Preliminary Ultra Low Latency Report

T5 WireDirect TCP WD-TOE, UDP WD-UDP, and RDMA iWARP

Abstract

Chelsio performed networking latency and message rate testing of its high-performance, low-latency WireDirect 10GbE server adapter based on the T5 ASIC generation. Chelsio WireDirect performs network processing at user-level and is binary compatible with existing applications that use TCP/UDP with BSD sockets. The test found that the Chelsio T520-LL-CR server adapter with Chelsio WireDirect UDP and TOE achieved ultra low latency for both UDP and TCP. In back-to-back testing, the server adapter achieved minimum UDP 6 byte latency 1.6us, and minimum TCP 48 byte latency 2us. The latency was very deterministic with 99% of the messages being delivered with UDP 6 byte latency of less than 2us and TCP 48 byte latency of less than 2.3us at near zero jitter. Chelsio measured the performance of TCP and UDP messaging using the widely used sockperf benchmark tool. Note that all numbers presented here are preliminary and have not been fully optimized.

Chelsio adapters also offer the RDMA iWARP solution to achieve low latency with kernel bypass and zero copy over an RDMA Verbs QueuePair interface. RDMA accomplishes this by offloading onto a "channel adapter" interface card that performs the tasks traditionally performed by the operating system during network transfers. The result is high throughput, low latency, and low CPU utilization message transfer. Chelsio measured the performance of RDMA messaging using the widely used *rdma_lat* and *rdma_bw* benchmark tools. The test found that the Chelsio T520-LL-CR server adapter with Chelsio RDMA iWARP achieved extremely low latency for RDMA messaging. In back-to-back testing, the server adapter achieved consistent minimum RDMA iWARP 24 byte latency 1.6us with the average latency also 1.6us.

The test platform used servers and processors typically deployed in the financial services industry today.

Low Latency Messaging

High Frequency Trading has transformed the investment landscape, accounting for over two thirds of current trading volume. But as traffic and complexity increase, so do the consequences of inefficiencies in the network and microbursts have become a big issue when dealing with data market feeds. This means that the capacity in messages per second that the networking interface can handle will determine if the required data is received in time to make the winning trade. Chelsio adapters achieve 4 to 5 times the message rate of competing adapters on the market at 6.2Mmps with RDMA iWARP, 3Mmps with UDP WD-UDP and 2.3Mmps with TCP WD-TOE. When reaction time counts for so much, why risk profitability by choosing slower network interface cards for a trading infrastructure? Chelsio now delivers TCP/UDP socket and RDMA iWARP connectivity with the lowest delay and the highest packet processing capacity available in a network interface card.

FUNDAMENTAL RDMA iWARP LATENCY FINDINGS

Chip: size	min	avg	max
T5 24 Byte	1.61502	1.64652	2.88298

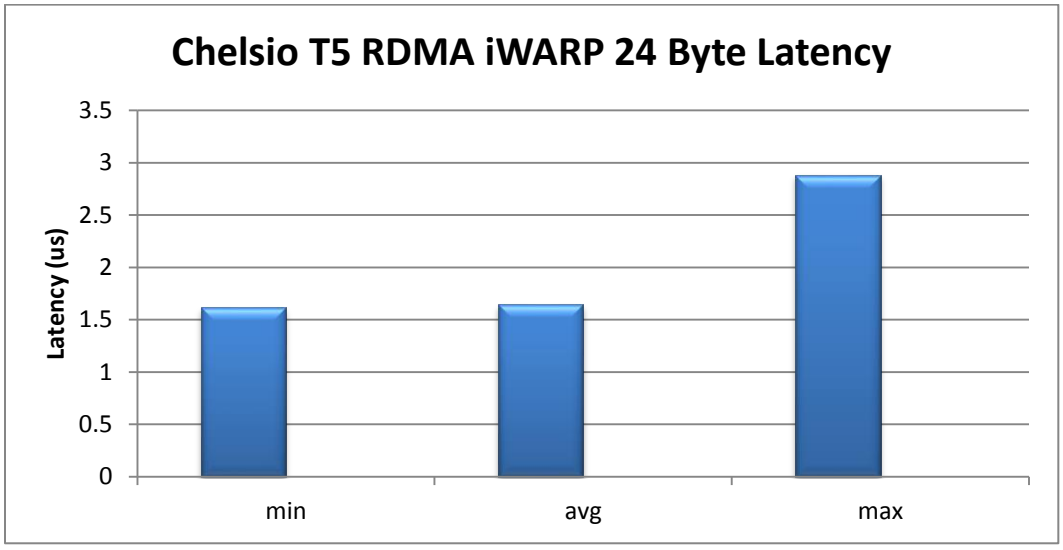

Exhibit 1: One way, T5 24 Byte RDMA Latency in microseconds

Exhibit 1 summarizes the RDMA latency test results. The Chelsio adapters' back-to-back minimal 24 byte latency is 1.6us. This latency was very deterministic, with an average for the 24 byte latency at 1.6us and near zero jitter.

FUNDAMENTAL RDMA iWARP MESSAGE RATE FINDINGS

Chip: size	avg	max
T5 24 Byte	6.22752	6.229353

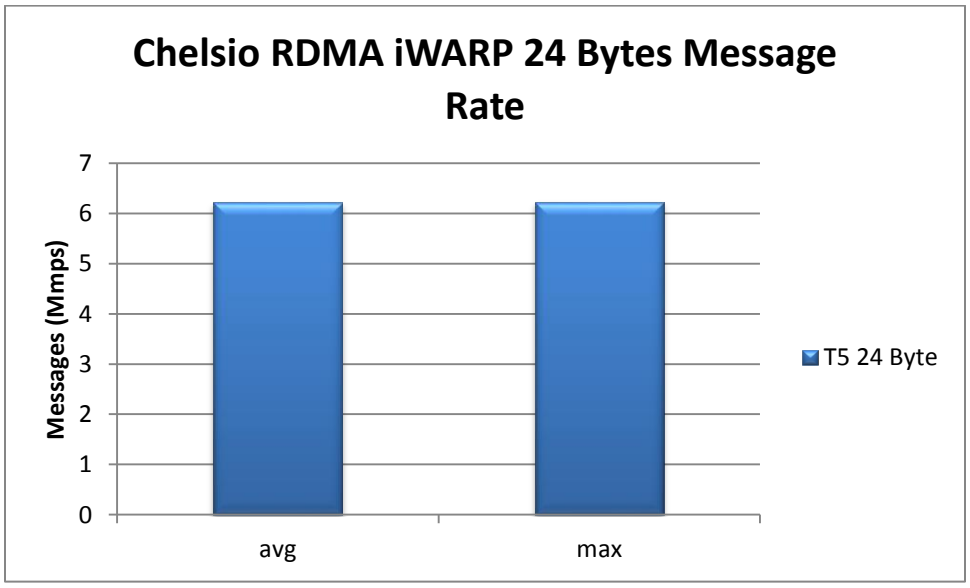

Exhibit 2: T5 24 Byte RDMA Message rate in millions of messages per second

Exhibit 2 summarizes the RDMA message rate test results. The Chelsio adapters' back-to-back message rate achieved of 6.2 Mmps.

FUNDAMENTAL UDP LATENCY FINDINGS

Chip: size	stddev	min	avg	99%ile	max
T5 6 Byte	0.057	1.687	1.77	2.026	2.25

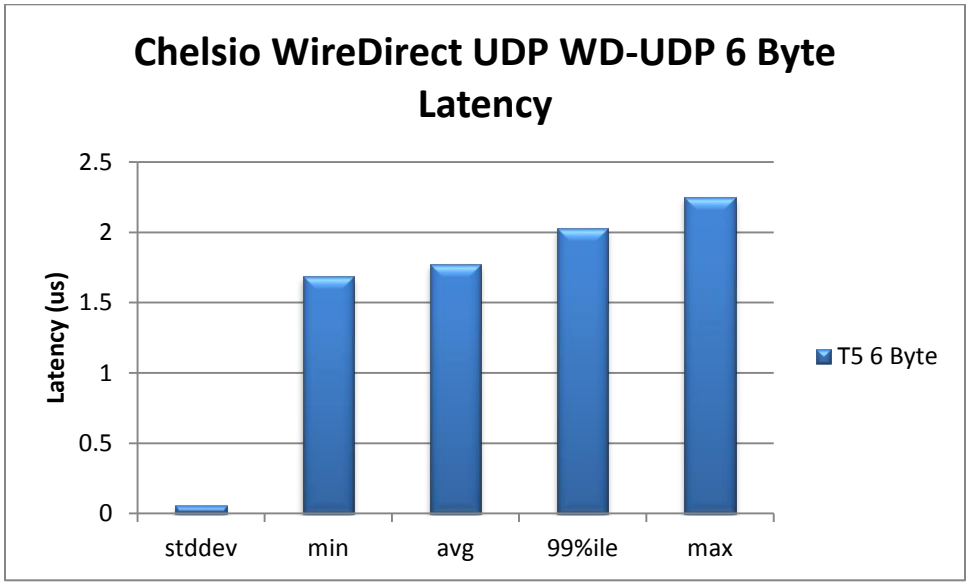

Exhibit 3: One way, T5 6 Byte UDP Latency in microseconds

Exhibit 3 summarizes the UDP latency testing. The Chelsio T5 adapters' back-to-back minimal 6 byte latency is 1.6us. This latency was very deterministic 99% of the 6 byte latency of 2us with near zero jitter.

FUNDAMENTAL UDP LATENCY FINDINGS AT MULTIPLE MESSAGE RATES

Chip: size	mps	stddev	min	avg	99%ile	max
T5 48 Byte	500000	0.064	2.064	2.231	2.37	2.47
T5 48 Byte	1000000	0.054	2.057	2.23	2.365	2.519
T5 48 Byte	1500000	0.052	2.058	2.223	2.356	2.672
T5 48 Byte	max	0.073	2.462	2.629	2.73	2.853

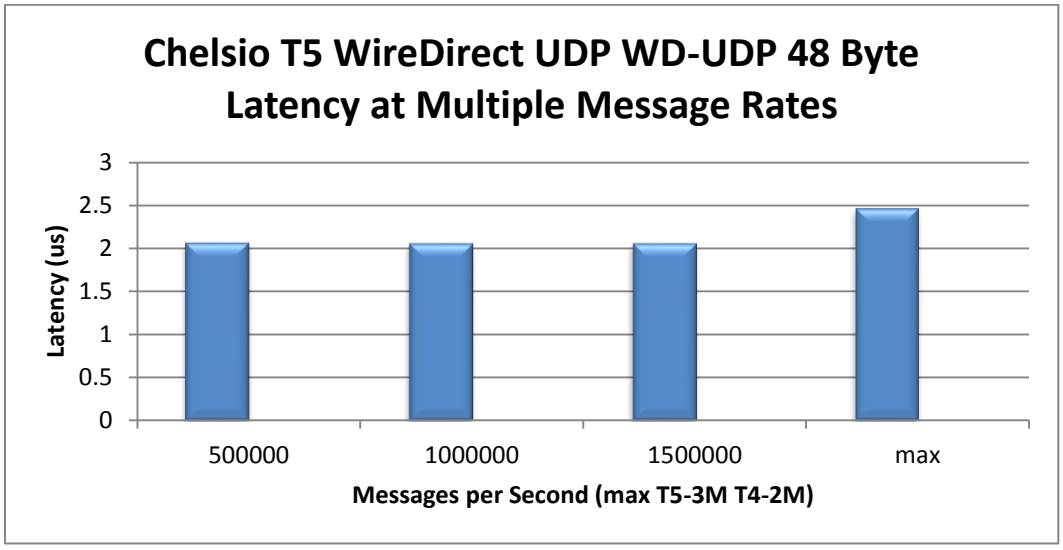

Exhibit 4: One way, T5 48 Byte UDP Latency in microseconds at Multiple Message Rates

Exhibit 4 summarizes the UDP latency message rate testing. The Chelsio adapters' latency for multiple message rates is very constant at about 2.1us with a max message rate of 3Mmps.

FUNDAMENTAL TCP LATENCY FINDINGS

Chip: size	stddev	min	avg	99%ile	max
T5 48 Byte	0.204	2.003	2.119	2.298	11.173

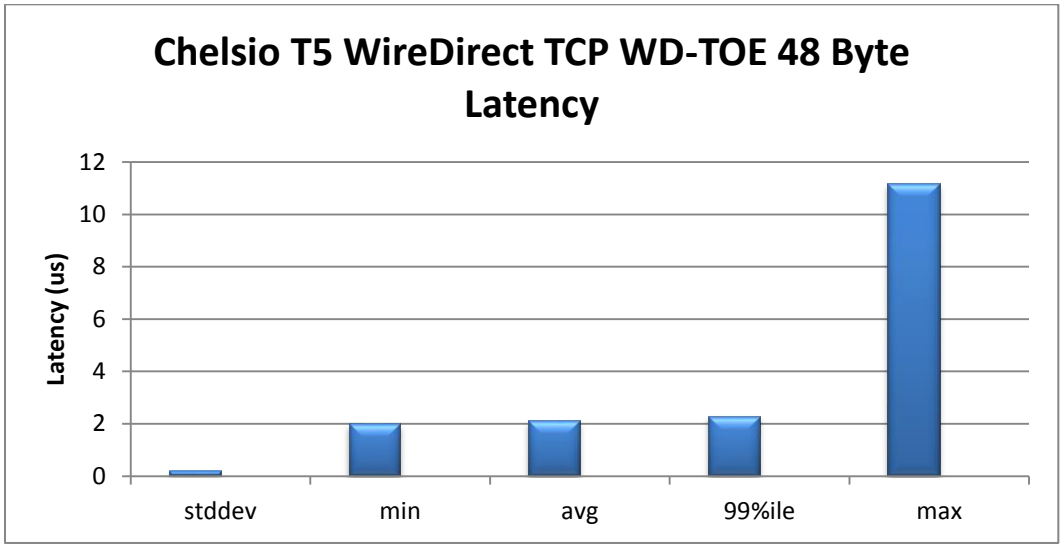

Exhibit 5: One way, T5 48 Byte TCP Latency in microseconds

Exhibit 5 summarizes the TCP latency testing. The Chelsio adapters' back-to-back minimal 48 byte latency of 2us. This latency was very deterministic with 99% of the 48 byte latency of 2.3us with near zero jitter.

FUNDAMENTAL TCP LATENCY FINDINGS AT MULTIPLE MESSAGE RATES

Chip: size	mps	stddev	min	avg	99%ile	max
T5 48 Byte	500000	0.058	2.06	2.192	2.337	3.523
T5 48 Byte	1000000	0.131	2.062	2.285	2.488	3.533
T5 48 Byte	1500000	0.049	2.101	2.259	2.396	3.359
T5 48 Byte	max	0.122	2.162	2.589	2.998	10.505

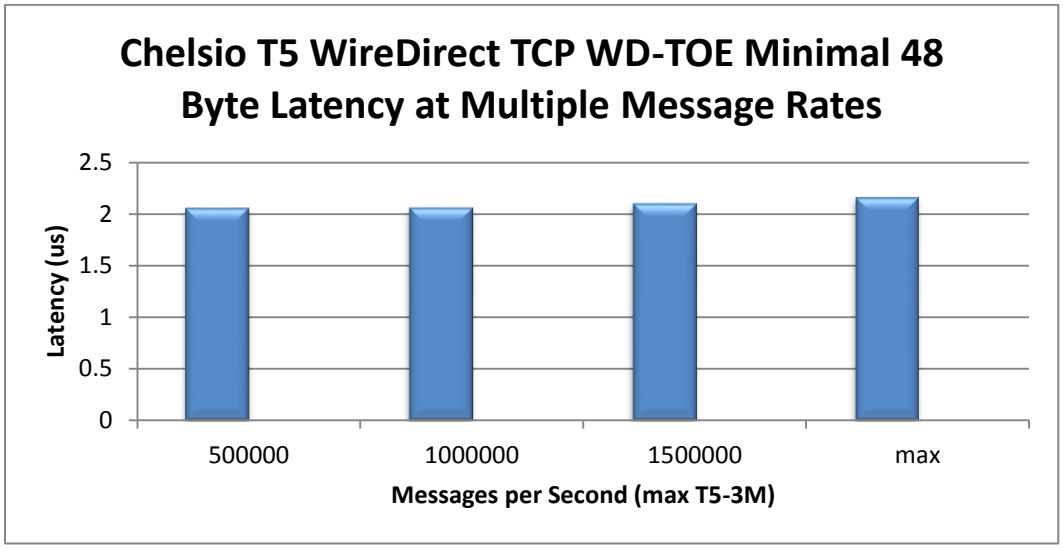

Exhibit 6: One way, T5 48 Byte TCP Latency in microseconds at Multiple Message Rates

Exhibit 6 summarizes the TCP latency message rate testing. The Chelsio adapters' latency for multiple message rates is very stable at about 2us with a max message rate of 2.3Mmps.

Conclusions

This paper presented the one-way, UDP/TCP 48 byte latency and RDMA 24 byte latency with a configuration using the Chelsio T5 ASIC based server adapter with Chelsio WireDirect and RDMA iWARP, with up to 3 million messages per second (mps) transmission rates for UDP/TCP and up to 6 million messages per second (mps) for RDMA iWARP.

The following was observed for T5 RDMA iWARP 24 byte message sizes:

- RDMA iWARP minimum latency was 1.6us
- RDMA iWARP average latency was 1.6us
- RDMA iWARP max message rate of 6.2Mmps

The following was observed for T5 UDP/TCP 6 and 48 byte message sizes:

- UDP latency minimum was 1.6us
- UDP latency average was 1.7us
- TCP latency minimum was 2us
- TCP latency average was 2.1us
- UDP 99th percentile latency was 2us
- UDP max message rate of 3Mmps
- TCP 99th percentile latency was 2.3us
- TCP max message rate of 2.3Mmps

Chelsio Adapters with WireDirect

The Chelsio WireDirect server adapters deliver unmatched message rates with low latency and jitter over standard Ethernet along with the lowest CPU utilization and power consumption, enabling the industry's best performance and scalability for financial services and other enterprise data centers. WireDirect performs network processing at user-level and is binary compatible with existing applications that use TCP/UDP with BSD sockets. Chelsio's network interface cards meet all the requirements for low latency High Frequency Trading operations. Chelsio's custom T5 ASIC offers protocol acceleration for UDP/TCP/RDMA/iSCSI/FCoE/NFS/CIFS. This makes the Chelsio adapter an ideal Unified Wire adapter, simultaneously accelerating processing for all protocols with the same card/driver/firmware. Using Chelsio's WireDirect application acceleration middleware in combination with Chelsio's TOE ASIC enables full operating system bypass dramatically reduces host processing overheads and enables high transaction rates while substantially reducing application latency with very low jitter.

Chelsio adapters also provide Remote Direct Memory Access (RDMA) iWARP API. RDMA is a networking paradigm that allows a user process on one computer to move messages over a network directly between its virtual memory and the virtual memory of a process on another computer without operating system intervention on either side of the data transfer. RDMA accomplishes this by offloading onto a "channel adapter" interface card the tasks traditionally performed by the operating system during network transfers. The result is high throughput, low latency, low CPU utilization message transfer. RDMA is supported by the Open Fabrics Alliance, which provides open-source software to efficiently utilize this technology from user space (the OFED stack). The OFED stack is in-boxed in the main Linux distributions, and is being ported to other operating systems.

###