

## Storage over IP: When Does Hardware Support help?

By Prasenjit Sarkar, Sandeep Uttamchandani, Kaladhar Voruganti

In Proceedings of FAST '03.

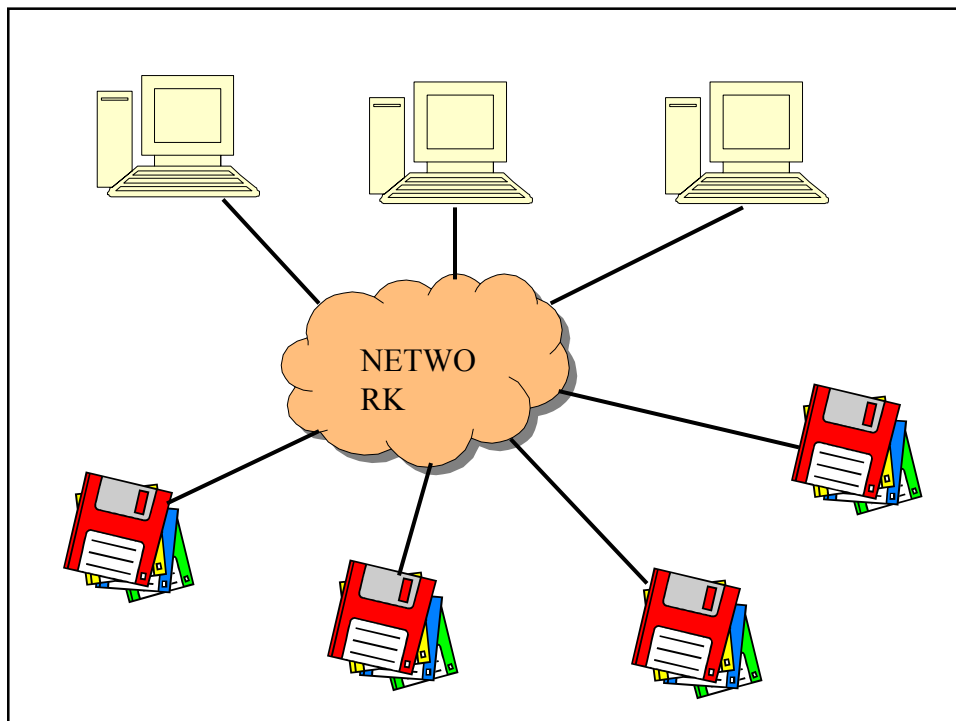
Presenter: Ben Wun  
Discussion Leader: ???

## Motivation

- Traditional model of storage attached to each host is becoming inadequate
  - Limited number of disks per host
  - Limited distance between host and storage
  - Lacks scalability

## Storage Area Networks

- Storage no longer coupled to hosts
- Storage devices provide block storage over a network with multiple hosts and multiple independent storage devices
- Existing solutions:
  - Infiniband
  - Fibre Channel



## Storage over IP

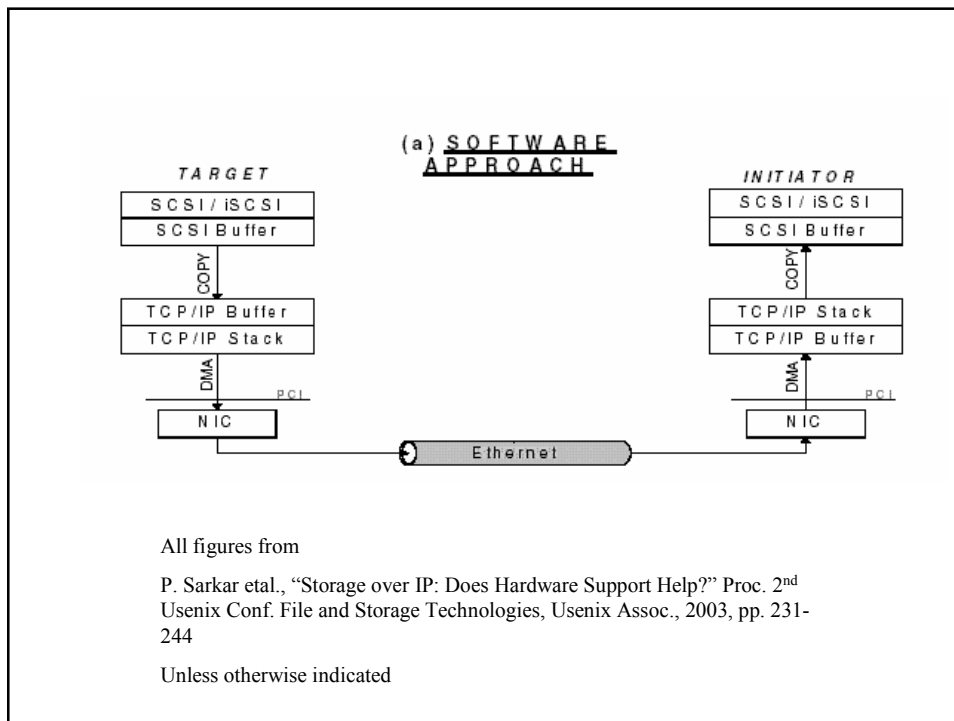
- Advantages of using TCP/IP
  - Homogenous networks
  - Established, well tested
  - Cheap
  - Scalable routing
  - Gigabit and better Ethernet provides enough bandwidth

## iSCSI

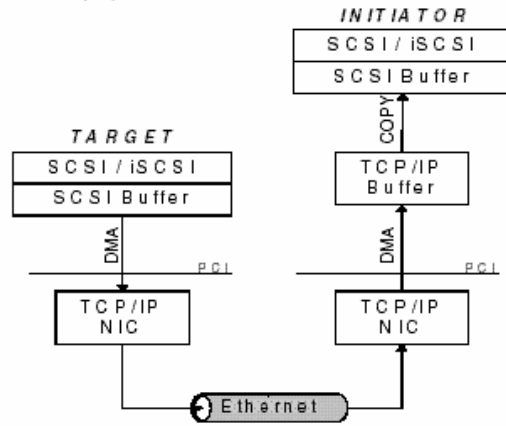
- One of several protocols for Storage over IP
- Internet SCSI
  - Send SCSI commands over TCP/IP

# Support for iSCSI

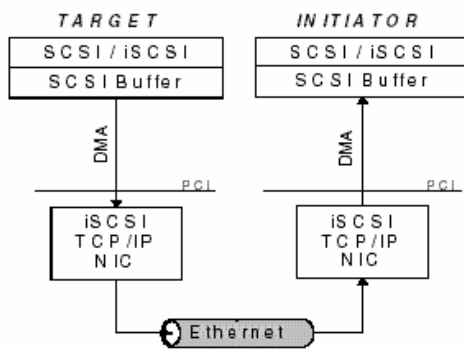
- Software
- TCP Offload Engine (TOE)
  - TCP/IP processing offloaded
  - Storage protocol on host
- Host Bus Adapter (HBA)
  - TCP/IP and storage protocol processing on special adapter



**(b) IOE APPROACH**



**(c) HBA APPROACH**



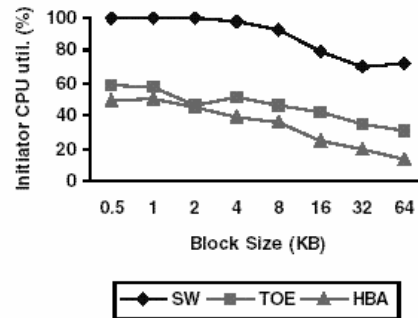
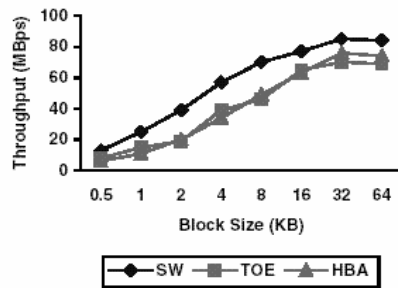
## Advantages of offloading

- TCP copy/checksum is expensive
  - Offloading reduces load on processor
- Interrupt and other system overheads reduced
- Fewer data copies
- CPU freed for application processing

## Test 1: Microbenchmark

- Tests iSCSI initiator reads
- iSCSI target uses software only approach
  - Authors made sure this was not the bottleneck

## Block Size Sensitivity



## Block Size Sensitivity

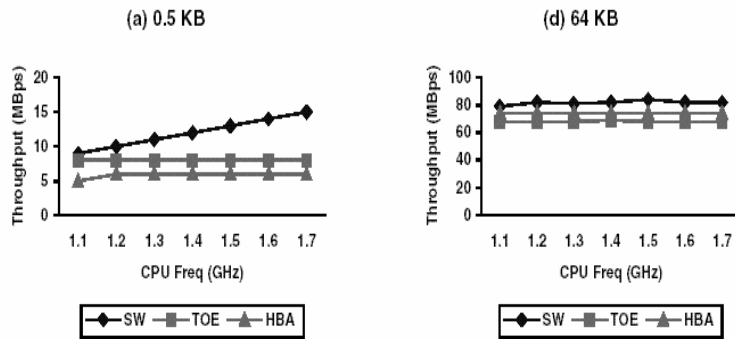
- Software offers the best throughput
- TOE and HBA offer better CPU utilization
- Comparative Metric: CPU utilization/throughput
  - Important for applications that are compute bound
  - Hardware approaches are a win in this respect, especially when per byte costs dominate

<b>Block Size (KB)</b>	<b>Latency (ms)</b>			
	<i>0.5</i>	<i>4</i>	<i>8</i>	<i>64</i>
<i>Software</i>	0.12	0.17	0.22	0.97
<i>TOE</i>	0.17	0.26	0.28	1.01
<i>HBA</i>	0.41	0.47	0.51	1.52

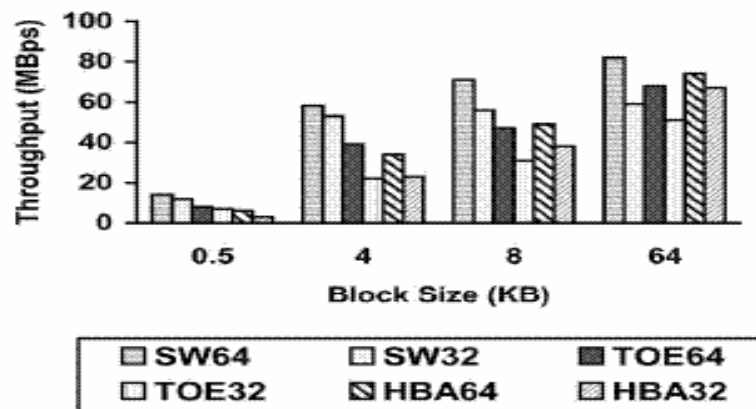
## Per Operation Overhead

	<b>Ops per second</b>	<b>Initiator CPU util. (%)</b>	<b>Initiator CPU util. per op (%)</b>
<i>Software</i>	8267	38	0.0046
<i>TOE</i>	5959	22	0.0036
<i>HBA</i>	2580	7	0.0027

## CPU speed sensitivity



## I/O Bus speed sensitivity



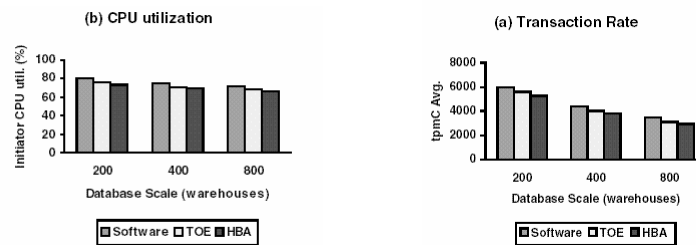
## Bottlenecks

- Memory is not
  - Changing memory speeds does not seriously affect results
- PCI can be
  - Especially in software approach
- CPU overhead for single operation is smaller for hardware than software
  - Implies driver processing not bottleneck
- Must conclude that slower offload engines are the main culprit

## TCP-C

- On-line transaction processing (OLTP)
- Measures number to transactions per minute
- It is CPU intensive, but software approach still wins
  - CPU is never fully utilized because number of connections is limited
  - Unfortunately, number of connections is not tunable

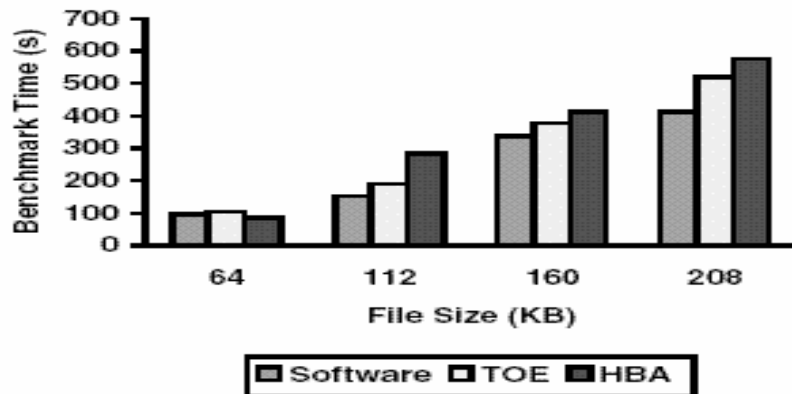
# TCP-C



# PostMark

- Simulates email, web-commerce, etc.
- Creates, reads and writes, and deletes files
- Software is best at large file sizes
  - Superior latency

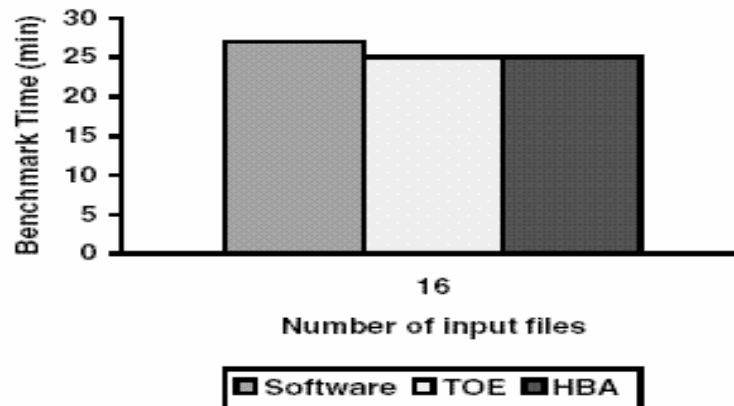
## Postmark



## TPIE-Merge

- Assesses I/O performance under high CPU load
- Offload approaches superior to software
  - High utilization/throughput ratio finally comes into play

## TPIE-Merge



## Analysis

- Software approach is clearly superior unless the application in question is highly compute bound
- The big bottleneck for hardware support is the lack of processing power on the offload engine
  - Faster offload cards require more power, cooling, money
  - May be more practical at 10Gbps or higher

## Speculative Improvements

- Multiple adapter systems
  - Aggregate processing power of multiple cards is greater
- Multi-Initiator Scenario
  - Offload of connection setup/teardown allows effective support of more connections

## Conclusion

- Current hardware support for storage area networks is not necessarily helpful
- Biggest problem is that offload engines are slower than the host processor
- We must reevaluate our approach for supporting high speed storage area networking

# Fibre Channel vs. iSCSI

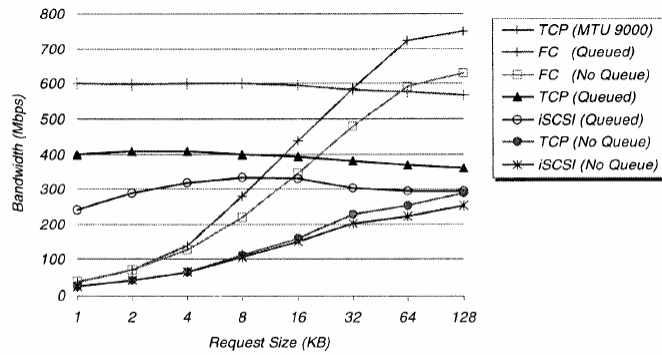


Figure 6. Fibre Channel target mode driver performance comparison with other methods (CPU 400MHz)

Figure From

Huseyin Simitci, Chris Malakapalli, Vamsi Gunturu. "Evaluation of SCSI over TCP/IP and SCSI over Fibre Channel Connections," *hoti*, p. 0087, The Ninth Symposium on High Performance Interconnects (HOTI '01), 2001.