The Applicability of the Universal HP-MPI to MPP LS-DYNA on Linux Platforms

Authors:

Yih-Yih Lin

Correspondence:

High Performance Computing Division
Hewlett-Packard Company
200 Forest Street
Marlborough, MA 01752
U.S.A.
1-404-774-5278
yih-yih.lin@hp.com

Keywords:
Interconnect
MPI
Performance

ABSTRACT

In addition to its parallel algorithm, MPP LS-DYNA for cluster computing derives its parallelism from the MPI library and the interconnect. MPI has always had a standard since its inception. On the other hand, there are diverse interconnects and switches in the market, each with unique low-level interface and performance considerations. This variety of interconnects has resulted in a difficult support burden for LSTC, the software developer, and also resulted in inconvenience for MPP LS-DYNA users. To address this difficulty in portability HP-MPI was invented. HP-MPI has enabled a single MPP LS-DYNA executable to work on most prevailing interconnects on a given Linux hardware platform, HP or non-HP. In this paper, the universal applicability of HP-MPI to diverse interconnects on Linux platforms will be verified by the MPP LS-DYNA simulation on the well-known 3-Vehicle-Collision model. Furthermore, it will be shown that the performance of HP-MPI is on a par with, and often better than, other MPI libraries. HP-MPI is ensured with the highest quality because every one of its release is qualified with a suite of 2000 tests. Thus we can safely claim that HP-MPI, being a supported product and requiring no end-users licensing, is the only MPI that MPP LS-DYNA users need on most Linux platforms.

INTRODUCTION

MPP LS-DYNA is an application based on Message Passing Interface (MPI). MPI has always had a standard since its inception. As shown in the author's two previous papers [1, 2], the performance of MPP LS-DYNA in a cluster is a function of computation and communication costs. The communication cost in MPP LS-DYNA is primarily determined by latency of the cluster's interconnect. Presently, diverse interconnects and switches with various latencies are offered by interconnect manufacturers. Although MPI implementations are provided by those manufacturers, each of their implementations is unique and not portable because each interconnect has a unique low-level (OS) interface, i.e. driver. LSTC is then forced to build and test MPP LS-DYNA with several different MPI implementations on the same Linux hardware platform, which has resulted in a difficult support burden for LSTC. HP-MPI was invented to help solve this difficulty in supporting diverse interconnects for MPP LS-DYNA , as well as other MPI applications.

The design goals of HP-MPI are two folds:

1. To enable a single application executable, like MPP LS-DYNA, to work on most prevailing interconnects on a given Linux hardware platform, HP or non-HP, thus making applications more portable.
2. To deliver as much of the performance of the underlying interconnect as possible.

In this paper, the well-known 3-Vehical-Collision model, as described in the URL http://www.topcrunch.org, is used to verify if the two goals have been achieved. The 3-Vehicle-Collision was cut short to 100 milliseconds from 150 milliseconds to reduce the author's investigation time. And the MPP LS-DYNA Version 970.5434a is used.

A Single MPP LS-DYNA Executable for Diverse Interconnects

The HP-MPI distribution for a particular Linux hardware platform supports all of the interconnect networks in a single (shared) library. As such, HP-MPI, used with MPP LS-DYNA, has several advantages over other MPI Implementations: As aforementioned, HP-MPI helps the portability of MPP LS-DYNA in Linux, where diverse interconnects are available. A single MPP LS-DYNA/HP-MPI executable will work on most prevailing interconnects for a given Linux hardware platform.

- Users of MPP LS-DYNA/HP-MPI never have to build the MPI themselves since HP-MPI is distributed as a library on a given hardware platform, while many other MPIs require users to build the library from the source.
- HP-MPI requires no end-user licensing and is supported by HP via the application developer.
- HP-MPI supports almost all prevailing interconnects, including:
  - InfiniBand VAPI or IT-API
  - Myrinet GM
  - Quadrics ELAN3/4
  - Any RDMA enabled GigE
  - GigE TCP/IP

Although the user can specify the kind of interconnect explicitly, HP-MPI can also automatically detect it. If a cluster has only a single interconnect, HP-MPI chooses the corresponding driver. On the other hand, if a cluster has multiple interconnects, HP-MPI detects and chooses the best interconnect available on each host, thus allowing multiple interconnects to coexist.

The author has successfully tested MPP LS-DYNA/HP-MPI releases with the interconnects listed above on Intel Itanium2 and AMD Opteron Linux clusters. For example, shown in Figure 1 are the timing results of MPP LS-DYNA/HP-MPI on the 3-Vehicle-Collision model on the cluster of HP rx1620, a 1.6 GHz Intel Itanium2 server, and on the cluster of HP DL145, a 2.2 GHz AMD Opteron server, with the InfiniBand IT-API and GigE TCP/IP.

Performance Comparison on HP-MPI and other MPIs

HP-MPI strives to deliver as much performance of the underlying interconnect as possible. HP-MPI works closely with interconnect providers to ensure that the best performance is delivered with their products.

In order to support different interconnects with the same code base, in HP-MPI a generic low-level abstraction layer is dynamically loaded to allow the MPI library to interface with the device (interconnect). This approach of HP-MPI has a device-layer interface further away from the user-interface than many other MPI implementations, such as MPICH; the approach of HP-MPI is believed to be more efficient than those of many other MPI implementations.

The focus of network-specific MPI implementation is generally on the inter-host, but current clusters generally comprise of hosts of multiple shared-memory processors.
Shared-memory transfer performance remains important to an MPI implementation,

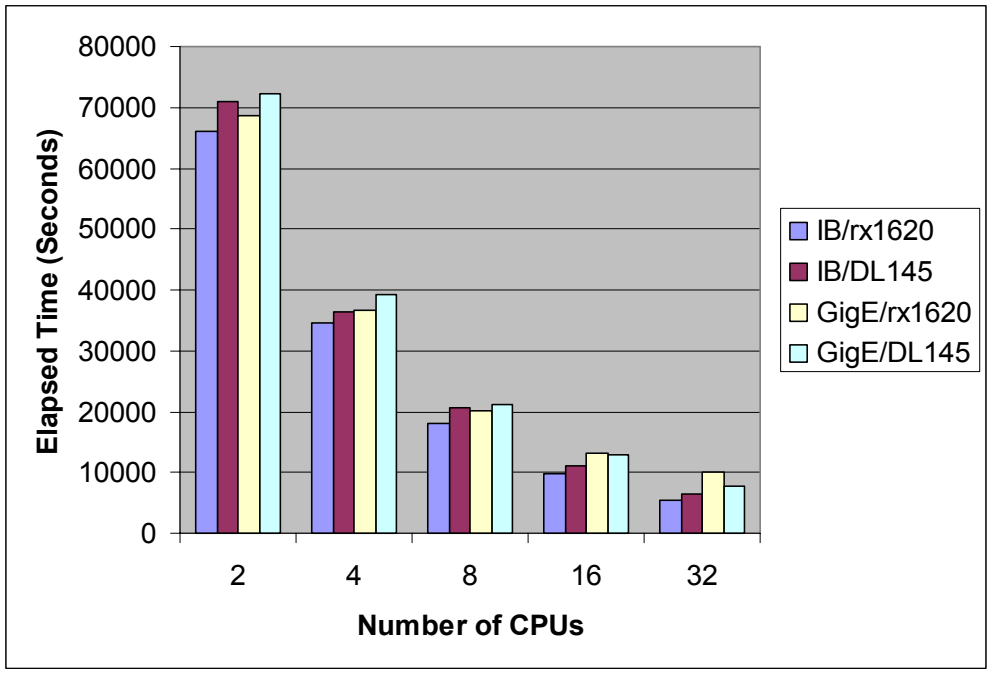

Figure 1. HP-MPI Timings with Infiniband and GigE on Itanium2 rx1620 server and Opteron DL145 server

and HP-MPI has a performance advantage in this area compared to many other MPI
implementations. Among several optimization techniques that HP-MPI has used are assembly coding for enhancing the interconnect's bandwidth and the so-called "fast path" for reducing the interconnect's latency.

To investigate whether such optimization efforts are fruitful for MPP LS-DYNA, the author has measured the 3-Vehicle-Collision timings on the same GigE cluster of Intel Itanium2, HP rx1620 (Figure 2) and on the same GigE cluster of AMD Opteron, HP DL145 (Figure 3). These two results show that from 2- to 32-CPUs, HP-MPI is faster in most cases than LAM-MPI and MPICH.

The Universality and the Quality of HP-MPI

In addition to Intel Itanium and AMD Opteron, HP-MPI supports both Linux platforms of 32- and 64-bit Intel Xeon. Not only HP-MPI supports diverse interconnects with a single library on those Linux platforms, but it also ensures that a single library supports the many flavors and versions of Linux, e.g. Red Hat 2.1, 3.0, 4.0, and SuSe SLES 9.x. Furthermore, many other server types and OS's, including PA-RISC/HP-UX, ALPHA/TRU64, Itanium/HP-UX are supported by HP-MPI with identical user interface.

HP-MPI is ensured with highest quality. Every release of HP-MPI is qualified with a suite of 2000 tests.

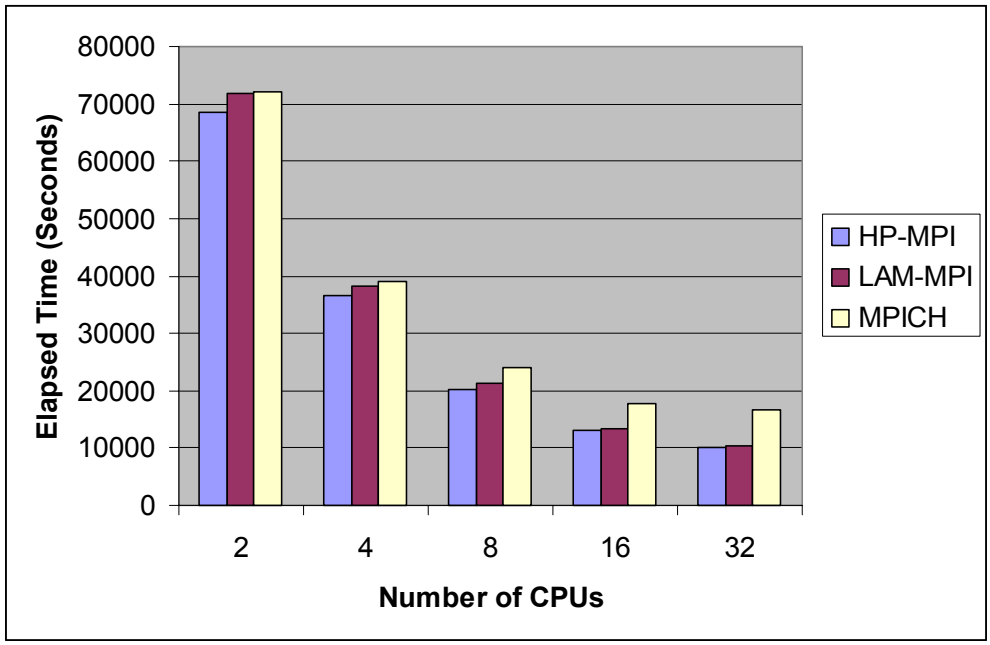

Figure 2. Timings of HP-MPI, LAM-MPI and MPICH on a GigE cluster of Intel Itanum2 rx1610 server

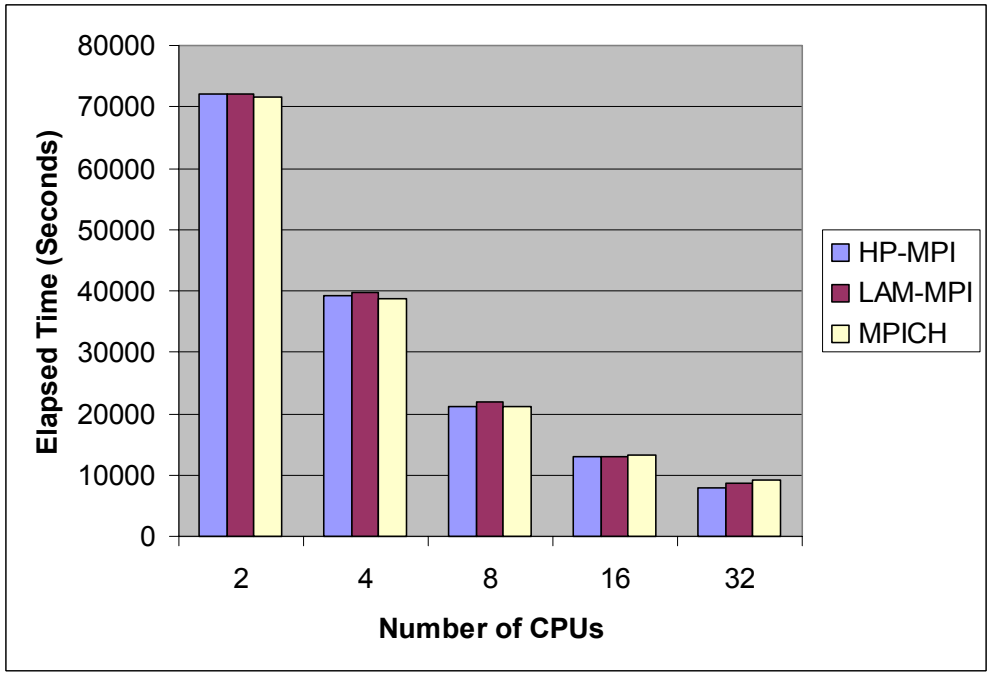

Figure 3. Timings of HP-MPI, LAM-MPI and MPICH on a GigE cluster of AMD Opteron DL145 server

Summary and Conclusions

We have established here that a single MPP LS-DYNA executable, using HP-MPI, works on diverse interconnects and switches, and with better performance than other MPI implementations. Every HP-MPI release is also ensured with the highest quality. Therefore, it is safe to claim that HP-MPI is the only HP-MPI that MPP LS-DYNA users need for most Linux platforms.

References

1. Yih-Yih Lin, "Determining the MPP LS-DYNA Communication and Computation Costs with the 3-Vehicle Collision Model and the InfiniBand Interconnect," 8th International LS-DYNA Users Conference 2004, 12-27—12-31.
2. Yih-Yih Lin, "A Quantitative Approach for Determining the Communication and Computation Costs in MPP LS-DYNA Simulations," FEA News, February 2004.