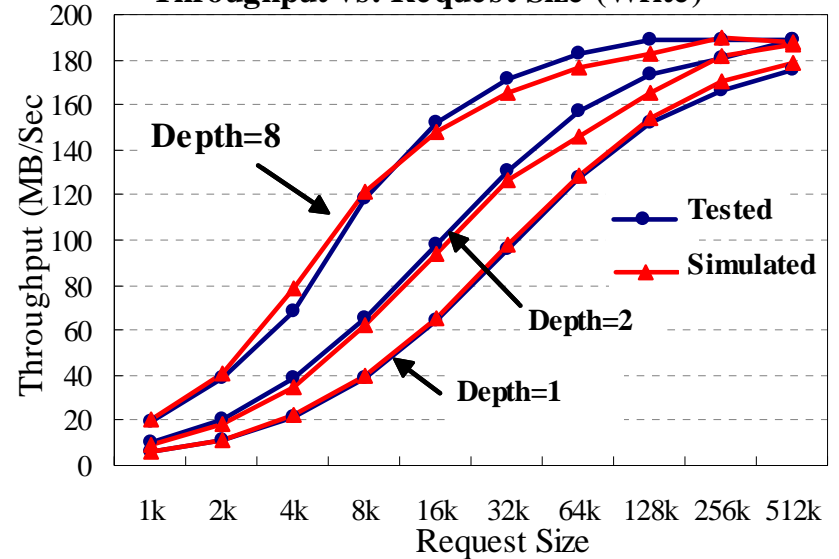


Maximum IOPS vs. Request Size (Read)
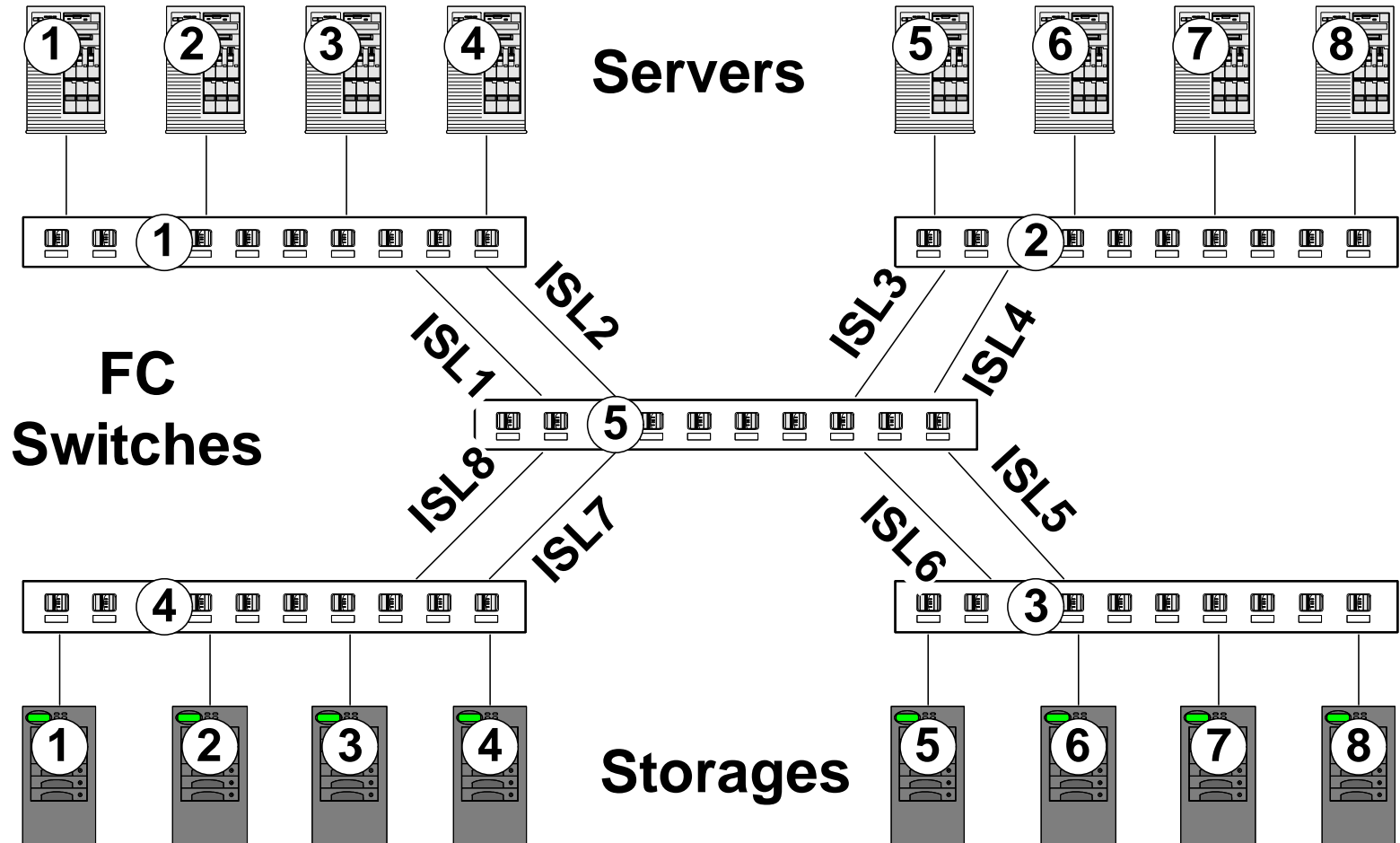
Maximum IOPS vs. Request Size (Write)

Throughput vs. Request Size (Read)

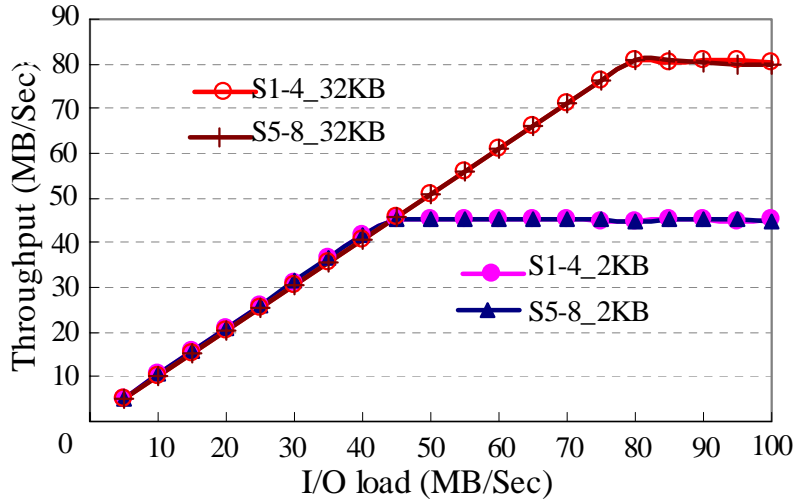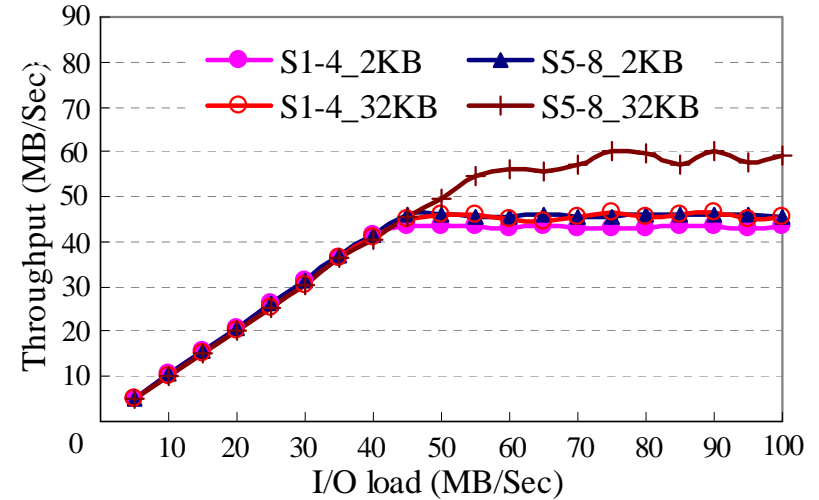Throughput vs. Request Size (Write)

# FC Network Simulation and Analysis



Servers

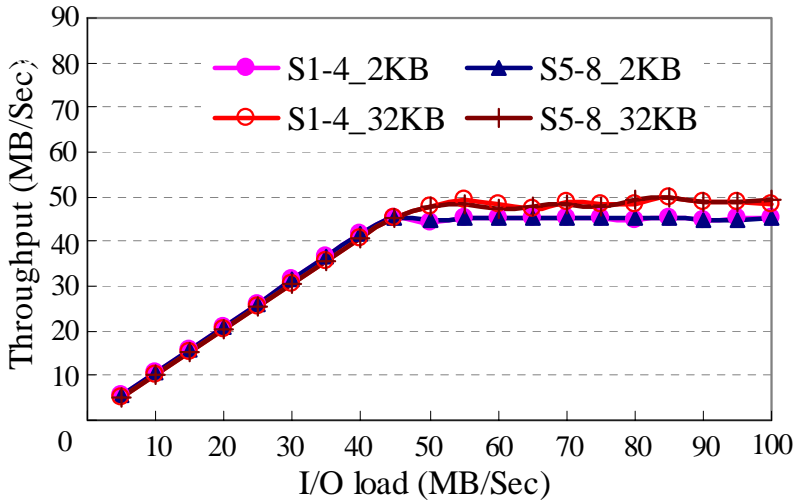FC Switches

ISL1 ISL2 ISL3 ISL4
ISL8 ISL7 ISL6 ISL5

Storages
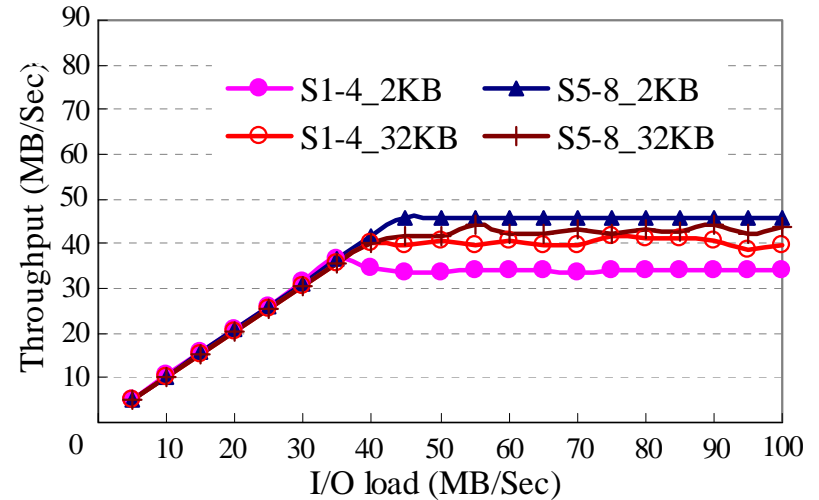
# Max Throughputs Impact

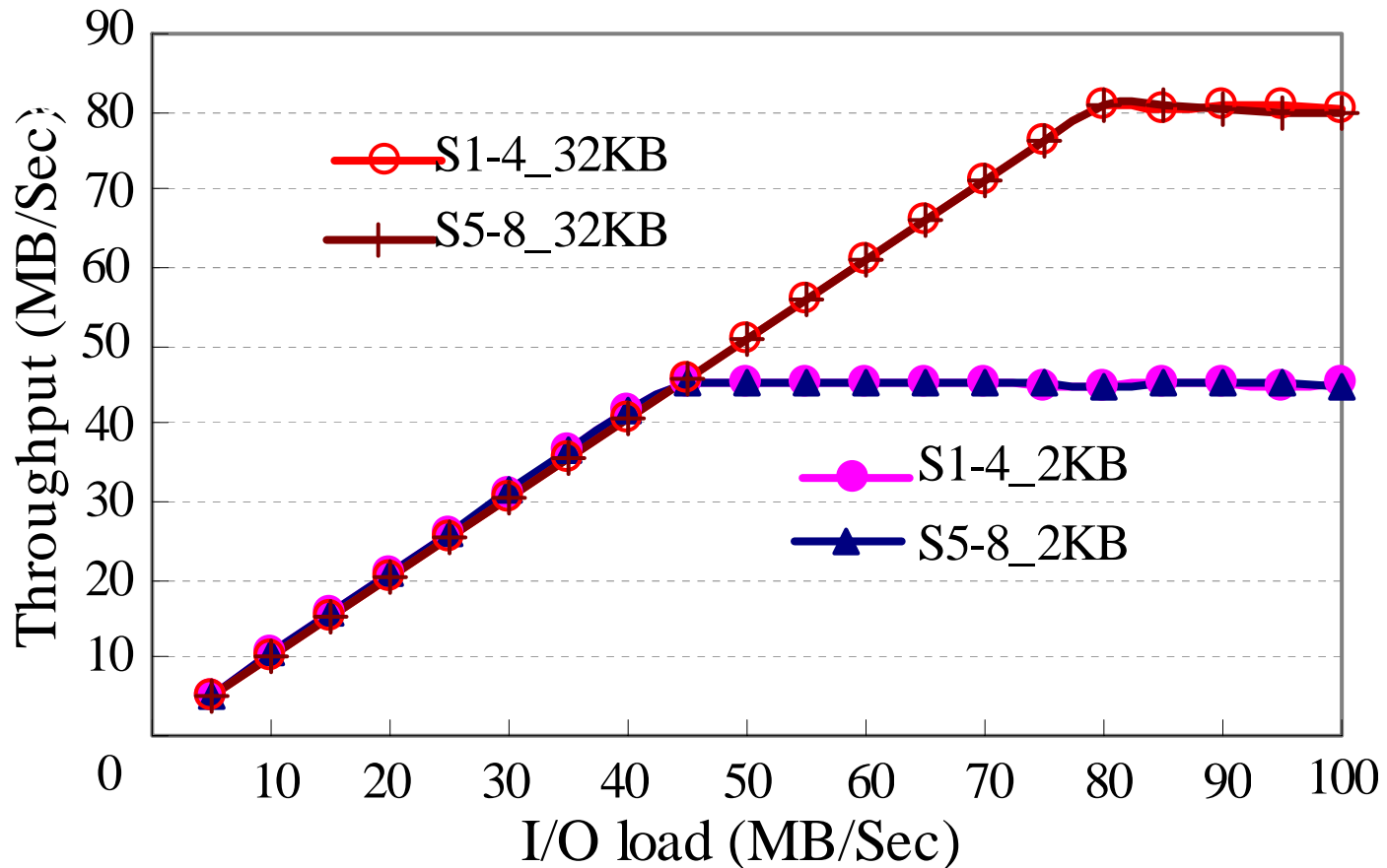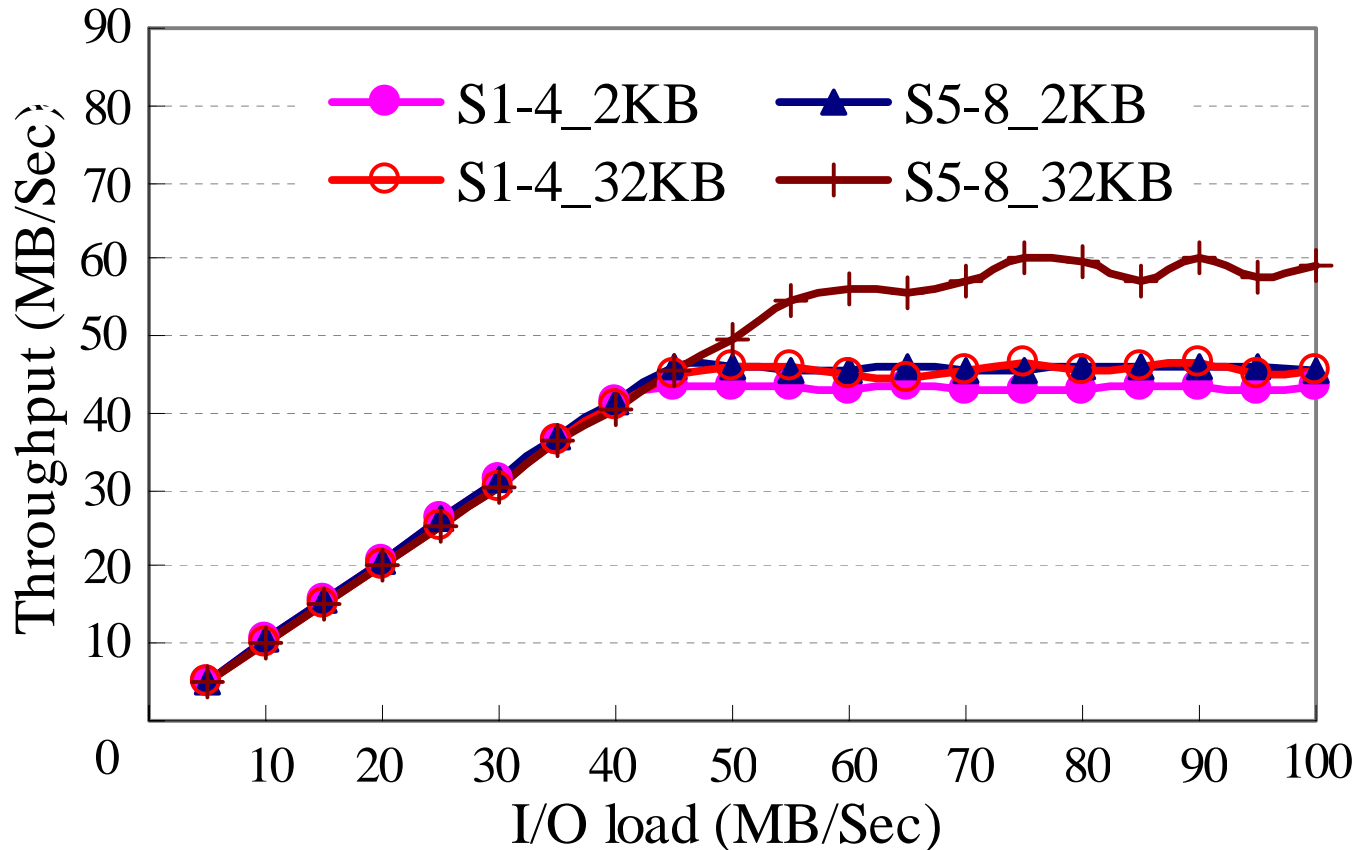a) No link failed

b) ISL1 failed

c) ISL8 failed

d) Both ISL1 and ISL8 failed
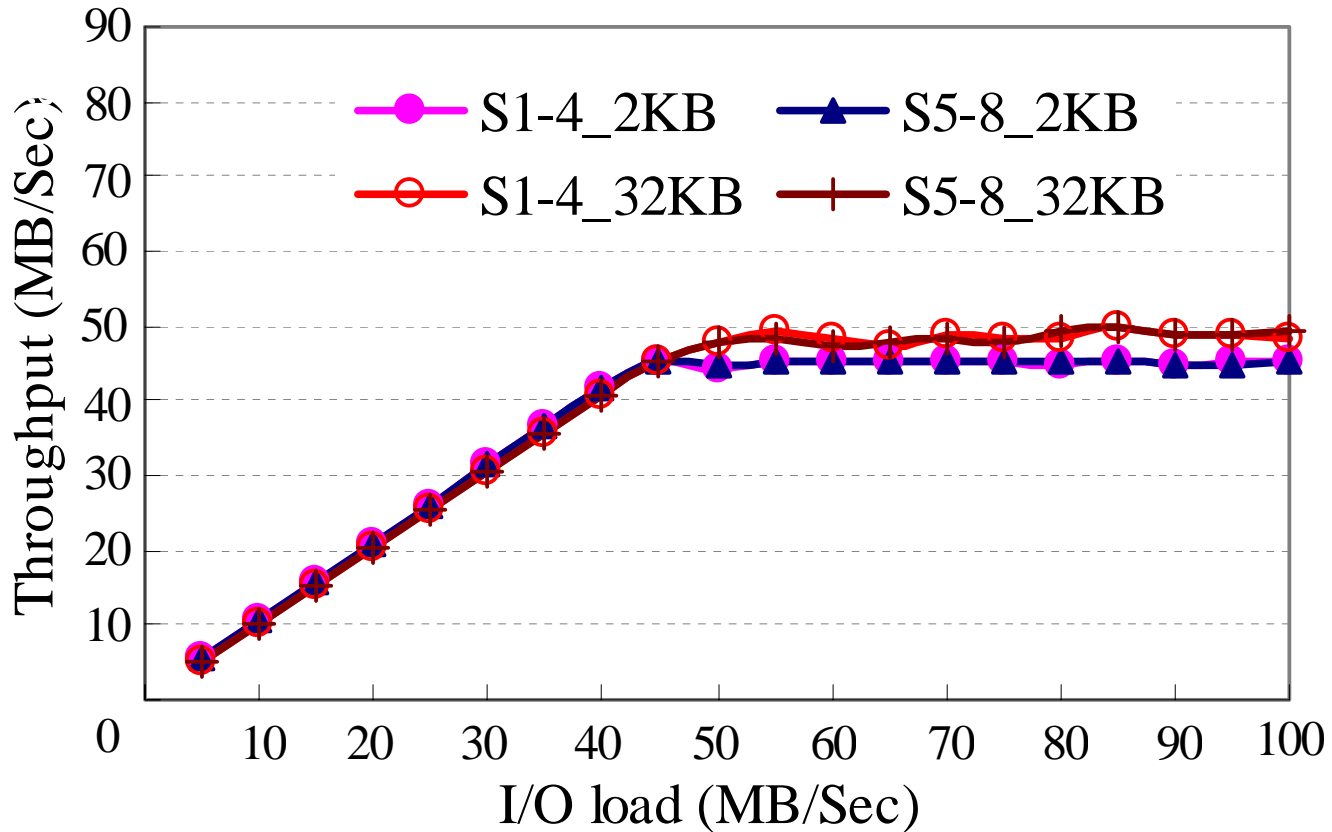
# Max Throughputs Impact



- **45MB/s for 2KB IO size, 80MB/s foe 32KB IO size.**
- **Performance is limited by network.**
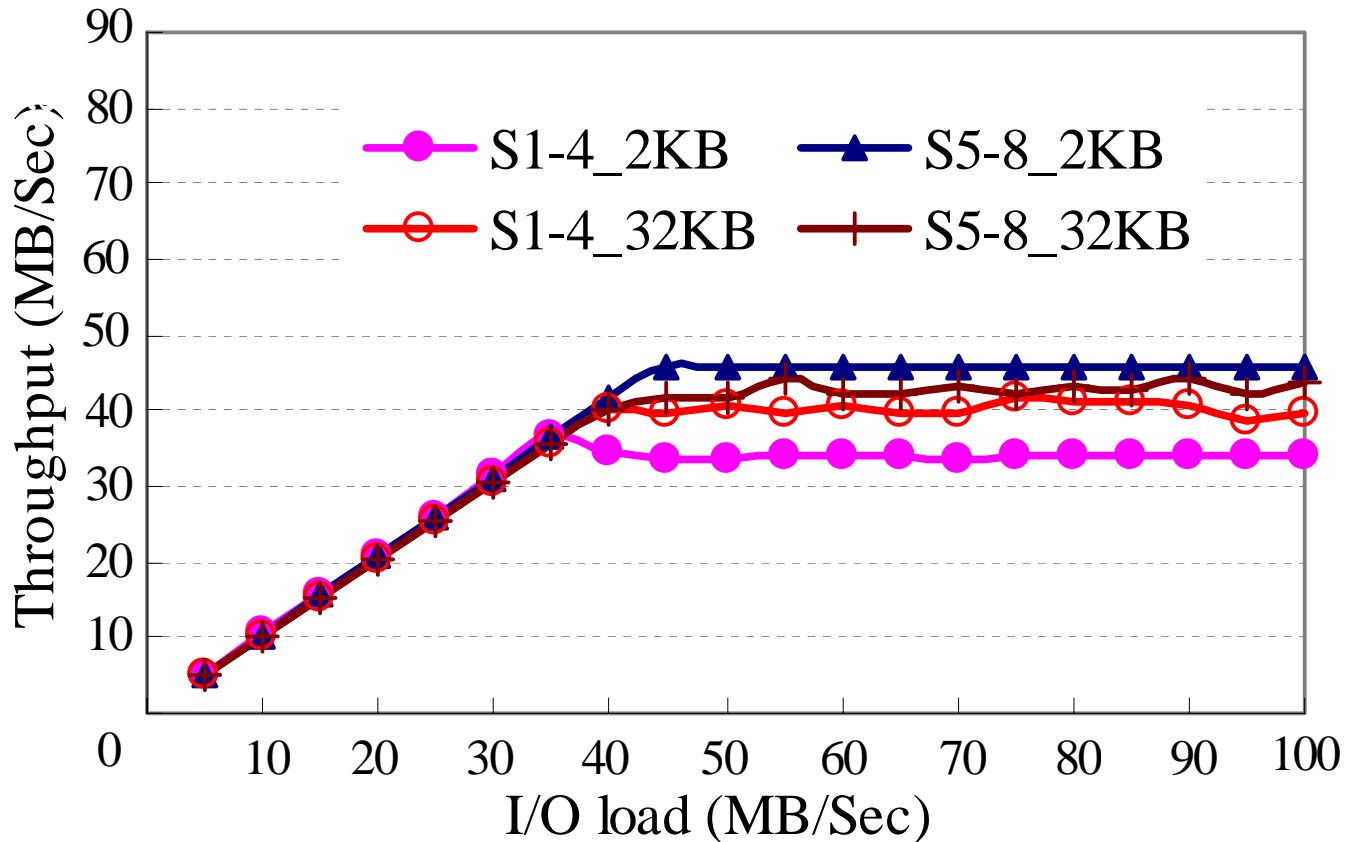
# Max Throughputs Impact



- **The asymptotic throughput of all servers drops.**
- **head-of-line blocking.**

# Max Throughputs Impact



•**The storage link failure decreases the FC network throughput.**
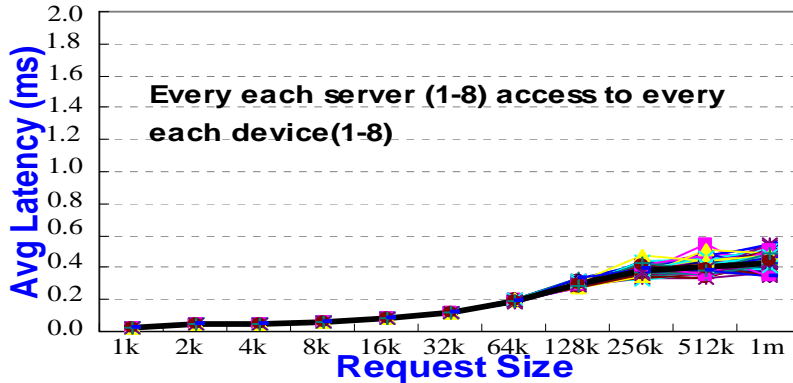
# Max Throughputs Impact



- **The link failure decreases the FC network throughput.**
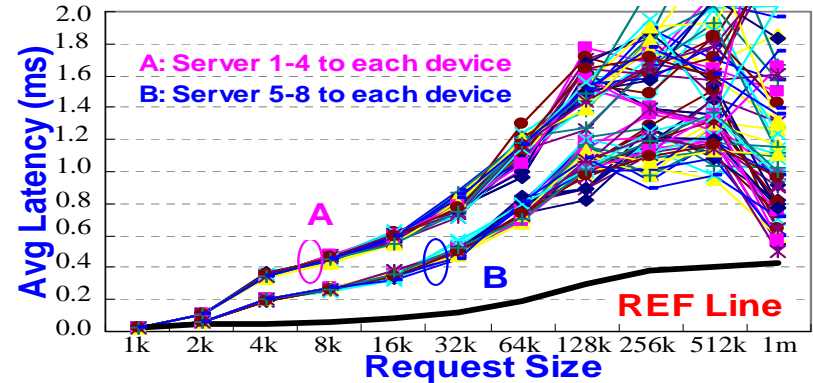- **The max throughputs are sensitive to network failure location.**
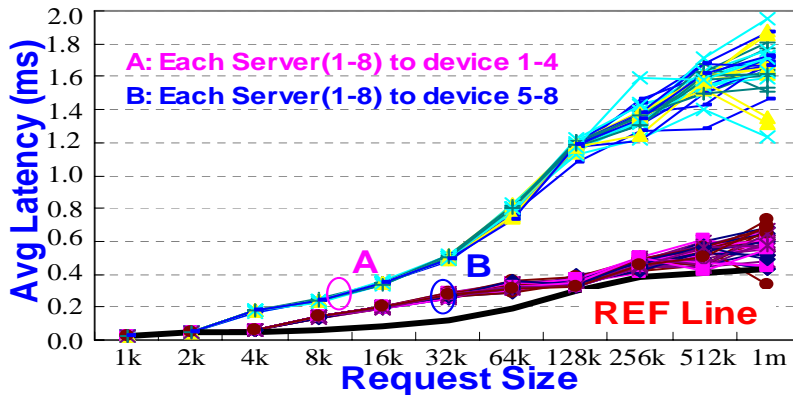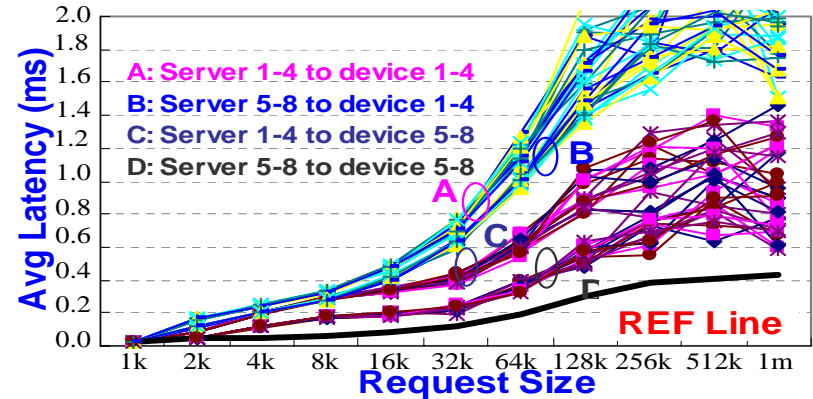
# Data Frame Latency Study



a) Case I: No link failure

b) Case II: ISL 1 failure

c) Case III: ISL8 failure

d) Case IV: Both ISL 1 and ISL 8 failure

- The frame latency increases rapidly after knee point
- The frame latency is sensitive to network failure location.

# Summary

✓ **SAN modeling**

   - **We have developed a simulation tool SANSim which supports FC frame level simulation.**

✓ **Model validation**

   - **We have conducted the experiments and compared experimental and simulated results. The results show that SANSim model is accurate with error range less than 3% for read operation.**

✓ **FC Network Simulation and Analysis**

   - **The link failure decreases the FC network throughput. The maximum throughputs are sensitive to network failure location.**

   - **The simulation results show that the performance loss can reach up to 50% when the servers runs with high workload.**

   - **The FC SAN network designers have to consider I/O workload requirements as well as the impact of network failure for critical application.**

# Future Work

- File system/OSD simulation

- Application simulation

- IP Storage networking module

- Emulator

- Remote storage

# Thank You

For more technical information, please contact:

Dr Zhu Yaolong

(email : Zhu_Yaolong@dsi.a-star.edu.sg)