An Approach for Hierarchical System Level Diagnosis of Massively Parallel Computers Combined With a Simulation-Based Method for Dependability Analysis

J. Altmann, F. Balbach, A. Hein

Institut für Mathematische Maschinen und Datenverarbeitung (IMMD) III,
Universität Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, Germany

Keywords: massively parallel computers, system level diagnosis, simulation-based analysis, scalable and object-oriented simulation models

Abstract: The primary focus in the analysis of massively parallel supercomputers has traditionally been on their performance. However, their complex network topologies, large number of processors, and sophisticated system software can make them very unreliable. If every failure of one of the many components of a massively parallel computer could shut down the machine, the machine would be useless. Therefore fault tolerance is required. The basis of effective mechanisms for fault tolerance is an efficient diagnosis.
This paper deals with concurrent and hierarchical system level diagnosis for a particular massively parallel architecture and with a simulation-based method to validate the proposed diagnosis algorithm. The diagnosis algorithm is presented and we describe a simulation-based method to test and verify the algorithms for fault tolerance already during the design phase of the target machine.

1. Introduction

1.1 Motivation and Related Work

Usually, the primary focus in the analysis of massively parallel supercomputers has traditionally been on their performance. However, their large number of processors and network components and their sophisticated system software can make them very unreliable. Large numbers of components increase the probability that one out of these numerous components fails. Taking into account, that applications for massively parallel computers require long execution times [7], the overall system would become useless if it fails as soon as one single component becomes faulty. Therefore, the parallel system has to include fault tolerance mechanisms that tolerate failures of subcomponents.