Portable SHMEMCache: A High-Performance Key-Value store on OpenSHMEM and MPI

Achievement: Developed high-performing Memcached, that leverages high-performing programming models such as OpenSHMEM and MPI and HPC networks such as Gemini/Aries and InfiniBand.

Significance and Impact: This work enables applications to use Memcached, a distributed caching framework, on HPC systems such as Titan and Rhea.

Research Details:

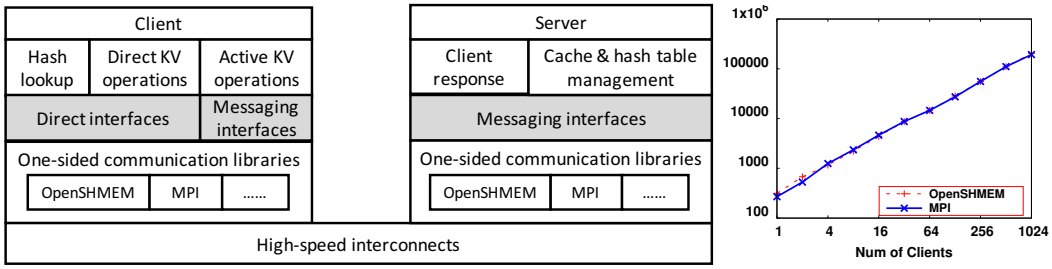
- Developed SHMEMCache an implementation of Memcached using OpenSHMEM and MPI.
- Demonstrated the advantages of using one-sided semantics, high-performing network, and HPC systems for Key-Value workloads.
- Systematically evaluated SHMEMCache on Titan to demonstrate the performance and scalability advantages.

Sponser/Facility: Work was performed with support from ORNL and DOD

PI and affiliation: Manjunath Gorentla Venkata from CSMD – Oak Ridge National Laboratory

Team:, Manjunath Gorentla Venkata (ORNL), Neena Imam (ORNL), Huansong Fu (FSU) and Weikuan Yu (FSU)

Publication: Huansong Fu, Manjunath Gorentla Venkata, Neena Imam, and Weikuan Yu . In Proceedings of the Fourth OpenSHMEM Workshop, Annapolis, MD, USA, August 2017.



(a) The Communication architecture of SHMEMCache (b) KV operation throughput with varying number of clients on Titan

Overview:

SHMEMCache is an implementation of Memcached using OpenSHMEM and MPI. This can leverage the OpenSHMEM one-sided operations (SHMEM_PUT and SHMEM_GET) for the KV operations such as SET and GET. Using OpenSHMEM's one-sided operations will take advantage of RDMA capabilities to transfer data, avoiding multiple copies between the user buffer and system buffer, and also the need for server involvement. SHMEMCache also provides novel solutions for data consistency issue, carries out cache management in a coarse-grained and lightweight manner, and scales well to more than one thousand machines. Further, SHMEMCache was extended to use MPI one-sided operations, in addition to OpenSHMEM operations.Our experimental results show that SHMEMCache can deliver very low latency

and high throughput KV operations at scale while still ensuring data consistency. The details of the design, implementation, and evaluation can be found in the publication.