

Topic 2: Performance Prediction and Evaluation (Introduction)

Adolfy Hoisie, Michael Gerndt, Shajulin Benedict,
Thomas Fahringer, Vladimir Getov, and Scott Pakin

Topic Committee

In recent years, a range of novel methodologies and tools have been developed for the purpose of evaluation, design, and model reduction of existing and emerging parallel and distributed systems. At the same time, the coverage of the term "performance" has constantly broadened to include reliability, robustness, energy consumption, and scalability in addition to classical performance-oriented evaluations of system functionalities. Indeed, the increasing diversification of parallel systems, from cloud computing to exascale, being fueled by technological advances, is placing greater emphasis on the methods and tools to address more comprehensive concerns. The aim of the Performance Prediction and Evaluation topic is to bring together system designers and researchers involved with the qualitative and quantitative evaluation and modeling of large-scale parallel and distributed applications and systems to focus on current critical areas of performance prediction and evaluation theory and practice.

The three papers selected for the topic area reflect the broadening perspective of parallel performance involving automatic comparison of performance traces, assessment of the uncertainty in performance prediction through simulation, and performance tuning for NUMA architectures.

The paper "Alignment-Based metrics for Trace Comparison" uses sequence-alignment algorithms to align events across two traces obtained from performance analysis tools to identify which events were added or removed or took a different amount of time to complete.

The uncertainty of Extreme-Scale HPC simulation is assessed in the paper "Validation and Uncertainty Assessment of extreme-Scale HPC Simulation through Bayesian Inference". The paper takes a statistical approach to quantify the uncertainty involved in predicting performance via simulation.

The third paper on "Dynamic thread Pinning for Phase-Based OpenMP Programs" investigate the impact of adjusting thread pinning (affinity) dynamically for each parallel region of an OpenMP program. The authors perform a dynamic analysis of memory accesses and calculate a gain function between pairs of threads, which is in turn used to generate a recursive partitioning of threads on shared memory spaces.