Performance Analysis and Testing of Storage Area Network

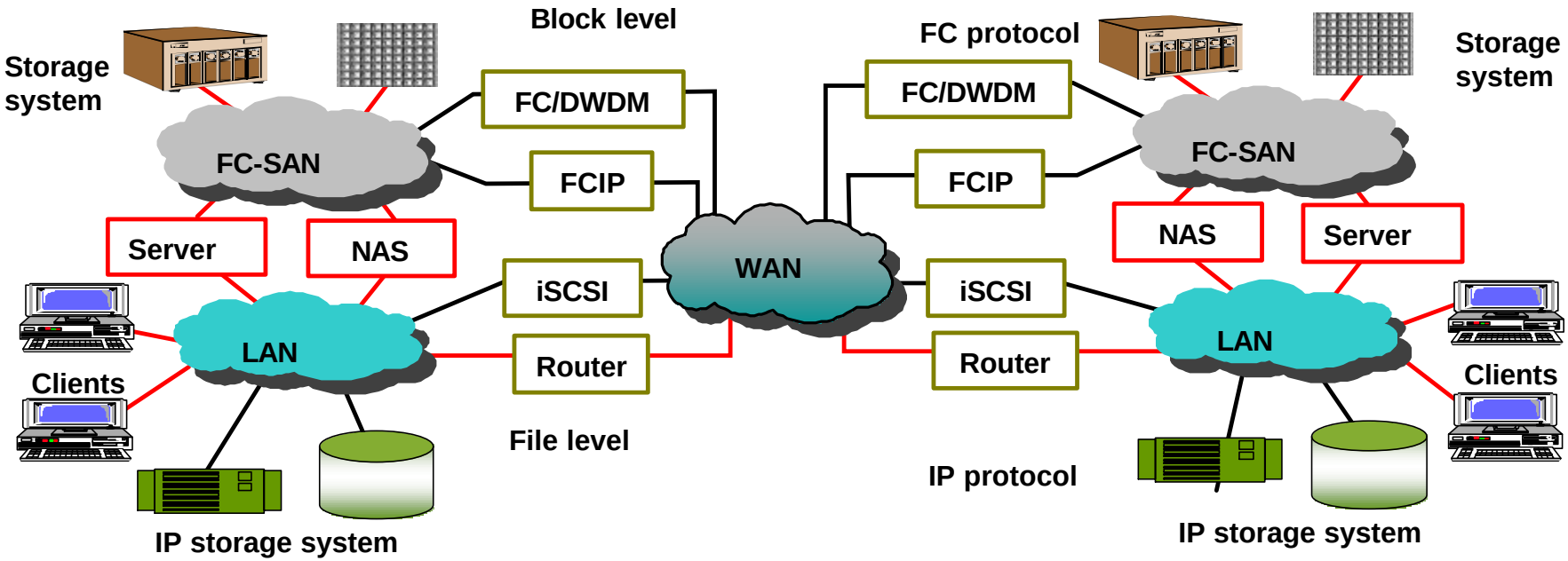
Yao-Long Zhu, Shu-Yu Zhu and Hui Xiong

Data Storage Institute, Singapore

Email: dsizhuyl@dsi.nus.edu.sg
http://www.dsi.nubs.edu.sg

Motivations

- What should we do to optimize the storage system directly connected to storage network instead of directly connected to server?
- How to compare IP storage, Fibre Channel, and InfinBand?
- Do we need new algorithms to replace the RAID technology which introduced in 1980s?
- How to evaluate and analyze the performance of the Storage Area Network easily and quickly?
- Modeling and simulation is a faster way to study these questions than implementation and testing



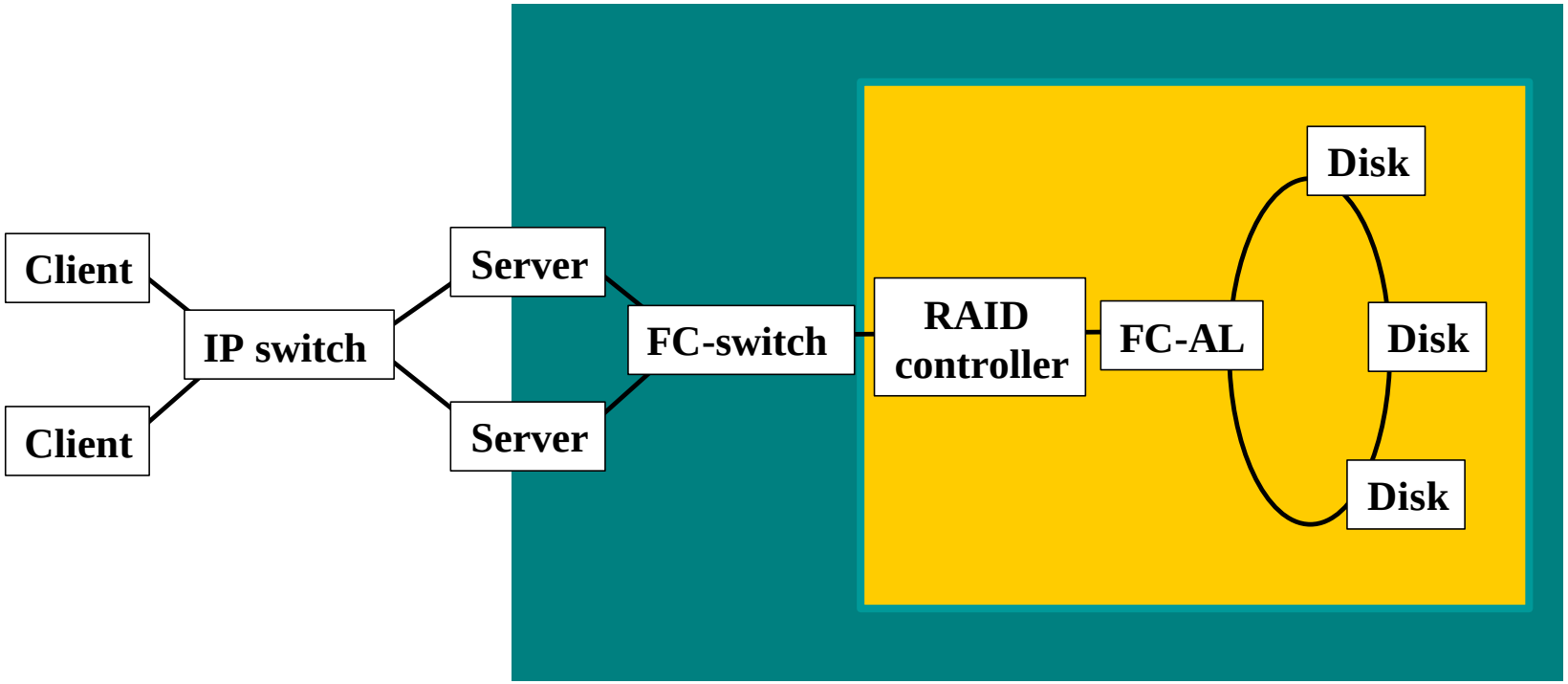
Key Points

Build up a Queuing Network Model for total SAN performance analysis from perspective of networking

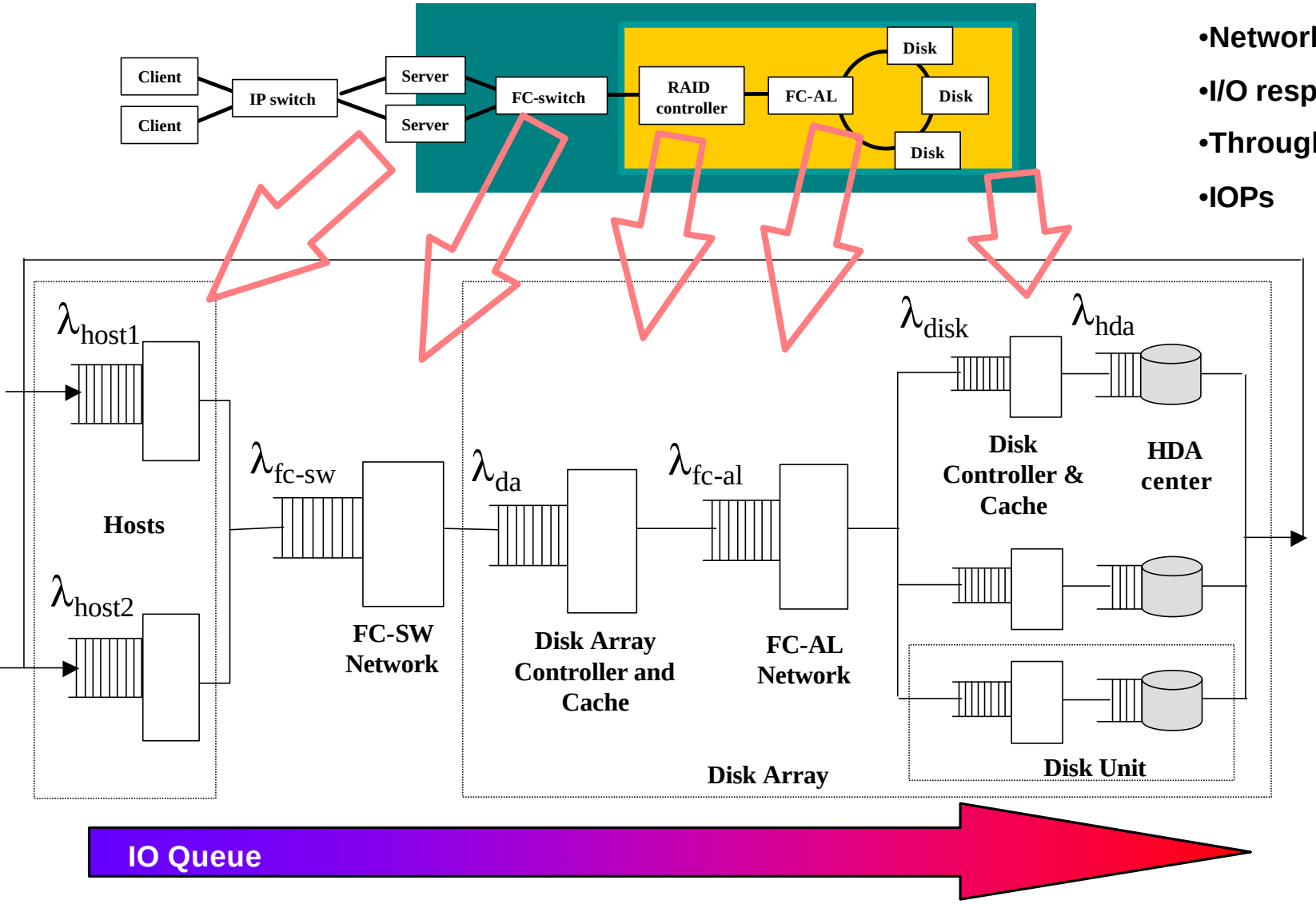
- **Analyze the effects of the I/O workload on the SAN performance**
- **Analyze the effects of disk cache and Fork/Join model**
- **Analyze the FC-AL's scheduling algorithms**

Comparison of the theoretical and experimental results

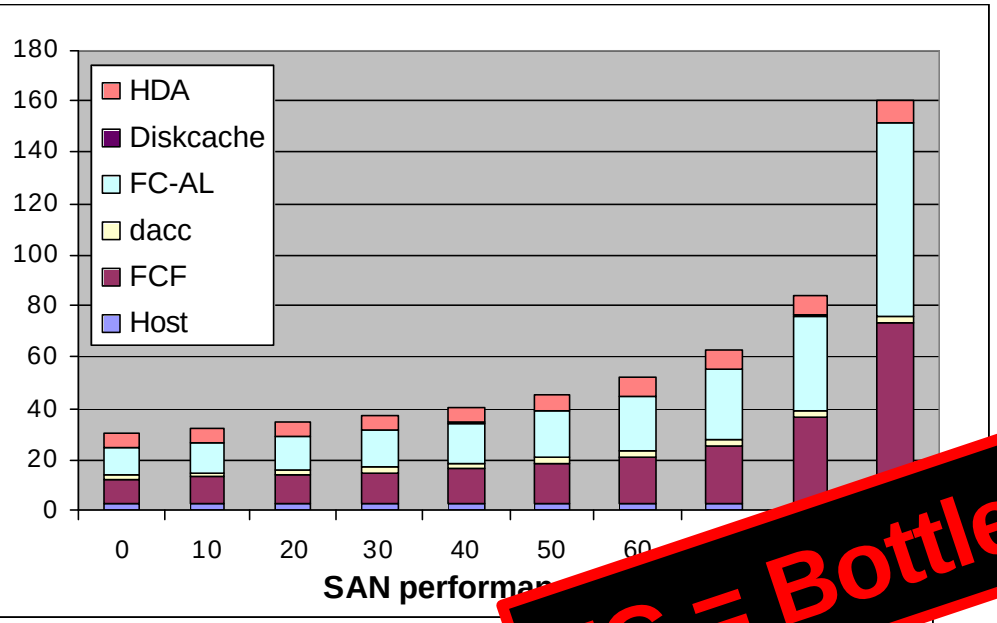
Modeling of SAN



Queuing Network Model



SAN Performance for Big Sequential I/O

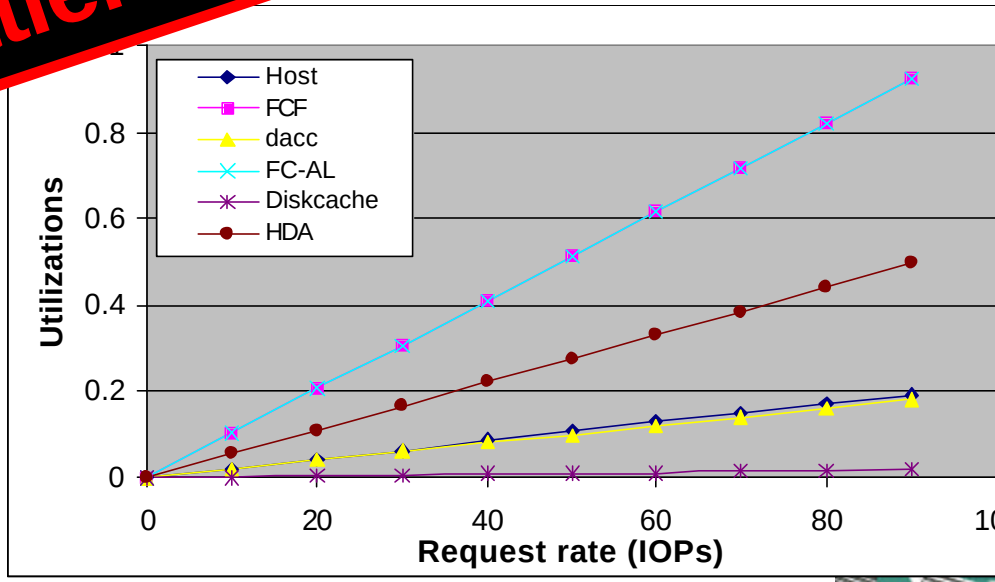


SAN performan

f service nodes vary with system
te for sequential 1Mb I/O size.

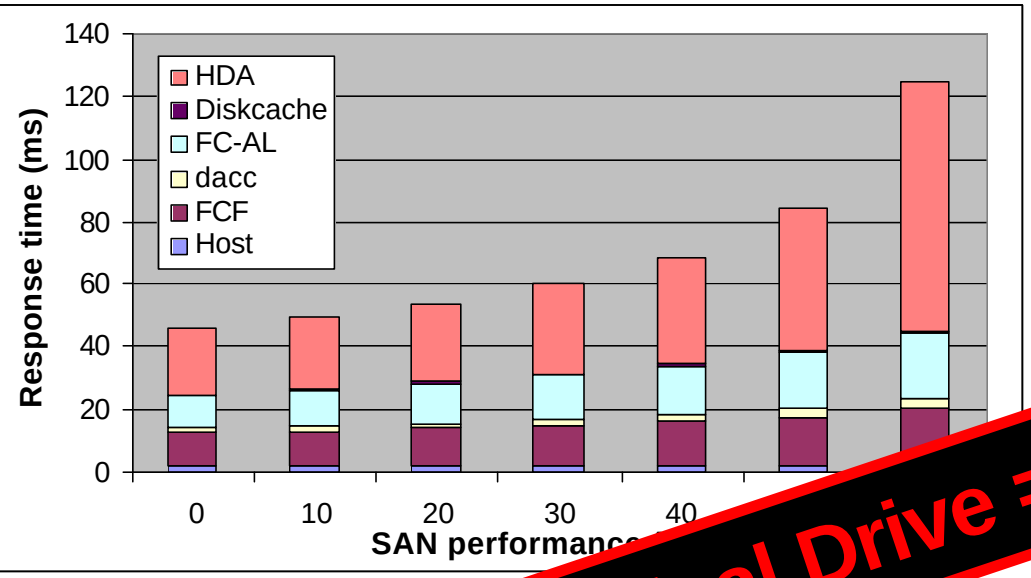
FC = Bottleneck !

System response time vari ystem throughput for sequential MB I/O size.



I/O workload: read
System configuration:
1 server+ 1 daccDACC + 5 FC-AL disks
HDA: Head-Disk Assembly (mechanical part)
dacc: Disk array controller and cache
FCF: FC fabric Switch

SAN Performance for Big Random I/O



System response time varies with system throughput for random 1MB I/O size.

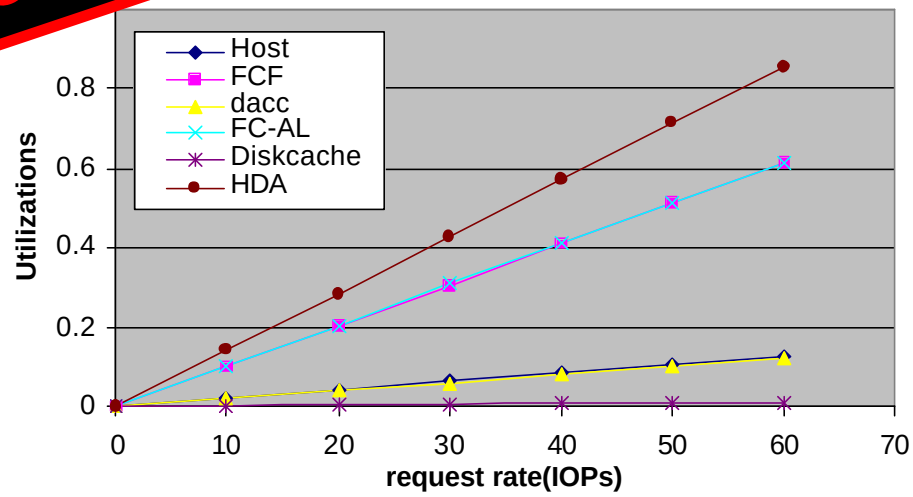
Physical Drive = Bottleneck !

HDA response time is t… …on (40%~60%) in the whole …se time distribution.

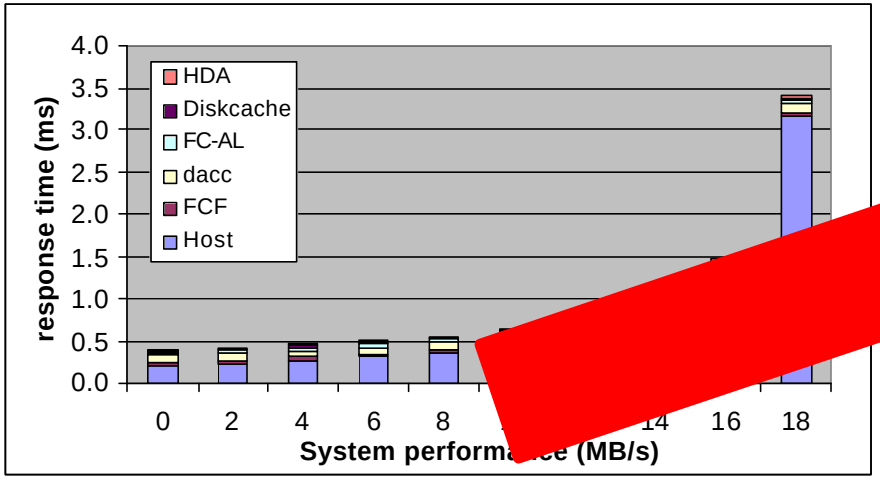
performance of

- **5 hard disks : 66MB/s**
- **15 Hard disks : 79MB/s**
- **25 Hard disks : 82MB/s**

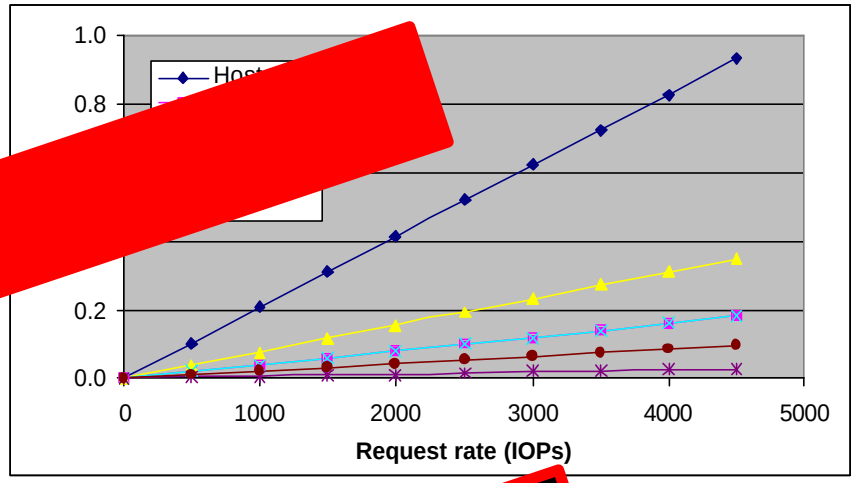
…ice nodes vary with …t rate for random 1MB I/O size



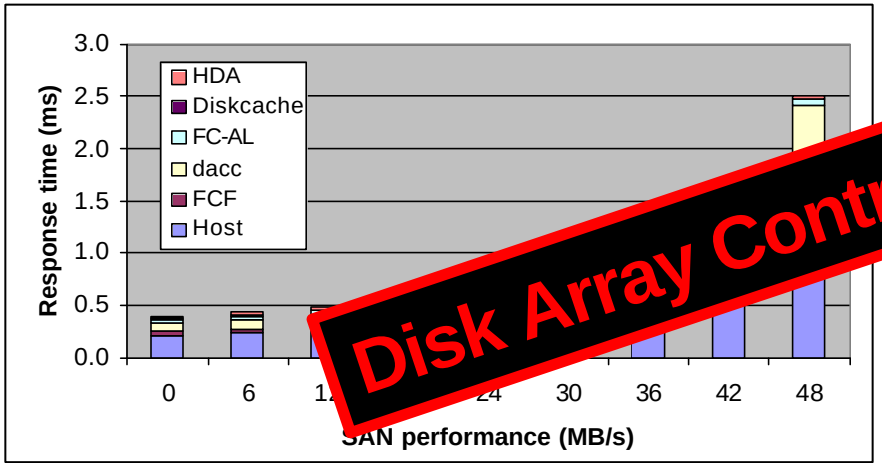
SAN Performance for Small Sequential I/O



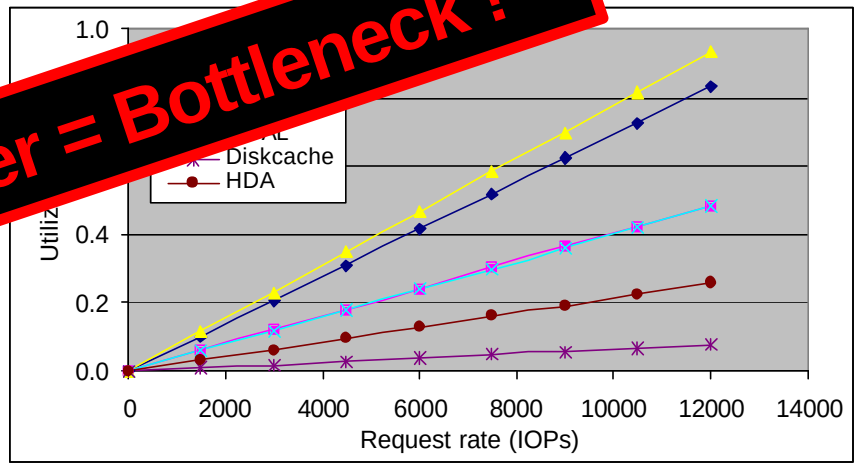
System response time varies with the throughput for sequential 4KB I/O (single user)



Utilizations vary with syst… …est rate for sequential 4KB I/O… …4500 IOPs)



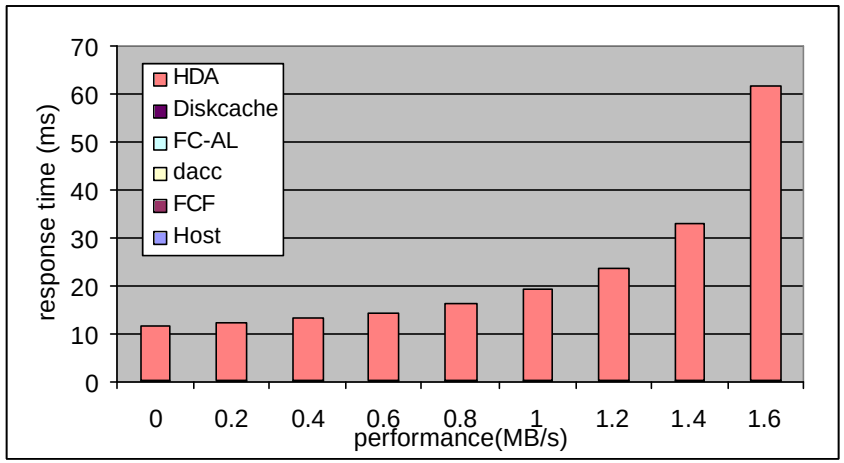
System response time varies with the throughput for sequential 4KB I/O (2 users).



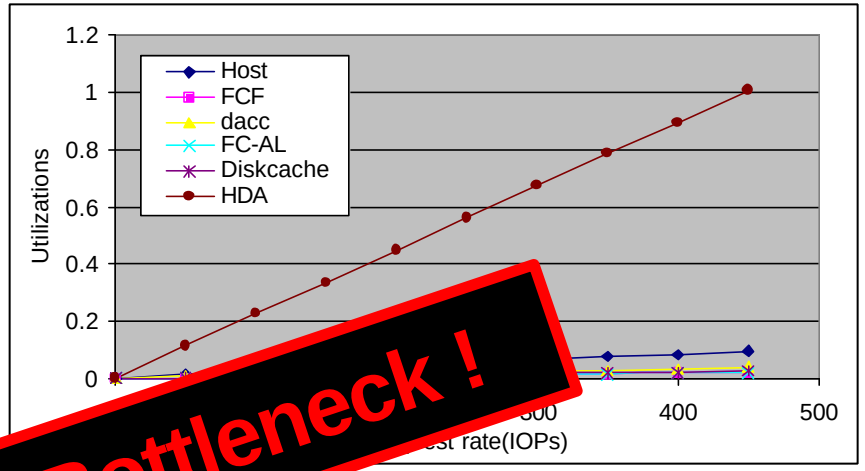
Utilizations varies with the system request rate for sequential 4KB I/O (2 users). (5600 IOPs)

Disk Array Controller = Bottleneck !

SAN Performance for Small Random I/O

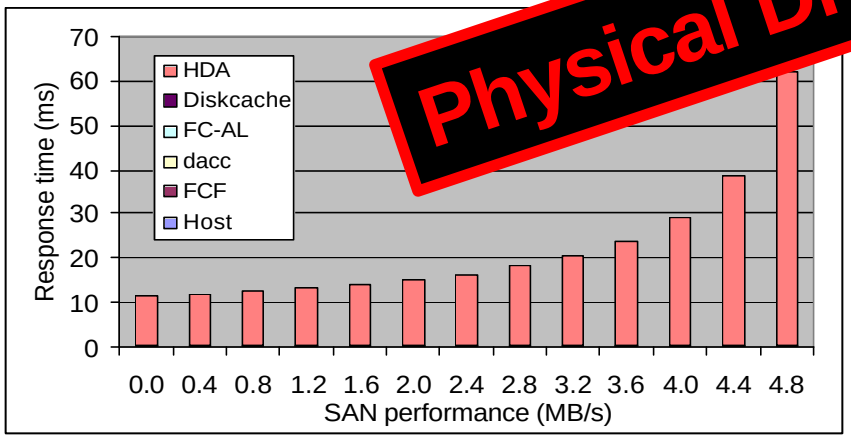


System response time varies with the throughput for random 4KB I/O (5 HDDs).



Physical Drive = Bottleneck !

ations vary with the system request rate for random 4KB I/O (5 HDDs).

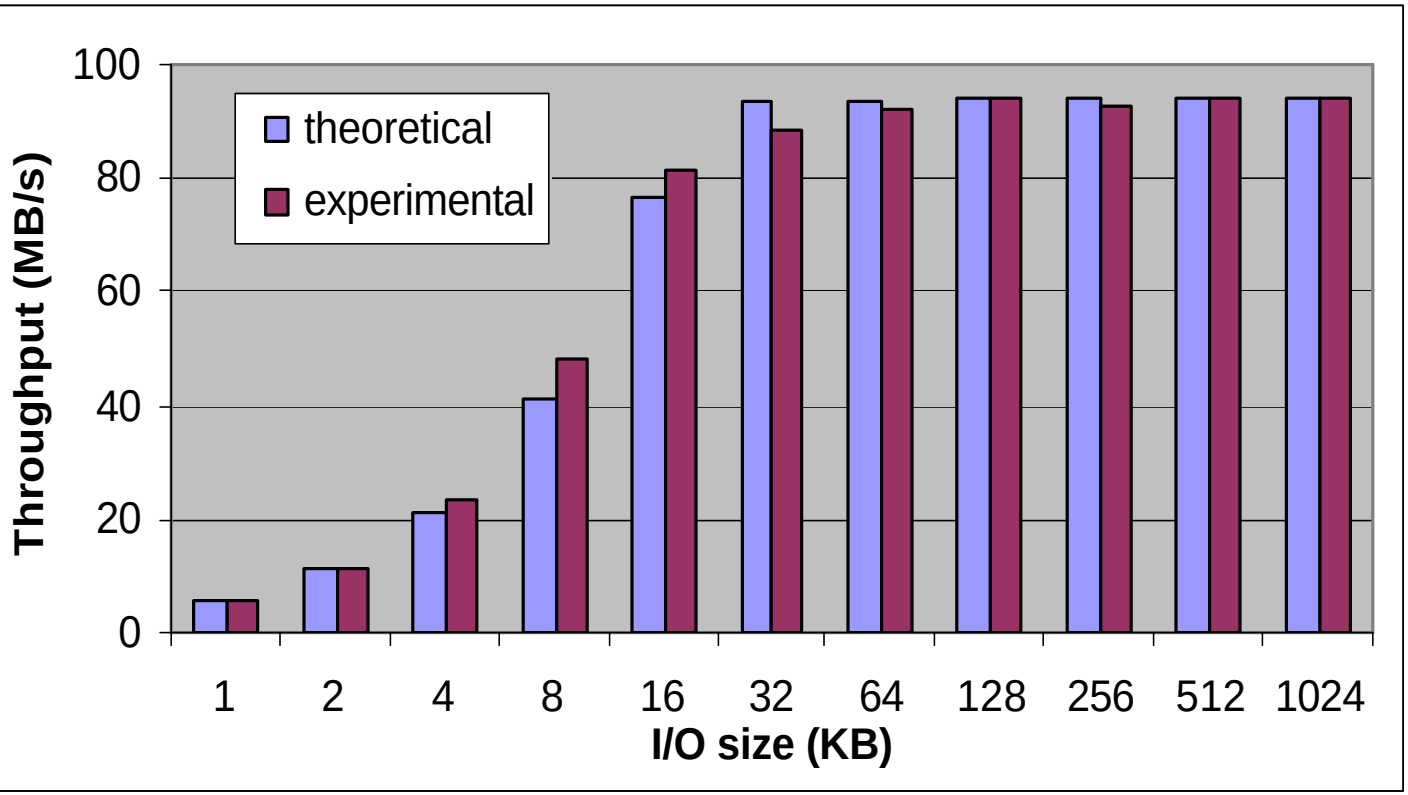


System response time varies with the throughput for random 4KB I/O (15 HDDs).

Performance of

5 hard disks	**:**	**420 IOPs**
15 hard disks	**:**	**1260 IOPs**
25 hard disks	**:**	**2100 IOPs**

Empirical Model Comparison

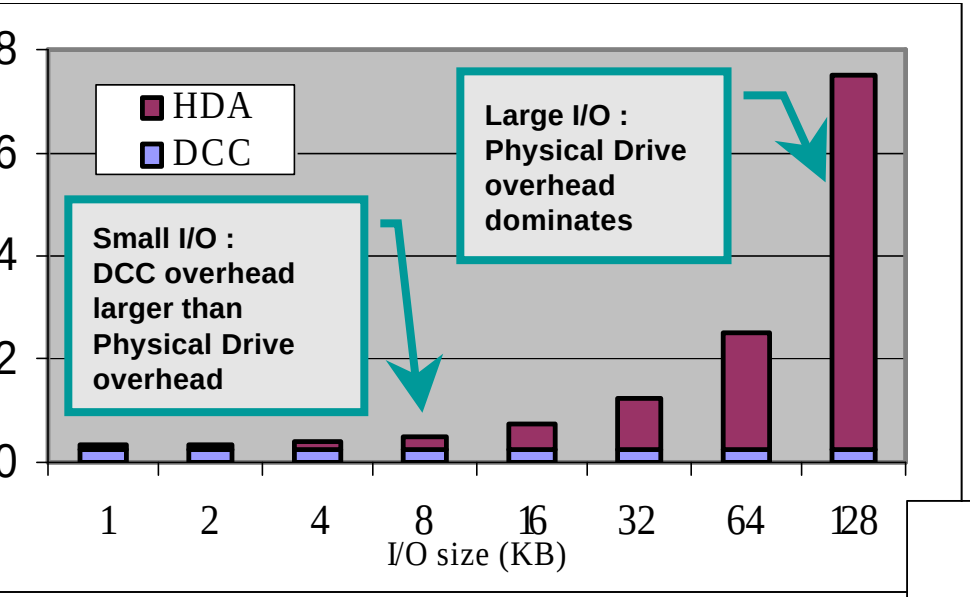


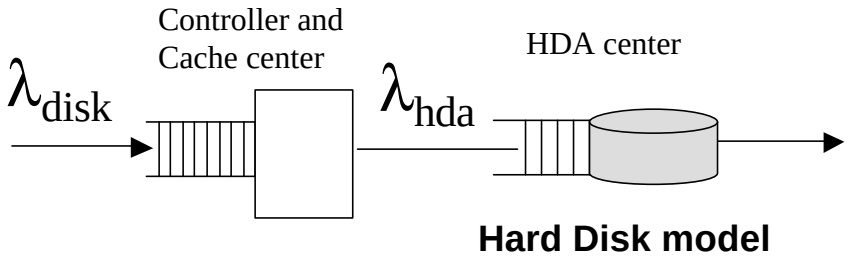
SAN performance varies with I/O size for sequential I/O workload

The tests are based on the multiple Pentium 733MHz hosts with 64bits and 66MHz PCI bus, HBA with Qlogic QLA 2200A, 1G FC switch with Brocade Silkwarm 2400, and a self-developed virtual FC Disk. IOMeter is used as the benchmark tool.

Both of the testing and theoretic results show that the FC network is the system bottleneck for big I/O size (>32KB), and the storage system controller overhead is the system limitation for small I/O size (<16KB).

Lesson #1: Disk Cache / Controller Performance

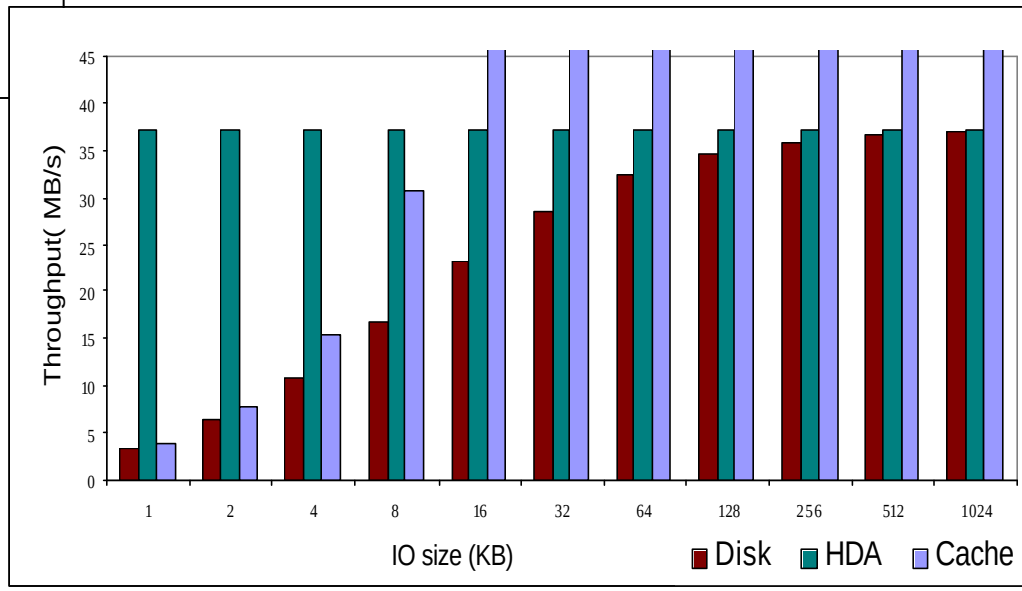




Hard Disk model

Bandwidths of the disk unit, DCC and HDA vary with IO size for single hard disk.

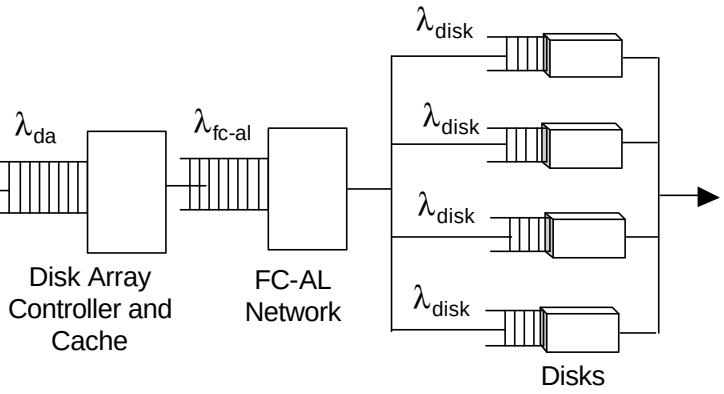
Response time varies with I/O size when the request rate is 200 IOPs for hard disk.

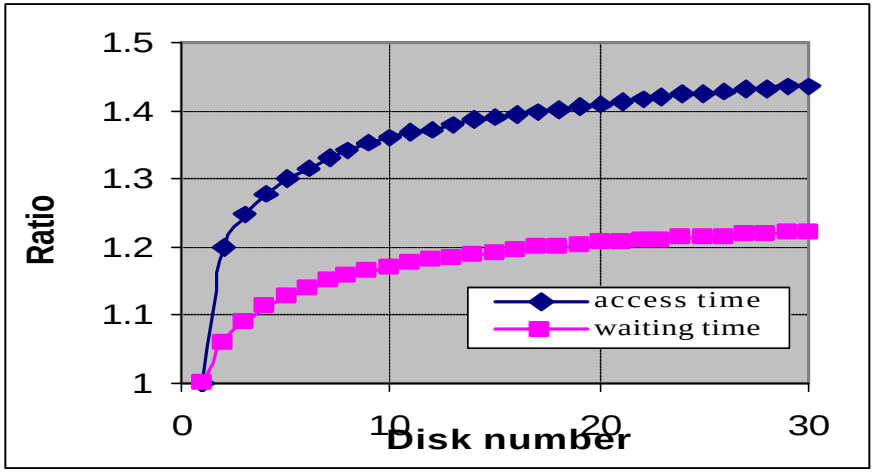


Single Disk – Cache MAY be bottleneck

Disk Array – Cache depends largely on Disk Stripe Size

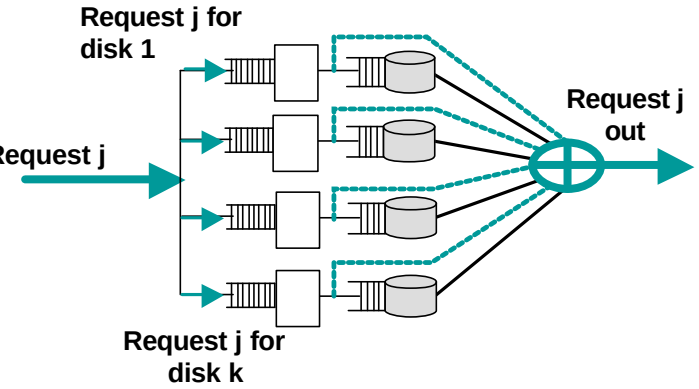
Lesson #2: Disk Array and Fork/Join Model



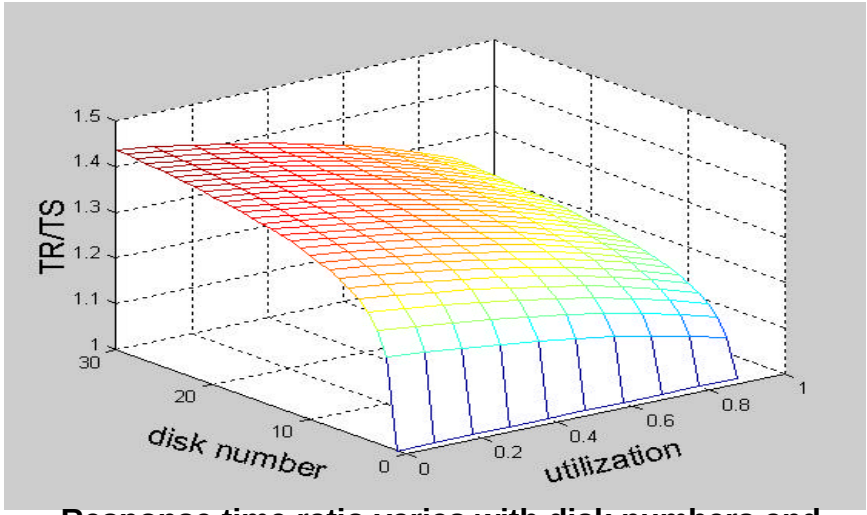


Fork/Join access time and queue waiting time ratio vary with disk numbers.

Fork/Join model

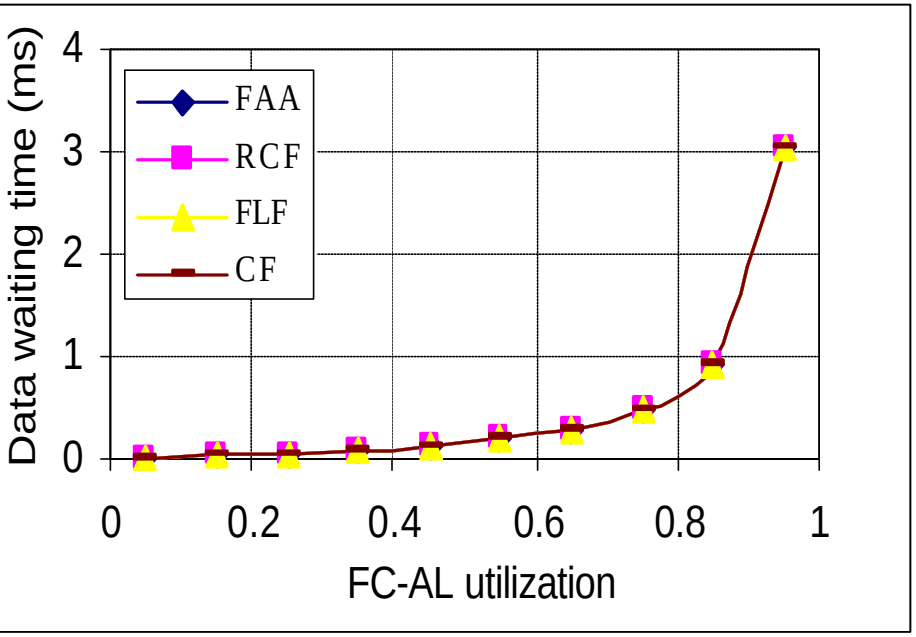


$$TR_k(l) = TS_k(0) + \frac{d * m_{disk}^2 * r_{disk} * E[TS_{disk}^2]}{m_{disk} - l_{disk}}$$

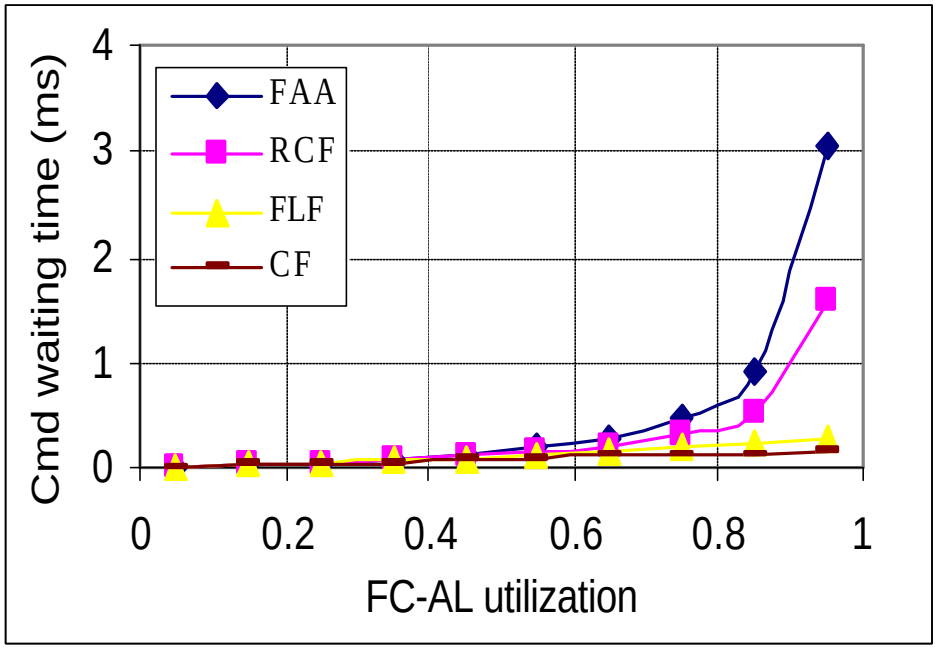


Response time ratio varies with disk numbers and system utilization.

Lesson #3: FC-AL Algorithms



Data average waiting time varies with FC-AL utilization.



Command average waiting time varies with FC-AL utilization

FAA: Fairness Access Algorithm

RCF: Read Command First

FLF: FL-port First

CF: Commands First

Lastly….

The queuing network model has been shown to be a useful tool for analyzing the overall SAN system performance

The model was also used to analyze our advanced SAN storage technology, DA^2

We missed the proceedings publication. However, this paper will be made available on the conference website!

More Questions? Visit our Poster Session

Thank you for your attention!