Implementation and Evaluation of iSCSI over RDMA

Ethan Burns and Robert Russell

{eaburns,rdr}@iol.unh.edu

University of New Hampshire InterOperability Laboratory
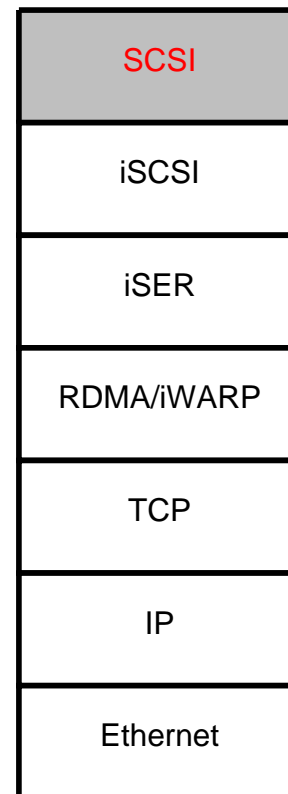121 Technology Drive, Suite 2
Durham, NH 03824-4716

Goal

Create an iSCSI implementation that makes use of Remote Direct Memory Access (iWARP) with the iSER extensions. Evaluate the performance of the implementation.

SCSI

Create an **iSCSI** implementation that makes use of Remote Direct Memory Access (iWARP) with the iSER extensions. Evaluate the performance of the implementation.

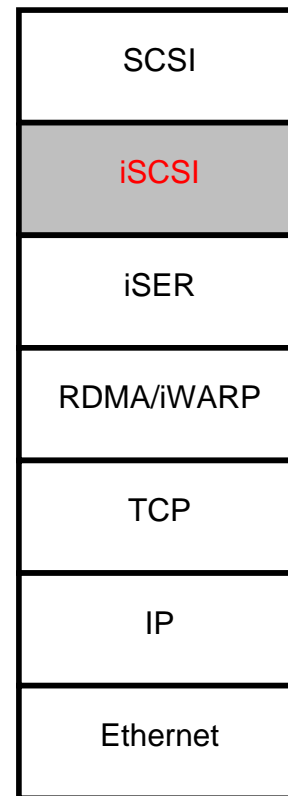| SCSI |
| :---: |
| iSCSI |
| iSER |
| RDMA/iWARP |
| TCP |
| IP |
| Ethernet |

SCSI

Create an **iSCSI** implementation that makes use of Remote Direct Memory Access (iWARP) with the iSER extensions. Evaluate the performance of the implementation.

☐ Small Computer System Interface
☐ Architecture for connecting peripheral devices to computers
☐ Client/Server:

  – **Initiator** (Client)
  – **Target** (Server)

☐ Traditionally an internal parallel SCSI bus
☐ Limitations on number of devices and cable length

iSCSI

Create an **iSCSI** implementation that makes use of Remote Direct Memory Access (iWARP) with the iSER extensions. Evaluate the performance of the implementation.

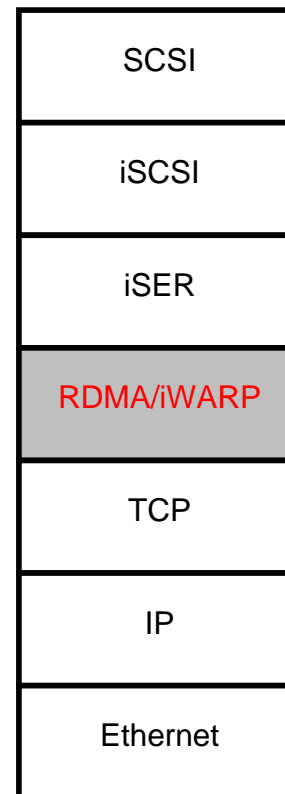| SCSI |
| --- |
| iSCSI |
| iSER |
| RDMA/iWARP |
| TCP |
| IP |
| Ethernet |

iSCSI

Create an **iSCSI** implementation that makes use of Remote Direct Memory Access (iWARP) with the iSER extensions. Evaluate the performance of the implementation.

☐ Internet Small Computer System Interface

– RFC3720 (2004)

☐ A solution to the scalability issues of traditional SCSI
☐ A transport for SCSI commands and data over TCP/IP
☐ Two phases

– Login Phase – for negotiating connection parameters
– Full Feature Phase – for data transfer

RDMA

Create an iSCSI implementation that makes use of **Remote Direct Memory Access (iWARP)** with the iSER extensions. Evaluate the performance of the implementation.

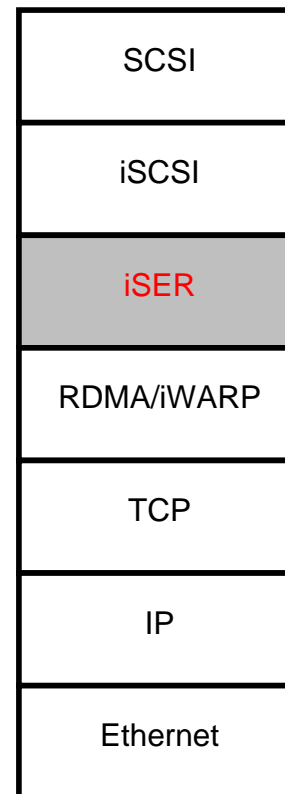| SCSI |
| --- |
| iSCSI |
| iSER |
| RDMA/iWARP |
| TCP |
| IP |
| Ethernet |

RDMA

Create an iSCSI implementation that makes use of **Remote Direct Memory Access (iWARP)** with the iSER extensions. Evaluate the performance of the implementation.

☐ Remote Direct Memory Access

☐ Typical CPU becomes bottleneck with 10GigE

- Data copying
- Network interrupts
- Packet processing

☐ Zero-copy data transfers

☐ Offloads network processing

☐ Makes full utilization of a 10GigE link

☐ iWARP protocol suite provides RDMA over TCP/IP

- RFC5040 (2007), RFC5041 (2007), RFC5044 (2007), ...

iSER

Create an iSCSI implementation that makes use of Remote Direct Memory Access (iWARP) with the **iSER extensions**. Evaluate the performance of the implementation.

SCSI
iSCSI
iSER
RDMA/iWARP
TCP
IP
Ethernet

iSER

Create an iSCSI implementation that makes use of Remote Direct Memory Access (iWARP) with the **iSER extensions**. Evaluate the performance of the implementation.

☐ iSCSI Extensions for RDMA

– RFC5046 (2007)

☐ Allow iSCSI to use RDMA hardware
☐ Translate and encapsulate iSCSI over RDMA
☐ Transition from streaming TCP to RDMA enabled

– Negotiate use of iSER during iSCSI negotiation phase
– Transition to RDMA mode before iSCSI data transfer phase

Implementation

Protocols

Protocols
SCSI
iSCSI
iSER
RDMA/iWARP
TCP
IP
Ethernet
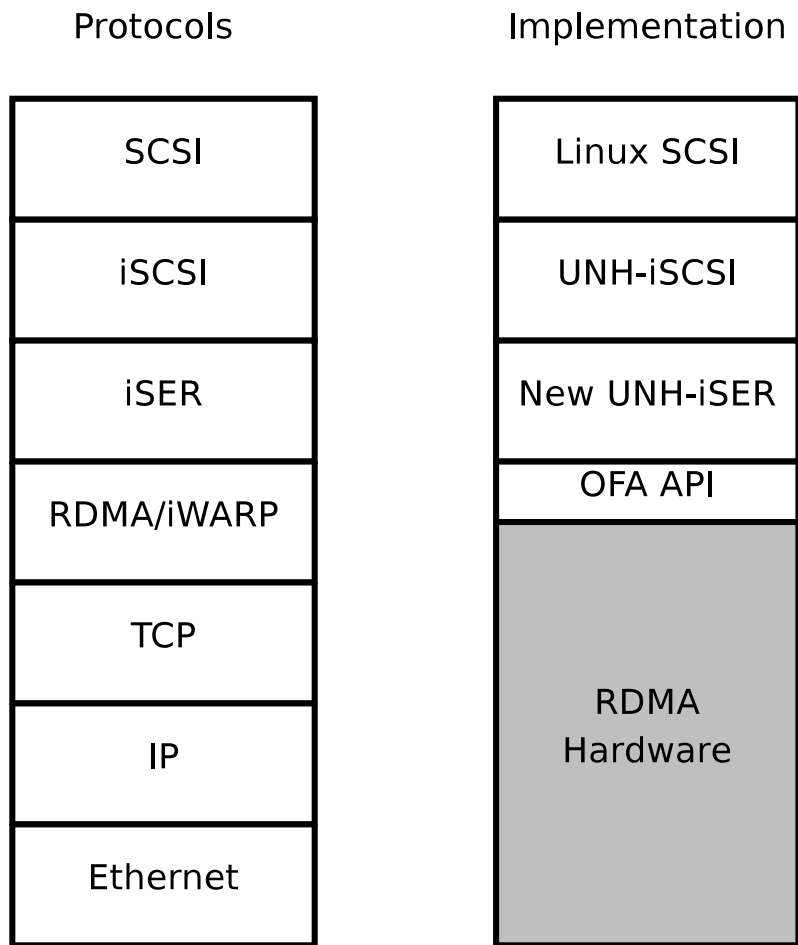
Implementation

Implementation
Linux SCSI
UNH-iSCSI
New UNH-iSER
OFA API
RDMA Hardware

Implementation

☐ Extend UNH-iSCSI to support the iSER extensions

    –   Set of Linux kernel modules

    –   Created and supported at UNH

☐ Use the OpenFabrics Alliance Stack

    –   Access to RDMA hardware

    –   Included in Linux kernel

    –   Provides a user-space API

☐ Create both a kernel-space and user-space solution

Issues Uncovered

Current RDMA hardware does not support TCP stream
transitioning

☐ Bring up connection in RDMA mode

☐ No run-time selection for iSER v.s. traditional iSCSI

☐ Additional iSER operational primitives for connection
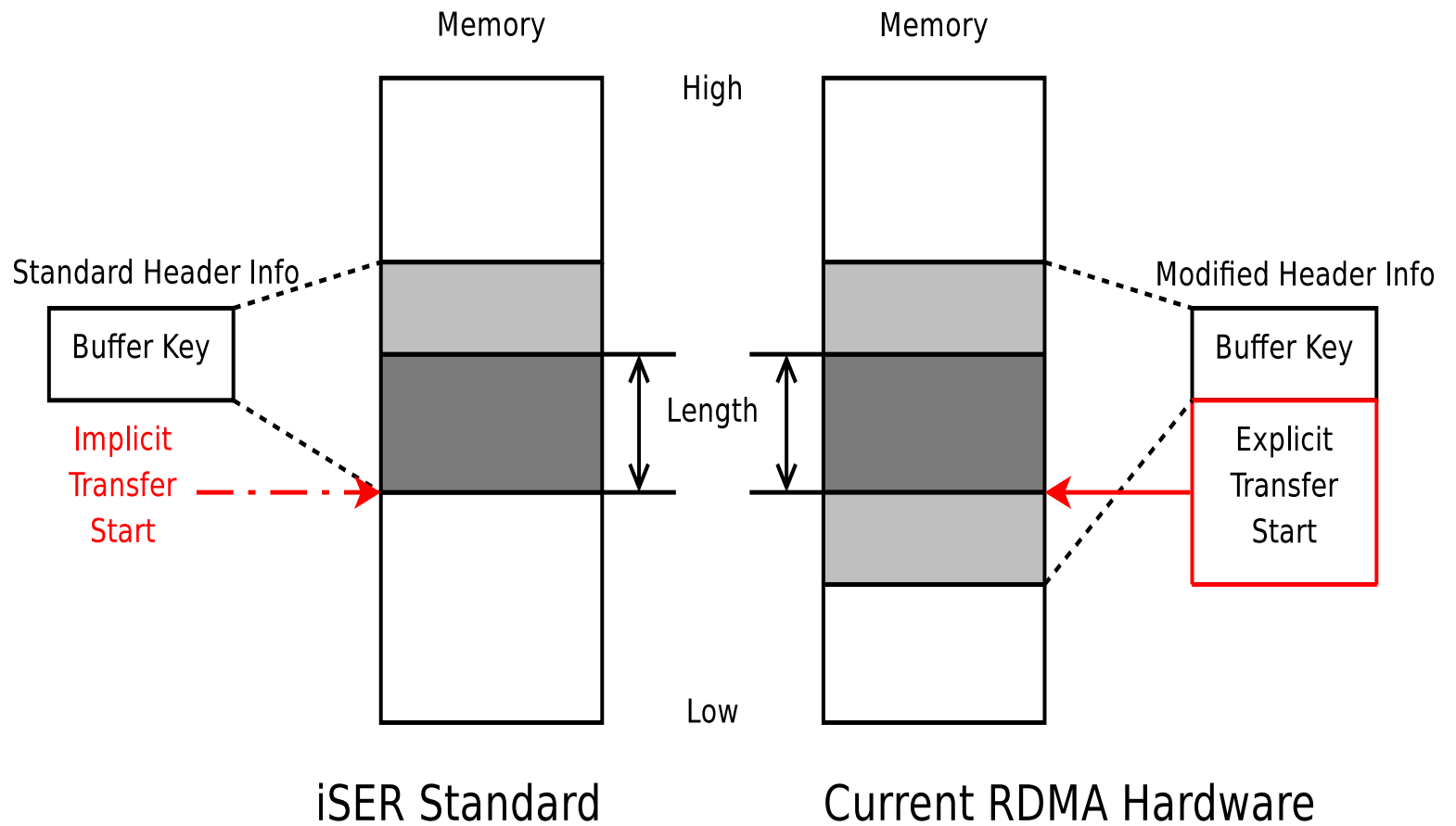establishment

Issues Uncovered

Standard iSER header for iWARP does not contain fields for all data required by current hardware.

☐ We added additional iSER header fields to advertise missing information

Issues Uncovered

Memory

Memory

High

Standard Header Info

Modified Header Info

Buffer Key

Buffer Key

Length

Implicit
Transfer
Start

Explicit
Transfer
Start

Low

iSER Standard

Current RDMA Hardware

Evaluation

☐ MEMORYIO mode (on the target)

☐ Four 2.6GHz Intel 64-bit cores

☐ 4GB main memory
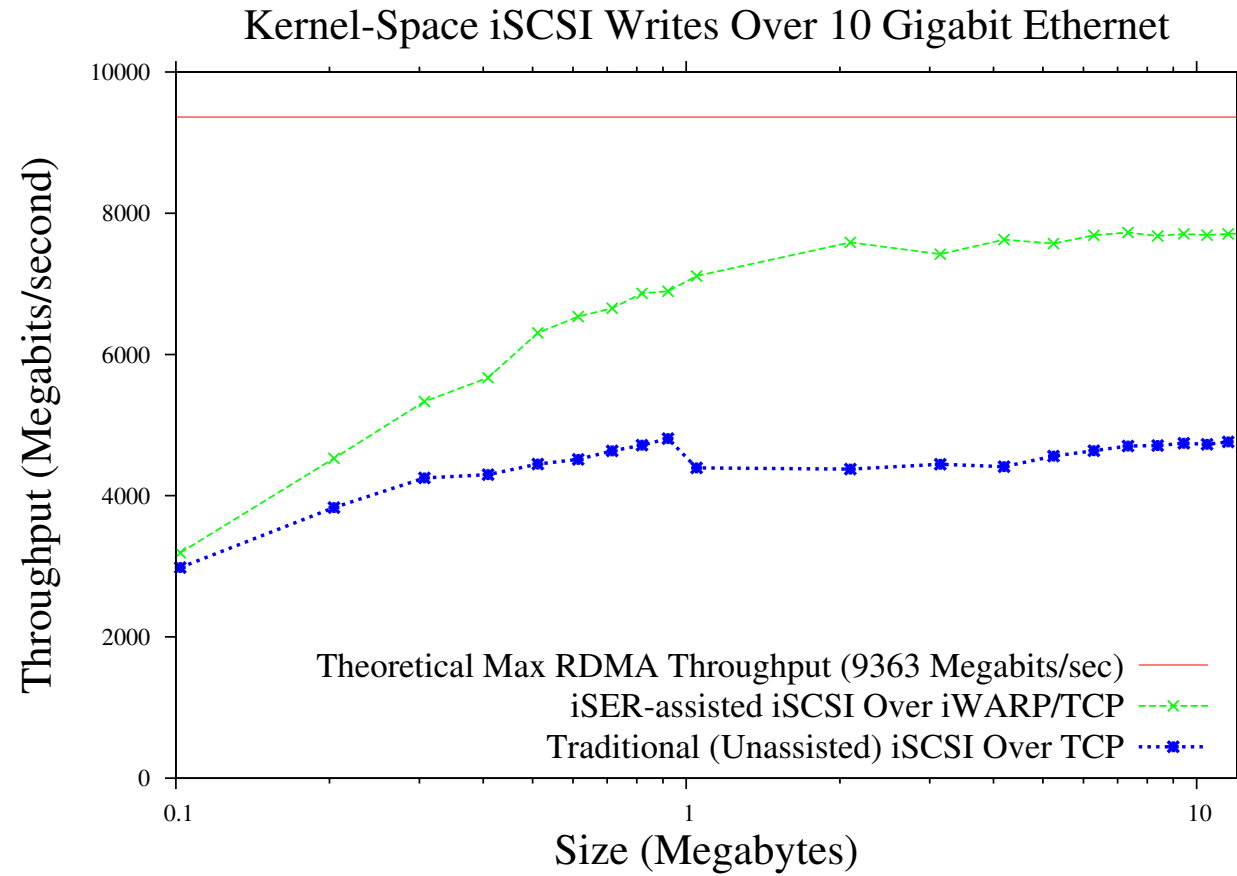
☐ Chelsio R310E-CXA 10Gigabit Ethernet iWARP adapters

Evaluation

Kernel-Space iSCSI Reads Over 10 Gigabit Ethernet

Evaluation

Kernel-Space iSCSI Writes Over 10 Gigabit Ethernet

Evaluation

User-Space iSCSI Writes Over 10 Gigabit Ethernet
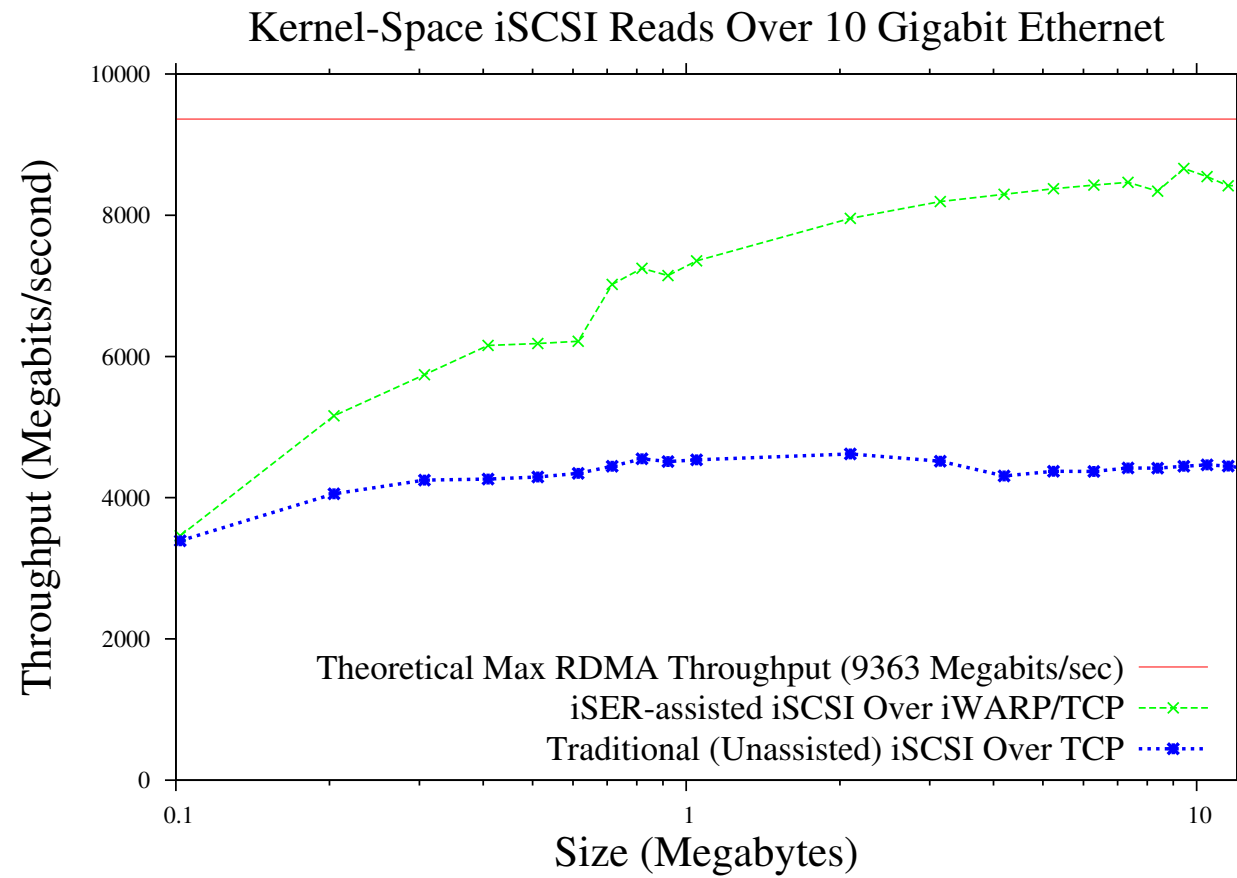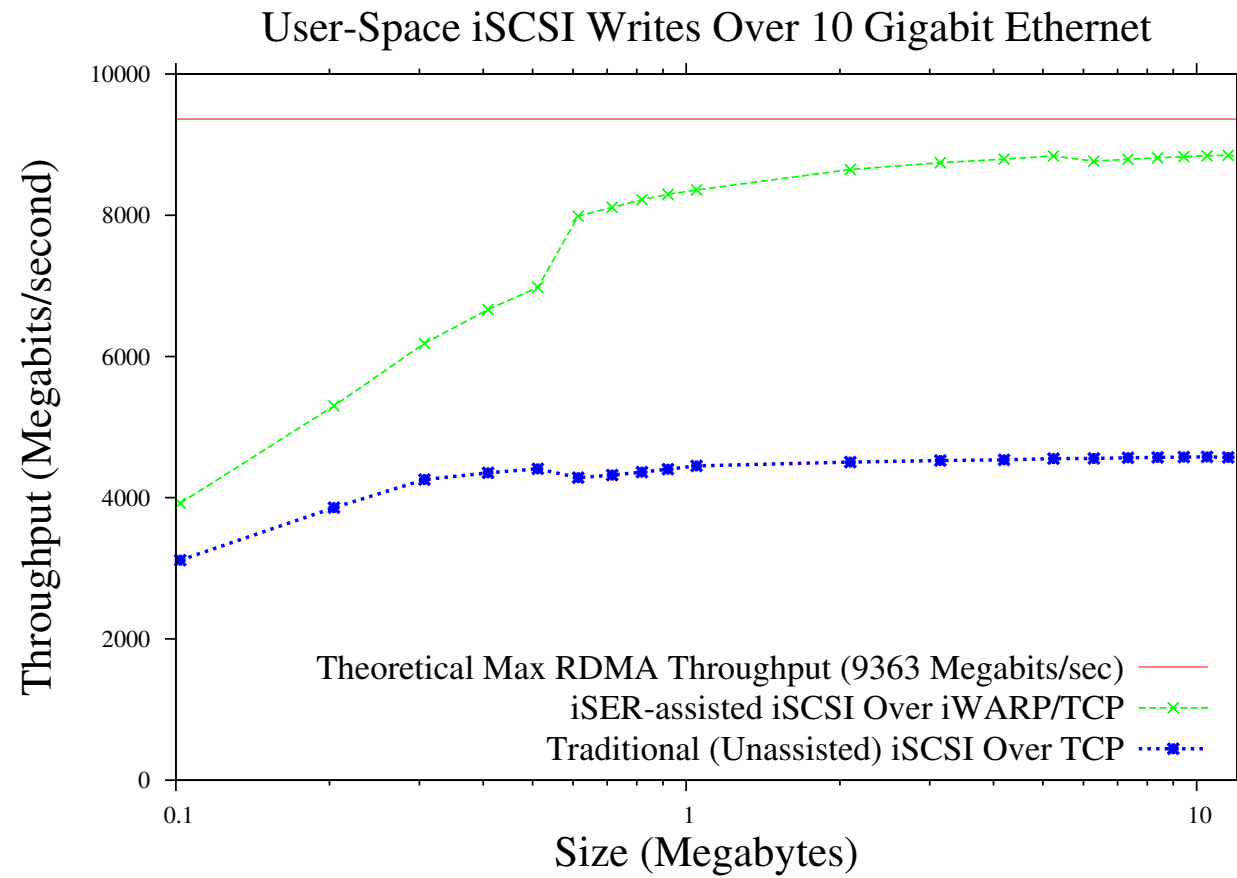
Future Work

☐ Further Performance Evaluation

– Response time
– CPU utilization

☐ Further Comparisons

– Infiniband
– TCP offloading
– iSCSI offloading

☐ iSCSI Parameters

– Immediate/Unsolicited data
– Multiple outstanding commands
– Multiple connections

Questions

Source Available at:
http://sourceforge.net/projects/unh-iscsi