

SMBDirect Latency on Windows Server 2012 R2

Realize Full SSD Storage Performance with T5 RDMA over Ethernet

Executive Summary

Chelsio's Terminator 5 ASIC RDMA over Ethernet (iWARP) implementation provides high performance, CPU offload and low latency with host stack bypass, while concurrently enabling a full suite of networking and storage features, including user space IO with WireDirect, full offload of TCP/IP and UDP/IP, iSCSI and FCoE, all traffic managed and firewalled.

Storage protocols like SMB are particularly well suited to using and benefiting from RDMA, because of their characteristics and performance requirements. This paper compares SMB latency results for regular NIC and RDMA, taken with Chelsio's T580-CR adapter. READ and WRITE tests at various IO size show a round trip latency of nearly 5 μ s and consistent performance under load. These ultra low latency results with RDMA allow realizing the performance potential of recent and future SSD devices.

Overview

Remote DMA (RDMA) is a technology that achieves unprecedented levels of **efficiency**, thanks to direct **memory-to-memory** communication, without CPU involvement or buffer copies. With RDMA enabled adapters, all packet and protocol processing required for communication is handled by the network adapter itself, necessarily in hardware for high performance. **iWARP** is the IETF standard for RDMA over Ethernet, which provides all the benefits of RDMA, including **bypassing the host stack**, **low latency** communications and **zero copy** data transfer. Furthermore, thanks to its robust TCP/IP foundation, iWARP is **routable** and a native client of private and public clouds, requiring no changes to the standard Ethernet infrastructure.

In an era of Big Data, massive datacenters, pervasive virtualization and focus on "green" operation and efficiency, RDMA use is rapidly gaining ground. Moreover, RDMA support is natively supported in today's major server operating systems. By providing high level, simplified communication abstractions, such integration further lowers the barrier to realizing the benefits of RDMA, and is further contributing to the acceleration in RDMA adoption. Specifically, Windows Server 2012 introduced **SMB Direct**, an RDMA transport for SMB that has shown significant performance benefits, with unprecedented performance in key applications, such as virtual machine migration.

This paper presents latency benchmarks results that demonstrate the low remote storage access latency made possible with SMB Direct. Such latency numbers allow exploiting the full potential of ultra **low latency SSD** drives, and in combination with the efficient high throughput made possible by iWARP, provide the next generation, scalable storage network over standard, **cost effective Ethernet**.

Test Results

The following graph compares 40GbE NIC and iWARP RDMA round trip latency numbers at various IO sizes, collected using the **sqlio** tool.

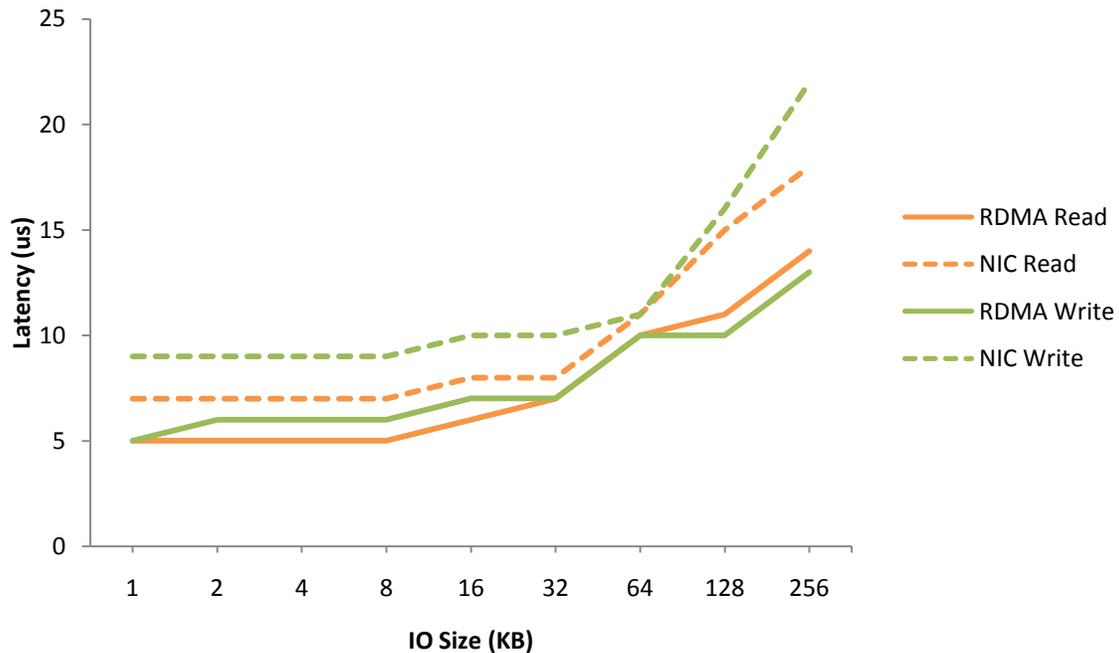


Figure 1 –RDMA and NIC Round Trip Latency vs. IO Size

The results show READ RTT latency for RDMA to be consistently below 5usec for IO sizes up to 8KB, about 33% lower than the NIC, and only reaches the base NIC latency when the IO size exceeds 32KB. Similarly, the WRITE RDMA RTT latency is in the 5-6usec for up to 8KB IO size, about 40% lower than the NIC, and only reaches base NIC latency at IO size of 64KB. Furthermore, the RDMA latency for RDMA increases with a desirably shallower slope than the NIC as the IO size is increased.

Test Configuration

The following sections provide the test setup configuration details.

Topology

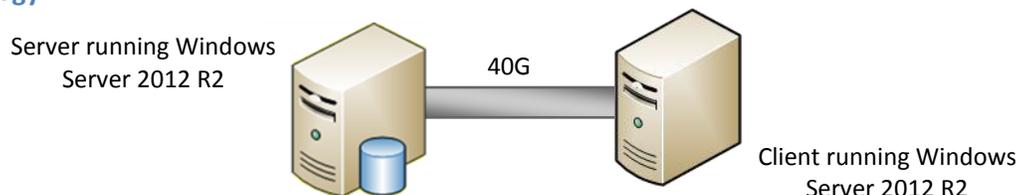


Figure 2 – Simple Back-to-Back Topology

Setup Configuration

The test configuration consists of 2 machines connected back-to-back using a single port, a Server and Client, each with 2 Intel Xeon CPU E5-1660 v2 8-core processors clocked at 3.40GHz and with 64 GB of RAM. A Chelsio T580-CR adapter was installed in each system with Windows Unified Wire software v5.0.0.33 and Windows Server 2012 R2 operating system. The standard MTU of 1500B is configured.

I/O Benchmarking Configuration

The **Microsoft sqlio v2.15** tool was used to assess performance. This test used sample IO sizes varying from 1KB to 256KB.

Command Used

```
# sqlio2.exe -T<100/0> -s20 -t1 -BN -LSi -dz testfile.dat -o40 -b<IO Size>
```

Conclusion

This paper compared the SMB latency performance for NIC and RDMA (SMB Direct) on Windows Server 2012 R2, running over Chelsio's T580 40Gbps Ethernet adapter. The SMB RDMA transport latency was near 5 μ s for READ and WRITE RDMA, and consistently outperformed the NIC even at large IO sizes.

Thanks to its hardware TCP/IP based foundation, T5's iWARP RDMA solution provides all the benefits of RDMA over standard Ethernet without need for special equipment or network configuration, while providing private and public cloud scale routability, congestion control and traffic management.

Related Links

[The Chelsio Terminator 5 ASIC](#)

[SMBDirect 40 GbE iWARP vs 56Gb Infiniband](#)

[Windows Server 2012 R2 SMB Performance](#)

[40Gb Ethernet: A Competitive Alternative to InfiniBand](#)