Electronic Transactions on Numerical Analysis.
Volume 15, pp. 211-224, 2003.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

FACTORS INVOLVED IN THE PERFORMANCE OF COMPUTATIONS ON BEOWULF CLUSTERS*

PAUL A. FARRELL † AND HONG ONG ‡

Abstract. Beowulf (PC) clusters represent ve platform for large scale scientific computations. In this paper, we discuss the effects of some possible configuration, hardware, and software choices on the communications latency and throughput attainable, and the consequent impact on scalability and performance of codes. We compare performance currently attainable using Gigabit Ethernet with that of Fast Ethernet. We discuss the effects of various versions of the Linux kernel, and approaches to tuning it to improve TCP/IP performance.

We comment on the relative performance of LAM, MPICH, and MVICH on a Linux cluster connected by a Gigabit Ethernet network. Since LAM and MPICH use the TCP/IP socket interface for communicating messages, it is critical to have high TCP/IP performance for these to give satisfactory results. Despite many efforts to improve TCP/IP performance, the performance graphs presented here indicate that the overhead incurred in protocol stack processing is still high. We discuss the Virtual Interface Architecture (VIA) which is intended to provide low latency, high bandwidth message-passing between user processes. Developments such as the VIA-based MPI implementation MVICH can improve communication throughput and thus give the promise of enabling distributed applications to improve performance. Finally we present some examples of how these various choices can impact the performance of an example multigrid code.

Key words. cluster computation, gigabit ethernet, TCP/IP performance, virtual interface architecture (VIA), MPI, MPICH, LAM, MVICH, NAS multigrid benchmarks

AMS subject classifications. 65Y05, 68M14, 68M12, 68M20, 68N99, 65Y20

* Received June 27, 2001. Accepted for publication October 19, 2001. Recommended by Ulrich Ruede.
†Department of Computer Science, Kent State University, Kent, Ohio 44242, U.S.A.
‡Department of Computer Science, Kent State University, Kent, Ohio 44242, U.S.A.
This work was supported in part by NSF CDA 9617541, NSF ASC 9720221, NSF ITR 0081324, by the OBR Investment Fund Ohio Communication and Computing ATM Research Network (OCARnet), and through the Ohio Board of Regents Computer Science Enhancement Initiative