World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:04, 2016

The Effective Use of the Network in the Distributed Storage

Authors : Mamouni Mohammed Dhiya Eddine

Abstract : This work aims at studying the exploitation of high-speed networks of clusters for distributed storage. Parallel applications running on clusters require both high-performance communications between nodes and efficient access to the storage system. Many studies on network technologies led to the design of dedicated architectures for clusters with very fast communications between computing nodes. Efficient distributed storage in clusters has been essentially developed by adding parallelization mechanisms so that the server(s) may sustain an increased workload. In this work, we propose to improve the performance of distributed storage systems in clusters by efficiently using the underlying high-performance network to access distant storage systems. The main question we are addressing is: do high-speed networks of clusters fit the requirements of a transparent, efficient and high-performance access to remote storage? We show that storage requirements are very different from those of parallel computation. High-speed networks of clusters were designed to optimize communications between different nodes of a parallel application. We study their utilization in a very different context, storage in clusters, where client-server models are generally used to access remote storage (for instance NFS, PVFS or LUSTRE). Our experimental study based on the usage of the GM programming interface of MYRINET high-speed networks for distributed storage raised several interesting problems. Firstly, the specific memory utilization in the storage access system layers does not easily fit the traditional memory model of high-speed networks. Secondly, client-server models that are used for distributed storage have specific requirements on message control and event processing, which are not handled by existing interfaces. We propose different solutions to solve communication control problems at the filesystem level. We show that a modification of the network programming interface is required. Data transfer issues need an adaptation of the operating system. We detail several propositions for network programming interfaces which make their utilization easier in the context of distributed storage. The integration of a flexible processing of data transfer in the new programming interface MYRINET/MX is finally presented. Performance evaluations show that its usage in the context of both storage and other types of applications is easy and efficient.

Keywords : distributed storage, remote file access, cluster, high-speed network, MYRINET, zero-copy, memory registration, communication control, event notification, application programming interface
Conference Title : ICIN 2016 : International Conference on Information Networking
Conference Location : Lisbon, Portugal
Conference Dates : April 14-15, 2016