CENTER FOR COMPUTATION & TECHNOLOGY
Interdisciplinary | Innovative | Inventive

LOUISIANA STATE UNIVERSITY

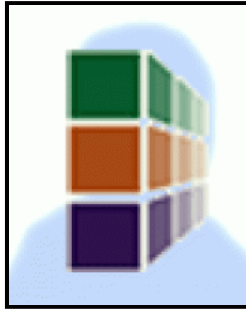### Events

Current Events
Lectures▼
Events Archive▼

CCT Colloquium Series

Predicting Bounds on the Batch Queuing Delay Experienced by User Jobs in Real Time

Rich Wolski, University of California, Santa Barbara

Associate Professor, Computer Science

Johnston Hall 338
October 06, 2006 - 11:00 am

Abstract:

In this talk, we present a new method for providing TeraGrid end-users with real-time predictions of the bounds on queuing delay individual jobs will experience when waiting to be scheduled to a machine partition. Predicting the delay users will experience while waiting for their jobs to be be scheduled is a problem that has been studied both by the academic and commercial HPC communities for some time. Our approach, based on a new statistical methodology, predicts bounds on the waiting time (upper or lower) that individual jobs will experience with quantified confidence measures. Thus the predictions made by this system constitute a statistical guarantee of best-case and worst-case waiting delay where the confidence measure quantifies the quality of the guarantee. We have implemented this new methodology as part of the Network Weather Service and deployed it on TeraGrid where it currently provides real-time bounds predictions. In the talk we will report on the effectiveness of the system which has been in operation as a prototype for approximately 8 months. We will discuss the methodology and its evaluation using batch-queue logs spanning 10 years at the NSF and open DOE supercomputer centers. We will also demonstrate the web interface to the system and make "live" predictions of TeraGrid delay bounds during the presentation from the web page located at http://nws.cs.ucsb.edu/batchq and we will detail the operation of a set of command-line tools that are portable among all ETF architectures. Our results show that it is possible to predict delay bounds with specified confidence levels for individual jobs in different queues, and for jobs requesting different ranges of processor counts and different maximum execution delays Using these predictions, users with roaming allocations or with allocations at multiple TeraGrid sites can choose the machine that is most likely to minimize turn-around time. Users can also determine the probability that a job will meet a specified deadline in a particular queue. Finally, the system is portable to all ETF architectures making it possible for users to consider the use of heterogeneous resources, and to predict which is most likely to impose the shortest waiting time for their jobs.

Speaker's Bio:

Rich Wolski is an Associate Professor in Computer Science at the University of California, Santa Barbara (UCSB). Having received his M.S. and Ph.D. degrees from the University of California at Davis (while he held a full-time research position at Lawrence Livermore National Laboratory) he has also held positions at the University of California, San Diego, and the University of Tennessee. He is currently also a strategic advisor to the San Diego Supercomputer Center and an adjunct faculty member at the Lawrence Berkeley National Laboratory. Dr. Wolski heads the Middleware and Applications Yielding Heterogeneous Environments for Metacomputing (MAYHEM) Laboratory which is responsible for several national scale research efforts in the area of high-performance distributed computing and grid computing.

Center for Computation & Technology
2003 Digital Media Center • Telephone: +1 225/578-5890 • Fax: +1 225/578-8957