Maximize 40GbE Bandwidth with RDMA over Converged Ethernet (RoCE)

Background

With 4K fast becoming the standard for digital projection in movie theaters, and TVs also now supporting 4K, consumer demand for higher-resolution content is on the rise. As a result, more content is being shot and edited in both 4K and 8K resolutions. This development in turn has created a need for high-bandwidth 4k and 8k video workflows with throughput requirements that exceed what is possible with 10 Gigabit Ethernet.

While 4K and 8K footage can be compressed for storage, media files are typically streamed uncompressed for editing, rendering and special effects work. The demand to support these high-bandwidth video streams is currently driving heavy interest in 40 Gigabit Ethernet.

Do all 40GbE products perform the same?

It is important to note that throughput is at least partially a function of latency, or the lack thereof. In order to achieve the 40Gb/s throughput required to support multiple uncompressed streams of 4K or 8K video, latency must be carefully managed since data transfer from computer memory through the CPU and Ethernet TCP/IP stacks introduces significant overhead.

While a latency effect was prevalent with 10GbE technology, most users were still able to achieve 800-900MB/s of throughput via TCP—enough for almost a full 4k-Full DPX stream. With 40GbE, however, the need for latency management increases exponentially. Without proper understanding of this limitation, a user expecting full throughput will be disappointed by the number of video streams the system can support, not to mention the discovery that their 40GbE connection lacks sufficient bandwidth for uncompressed 8K video.

How can I manage latency with 40GbE?

Remote Direct Memory Access (RDMA) is a feature designed to enable memory-to-memory data transfers. It has the benefit of eliminating the intermediate step of utilizing the CPU to transfer data, which decreases latency (and, as a result, increases throughput) while freeing up the CPU to conduct other tasks. There are actually two RDMA iterations on the market: RDMA over Converged Ethernet (RoCE) and iWARP. In addition to bypassing the CPU during data transfers, the more common version, RoCE, also bypasses the TCP/IP network stacks to further reduce latency and enhance throughput (see Figure 1). The second version, iWARP, operates on similar principles as RoCE, but adds several crucial nanoseconds of latency since it does not bypass the TCP/IP stack when transmitting data.

In a study conducted by NASA's Advanced Supercomputing Division (Boscia & Sidhu, 2013) comparing 40GbE RDMA to 40GbE TCP, a sustained rate of 21Gb/s was the best case scenario for a 40GbE card with a measured 6-7 microseconds of latency over TCP. A 40GbE card configured with RoCE, in contrast, dropped latency below 1 microsecond. This yielded a sustained rate of up to 39Gb/s — a nearly 100% performance
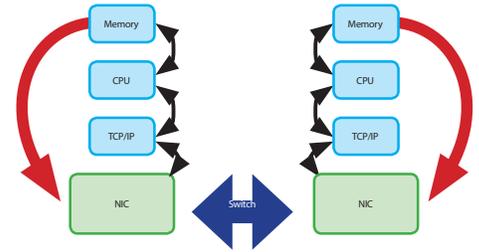

Figure 1(Red = RoCE transfer; black = TCP transfer)

improvement. What is the practical upshot of this increased bandwidth utilization? The ability to enhance 4K or 8K video workflow efficiency by supporting additional streams of video.

How do I implement RDMA/RoCE?

Regardless of how familiar you may be with RDMA/RoCE, the actual implementation and utilization of the protocol is not complicated and requires no major infrastructure upgrades. Put simply, implementing RoCE for applications such as uncompressed 8K data transfer requires a 40GbE NIC with RoCE drivers, a DCB Ethernet switch and storage/servers featuring one of the numerous file transfer protocols that support RoCE (NFS RDMA and SMB RDMA, for example). As long as these requirements are met, taking full advantage of 40GbE bandwidth becomes relatively simple.

Video Format	Bandwidth per stream	Streams per 40GbE TCP	Streams per 40GbE RoCE
4K DPX	6.4Gb/s	3	6
4K-Full DPX	9.2Gb/s	2	4
4K-Full EXR	14.7Gb/s	1	2
8K DPX	23.4Gb/s	0	1
8K EXR	28.2Gb/s	0	1

The impact of RoCE for 40GbE vs. TCP in the number of supported video streams

Maximize 40GbE Bandwidth with RDMA over Converged Ethernet (RoCE)

Technical Features

- Single– and dual–port configurations (Quad-port available in 2016)
- QSFP+ connectors (QSFP modules included with –000 models)
- Up to 40Gb/s throughput per port
- High-performance x8 PCIe 3.0 bus
- Low profile form factor
- Driver support for Windows® , Linux and Mac® operating systems (Mac support coming in 2016)
- RDMA over Converged Ethernet (RoCE) enables industry-leading low-latency (1us MPI ping latency) and decreases CPU utilization
- TCP/UDP/IP hardware-based stateless offloads
- SR-IOV technology dedicates adapter resources for VMs within servers
- Guaranteed bandwidth and low-latency services
- Energy Efficient Ethernet enables lowest power draw on the market for a 40GbE NIC
- Hardware-based I/O virtualization

ATTO FastFrame 40GbE NICs with RoCE

ATTO Technology's single- and dual-port 40GbE NICs feature industry-leading latency management with RoCE. With FastFrame 40GbE NICs, users can expect significantly reduced CPU utilization/latency and uncompromised 40GbE bandwidth. Given ATTO's long experience in developing solutions optimized for media & entertainment, our FastFrame 40GbE NICs provide the premier Ethernet connectivity option for demanding 4k and 8k video workflows.

About ATTO Technology

With a focus toward markets that require higher performance, ATTO manufactures Host Bus Adapters, RAID adapters, network adapters, storage controllers, Thunderbolt™-enabled Desklink™ devices, switches and software. ATTO solutions provide a high level of connectivity to interfaces including Fibre Channel, SAS, SATA, iSCSI, Ethernet, FCoE and Thunderbolt. ATTO distributes its products worldwide through Original Equipment Manufacturers (OEMs), Systems Integrators and Value Added Resellers (VARs). www.attotech.com