Introduction to HyperSCSI

or **"Designing a network storage protocol"**

By Patrick Khoo and Wilson Wang

MCSA
Modular Connected Storage Architecture

DSI SINGAPORE

Why Not?

- We are building Ethernet-based storage
- Instinct says use TCP/IP because it has everything we need for networking
- "We're stuck with TCP and the higher levels because of Ethernet's inability to handle packet management and stuff." – Mass Storage Conference 2002
- Solution: Focus on solving Ethernet's inability for storage, not TCP's shortcomings (eg TOEs) and do so without "rebuilding" TCP/IP
- Can it be done? Yes, HyperSCSI demonstrates that this can be done

HyperSCSI

The transmission of SCSI commands & data across a network

- Support for various storage devices & interface technologies, and existing applications & hardware (currently includes SCSI, IDE and USB devices)
- Runs on raw Ethernet (100Mbit/s, GE and GE+Jumbo Frames) as well as IP-based infrastructure
- Supports device specific options including encryption and access controls
- Support for multiple redundant load-balancing with automatic fail-over links
- Support for plug-and-play active device discovery functions
- Provides in-band management functions
- Easy to deploy and use, but provides users with plenty of choices

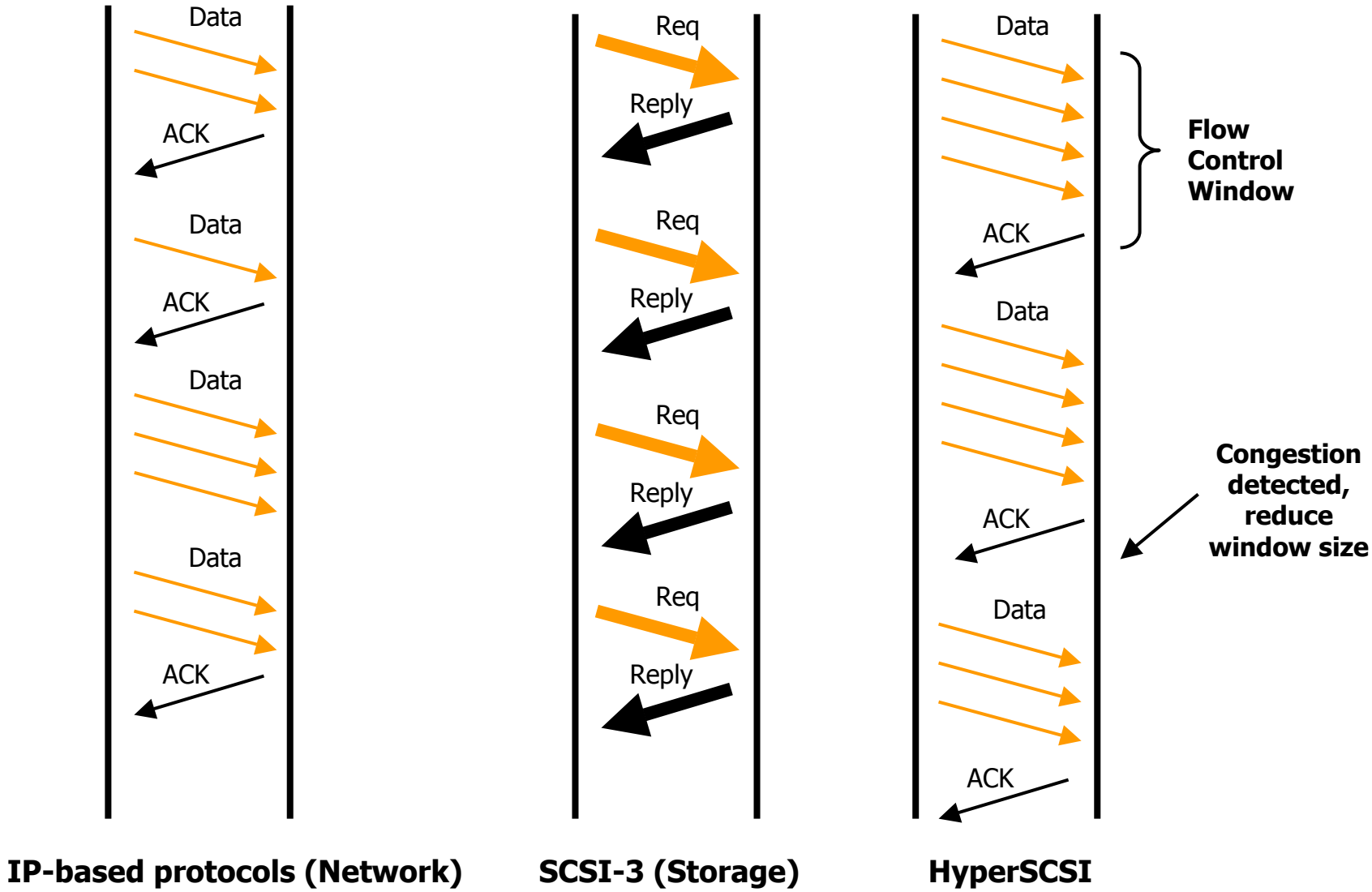
Delivering SCSI

	SAR	Flow Control	Delivery	Reliability	Wide-area Connectivity	Process Complexity	Channel Efficiency
Ethernet	No	No	Out of Order	Best Effort	No	Low	Highest
IP	No	No	Out of Order	Best Effort	Yes	Medium	High
UDP/IP	No	No	Out of Order	Best Effort	Yes	High	Medium
TCP/IP	Yes	Yes	In Order	Guaranteed	Yes	Highest	Low

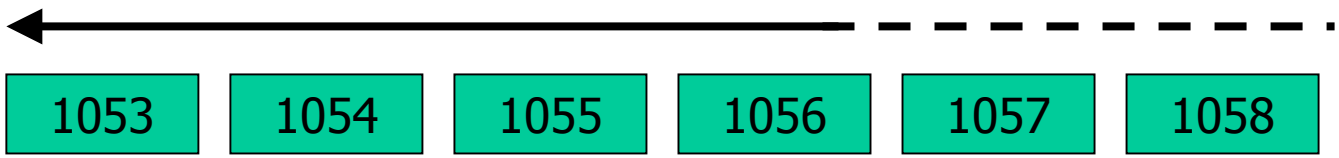
Problems with Ethernet include:

- No flow / congestion control
- Undeterministic transmission
- Best effort transmission / reliability
- Small frame size
- Lack of security
- No wide-area capability

Flow / Congestion Control & SAR

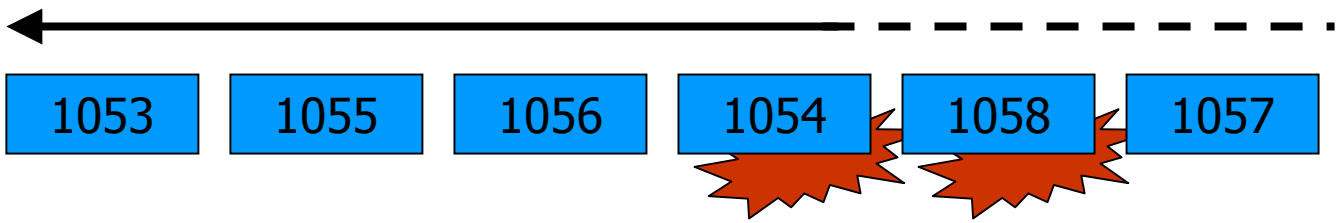


Deterministic Data Delivery



Network Data Stream

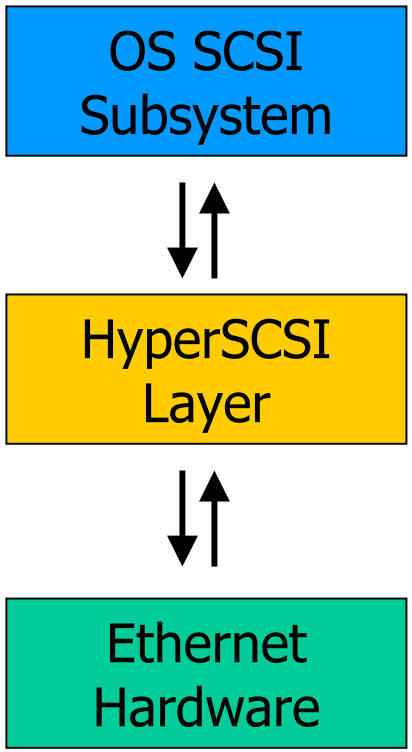
- Transfers must be in-order, otherwise, command completion, security will be compromised



SCSI Block Transfers

- SCSI blocks can be transferred out-of-order
- Storage protocols do not need to guarantee in-order delivery of commands

Best Effort Reliability

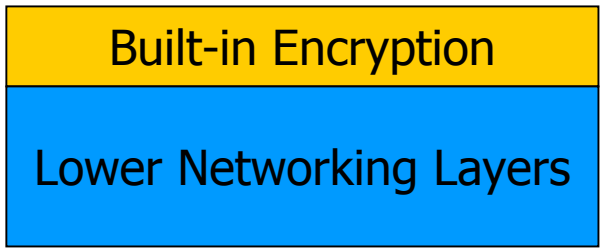


Check and Verify SCSI Command and Data Blocks

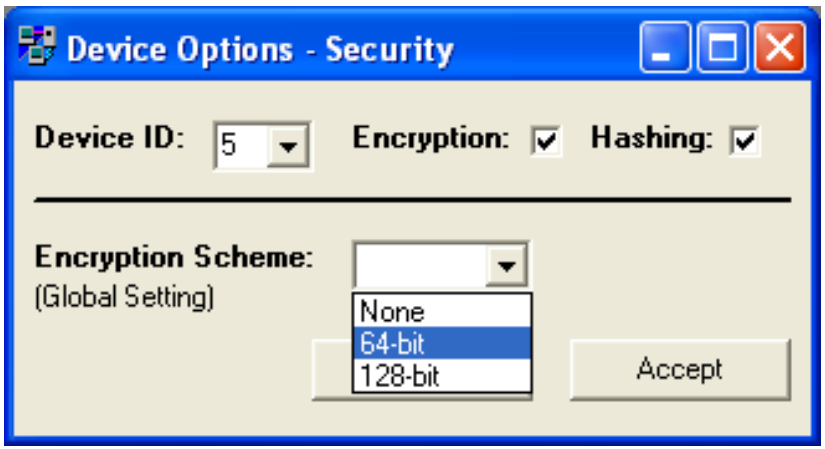
Provide SAR, Flow Control and Packet Retransmission functions

Guarantee data accuracy and correctness

Encryption and Security



- Security built into protocol itself provides more granularity, higher performance and is easier to deploy

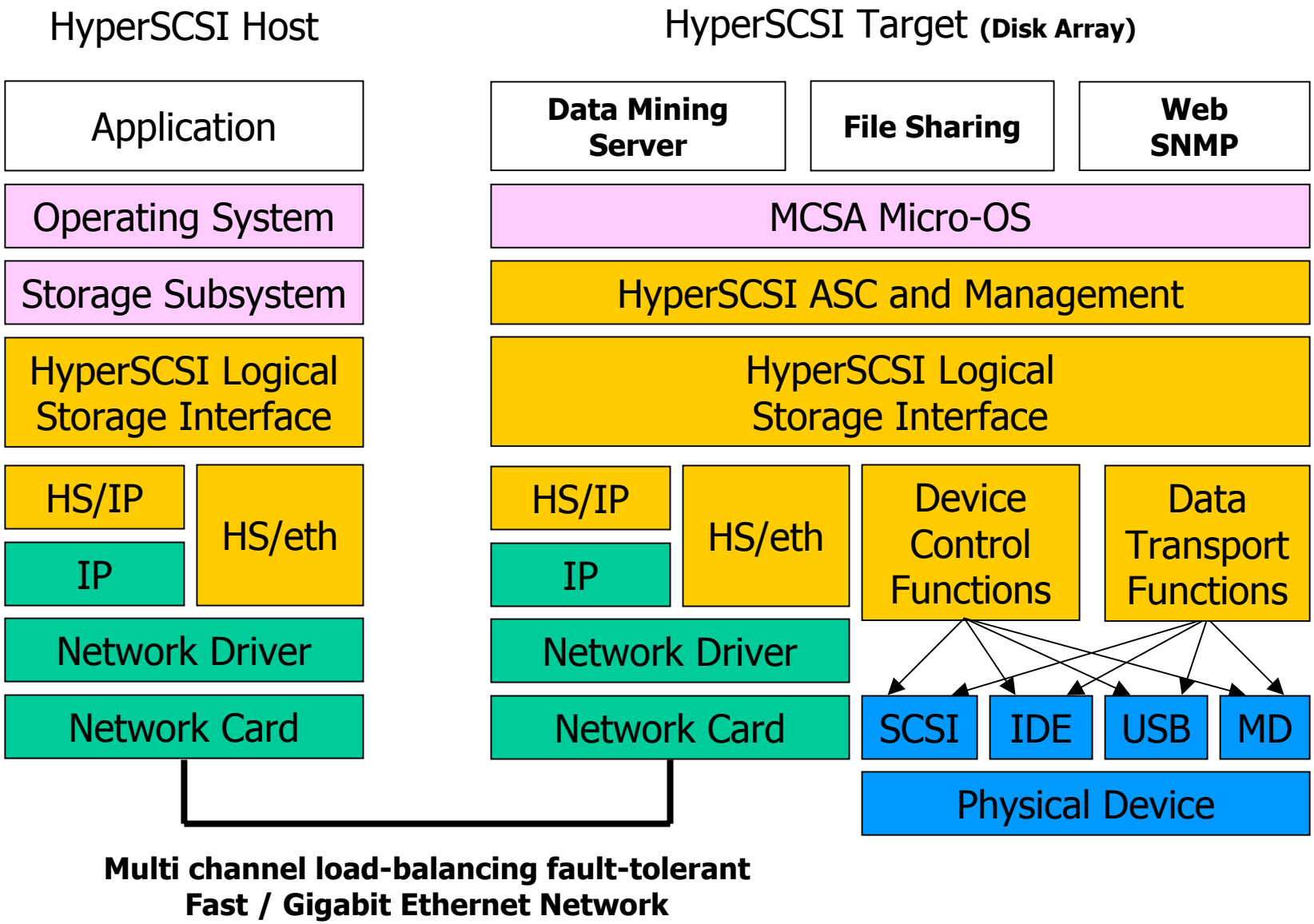


Wide-area Support

Two implementations of HyperSCSI in development

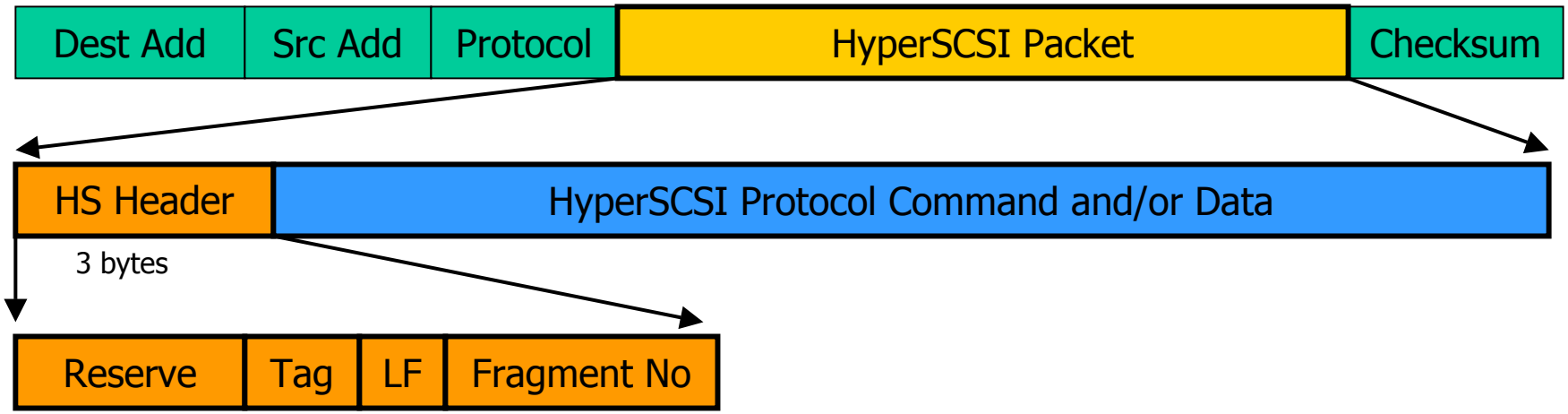
- **HS/eth** – HyperSCSI over Ethernet. IEEE Ethernet Type Field number: #889a (3 Dec 2001)
- **HS/IP** – HyperSCSI over IP. IANA User Port number: 5674/tcp and 5674/udp (1 Feb 2002)
- HS/IP and HS/eth can be deployed together in the same network environment, or separately. They function independently of each other, but all clients and servers have the capability to understand both. It is only an "on/off" switch for the user.

HyperSCSI Architecture

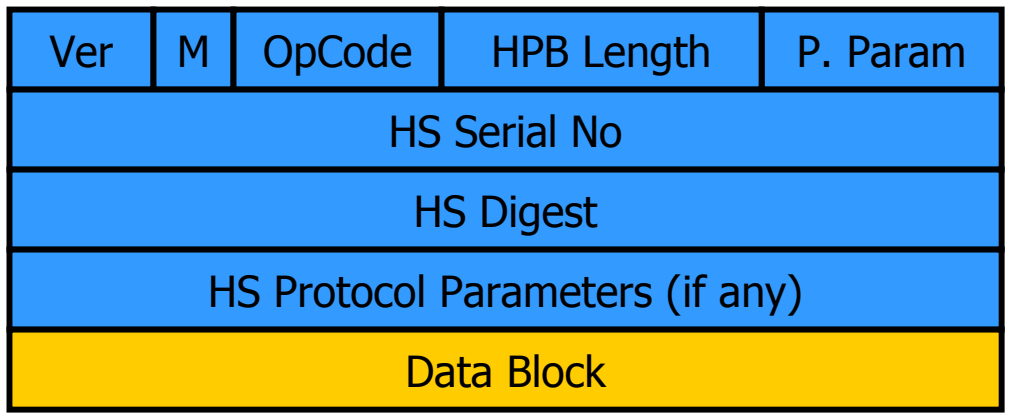


HyperSCSI Protocol

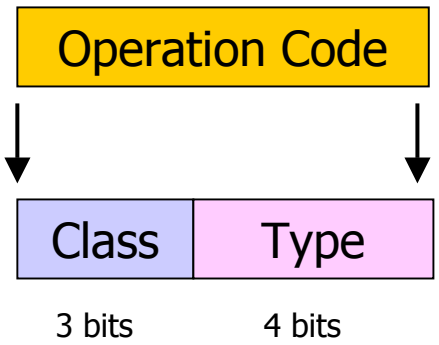
HyperSCSI Packet Framing / Encapsulation



HyperSCSI Command and Data Block

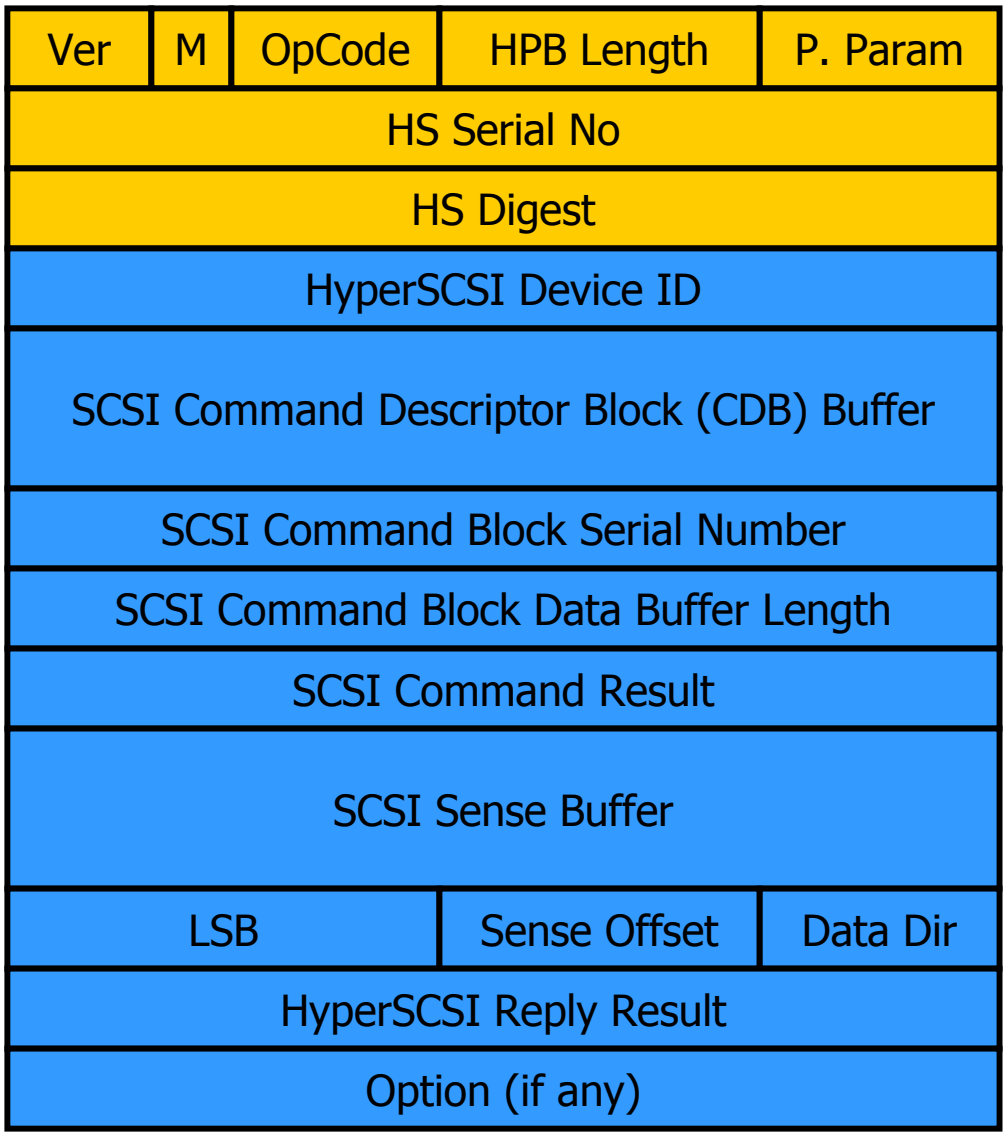


HyperSCSI Operation Codes



Class		Type	Name
Command Block Encapsulation	0x00	0x00	HCBE_REQUEST
		0x01	HCBE_REPLY
Connection Control	0x01	0x00	HCC_DEVICE_DISCOVERY
		0x01	HCC_ADN_REQUEST
		0x02	HCC_ADN_REPLY
		0x03	HCC_DISCONNECT
Flow Control	0x02	0x00	FC_ACK_SNR
		0x01	FC_ACK_REPLY
Multi-Channel	0x03	0x00	HMC_ADDR_REPORT
		0x01	HMC_ADDR_REPLY
		0x02	HMC_LOCAL_REQUEST
		0x03	HMC_LOCAL_REPLY
		0x04	HMC_REMOTE_REQUEST
		0x05	HMC_REMOTE_REPLY

HyperSCSI Command Block Example

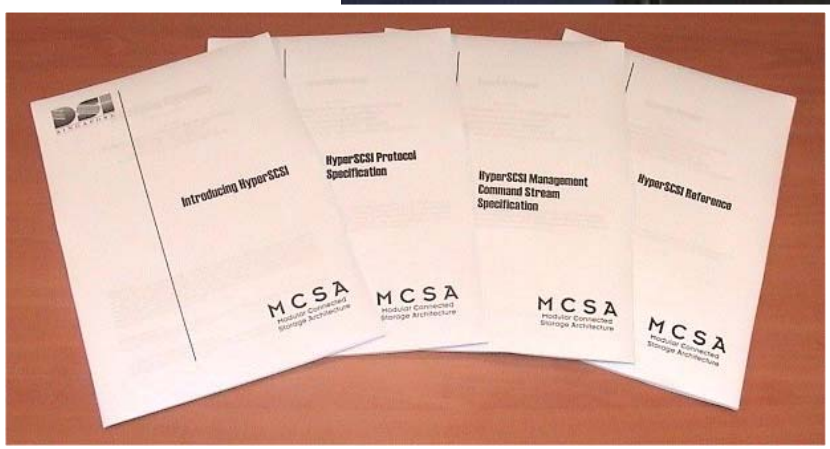


OpCode in this example is either 0x00 or 0x01, **HCBE_REQUEST** or **HCBE_REPLY**

HyperSCSI Protocol Parameters

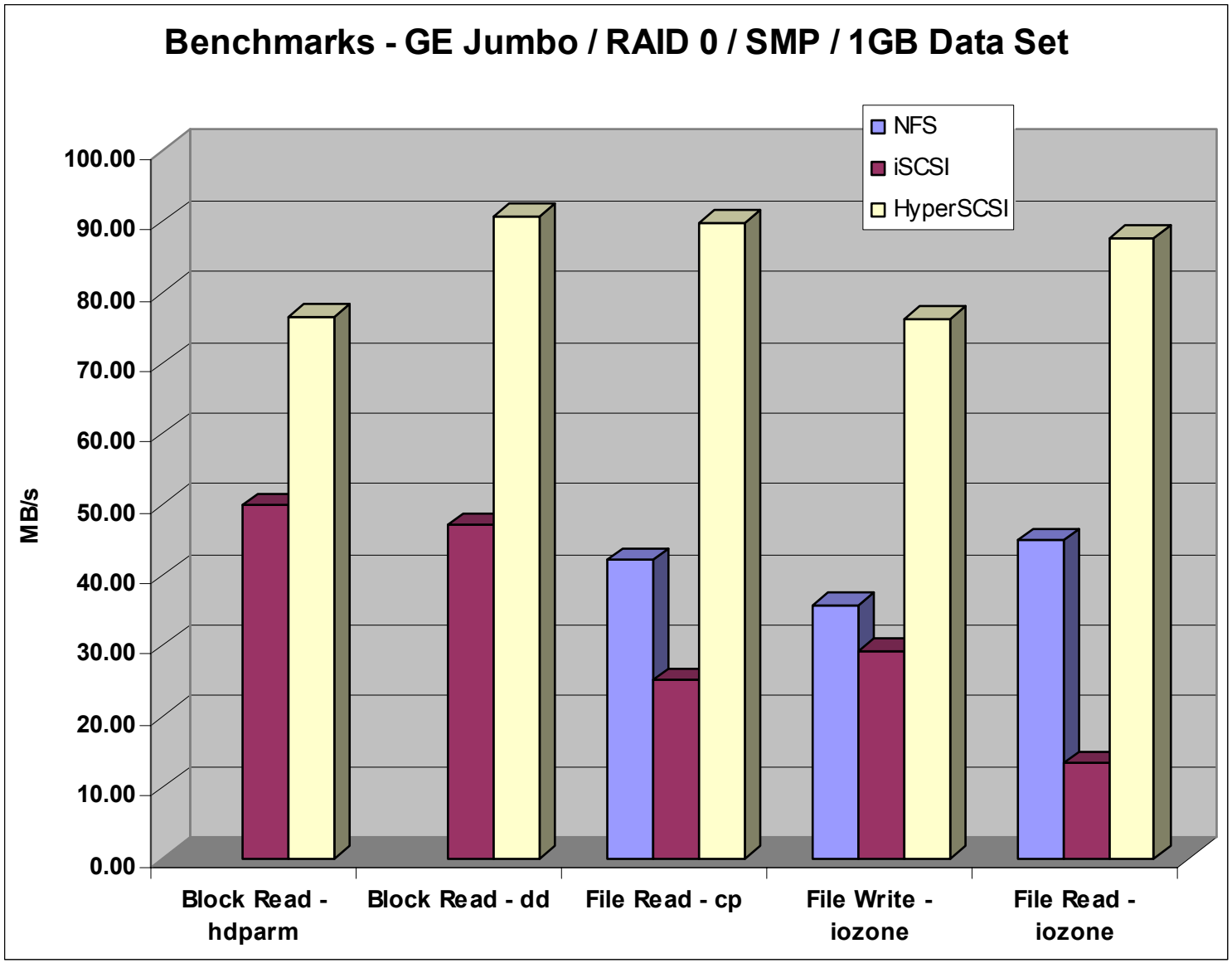
Development Progress

Prototype Demo and Test Area

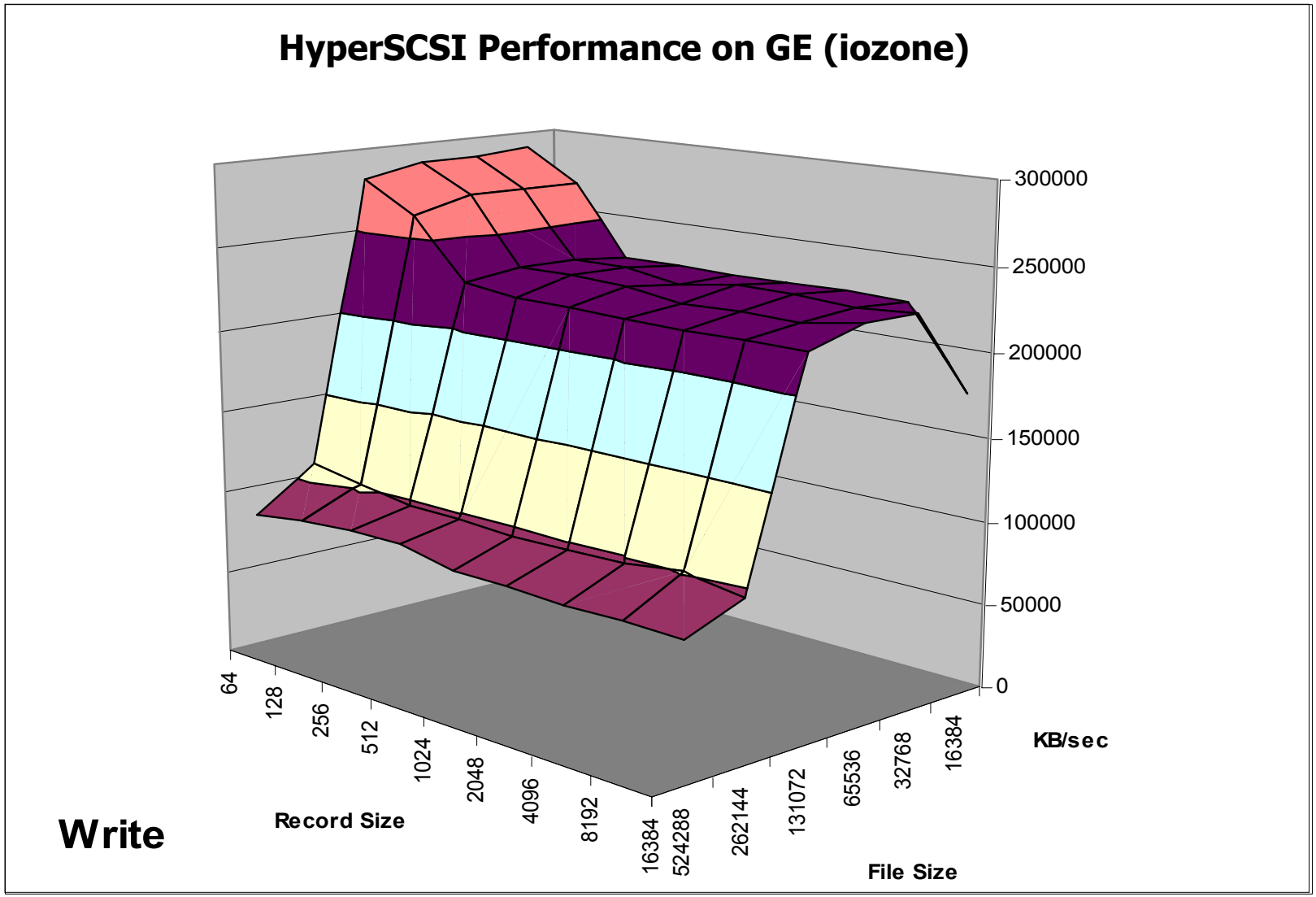


HyperSCSI Documentation

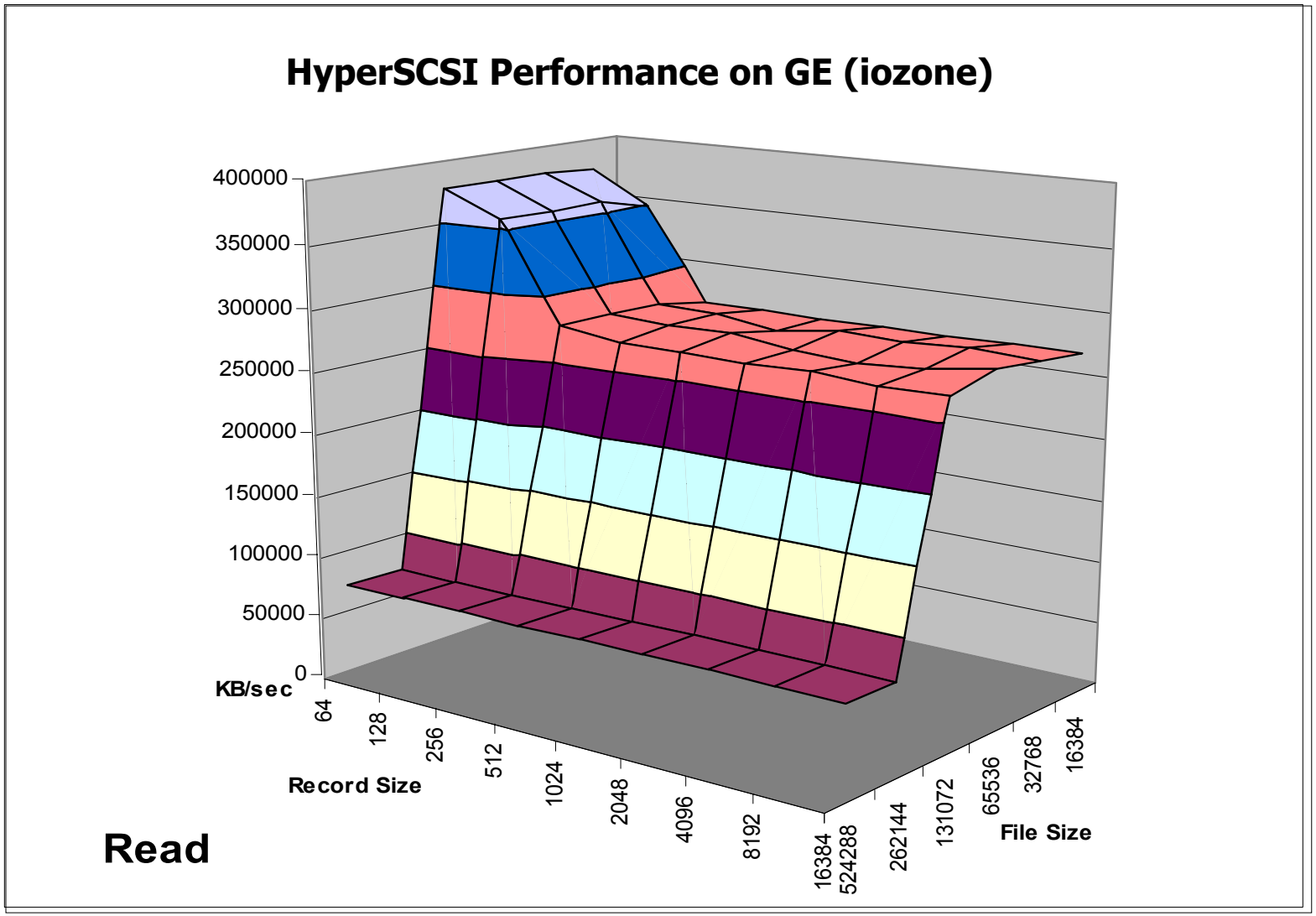
HyperSCSI Block & File Access



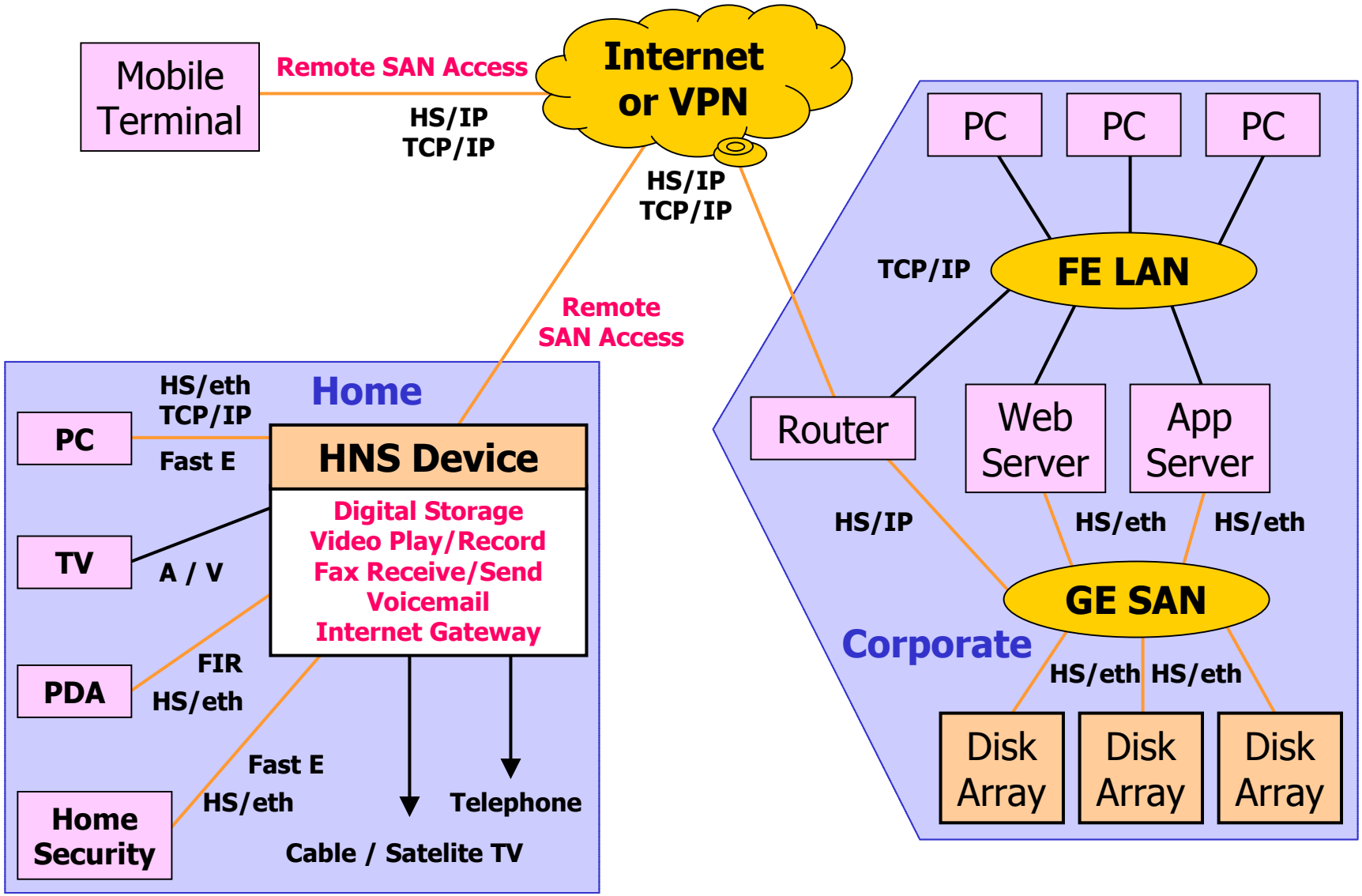
HyperSCSI Performance Profiling



HyperSCSI Performance Profiling



HyperSCSI in Action



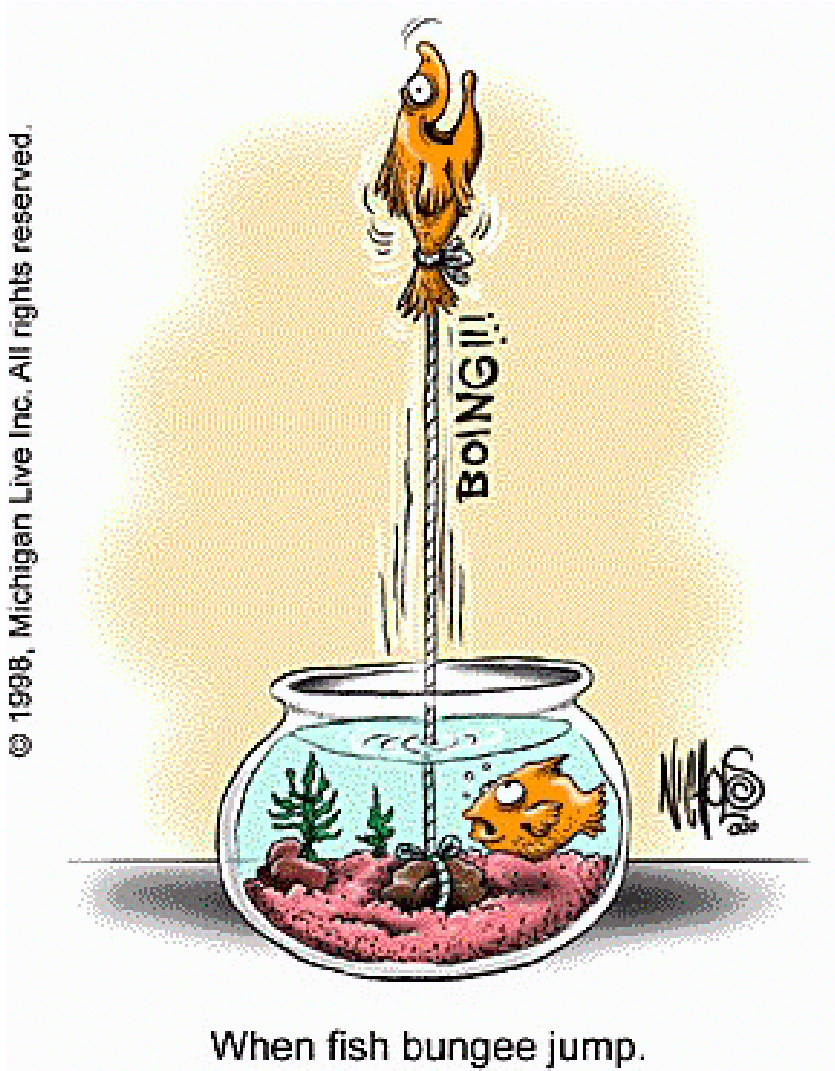
Conclusion

- HyperSCSI manages to provide Ethernet-based storage without rebuilding TCP/IP
- HyperSCSI proves that pure Ethernet (without TCP/IP) can be a viable alternative to building Ethernet-based network storage

- For more information about HyperSCSI and implementing it for your own use, please see our website

Thank You

http://nst.dsi.nus.edu.sg/mcsa/



MCSA
Modular Connected Storage Architecture

DSI SINGAPORE