

iSCSI Heritage and Future

iSCSI Built-in RDMA Compared to iSER over iWARP and IB

Executive Summary

iSCSI, the Internet SCSI standard, is the leading Ethernet SAN protocol, with native initiator support integrated into all the major operating systems and built-in RDMA for high performance offloaded implementations. This paper dispels the myth that high performance iSCSI requires using a new protocol (e.g, iSER) that runs over RDMA, and breaking compatibility with the existing iSCSI ecosystem.

By comparing the performance of Chelsio's hardware offloaded iSCSI with iSER over Ethernet and FDR InfiniBand, this paper shows that there is no need to forgo iSCSI's heritage to achieve high performance. The results demonstrate that iSCSI at 40Gbps provides virtually the same performance in throughput and CPU utilization as iSER, thanks to built-in RDMA that is optimized for SCSI storage. iSCSI preserves existing equipment without requiring a fabric overhaul, additional acquisition and management costs, with a high performance option thanks to hardware offload, and a strong roadmap to 100Gbps and beyond. iSER, on the other hand, adds unnecessary layers, and is merely a way for RDMA fabrics to offer iSCSI emulation that is incompatible with the installed software base.

Overview

The Terminator 5 (T5) ASIC from Chelsio Communications, Inc. is a fifth generation, high-performance 2x40Gbps/4x10Gbps, unified wire engine which offers PDU iSCSI offload capability in protocol acceleration for both file and block-level storage (iSCSI) traffic. Furthermore, iSCSI support is part of a complete, fully virtualized unified wire offload suite that includes FCoE, RDMA over Ethernet, TCP and UDP sockets and user space I/O.

By leveraging Chelsio's proven TCP Offload Engine (TOE), offloaded iSCSI over T5 enjoys a distinct performance advantage over regular NIC, as well as superior data integrity protection. In fact, T5 includes enhanced data integrity protection for all protocols, and particularly so for storage traffic, including full end-to-end T10-DIX support for both iSCSI and FCoE, as well as internal datapath CRC, and ECC-protected memory. Unlike iSER, iSCSI comes with superior initiator compatibility making it an ideal choice for high performance storage and data center scale storage networking solutions.

Test Results

The following graphs compare the single port iSCSI READ and WRITE throughput and CPU usage numbers with iSER over iWARP RDMA (Chelsio), IB FDR-10 and IB-FDR (both Mellanox). The numbers were obtained by varying the I/O sizes using the **iozone** tool.

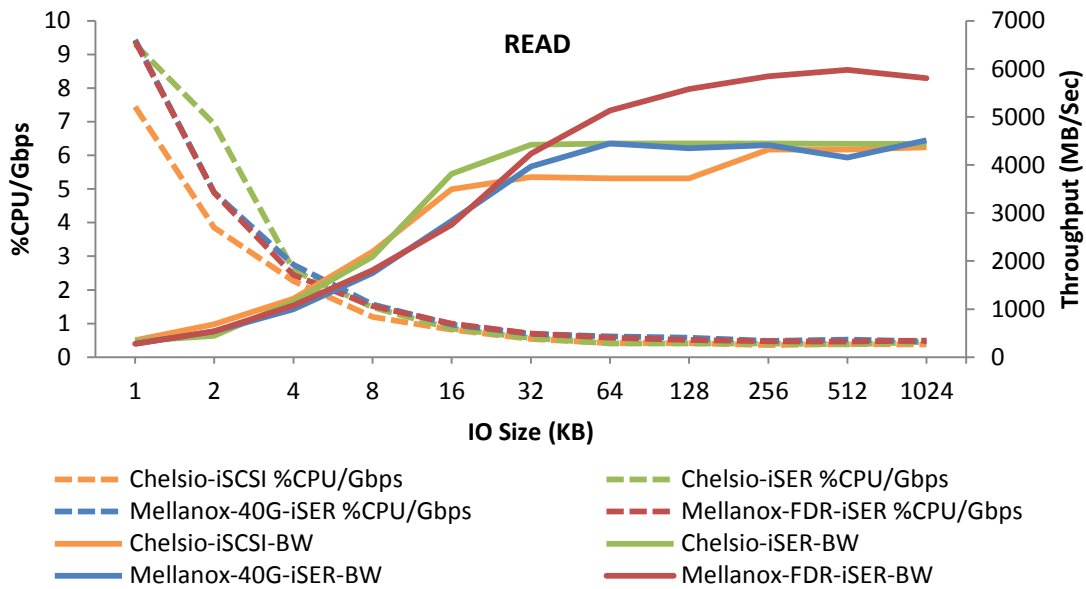


Figure 1 – READ Throughput and %CPU vs. IO size

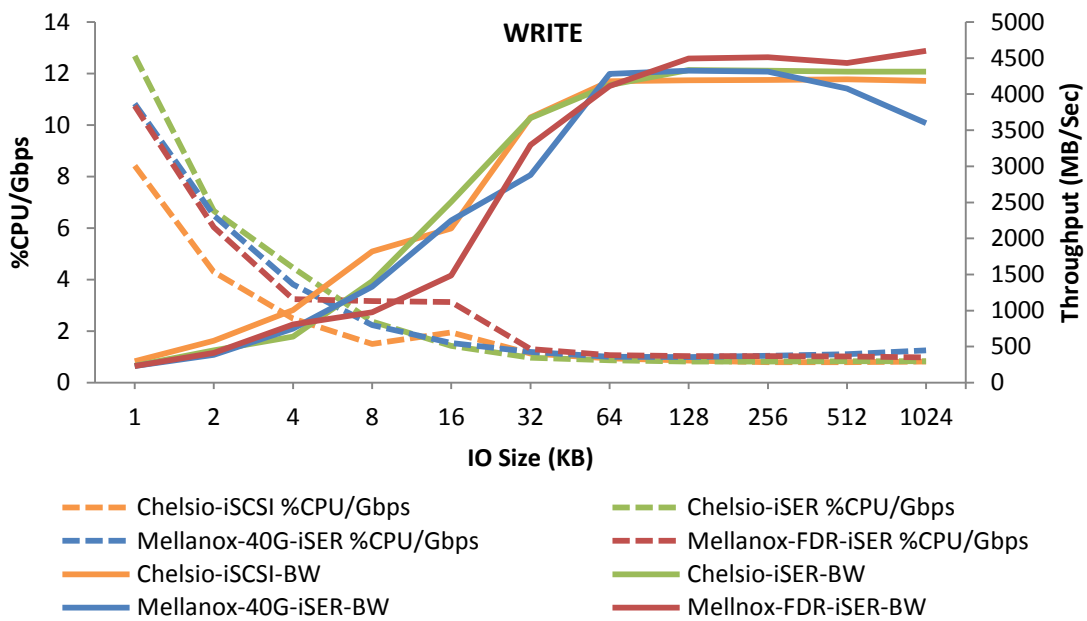


Figure 2 – WRITE Throughput and %CPU vs. IO size

While both Chelsio iSCSI and iSER provide performance that is superior or comparable to InfiniBand across most of the range of interest, the results show that iSCSI at 40Gbps provides the highest throughput in the particularly interesting small IO range, up to 16KB. Furthermore, iSCSI's CPU usage per Gbps is considerably lower than all iSER alternatives, confirming the higher efficiency operation of the layer optimized design. It is noteworthy that IB FDR only achieves large IO throughput that exceeds 40Gbps in the READ direction, and its FDR-10 40Gbps solution exhibits poor performance at large IO sizes.

Test Configuration

The following sections provide the test setup and configuration details.

Topology

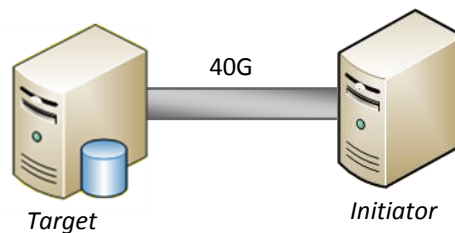


Figure 3- Simple Back-to-Back Test Topology

Network Configuration

The network configuration consists of a target storage array connected back-to-back with an initiator machine using a single port. Standard MTU of 1500B is used.

- **The storage array** is configured with 2 Intel Xeon CPU E5-2687W v2 8-core processors running at 3.40GHz (HT enabled) with 64 GB of RAM.

In the Chelsio iSCSI and iSER over iWARP RDMA setup, a T580-CR adapter is installed with TGT target driver and RHEL 6.5 (3.16.0) operating system.

In the InfiniBand setup, a Mellanox MCX353A-FCBT Connect-X adapter is installed with Mellanox OFED driver v2.2-1.0.1 and RHEL 6.5 operating system.

- **The initiator machine** is configured with an Intel Xeon CPU E5-1660 v2 6-core processor running at 3.70GHz (HT enabled) with 64 GB of RAM.

In the Chelsio iSCSI and iSER over iWARP RDMA setup, a T580-CR adapter is installed with open iSCSI Initiator and RHEL 6.5 (3.16.0) operating system.

In the InfiniBand setup, a Mellanox MCX353A-FCBT Connect-X adapter is installed with open iSCSI Initiator and RHEL6.5 operating system.

Storage Topology and Configuration

The storage array is configured with 16 targets, each configured with 1 ramdisk LUN.

I/O Benchmarking Configuration

iozone was used to assess the I/O capacity of the configuration. The I/O sizes used varied from 1KB to 1024KB with an I/O access pattern of random READs and WRITES.

Commands Used

```
[root@host]# iozone -i 0 -M -c -e -w -I -+u -s 2048000 -t 16 -r <IO-size> -+m /tmp/IOzone.config
```

```
[root@host]# iozone -i 1 -M -c -e -w -I -+u -s 2048000 -t 16 -r <IO-size> -+m /tmp/IOzone.config
```

Conclusion

This paper provided performance results for Chelsio's offloaded iSCSI solution running over Chelsio's T5 40Gbps Ethernet ASIC. The benchmark results show that:

- Chelsio's T5 delivers superior iSCSI SAN performance using its T580-CR Unified Wire Network adapter over the range of small I/O sizes of interest (up to 16KB), and comparable performance to the other 40Gbps fabrics for large I/O.
- The iSCSI solution provides significantly higher efficiency in CPU utilization per Gbps of throughput.

These results show that, with its built-in RDMA, iSCSI can achieve high performance without the need for a new fabric that breaks compatibility with the existing large installed base, and introduces inefficiencies that reflect in increased system utilization and cost. The same T5 engine that can process iSCSI at 40Gbps scales to 100Gbps speeds and beyond, guaranteeing a solid performance roadmap for the protocol.

Finally, being part of Chelsio's Unified Wire Ethernet solution, T5's iSCSI implementation provides the same combination of uncompromising performance and feature rich solution as the rest of the offloaded protocols.

Related Links

[The Chelsio Terminator 5 ASIC](#)

[iSCSI at 40Gbps](#)

[High Performance iSCSI for Virtual Machines](#)

[TCP Offload at 40Gbps](#)