Low Latency Performance at 100GbE

Chelsio iWARP RDMA, WireDirect (WD-TOE/TCP/UDP) Solution

Executive Summary

Chelsio recently released its next generation Terminator 6 (T6) ASIC, a 1/10/25/40/50/100GbE Unified Wire controller. This paper provides an overview of Chelsio 100GbE iWARP RDMA and WireDirect (WD) TCP/TOE/UDP solution, demonstrating low network latency, high message rate compared to the latest Competitor adapters. T6 ASIC is capable of delivering sub-µsec latency and provides much needed acceleration for the High Performance Computing (HPC), High Frequency Trading (HFT) and other financial services applications.

The Chelsio iWARP RDMA Solution

Chelsio's Terminator 6 ASIC offers a high performance, robust fourth-generation implementation of RDMA (Remote Direct Memory Access) over 1/10/25/40/50/100Gb Ethernet – iWARP, a plug-and-play, scalable, congestion controlled and traffic managed fabric, with no special switch and configuration needed. T6 enables a unified wire for LAN, SAN and cluster applications, built upon a high bandwidth and low latency architecture along with a complete set of established networking, storage and cluster protocols operating over Ethernet (TCP, UDP, iSCSI, iSER, SMB3.x, iWARP, NVMe-oF, FCoE, NFS/Lustre over RDMA).

Chelsio supports several Message Passing Interface (MPI) implementations through integration within the Open Fabrics Enterprise Distribution (OFED), which has in-boxed Terminator drivers since release 1.2. Continuing the performance curve established by previous generations of Terminator, T6 delivers sub-µsec latency to enable a high performance end-to-end RDMA solution that meets or exceeds the fastest FDR/EDR InfiniBand speeds in real-world applications.

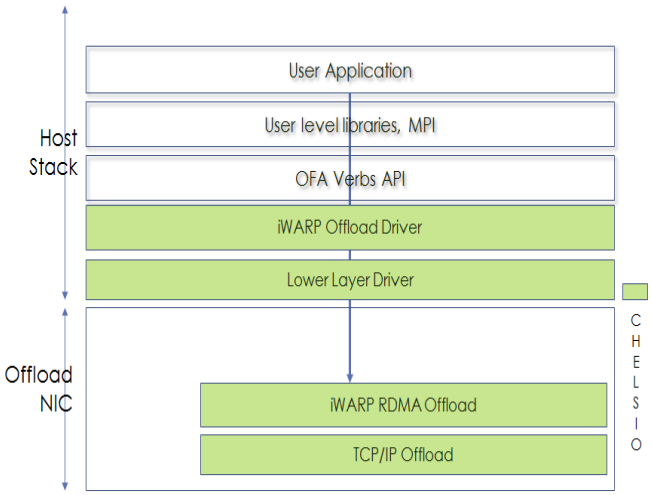


Figure 1 – Chelsio iWARP Stack Overview

The Chelsio WireDirect (WD) Solution

WireDirect provides user space applications direct access to Ethernet Packets on the wire via the Chelsio T5/T6 ASIC, without having to go through any of the I/O layers of the host computing stack. It enables a 100 percent kernel bypass and zero copy with packet polling for data to and from the network wire. In the heart of the WireDirect implementation, WD-QP (Queue Pair) API is the core component which provides an asynchronous RDMA Verbs-like full user mode Ethernet interface and enables zero-copy, low latency Ethernet I/O option for the applications needing low

latency/high packet rates. WD-TCP, WD-TOE, and WD-UDP are other WireDirect modules, built on top of WD-QP to provide seamless I/O acceleration to applications.

Additionally, Low Latency WireDirect I/O fabric also allows bi-directional (egress + ingress) on-the-wire packet tracing, packet capture and packet filtering by bypassing the host OS kernel and I/O stack and going directly to user space that packet capture applications like tcpdump, wireshark, and ethereal can use.

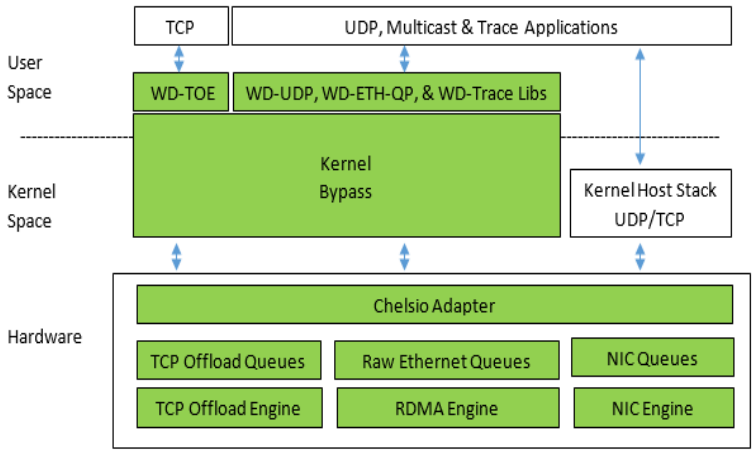


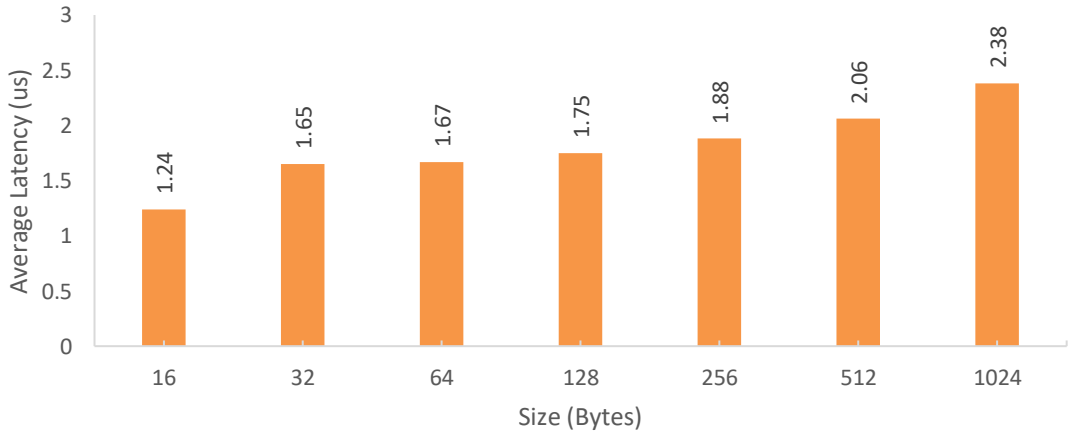
Figure 2 – Chelsio WireDirect Stack Overview

High Frequency Traders always seek the lowest latency out of networking cards so that they can take action on a trade quicker than their competition. The average latency accurately shows what the expected latency of transfer as min latency would be a small subset of what the latency of a transfer could achieve. In this paper, we compare the average latency of Chelsio T6 Adapters with latest Competitor adapters.

Test Results

Chelsio iWARP RDMA Performance:

HFT applications that use RDMA will see the lowest latency since the adapters have built in hardware for avoiding memory copies by placing data directly into application buffers and bypassing the kernel to avoid CPU intervention with drivers that poll from user space. RDMA is a bookend technology that can be used in HFT scenarios where RDMA adapters are used on both sides of the connection.

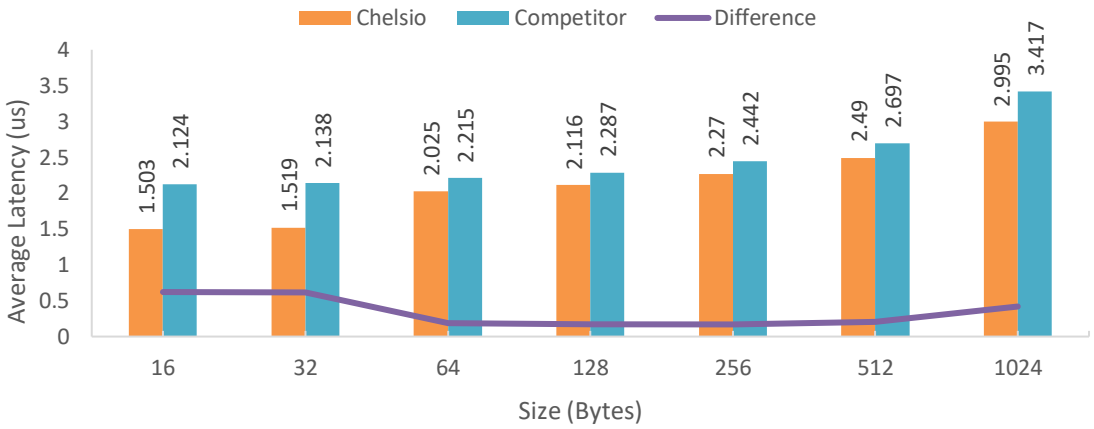


Chelsio Lowest RDMA Latency 1.24us, .263us/.418us less than UDP/TCP
Chelsio Highest RDMA Message Rate is 5.9M, 1.7M/3.1M more than UDP/TCP

Figure 3 – Chelsio iWARP RDMA Latency and Message Rate

Chelsio WireDirect TCP Offload (WD-TCP/TOE) Performance:

HFT applications that use TCP will see ultra-low latency by using WireDirect (WD) to expose Chelsio's TOE (TCP Offload Engine) ASIC queues to poll from user space, avoid memory copies with direct data placement (DDP) and bypass the kernel. The full TCP stack is offloaded: including connection establishment, teardown, retransmission, timers and states of all connections. WD-TOE accelerates read/write, readv/writev, send/recv, sendmsg/recvmsg, sendto/recvfrom and select/poll/epoll. WD-TOE allows single ended TCP socket connection acceleration that is required in a lot of HFT scenarios.

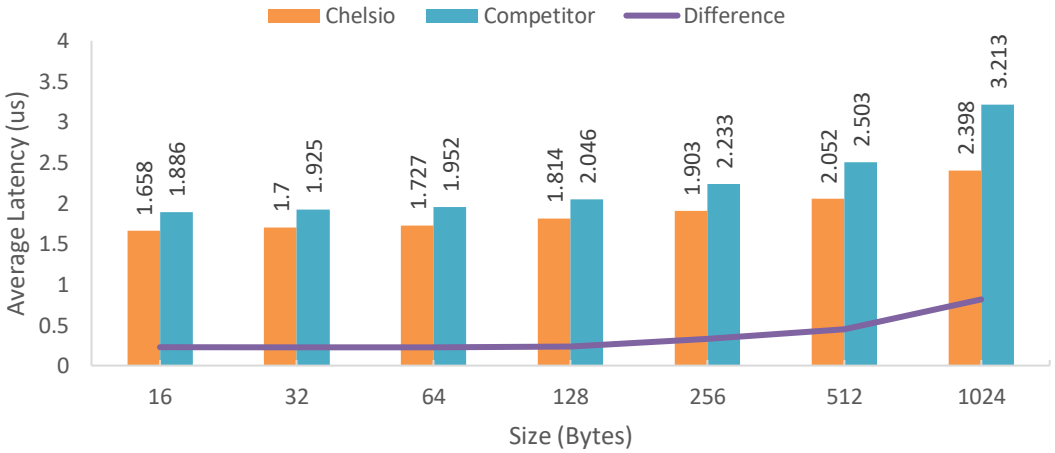


Chelsio TCP Latency is Lower than Competition by .171us to .621us
Chelsio TCP Message Rate of 2.8M is 190K Higher than Competition

Figure 4 – TCP Latency and Message Rate

Chelsio WireDirect UDP (WD-UDP) Performance:

HFT applications that use UDP and Multicast will see ultra-low latency by using WireDirect (WD) to expose Chelsio's Raw Ethernet queues to poll from user space, avoid memory copies and bypass the kernel. The WireDirect user space UDP stack accelerates read/write, readv/writev, send/recv, sendmsg/recvmsg, sendto/recvfrom and select/poll/epoll. WD-UDP allows single ended UDP socket connection acceleration that is required in a lot of HFT scenarios.



Chelsio UDP Latency is Lower than Competition by .225us to .815us
Chelsio UDP Message Rate of 4.25M is 660K Higher than Competition

Figure 5 – UDP Latency and Message Rate

Test Setup

Topology

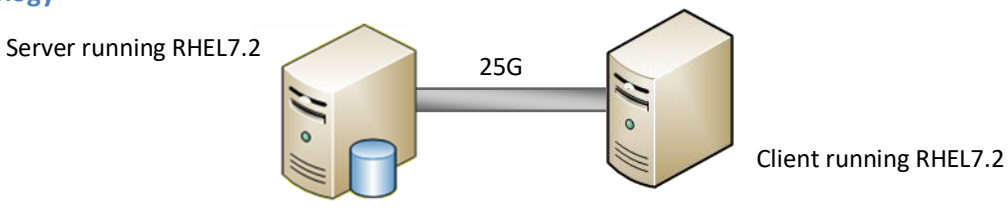


Figure 6 – Simple Back-to-Back Test Topology

Network Configuration

The test setup consists of 2 machines connected back-to-back using a single port, a Server and Client, each configured with 1 Intel Xeon CPU E5-1620 v4 4-core processor (HT enabled) clocked at 3.50GHz, 16GB of RAM and RHEL7.2 Operating System. Chelsio's T6225-LL-CR is installed in each machine. MTU of 1500B is used.

Conclusion

This paper showcases the significant performance benefits of Chelsio T6 based 100G iWARP RDMA and WireDirect solution. T6 solution is capable of delivering a sub-µsec latency and provides much needed acceleration for the High Performance Computing (HPC), High Frequency Trading (HFT) and other financial services applications. The results show that T6 adapters deliver:

- The lowest Average RDMA Latency of 1.24 µsec.
- Significantly lower Average TCP and UDP latency than the competitor using WireDirect kernel bypass technology over a range of useful packet sizes.

Chelsio adapters are the best solution to achieve the lowest latencies for RDMA/TCP/UDP that are required by the HFT market companies to ensure their trades are faster than the competition.

Related Links

The Chelsio Terminator 6 ASIC
Chelsio T6 Latency Demonstration
Chelsio T520-LL-CR vs. Solarflare SFN7122F
High Frequency Trading Report.
Preliminary Ultra Low Latency Report
Packet Rate Performance Report