Validation of SNAP I/O Performance on Dual-Socket PowerEdge Rack Servers

Tech Note by

Matt Ogle
Mike Darby
Rich Hernandez

Summary

Using Non-SNAP IO communication paths for one-NIC dual-socket servers increases UPI overhead, which slows down bandwidth and increases latency for CPU applications. Resolving this by adding another NIC card will increase solution TCO.

The adoption of SNAP I/O allows a dual-socket server to bypass traversing the UPI lanes when using one-NIC configurations, ultimately increasing performance and TCO for one-NIC dual socket solutions.

This DfD will measure the performance readings of SNAP I/O against two Non-SNAP I/O configurations to demonstrate how using SNAP I/O can increase bandwidth, reduce latency and optimize user TCO.

SNAP I/O Value Proposition

Dual-socket servers offer ample compute power to meet the needs of a wide range of workloads. However, if the network adapters in the system are unbalanced, users may be at risk of creating a bottleneck that will reduce bandwidth and increase latency. SNAP I/O is a solution which leverages Mellanox Socket Direct technology to balance I/O performance without increasing the TCO. By allowing both CPUs to share one adapter, data can avoid traversing the UPI inter-processor link when accessing remote memory.

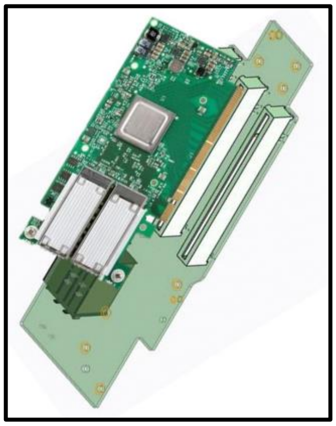

Figure 1: SNAP I/O Card

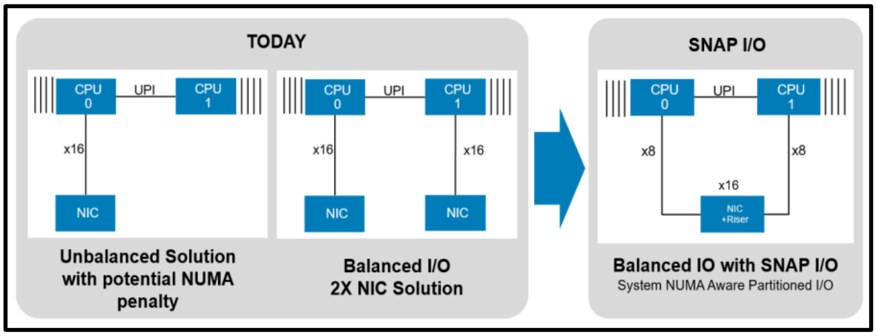

Figure 2: Comparing an unbalanced one-NIC solution and a balanced two-NIC solution to a SNAP I/O one-NIC solution. The SNAP I/O solution on the right allows CPU 0 and 1 to communicate to their corresponding NIC card without traversing the UPI channels, therefore reducing latency/TCO and freeing up UPI bandwidth for applications

As seen in Figure 2, the unbalanced configuration has CPU 0 in direct communication with the NIC through a PCIe x16 slot, while CPU 1 must traverse the UPI channel to CPU 0 first before it can communicate with the NIC. This data travel path adds latency overhead when traversing the UPI channel and can impact total bandwidth at high speeds. One solution to this is to have an additional NIC card connected directly to CPU 1, but this solution will introduce a 2x cost multiplier, including a 2nd NIC card, cable and switch port. Rather than doubling NIC and switch costs, Dell SNAP I/O can bridge the two sockets together by splitting the PCIe x16 bus into two x8 connectors and allowing the OS to see it as two NICs.

Test Scope and Configurations

To characterize performance variances, two testing devices were configured (see Figure 3). The SNAP I/O configuration used the PowerEdge R740 while the unbalanced one-NIC configuration and balanced two-NIC configuration used the PowerEdge R740xd. Aside from the chassis form factor and SNAP I/O riser, both pieces of apparatus were configured identically so the comparison was apples-to-apples.

Two test platforms were used to measure network bandwidth, latency, UPI utilization and CPU utilization. The first set of tests measured performance for an OS test scope, including benchmarks like iperf, qperf, Pcm.x and top. The second set of tests measured performance for a Docker test scope, including benchmarks like iperf3 and qperf.

	SUT1	SUT2
Product Model	R740	R740XD
IP	iDRAC: 100.109.34.177 Host: 100.109.34.180	iDRAC: 100.109.33.77 Host: 100.109.34.182
Chassis/Sled	8X2.5" RSR 1B(+2A)+3A	12X3.5" RSR 1B(+2A)
Processor	2*6142	2*6142
Memory	12*16GB RDIMM	12*16GB RDIMM
Drive 1	1*960GB SATA S4500	1*960GB SATA S4500
RAID Controller	1*H330 mini	1*H330 mini
RAID Config	Non-RAID	Non-RAID
PCIe Card 1	Mellanox ConnectX-5 Ex 100 GbE QSFP Adapter(71C1T) OR Mellanox ConnectX-5 Dual Port 25GbE SFP28 PCIe Adapter(V5DG9)	Mellanox ConnectX-5 Ex 100 GbE QSFP Adapter(71C1T) OR Mellanox ConnectX-5 Dual Port 25GbE SFP28 PCIe Adapter(V5DG9)
Slot request for PCIe card 1	Slot 6 (2A slot4 CPU2 Port2-ABCD) SNAPI Riser:	Slot 6 (2A slot4 CPU2 Port2-ABCD)
NIC (NDC/LOM)	BRCM GbE 4P 5720-t rNDC	BRCM GbE 4P 5720-t rNDC
GPU	NA	NA
PSU	RPS 750W	RPS 750W
OS	CentOS 7.5	CentOS 7.5
BIOS or IDRAC customization	Test BIOS	Test BIOS
Commodity FW Version	16.23.02.44 for 100G NIC	16.23.02.44 for 100G NIC

Figure 3: Table displaying the two pieces of apparatus used for testing

Performance Comparisons

Latency

Figure 4 used the OS-level qperf test tool to compare the latency of the SNAP I/O solution against two benchmarks; the first being the NIC connected to the PCIe bus *local* to the CPU, and the second being the *remote* CPU that must cross the UPI to connect to the NIC. The graph shows that for both 100GbE and 25GbE NICs, the SNAP I/O latency is reduced by more than *40%* compared to the latency experienced by the remote CPU accessing the single NIC.

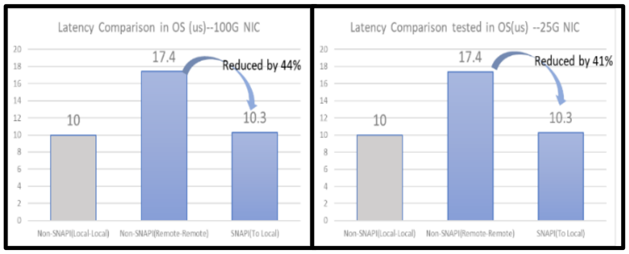

Figure 4: OS latency (in µs) of various configurations; local CPU, remote CPU and SNAP I/O

Figure 5 compares the latency of the SNAP I/O solution against the same two configurations in the docker environment. Like Figure 3, the graphs show that the latency of the SNAP I/O solution has reduced by more than *40%* compared to the latency experienced by the remote CPU.

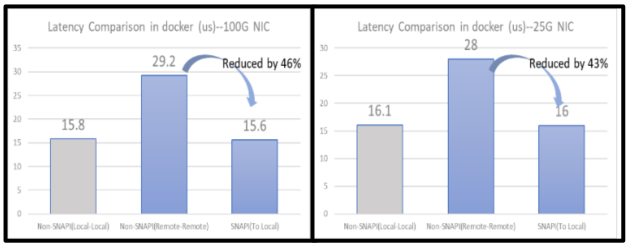

Figure 5: Docker latency (in µs) of various configurations; local CPU, remote CPU and SNAP I/O

Bandwidth

Figure 6 to the right compares the bandwidth of the SNAP I/O against the same two configurations by applying 5 stream memory tests to ensure there is enough UPI traffic for accurate iperf bandwidth testing. The graphs show that for 100G NICs, the bandwidth of the SNAP I/O solution compared to the bandwidth of the remote CPU has improved by *24%* for OS testing and by *9.2%* for docker testing.

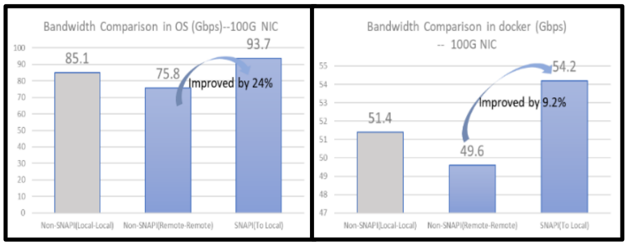

Figure 6: OS/Docker bandwidth (in µs) of various configurations; local CPU, remote CPU and SNAP I/O

UPI Utilization

UPI traffic exists because the CPUs are communicating tasks to each other, constantly working to keep up with user requests. SNAP I/O relieves the UPI of additional overhead by supplying a direct path to both CPUs that doesn't require UPI traversing, therefore freeing up UPI bandwidth. It should come as no surprise that SNAP I/O UPI traffic loading utilization is as low as 7%, while standard riser UPI traffic loading utilization is at 63%.

Non-SNAPI(Remote-Remote)	63%
SNAPI(To Local on receive side)	7%

Figure 7: Comparison of UPI traffic loading percentages

CPU Utilization

While iperf was running for latency/bandwidth testing, the CPU utilization was monitored. As we can see in Figure 8, the SNAP I/O and Non-SNAPI *sender-remote* utilization are identical, so SNAP I/O did not have any impact here. However, the *receiver-remote* utilization underwent a significant improvement, seeing the Non-SNAPI configuration reduce from 55% use to 32% use when configured with SNAP I/O. This is due to the even distribution of TCP streams reducing the average cache miss count on both CPUs.

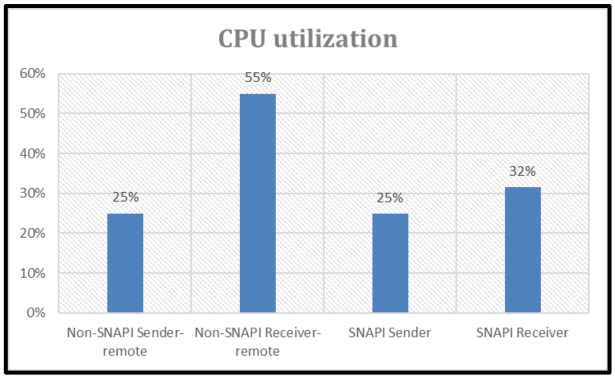

Figure 8: Bar graphs comparing CPU utilization of sender and receiver remotes for non-SNAP I/O and SNAP I/O configurations

Who Will Benefit from SNAP I/O

Using SNAP I/O to improve latency is most useful when the total cost of ownership (TCO) is priority, while maximum bandwidth and card-level redundancy are not. Customers using a 100GbE NIC that need more than 50Gb/s per CPU, or require two-card redundancy, may consider using a two-card solution to achieve the same latency. SNAP I/O should be used in environments where low latency is a priority and single-NIC bandwidth is unlikely to be the bottleneck. Environments such as containers and databases will thrive with SNAP I/O configured, whereas virtualization environments are not yet compatible with the SNAP I/O riser.

Conclusion

Dual-socket servers using a Non-SNAP I/O riser configuration may suffer from unbalanced I/O or a higher TCO. Having data travel from the remote socket across the UPI channel to reach the NIC introduces additional overhead that can degrade performance.

SNAP I/O solution provides an innovative riser that allows data to bypass the UPI channel, achieving a direct connection to a single NIC for two CPUs. As seen throughout this tech note, using a direct connection will deliver higher network bandwidth, lower latency, lower CPU utilization and lower UPI traffic. Additionally, the SNAP I/O solution is more cost-effective than purchasing a second NIC, cable and switch port.

PowerEdge DfD Repository
For more technical learning

Contact Us
For feedback and requests

Follow Us
For PowerEdge news

Copyright © 2020 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries