ABSTRACT OF THE DISSERTATION

Hierarchically Managed Reliable
Distributed Shared Memory Multicomputers

by

G. Janakiraman

Doctor of Philosophy in Computer Science

University of California, Los Angeles, 1997

Professor Yuval Tamir, Chair

Distributed Shared Memory (DSM) Multicomputers offer the advantages of the shared memory programming model on high-performance, scalable hardware. DSM implementations must not limit scalability or introduce performance bottlenecks. The reliability requirements of large DSM multicomputers can only be met using fault tolerance techniques.

Most DSM implementations manage memory and coherency at a single granularity of fixed size pages, resulting in unnecessary network traffic due to false-sharing, or of cache blocks requiring unacceptably large mapping tables to map large address spaces. We propose hierarchical management as a solution where mapping and transfer occur at the block level only when the page is actually shared. When pages are not shared, they incur the same mapping overhead as uniform page-level schemes. Performance evaluation based on detailed execution-driven simulation has shown that hierarchical management significantly outperforms page-level management, and performs nearly as well as block-level management, while using much smaller mapping tables, on real applications.

Error recovery schemes previously proposed for DSM systems checkpoint or log data at every process communication, leading to unnecessarily high checkpointing frequency or logging overhead, and associated network traffic. Such approaches are not suitable for scalable systems intended for communication-intensive applications. Efficiently implementing the lower overhead error recovery approach based on coordinated checkpointing and rollback on DSM multicomputers presents several challenges. We solve these problems and implement this approach for the DSM multicomputer. Our performance evaluation indicates that coordinated checkpointing and rollback incurs less checkpoint traffic than recovery schemes previously proposed for DSM systems.