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

Goal
SCSI
iSCSI
RDMA
iSER

Issues Uncovered
Evaluation
Future Work
Questions

Protocols

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

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**
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
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

- Goal
- SCSI
- iSCSI
- RDMA
- iSER
- Implementation
- Evaluation
- Future Work
- Questions

![Diagram showing differences between iSER Standard and Current RDMA Hardware](image)

- Standard Header Info
- Buffer Key
- Implicit Transfer Start
- Modified Header Info
- Explicit Transfer Start

Memory

High

Length

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

- Goal
- SCSI
- iSCSI
- RDMA
- iSER
- Implementation
- Issues Uncovered
- Evaluation
- Future Work
- Questions

Kernel-Space iSCSI Reads Over 10 Gigabit Ethernet

Throughput (Megabits/second)

Size (Megabytes)

Theoretical Max RDMA Throughput (9363 Megabits/sec)

iSER-assisted iSCSI Over iWARP/TCP

Traditional (Unassisted) iSCSI Over TCP
Evaluation

Goal
SCSI
iSCSI
RDMA
iSER
Implementation
Issues Uncovered
▶ Evaluation
Future Work
Questions

Kernel-Space iSCSI Writes Over 10 Gigabit Ethernet

Throughput (Megabits/second) vs Size (Megabytes)

- Theoretical Max RDMA Throughput (9363 Megabits/sec)
- iSER-assisted iSCSI Over iWARP/TCP
- Traditional (Unassisted) iSCSI Over TCP
Evaluation

Throughput (Megabits/second) vs. Size (Megabytes)

- Theoretical Max RDMA Throughput (9363 Megabits/sec)
- iSER-assisted iSCSI Over iWARP/TCP
- Traditional (Unassisted) iSCSI Over TCP

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