

Scalable Memcached design for InfiniBand Clusters using Hybrid Transports

Jithin Jose¹, Hari Subramoni¹, Krishna Kandalla¹, Md. Wasi-ur-Rahman¹, Hao Wang¹,
Sundeepp Narravula², and Dhabaleswar K. Panda¹

¹ *Department of Computer Science and Engineering,
The Ohio State University*

{jose, subramon, kandalla, rahmanmd, wangh, panda}
@cse.ohio-state.edu

² *Yahoo! Inc.*

Sunnyvale, California

{sundeeppn}
@yahoo-inc.com

Memcached is a general-purpose key-value based distributed memory object caching system. It is widely used in data-center domain for caching results of database calls, API calls or page rendering. An efficient Memcached design is critical to achieve high transaction throughput and scalability. Previous research [1] in the field has shown that the use of high performance interconnects like InfiniBand can dramatically improve the performance of Memcached. The Reliable Connection (RC) is the most commonly used transport model for InfiniBand implementations. Using this transport exclusively, however, has been shown to have scalability issues due to high memory consumption per connection. This requirement may not be favorable for middlewares like Memcached, where the server is required to serve thousands of clients. The Unreliable Datagram (UD) transport could potentially offer higher scalability, but has several other limitations, which need to be efficiently handled.

In this context, we introduce a hybrid transport model which takes advantage of the best features of RC and UD to deliver scalability and performance higher than that of a single-transport. To the best of our knowledge, this is the first effort aimed at studying the impact of using a hybrid of multiple transport protocols on Memcached performance. We present comprehensive performance analysis using micro-benchmarks, application benchmarks and realistic industry workloads. Our performance evaluations reveal that our Hybrid transport delivers performance comparable to that of RC, while maintaining a steady memory footprint. Memcached Get latency for 4 byte data size, is $4.28\mu s$ and $4.86\mu s$ for hybrid and RC transports, respectively. This represents a factor of twelve improvement over the performance of SDP. In evaluations using Apache Olio benchmark with 1,024 clients, Memcached execution time using RC, UD and hybrid transports are 1.61, 1.96 and 1.70 seconds, respectively. Further, our scalability analysis with 4,096 client connections reveal that our proposed hybrid transport achieves good memory scalability.

REFERENCES

- [1] J. Jose, H. Subramoni, M. Luo, M. Zhang, J. Huang, M. W. Rahman, N. S. Islam, X. Ouyang, H. Wang, S. Sur, and D. K. Panda, "Memcached Design on High Performance RDMA Capable Interconnects," in *International Conference on Parallel Processing (ICPP)*, Sept 2011.