Commodity Storage Networks

GENROCO

Stephen W. Bailey
GENROCO, Inc.
www.genroco.com

Network File Service

Clients

File Servers

TCP/IP

TCP/IP

TCP/IP

TCP/IP

TCP/IP

Ethernet/
802.3
Fabric

Parallel SCSI
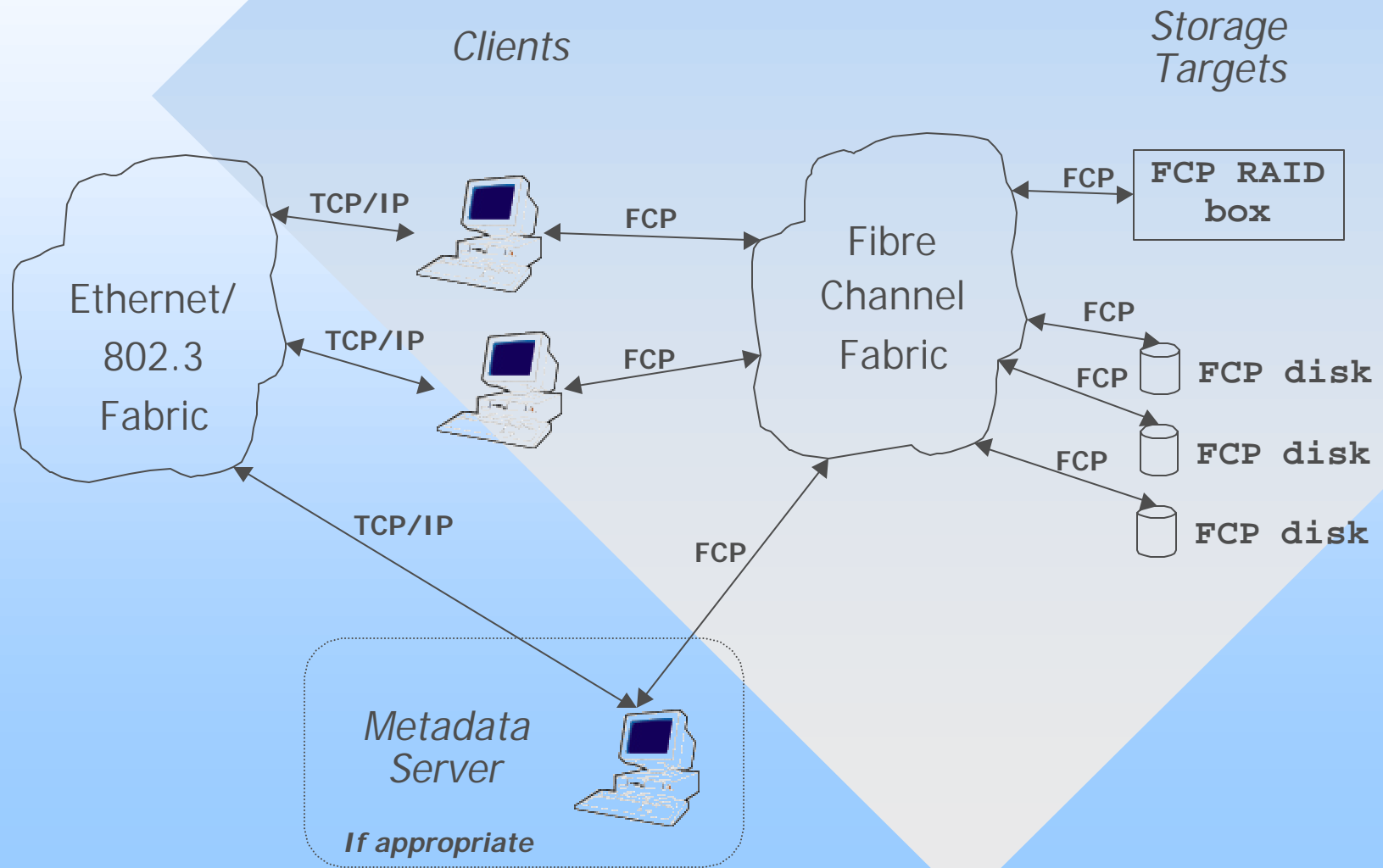
Parallel SCSI
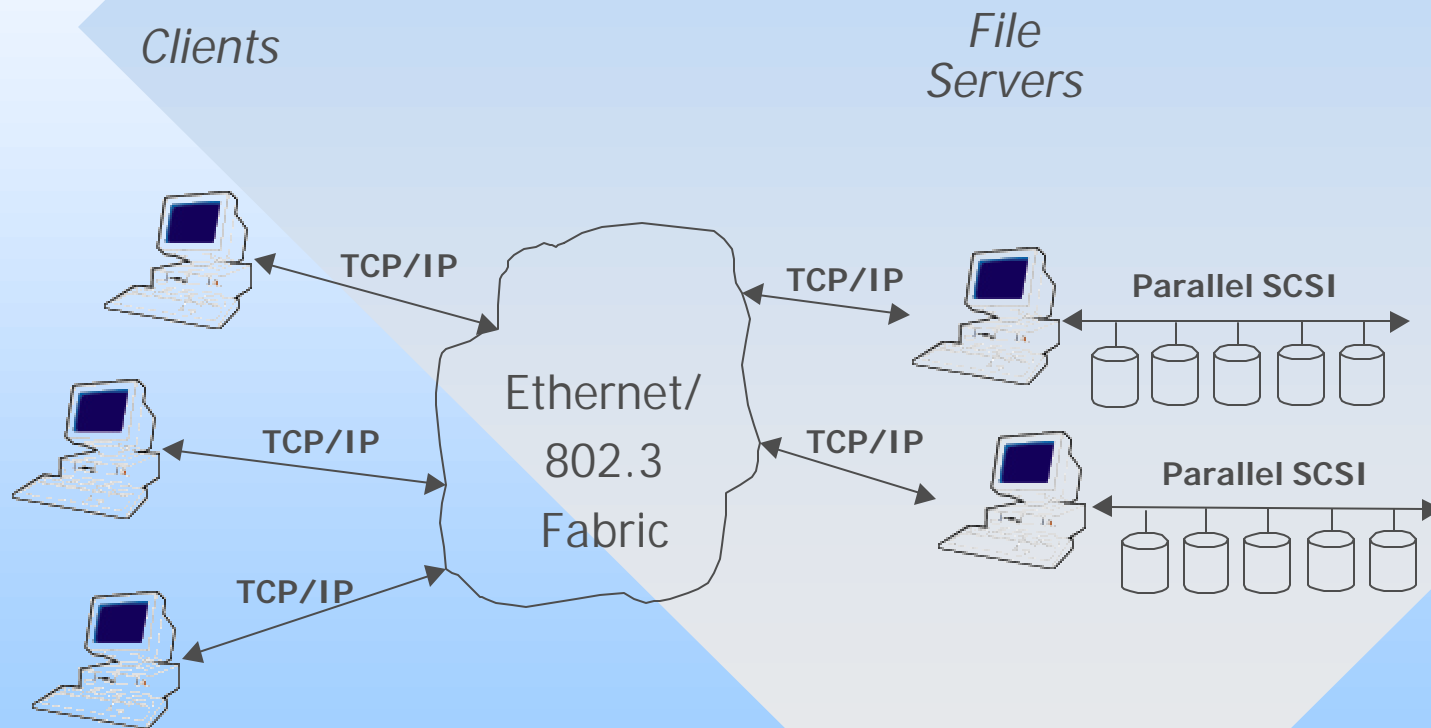
Why Not TCP/IP?

Implemented in software

- Data copies

- High host overhead:
    - 10 Mbit/s @ 1500 byte packets, 670 packets/s
    - 1 Gbit/s @ 1500 byte packets, 67000 packets/s

FC Storage Network

Clients

Storage Targets

Ethernet/ 802.3 Fabric

TCP/IP

FCP

TCP/IP

FCP

Fibre Channel Fabric

FCP → FCP RAID box

FCP → FCP disk

FCP → FCP disk

FCP → FCP disk

TCP/IP

FCP

Metadata Server

If appropriate

What's Wrong?

Two infrastructures means:

- At least twice the cost

- Less than twice the performance

Could run TCP/IP on FC but:

- FC TCP/IP is not widely supported (yet)

- Ethernet is inexpensive, ubiquitous, offers rapid compatible growth

FCP For Ethernet?

No, need to cast the net wider:

- TCP/IP has momentum, solves most current problems

- additional hardware must offer compelling advantages

Instead:

- build storage on protocol with compelling advantages

ST To The Rescue

Scheduled Transfer Protocol (ST) has wide design goals:

- protocol upon which upper layer protocols are built

- hardware data steering

- low latency

- asynchronous I/O support

- existing TCP/IP based protocol acceleration

- many media (Ethernet, FC, HiPPI, GSN, IP)

ST versus VI

- ST specifies wire protocol

- VI specifies software interface

- ST and VI hardware assists are very close

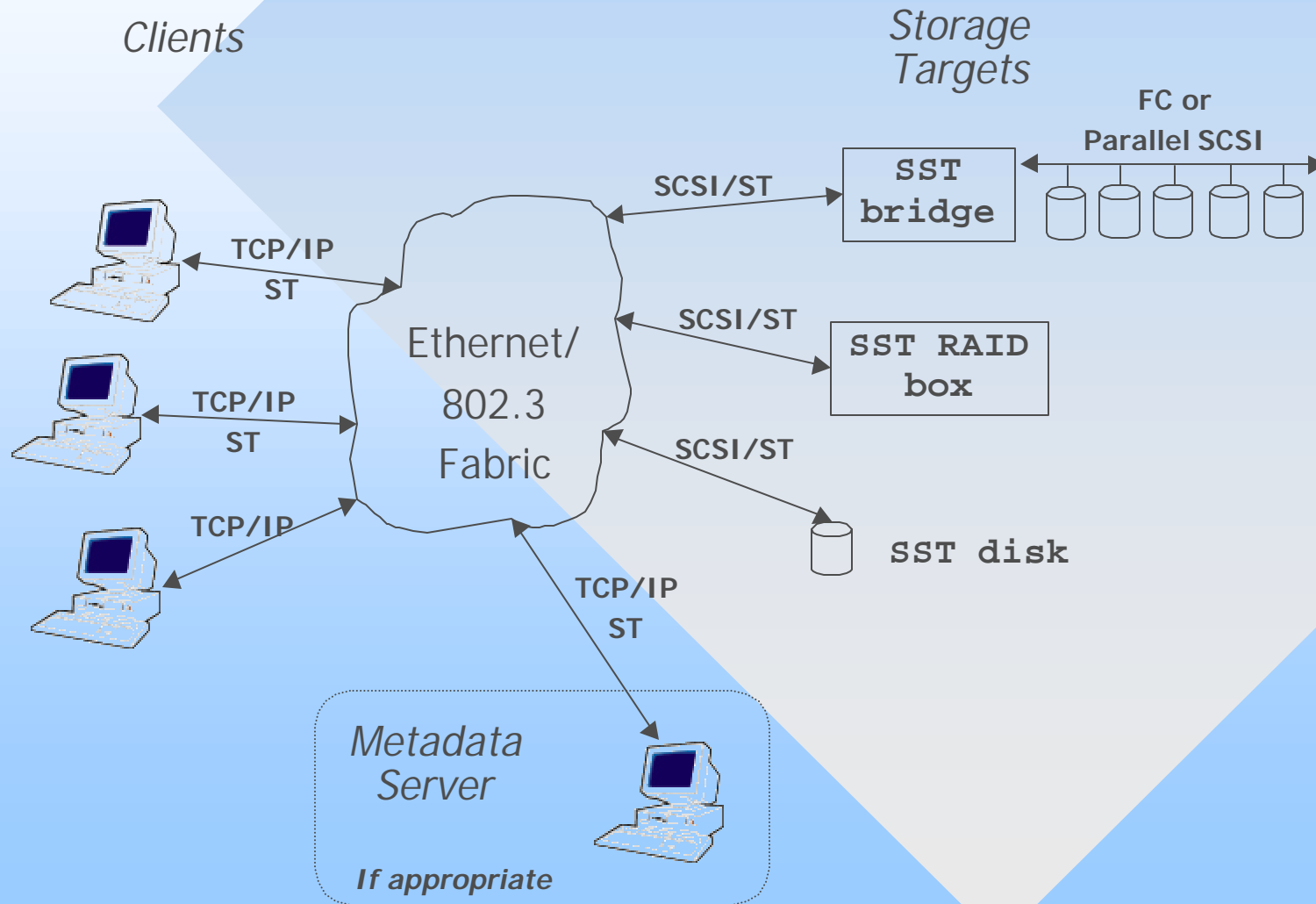- Software clients of ST and VI layers are very close

The Protocol Picture

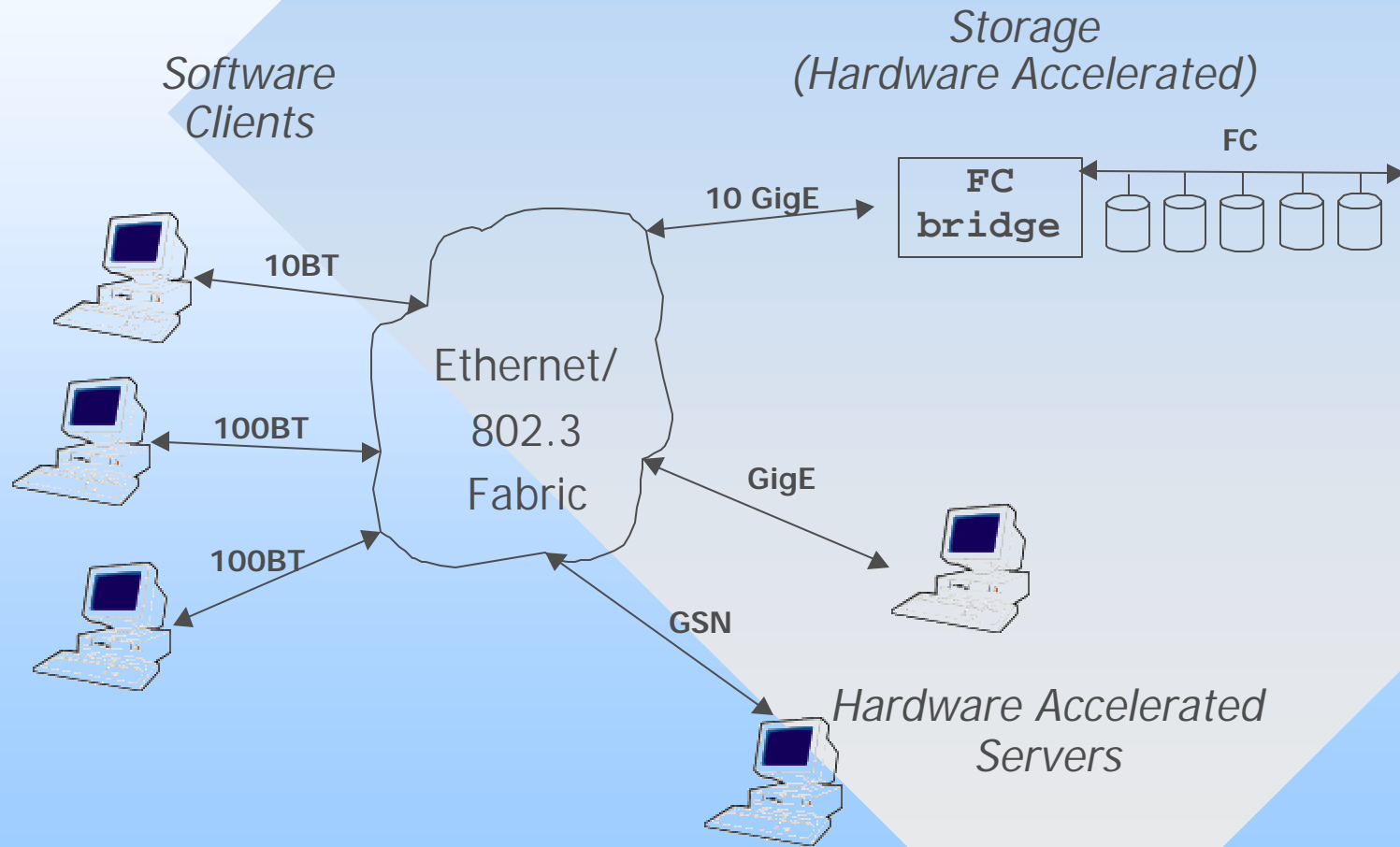| SCSI | FTP | NFS | . . . | *Software* |
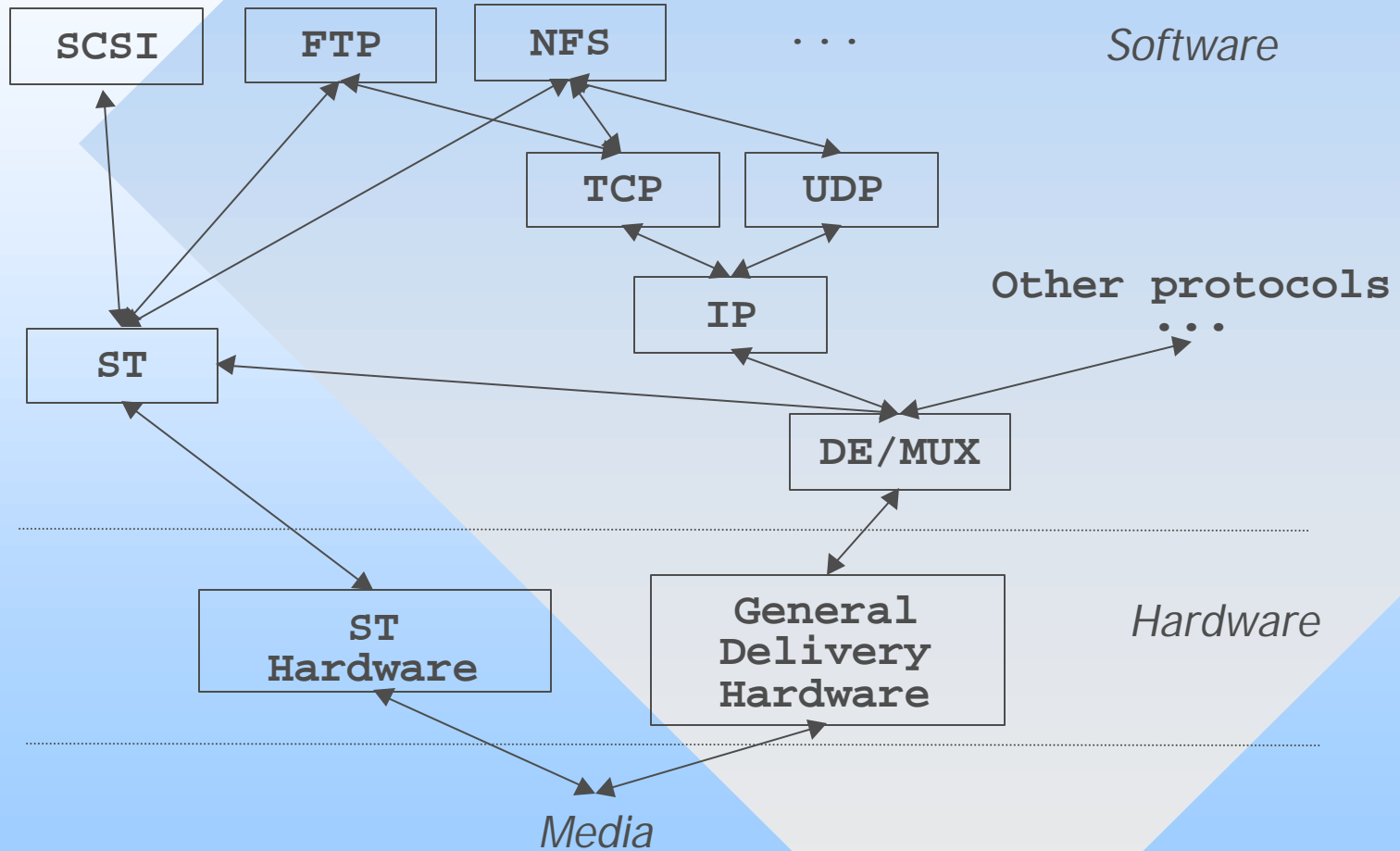
SCSI On ST (SST)

SST specification maps:

- Commands

- Data transfers

- Status

using stock ST hardware acceleration

Single Fabric Storage Network

Clients

Storage Targets

FC or
Parallel SCSI

SCSI/ST → SST bridge

TCP/IP ST

TCP/IP ST

TCP/IP

Ethernet/ 802.3 Fabric

SCSI/ST → SST RAID box

SCSI/ST → SST disk

TCP/IP ST

Metadata Server

If appropriate

The Role Of Software

Software Clients

Storage (Hardware Accelerated)

FC

| FC bridge |

10 GigE

10BT

100BT

Ethernet/ 802.3 Fabric

GigE

100BT

GSN

Hardware Accelerated Servers

What's Missing?

- Congestion avoidance

- Security

(oh, is that all?)

Why Internet Storage?

(Blech, all those 56 kilobit lines)

Because of infrastructure scale economies!

Tends towards:

- Single network attachment per host

- Independently administered infrastructure

What's At Stake?

Implement:

- Data steering

- Congestion control

- Security

in hardware and own storage & general networking

Storage is the lever

Congestion Is Tough

Core hardware issues:

- Complex transmit pacing

- Complex receive state

System-level issues:

- Vast buffering requirements

- Complex data source & sink operation scheduling
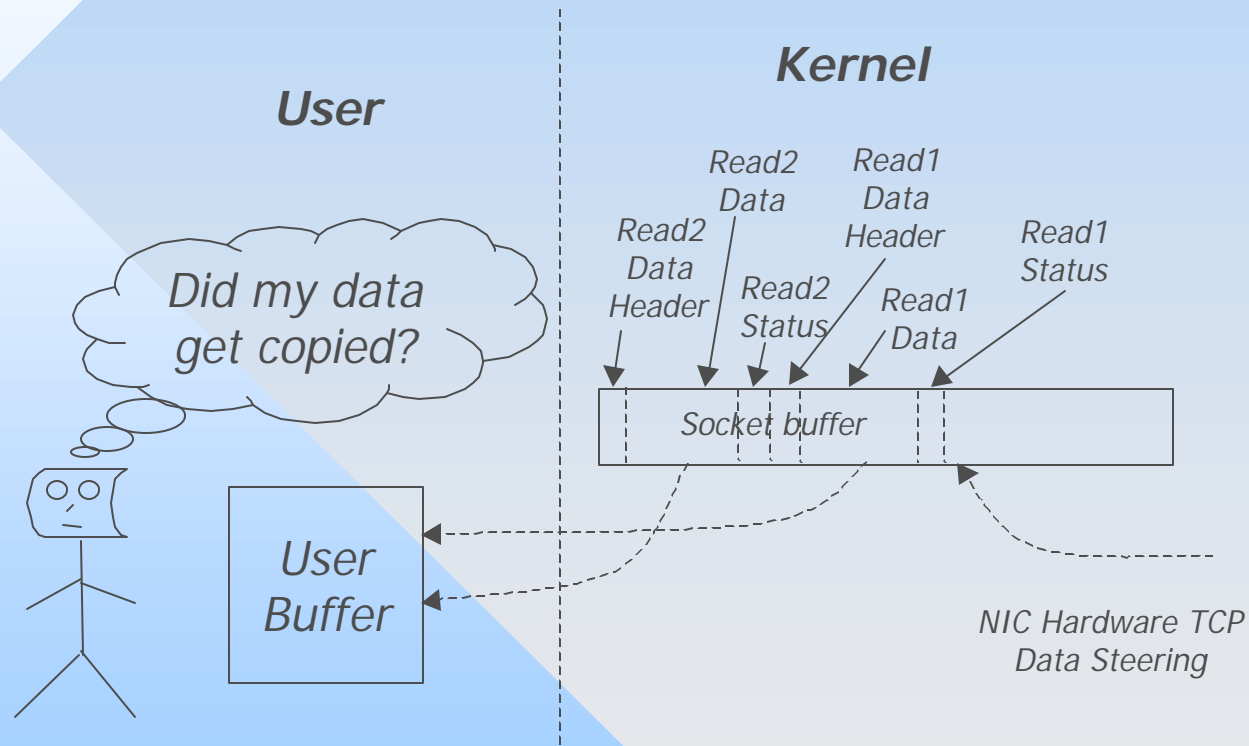
Who's On Deck?

Factions of iSCSI:

- SEP (Adaptec SCSI on TCP)

- Current draft (IBM SCSI on TCP)

- SCSI on SCTP (Sanlight)

- FC on IP (Lucent)

- SCSI on VIA on TCP (Cisco/Giganet)

Is Hardware TCP Enough?

No.

User

Did my data get copied?

User Buffer

Kernel

Read2 Data Header

Read2 Data

Read2 Status

Read1 Data Header

Read1 Data

Read1 Status

Socket buffer

NIC Hardware TCP Data Steering

But packet aggregation & transmit pacing still help

FCP For TCP?

Could build storage protocol:

- directly on TCP (FCP for TCP)

- on intermediate hardware accelerated muxing/demuxing protocol on TCP (ST or VI on TCP)

SCSI On ST (or VI) On TCP

Advantages:

- Hardware independent of storage protocol

- Hardware supports multiple protocols

- Storage protocol is transport independent

When Can We Have It?

Don't put your non-iSCSI gear on Ebay yet:

- TCP hardware acceleration under development

- SCSI on TCP protocol under development

- Hardware mux/demux protocol (ST or VI) on TCP under development

Flow-controlled SANs are the only storage networks today

(GSN, FC, flow controlled GigE InfiniBand)

What Can I Do?

Stuff the ballot box in the iSCSI IETF WG:

www.ece.cmu.edu/~ips

Contact your favorite vendor:

`www.genroco.com`

`sales@genroco.com`

Carl Pick `carl@genroco.com`
Don Woelz `don@genroco.com`
Alberto Guglielmi `alberto@genroco.com`