PARALLEL I/O SOLUTIONS FOR LARGE-SCALE PARTITIONED SOLVER SYSTEMS

By

Jing Fu

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: COMPUTER SCIENCE

Approved:

Christopher D. Carothers, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York

April 2010
(For Graduation May 2010)

With the development of high-performance computing, I/O issues have become the bottleneck for large scale scientific applications. This thesis investigates scalable parallel I/O solutions for one specific kind of applications - massively parallel partitioned solver systems. Typically such systems have synchronized "loops" and will write data in a well defined block I/O format consisting of a header and data portion. Our target use for such an parallel I/O subsystem is *checkpoint-restart* where writing is by far the most common operation and reading typically only happens during either initialization or during a restart operation because of a system failure.

We compare four parallel I/O strategies: 1 POSIX File Per Processor (1PFPP), a synchronized parallel IO library (syncIO), "Poor-Man's" Parallel I/O (PMPIO) and a new "reduced blocking" strategy (rbIO). Performance tests using real CFD solver data from PHASTA (an unstructured grid finite element Navier-Stokes solver) show that the syncIO strategy can achieve a read bandwidth of 6.6GB/Sec on Blue Gene/L using 16K processors which is significantly faster than 1PFPP or PMPIO approaches. The serial "token-passing" approach of PMPIO yields a 900MB/sec write bandwidth on 16K processors using 1024 files and 1PFPP achieves 600 MB/sec on 8K processors while the "reduced-blocked" rbIO strategy achieves an actual writing performance of 2.3GB/sec and *perceived/latency hiding* writing performance of more than 21,000 GB/sec (i.e., 21TB/sec) on a 32,768 processor Blue Gene/L. Further more, syncIO achieves a read performance of 45GB/sec and a write performance of 27 GB/sec with aligned data; rbIO achieves an actual write performance of 18 GB/sec and a perceived write performance of 167 TB/sec on a 131, 072 processor Blue Gene/P.