WHITE PAPER
Internet Wide Area RDMA Protocol (iWARP)
NetEffect 10 gigabit Ethernet
Server Cluster Adapters
Technical and High-Performance Computing

Intel® Ethernet 10 Gigabit iWARP LAMMPS Performance Study

EXECUTIVE SUMMARY

The *Intel® iWARP Ethernet Performance Series* of test results intends to help close the gap between real user requirements and the micro-benchmarks promoted by other RDMA vendors. Each paper in the series demonstrates the real-world performance of Intel iWARP on an industry standard application.

This paper reports on LAMMPs performance testing performed by the Research Computing and Cyberinfrastructure unit of Information Technology services at Penn State.

Julie Cummings
Intel Corporation

Introduction

RDMA enables direct, zero-copy data transfer between RDMA-capable server adapters and application memory, removing the need in Ethernet networks for data to be copied multiple times to operating system data buffers. The mechanism is highly efficient and eliminates the associated processor-intensive context switching between kernel space and user space. HPC applications can therefore reduce latency and perform message transfer very rapidly and consistently by directly delivering data from application memory to the network.

Both iWARP and InfiniBand use RDMA and a common API for HPC applications, however iWARP enables the use of RDMA over the familiar Ethernet fabric. Because iWARP runs over Ethernet TCP/IP, it enables both application and management traffic to operate over a single wire.

This paper reports on LAMMPs performance testing performed by the Research Computing and Cyberinfrastructure unit of Information Technology services at Penn State to identify how well iWARP fabrics support workloads on widely used high performance computing applications compared to InfiniBand.

iWARP Features

Unlike InfiniBand, iWARP is an extension of conventional Internet Protocol (IP), so standard IT management tools and processes can also be used to manage the traffic and resources associated with iWARP, which implements the following key performance features:

- **Kernel-Bypass:** Enabling applications to interface directly to the Ethernet adapter removes the latency of the OS and the expensive CPU context switches between kernel-space and user-space.
- **Direct Data Placement:** Writing the data directly into user space eliminates the need for wasteful, intermediate buffer copies, thus reducing processing latency and improving memory bandwidth.
- **Transport Acceleration:** The TCP/IP and iWARP protocols are accelerated in silicon vs. host software stacks, thereby freeing up valuable CPU cycles for application compute processing.

iWARP Benefits

HPC applications can use iWARP technology with NetEffect™ Ethernet Server Cluster Adapters from Intel to provide a high-performance, low-latency Ethernet-based solution. By making Ethernet networks suitable for these high-performance clustering implementations, iWARP provides a number of benefits:

- **Fabric consolidation.** With iWARP technology, LAN and RDMA traffic can pass over a single wire. Moreover, application and management traffic can be converged, reducing requirements for cables, ports, and switches.
- **IP-based management.** Network administrators can use standard IP tools to manage traffic in an iWARP network, taking advantage of existing skill sets and processes to reduce overall cost and complexity.
- **Native routing capabilities.** Because iWARP uses Ethernet and the standard IP stack, it can use standard equipment and be routed across IP subnets using existing network infrastructure.
- **Existing switches, appliances, and cabling.** The flexibility of using standard TCP/IP Ethernet to carry iWARP traffic means that no changes are required to Ethernet-based network equipment.

iWARP vs. Infiniband: LAMMPS

LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) is a classical molecular dynamics code that models an ensemble of particles in a liquid, solid, or gaseous state. It can model atomic, polymeric, biological, metallic, granular, and coarse-grained systems using a variety of force fields and boundary conditions. The application can model systems with only a few particles up to several billion.

In the most general sense, LAMMPS integrates Newton's equations of motion for collections of atoms, molecules, or macroscopic particles that interact via short- or long-range forces with a variety of initial and/or boundary conditions. For computational efficiency, LAMMPS uses neighbor lists to keep track of nearby particles. The lists are optimized for systems with particles that are repulsive at short distances, so that the local density of particles never becomes too large. On parallel machines, LAMMPS uses spatial-decomposition techniques to partition the simulation domain into small 3D sub-domains, one of which is assigned to each processor. Processors communicate and store "ghost" atom information for atoms that border their sub-domain. LAMMPS is most efficient (in a parallel sense) for systems whose particles fill a 3D rectangular box with roughly uniform density.

LAMMPS is designed to be easy to modify or extend with new capabilities, such as new force fields, atom types, boundary conditions, or diagnostics. LAMMPS is a freely-available open-source code, distributed under the terms of the GNU Public License[1]. The current version is written in C++. Earlier versions were written in F77 and F90. LAMMPS was originally developed under a US Department of Energy (DOE) Cooperative Research and Development Agreement between two DOE labs and three companies. It is distributed by Sandia National Labs[2].

LAMMPS runs efficiently on single-processor desktop or laptop machines, but it is designed for parallel computers. It will run on any parallel machine that compiles C++ and supports the MPI[3] message-passing library. This includes distributed- or shared-memory parallel machines and Beowulf-style clusters.

For more information, see the LAMMPS FAQ page[4].

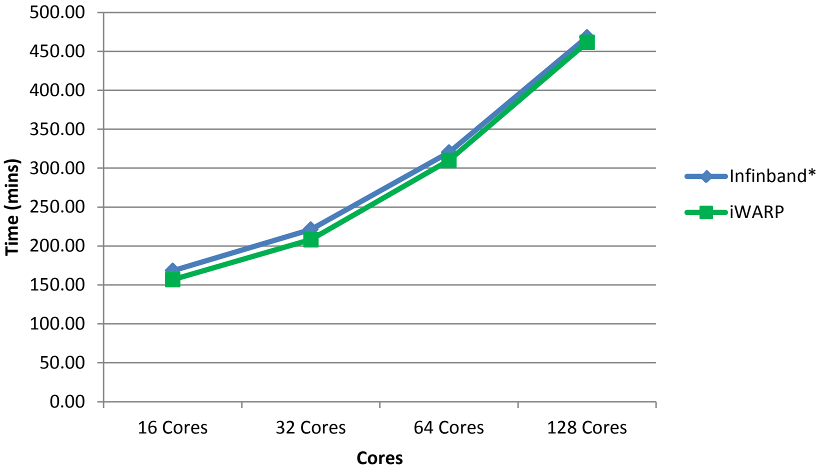

Figure 1. LAMMPS iWARP versus InfiniBand* performance-testing results (lower y-axis figures are better).

Test Scenario

The lithium-ion batteries used in cell phones and laptop computers are based on a liquid electrolyte in which a lithium salt is dissolved, and lithium is the cation that is transferred across the electrolyte during charge and discharge. Replacing the liquid electrolyte with a polymer based "solid" electrolyte, termed "solid polymer electrolyte" offers advantages in weight, size, flexibility, safety, and end-of-life disposal. However, the conductivity of these electrolytes falls short of required standards. The study of cation transport in solid polymer electrolytes is very important for overcoming this challenge.

While experimental techniques provide information on the diffusion coefficient, polymer segmental relaxation, and the content of mobile ions, it is difficult to determine a transport mechanism from these measurements. This testing uses molecular dynamics simulation to study ion transport and backbone mobility of a polyethylene oxide-based single-ion conductor for potential lithium ion battery application. In single-ion conductors, or ionomers, the anion is incorporated in the polymer chain. The conductivity then arises exclusively from the cation, which can eliminate unwanted buildup of anions on the electrodes. The simulation contains 27 molecules with a total number of atoms close to 6,000. Although this is a modest size, observation of cation dynamics into the diffusive regime requires simulation runs up to 500 ns, depending on the cation identity, the anion identity, and the temperature.

Conclusion

By providing realistic application performance instead of micro-benchmark test results, this report illustrates the danger of relying solely on synthetic benchmarking when evaluating networking options. HPC workloads behave much differently than, for example, a half round-trip latency test: using multiple connections in a switched environment with non-uniform I/O patterns.

The real-world application results shown in this report show the viability of iWARP Ethernet as an alternative to discrete, proprietary fabrics for HPC workloads. The fundamental advantages of a converged Ethernet network combined with easier IP-based management and native routing capability make iWARP a compelling solution for HPC use cases.

TEST ENVIRONMENT

All tests were performed by the Research Computing and Cyberinfrastructure unit of Information Technology services at Penn State.

The application software under test was LAMMPS 15 Jan 2010. (The results are shown in Figure 1.)

The test environment consisted of the following:

Servers

- Dell PowerEdge* R710 Server
- Two Intel® Xeon® processors X5560
- 48 GB RAM

Network Adapters

- 10 Gbps iWARP-enabled NetEffect Ethernet Server Cluster Adapter from Intel
- Mellanox Connect-X MT26428 QDR InfiniBand* Host Channel Adapter

System Software

- Red Hat Enterprise Linux* 5.6
- OpenFabrics Enterprise Distribution* 1.5.2
- OpenMPI 1.4.2

Switches

- iWARP: Arista 7148SX* with Jumbo Frames enabled
- InfiniBand: Mellanox MTS3600*

For more information on Intel® iWARP, please visit:

www.intel.com/go/ethernet

[1]1 http://www.gnu.org/copyleft/gpl.html.

2 http://www.sandia.gov/.

3 http://www-unix.mcs.anl.gov/mpi.

4 http://lammps.sandia.gov/FAQ.html.

IINFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Printed in USA PDF/TAR/SWU ♻ Please Recycle 326728-001US