Through the Looking-Glass: From Performance Observation to Dynamic Adaptation

Allen D. Malony
Department of Computer and Information Science
University of Oregon
malony@cs.uoregon.edu

Abstract
Since the beginning of ``high-performance" parallel computing, observing and analyzing performance for purposes of finding bottlenecks and identifying opportunities for improvement has been at the heart of delivering the performance potential of next-generation scalable systems. Interestingly, it is the ever-changing parallel computing landscape that is the main driver of requirements for parallel performance technology and the improvements necessary beyond the current state-of-the-art. Indeed, the development and application of our TAU Performance System over many years largely follows an evolutionary path of addressing measurement and analysis problems in new parallel machines and programming environments. However, the outlook to future parallel systems with high degrees of concurrency, heterogeneous components, dynamic runtime environments, asynchronous execution, and power constraints suggests a new perspective will be needed on the role of performance observation and analysis in respect to tool technology integration and performance optimization methods. The reliance on post-mortem analysis of application-level ("1st person") performance measurements is prohibitive for exascale-class machines because of the performance data volume, the primitive basis for performance data attribution, and the fundamental problem of performance variation that will exist. Instead, it will be important to provide introspection support across the exascale software stack to understand how system ("3rd person") resources are used during execution. Furthermore, the opportunity to couple a global performance introspection capability (a "performance backplane") with online performance decision analytics inspires the concept of an autonomic performance system that can feed back policy-based decisions to guide the computation to better states of execution. The talk will explore these issues by giving a brief retrospective on performance tool evolution, setting the stage for current research projects where a new performance perspective is being pursued. It will also speculate on what might be included in next-generation parallel systems hardware, specifically to make the exascale machines more performance-aware and dynamically-adaptive.

Categories and Subject Descriptors:
C.2.0 [**Computer Systems Organization**]: Performance of systems.

Keywords
High-performance computing; runtime environments; optimization methods

Bio

Dr. Allen D. Malony is a Professor in the Department of Computer and Information Science at the University of Oregon (UO) where he directs parallel computing research projects, notably the TAU parallel performance system project. He has extensive experience in performance benchmarking and characterization of high-performance computing systems, and has developed performance evaluation tools for a range of parallel machines during the last 25 years. His research interests also include computational science and neuroinformatics. Malony was awarded the NSF National Young Investigator award, was a Fulbright Research Scholar to The Netherlands and Austria, and received the prestigious Alexander von Humboldt Research Award for Senior U.S. Scientists by the Alexander von Humboldt Foundation. He is funded by the Department of Energy, the National Science Foundation, and the Department of Defense. Malony is the Director of the UO Neuroinformatics Center and the CEO of ParaTools, Inc., which he founded with Dr. Sameer Shende in 2004.

HPDC'15, June 15–20, 2015, Portland, Oregon, USA.
ACM 978-1-4503-3489-1/15/06.
http://dx.doi.org/10.1145/2749246.2749247