Windows Guest RDMA at 100Gbps

Networking to the VM at Bare Metal Speeds & Latencies

Introduction

Today, RDMA is known as the premier and superior form of network data transfer. It's use maximizes network bandwidth by enabling zero copy technology, which allows for full link speeds at very low latencies and very low CPU utilization. Microsoft was an early adopter supporting RDMA in their Windows Operating Systems since 2012. Since then they have grown RDMA support to an ever larger set of functional applications. Microsoft has recently introduced Guest RDMA, an enhancement that brings RDMA to the Virtual Machine (VM), again reaffirming their technological lead. Chelsio has been working closely with Microsoft on Guest RDMA and this paper describes the fruits of that labor by highlighting performance benchmarks at 100 Gb/s on Chelsio's T62100-CR adapters. Before the results are presented, it is useful to first walk through the history and features of RDMA on the Windows OS.

RDMA at Microsoft

iWARP and Microsoft – Chelsio has been working with Microsoft closely from the beginning to bring RDMA to life with iWARP, which is RDMA over TCP/IP. RDMA has three primary flavors: iWARP, RoCE, and Infiniband (IB). Of the three, **Microsoft recommends iWARP** over RoCE and IB for its superior ease-of-use capabilities, very low support needs, ability to route over long distances, all while maintaining a performance edge.

SMB Direct Storage – The initial use case of RDMA into the Microsoft ecosystem came with the introduction of SMB 3.0 and in particular, SMB Direct. This was introduced by Microsoft in Windows Server 2012 as an extension of Server Message Block (SMB), the popular storage network file sharing protocol.

Ease of Use with SMB Direct – With the use of Multichannel, SMB Direct detects the use of RDMA on both ends of the networking wire. If only one of two sides has an RDMA Network Interface Card (R-NIC), the protocol defaults to regular SMB with storage transferred over regular TCP/IP. If both network adapters at each end node are in fact R-NICs, RDMA is used. This is a subtle but transformative feature bringing seamless ease-of-use of RDMA to the masses. This is the first time that the use of RDMA, especially with iWARP, was truly plug-and-play.

SMB Direct & the Hyper-V Virtual Switch – Microsoft introduced in Windows Server 2016, support for SMB Direct with the Hyper-V Virtual Switch. This technology allows SMB Direct to be bound to the virtual switch simultaneously with regular networking traffic. This was the beginning for Microsoft to introduce Hyper-converged clusters that contain both storage and compute nodes. This combination of technologies makes up the networking protocol core of **Storage Spaces Direct (S2D)**. Through the **WSSD** program, Microsoft enables S2D through partners for hyper-converged appliances in the market.

Other RDMA Implementations on Windows – With the introduction of RDMA at Microsoft, the underlying technology called NDK enabled SMB Direct. It also allows for other RDMA derivatives, of which Chelsio has implemented three.

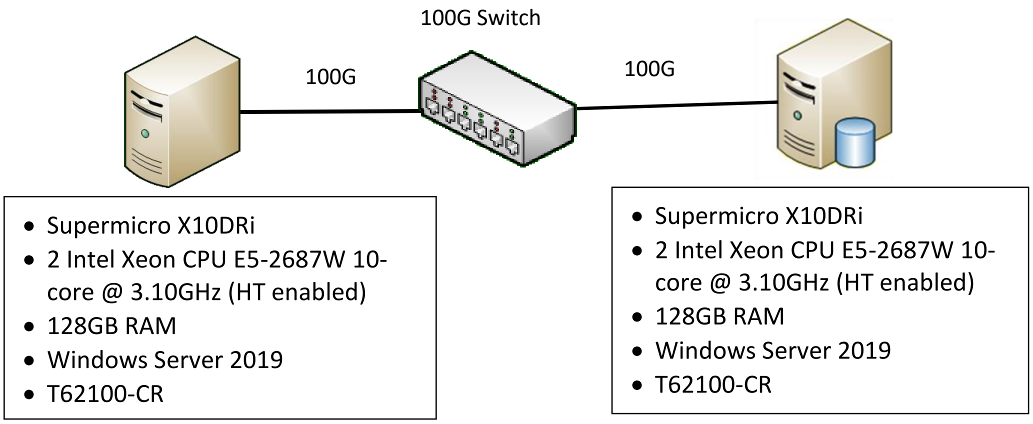
- In storage, NVMe over Fabrics **(NVMe-oF)** uses RDMA as the fabric and Chelsio is a large player in the market on Linux. Chelsio is now shipping an NVMe-oF initiator driver for use on Windows Server 2016/19 and Windows 10.
- Chelsio is also shipping today an initiator driver for iSCSI over RDMA **(iSER)**, which is a high performance block storage protocol.
- Finally, Chelsio today supports Network Direct **(ND)**, an RDMA technology used typically in massively parallel applications.

Guest RDMA

With Windows Server 2019 Microsoft introduces Guest RDMA, which allows for RDMA and SMB Direct to work through SR-IOV to the Virtual Machine (VM). It enables a direct network data path through the R-NIC to the VM allowing RDMA to perform its zero copy, low latency, and low CPU magic. For applications that require such network and storage performance, it now doesn't matter if they're running in the host Windows OS or in a VM based Windows OS. Chelsio is proud to have partnered with Microsoft to again be on the leading edge of RDMA with support for Guest RDMA on all models of shipping adapters.

Test Configuration

The setup consists of two physical systems connected through a 100GbE switch using a single port on each system. One system was running a single VM and configured with Guest RDMA. The other system ran RDMA on the host system. MTU of 9000 bytes was used. The OS on both was Windows Server 2019 Build 17686 throughout and the latest Chelsio Unified Wire driver was used.



- Supermicro X10DRi
- 2 Intel Xeon CPU E5-2687W 10-core @ 3.10GHz (HT enabled)
- 128GB RAM
- Windows Server 2019
- T62100-CR

- Supermicro X10DRi
- 2 Intel Xeon CPU E5-2687W 10-core @ 3.10GHz (HT enabled)
- 128GB RAM
- Windows Server 2019
- T62100-CR

Figure 1 – Test Setup

Storage Topology, Test Tool, and Command Line

A Shuntfilter was used for storage in the setup. The Diskspd Tool v2.0.17a was used with the following command lines:

```
WRITE:diskspd.exe -b$i -w100 -r -o64 -W5 -C5 -d30 -t32 -Sh <file>
READ:diskspd.exe -b$i -w0 -r -o64 -W5 -C5 -d30 -t32 -Sh <file>
```

Test Results

The RDMA write IOPS and throughput performance are shown below. As can be seen, write throughput reaches 90 Gb/s or greater from IO sizes of 32K bytes and higher. The IOPS performance reaches about 450K IOPS from IO sizes of 16K bytes and lower. For both, the number of connections (ranging from 1 to 16) doesn't affect results greatly. The reads (not published here), yields similar performance results as writes.

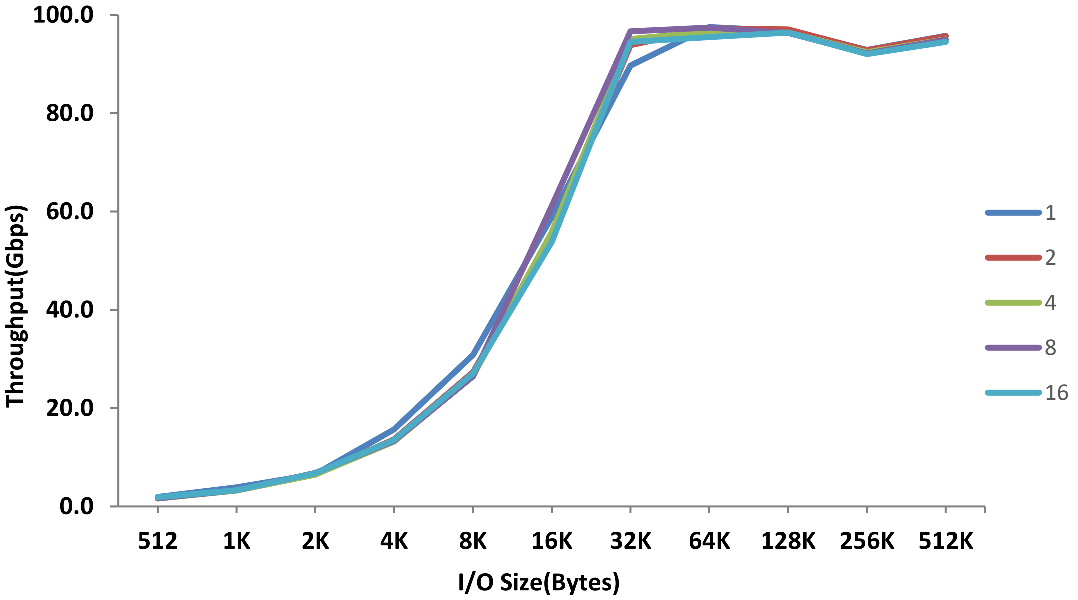


Figure 2 – Guest RDMA WRITE Throughput vs. I/O Size

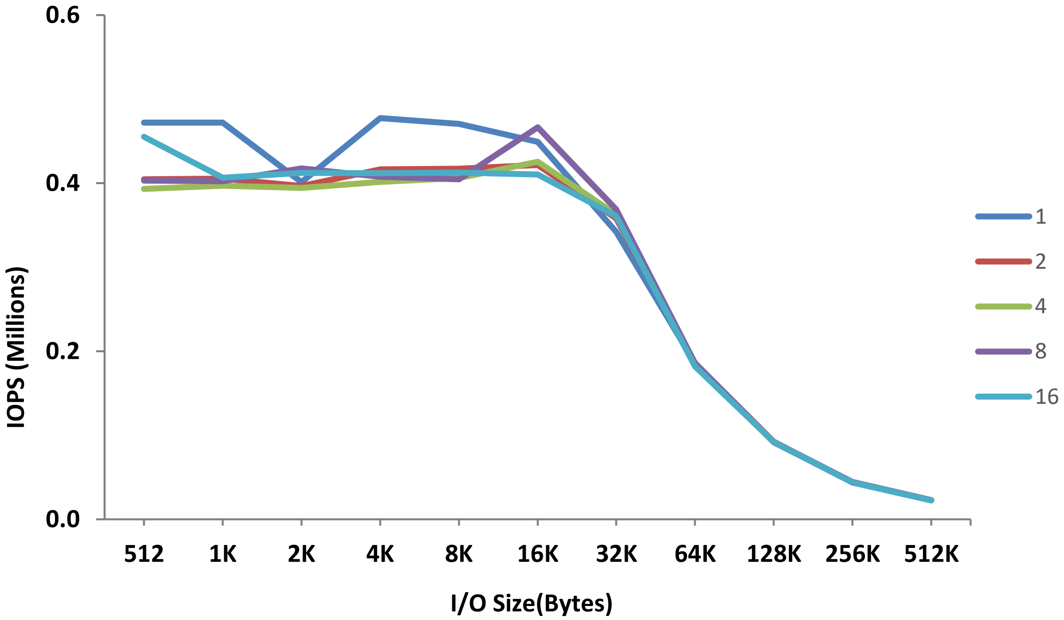


Figure 3 – Guest RDMA WRITE IOPS vs. I/O Size

The results above were achieved with a CPU utilization of ~1% on the server side and ~5% on the client (VM) side for 512Kbytes IO size. This highlights the value of RDMA offload, where high IOPs and high throughput can be achieved with low CPU utilization. This in turn frees up the server node for application use.

Conclusion

Chelsio has become an RDMA leader on Microsoft Windows Server 2016/2019 and Windows 10. This paper walked through the RDMA features and benefits from Microsoft and specifically the iWARP implementations from Chelsio. Highlighted is a new RDMA feature introduced in Windows Server 2019 called Guest RDMA. Shown here was the performance and benefits of Guest RDMA over 100 Gb/s adapters. RDMA on Windows operates on the full range of Chelsio adapters from 1 Gb/s to 100 Gb/s and everything in between.

Related Links

https://www.chelsio.com/nic/unified-wire-adapters/
https://www.chelsio.com/wp-content/uploads/resources/t5-10g-s2d-xio.pdf
https://www.chelsio.com/wp-content/uploads/resources/t6-100gb-s2d-microsoft.pdf
https://www.chelsio.com/wp-content/uploads/resources/uwire-cna-windows.pdf