OpenSHMEM Activities @ ORNL

Manjunath Gorentla Venkata for the Languages Team and Collaborators
Team

• ORNL
  – Neena Imam (Director)
  – Manjunath Gorenla Venkata / Manju (Team Lead)
  – Mathew Baker
  – Swen Boehm
  – Swaroop Pophale
  – Thomas Naughton
  – Eric Mitchell

• DoD
  – Nick Park
  – Bryant Lam

• NVIDIA
  • Sreeram Potluri
  • Anshuman Goswami
  • CJ Newburn

• Florida State University
  • Weikuan Yu and Students

• Paratools
  • Sameer Shende
  • John Linford

Team Website: http://www.csm.ornl.gov/openshmem/index.html
OpenSHMEM: Specification, Research, and Implementation

Overview

- Languages team at ORNL drives the development of the OpenSHMEM programming model and its ecosystem. The effort is focused in four areas:
  - Specification development
  - OpenSHMEM implementation
  - Research and exploratory topics
  - Outreach

Specification

- OpenSHMEM 1.4 & Next specification: Developed OpenSHMEM thread safety proposal, and merged request proposal

R&D

- OpenSHMEM-X: Added various capabilities including alltoall and alltoallv collective operations, multithread support, OpenSHMEM contexts, explicit RMA operations, and merged request handles
- OpenSHMEM-X uses PMIx for scalability, performance and stability
- OpenSHMEM Benchmarks: Developed various OpenSHMEM application kernels and benchmarks including multithreaded Graph 500, latency, bandwidth, SSCA, GUPs, SSSP, and YCSB
- SHMEMCache – Memcached implementation, which uses OpenSHMEM for communication
- NVSHMEM – OpenSHMEM for GPUs connected via PCIe, NVLINK or InfiniBand
- Improvements to TAU ecosystem for OpenSHMEM programmers with support for Callsites, OTF2 and simplified user interface
- Design, evaluation, and results are available as papers

Software Artifacts

- OpenSHMEM-X
- ORNL OpenSHMEM Benchmark Suite (OSB)
- OpenUCX: Unified Communication X
- NVSHMEM: OpenSHMEM for GAS Systems
- SHMEMCache: OpenSHMEM based Memcache

Outreach Activities

Team Website: http://www.csm.ornl.gov/openshmem/index.html
Team GitHub: https://github.com/orgs/ornl-languages/
4.3 Achieving Higher Bandwidth for Smaller Messages with merged request based RMA Operations and Multiple Threads

In this experiment, we are measuring the aggregated bandwidth achieved with multiple threads and merged request based RMA operations. Also, for the experiments a modified version of the OSU OpenSHMEM benchmarks, that was extended to support multi-threading with OpenMP, was used. The experiments were conducted on two nodes with one PE placed on each node.

Figure 5 shows the aggregated bandwidth as a function of the message size. The message size is plotted on the x-axis, and the bandwidth is plotted on the y-axis. All experiments but the single threaded experiment saturate the network with a sufficiently large message size. The plot for one and two threads shows a dip in the bandwidth when the message size exceeds 8 kB. This is due to a protocol switch in the networking layer. If more than 2 threads are used, the impact of the protocol switch is mitigated by the additional threads. From the figure, we can observe that the merged request based RMA operations help to achieve higher bandwidth for multithreaded OpenSHMEM.

4.4 Communication / Computational overlap

In the experiment, we measure the overlap achieved by the merged request based RMA operations using a modified version of the COMB benchmark [7]. The modifications include support for OpenSHMEM and merged request based RMA operations. The benchmark supports two modes, Post-Work-Wait and Post-Work-Poll modes of operation. The Post-Work-Wait is used to measure the largest uninterrupted computation-communication overlap duration, and we use this mode for the experiments in the paper.

The benchmark has three steps:
OpenSHMEM Benchmarks Developed

Developed OpenSHMEM benchmarks
- Multithreaded Latency and Bandwidth
- Multithreaded SSCA#1
- Multithreaded and Context-based GUPs
- Multithreaded Graph 500
- OpenSHMEM based SSSP (Bellman-Ford and Dijkstra)
- OpenSHMEM based YCSB
Acknowledgements

• Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy.

• Department of Defense