SUMMARY OF Ph.D. DISSERTATION

School Science for Open and Environmental Systems	Student Identification Number	SURNAME, First name OTSUKA, Tomohiro

Title

A Study on Building Large-scale Cluster Interconnects using VLAN Ethernet

Abstract

PC clusters, which are built by connecting commodity PCs, are the mainstream architecture in recent high-performance computing (HPC) platforms. In particular, systems that adopt highly cost-effective, standard Ethernet for their interconnects become and certainly continue to be the majority among clusters.

On the other hand, most clusters using Ethernet adopt simple trees as their network topologies, since Ethernet cannot include loops in its topology. However, a tree topology has a disadvantage in traffic congestion at its root, which may degrade overall performance especially in large-scale clusters.

VLAN-based routing method, which makes it possible to adopt various topologies including loops in Ethernet by applying IEEE 802.1Q tagging VLAN technology, is a solution to this problem. At present, however, it is difficult to apply the existing VLAN-based routing method to large-scale clusters due to the following two problems; the capability of system software to deal with VLAN tags, and the increase in required number of VLANs.

The goal of this research is to develop a technology for building large-scale clusters using Ethernet by solving these problems. For this purpose, two methods improving the existing VLAN-based routing method are proposed and evaluated in this dissertation. Also, possibility of applying these methods to large-scale clusters as well as comparison of them with other methods is discussed.

The first method called “switch-tagged method” makes it possible to employ VLAN-based routing when communication libraries do not support VLAN tagging, by inserting VLAN tags into Ethernet frames not at hosts but at switches. The second method called “VLAN renaming method” reduces required number of VLANs to the number of ports in a switch or less, by using VLANs only inside switches for determining output ports of frames tagged at input ports. In addition to these proposals, this dissertation also discusses determining factors in the performance of Ethernet with the VLAN-based methods, especially the deadlock problem, as well as methodology of VLAN assignment and selection of routing algorithms on typical topologies.

As a result of performance evaluations using a cluster with 32 hosts and 16 switches, the proposed methods introduce almost no overhead, and topologies built by using the methods achieve over 88% performance of an ideal flat topology in all applications. Moreover, it is ascertained that the proposed methods are applicable to larger networks than the original VLAN-based routing method, through generalizing required number of VLANs to obtain the maximum system scale when typical topologies are adopted by using these methods. In conclusion, it is possible to build large-scale and high-performance clusters using Ethernet by applying the proposed methods.