Abstract

Today, a cluster computing has become a mainstream of high-performance computing platform as cost-effective and practical resource and is replacing conventional super computers. In general, a high-end cluster computing is deployed on PCs gathered in a small space like racks and they are connected to each other by System Area Network (SAN), such as Myrinet, or high-performance Ethernet.

RHiNET is an original network to provide such a high-end cluster computing by utilizing surplus computation power of PCs used in offices for daily jobs. RHiNET achieves the high-end cluster computing by supporting both SAN-like high-performance communication and Ethernet-like connectivity.

Martini is a network interface controller developed for RHiNET. To obtain high-performance communication, Martini provides simple RDMA-based communication schemes by hardware. Also, Martini has some experimental features such as low-latency packet sending mechanisms called "On-the-fly (OTF)," a new methodology for cooperation of hardware and software called "Taking Over (TO)." In the research, core hardware logic of Martini and related low-level software has been implemented. Also, a system using Martini has been built up and evaluated. The results of basic performance evaluation show that Martini achieves 470Mbyte/s maximum bidirectional throughput and 1.74 μsec minimum latency. The latency of Martini is comparable with other cutting-edge network controllers and an advantage of hardware implemented RDMA is shown. Also, to validate an effectiveness of TO mechanism, a new communication scheme which uses it has been implemented. The evaluation results indicate that TO mechanism makes communication performance better than pure software processing.

Performance of Martini under the system has been evaluated by porting existing cluster system software called "SCore." A low-level communication library of SCore, called "PM," requires message communication but is not provided by Martini. Thus, the message communication has been implemented by software using RDMA-based communication schemes of Martini. As the evaluation results on 16-node RHiNET-2 system, speed-ups were achieved according to number of nodes on an application level. Also, it was found that the latency of message communication is enlarged when the size of system becomes large. This is mainly caused by the implementation with simple RDMA-based communication schemes of Martini.