

A Systematic Multi-step Methodology for Performance Analysis of Communication Traces of Distributed Applications based on Hierarchical Clustering

Gaby Aguilera ¹, Patricia J. Teller ¹, Michela Taufer ¹, and Felix Wolf ²

¹ University of Texas-El Paso
El Paso, TX 79968 USA
{maguilera, pteller, mtaufer}@utep.edu

² Forschungszentrum Jülich
52425 Jülich, Germany
f.wolf@fz-juelich.de

Abstract

Often parallel scientific applications are instrumented and traces are collected and analyzed to identify processes with performance problems or operations that cause delays in program execution. The execution of instrumented codes may generate large amounts of performance data, and the collection, storage, and analysis of such traces are time and space demanding. To address this problem, this paper presents an efficient, systematic, multi-step methodology, based on hierarchical clustering, for analysis of communication traces of parallel scientific applications. The methodology is used to discover potential communication performance problems of three applications: TRACE, REMO, and SWEEP3D.